UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS DE LA INFORMACIÓN

Departamento de Comunicación Audiovisual y Publicidad II

LA IDENTIFICACIÓN DE LOCUTORES EN EL ÁMBITO FORENSE

MEMORIA PRESENTADA PARA OPTAR AL GRADO DE DOCTOR POR

Carlos Delgado Romero

Bajo la dirección del Doctor:
Francisco García García

Madrid, 2001

ISBN: 84-669-2185-0

RESUMEN

*Desde comienzos de los años sesenta se viene practicando sistemáticamente, tanto por parte de expertos o instituciones con carácter público como privado, lo que ha dado en llamarse identificación de locutores con propósitos forenses. Hasta prácticamente esta última década, y debido fundamentalmente al carácter multidisciplinar de esta técnica de investigación, dos corrientes metodológicas (ingenieros-fonetistas) han compartido protagonismo en el desarrollo de la misma. Afortunadamente, antes de cruzar la emblemática referencia del año 2000, y como consecuencia del trabajo entre expertos desarrollado en distintos ámbitos internacionales, dicha dicotomía preponderante dejó de tener sentido en favor de una nueva perspectiva de estudio, que en 1.999 fue señalada como la opción más idónea para el actual entorno de identificación forense del locutor. Nos estamos refiriendo a los que denominamos "**Métodos Combinados**", o lo que es lo mismo, aquellos métodos cuyos enfoques básicos de estudio se proyectan y "combinan" a través de tres sistemas de análisis : perceptivo-auditivo, acústico y fonético-lingüístico. Es decir, por una parte, los métodos combinados no representan algo más que la voluntad de no ignorar alguna de las que en la actualidad son consideradas aproximaciones básicas con mayor nivel de fiabilidad y, por otra, la insoslayable necesidad de utilizar dichas opciones de estudio de una forma inter-relacionada. Asimismo, es conveniente aclarar, que dentro de esta propuesta combinada queda contemplada la utilización de aplicaciones complementarias de cálculo, análisis o reconocimiento, ya sean de carácter semiautomático o automático.*

En este sentido, nuestro trabajo introducirá un novedoso modelo combinado, en el que se conjugan las tres opciones clásicas y un prototipo de reconocimiento automático basado en la modelación de clases fonéticas por mezclas de gaussianas (GMM). Dicho modelo será desarrollado en un caso práctico en el que se incluirán otras aportaciones científicas que conformarán la columna vertebral del presente estudio. Estamos hablando, en primer lugar, de la definición de un sistema de evaluación de parámetros, a través del cual éstos quedarán totalmente estructurados y referenciados respecto a sus posibles circunstancias y contextos de comparación.Y, en segundo lugar, a la ubicación de dichos objetos de estudio en cada uno de los sistemas de análisis propuestos para su óptima apreciación.

Adicionalmente, efectuaremos un estudio histórico de la técnica forense de identificación de locutores analizando su problemática, sus expertos, sus distintas perspectivas metodológicas y el estado de la cuestión en la actualidad.

Por último, plantearemos una nueva opción de análisis denominada APRES, cuyo objetivo fundamental será otorgar un alto nivel de objetividad a uno de los sistemas de análisis que integran nuestro modelo.

ABSTRACT

Since the early sixties, the so-called Forensic Speaker Identification (F.S.I.) has been practised by both public and private experts or institutions. Up to the last few years, two main methodological approaches (engineering/phoneticians) have shared leadership and controversy in developing this investigation technique. However, due to different studies recently undertaken by our vanguard forensic international community, a common methodological alternative has been agreed upon. We refer to those named by us as ***"combined methods"****. Such a name, comes from its more remarkable aspect, or similarly, the combined methods will always come framed by the combination of the three basic F.S.I. analysis systems: auditory-perceptive analysis, the acoustic analysis and the phonetic-linguistic analysis****.***

This new conception does not exclude in any way the supplementary analysis represented by the automatic and semiautomatic speaker recognition systems. On the contrary, we must be broadly open to these kinds of options, although, at present they can only be appreciated as an additional element to the three main systems already mentioned.

In this way, our work will introduce a new combined model, integrated by the three basic F.S.I. systems and a complementary speaker automatic recognition prototype, Gaussian Mixture Models-based (GMM). This model will be tested through a practical case involving other scientific contributions to be considered the core of this investigation. Thus,on the one hand a system for assessment of parameters in their varied forensic comparative contexts, is defined. On the other, such parameters will be categorized and located for an accurate evaluation in those that are considered by us to be more suitable analysis systems. Additionally, some F.S.I. historical records, which will mention their experts, their methodological perspectives and state of the art, will be introduced.

Finally, a new analysis tool (APRES) will be presented, and this fundamentally will aim to improve the performance of some perceptive tasks with the utmost thoroughness and effectiveness that forensic speaker identification technique demands today.

INTRODUCCIÓN GENERAL

Cualquiera de nosotros es capaz de identificar una voz conocida a través de un teléfono. Pero también puede ocurrir que nuestro proceso de percepción auditiva pueda jugarnos malas pasadas y hacernos confundir entre sí voces que nos resultan muy familiares.

Partiendo de este simple planteamiento, podemos ser conscientes -aunque sólo sea a nivel perceptivo- de lo fácil o difícil que puede llegar a ser el hecho de asociar una locución concreta con el sujeto emisor que la produjo.

Pero dejando a un lado las referencias domésticas, y situándonos en un entorno de investigación científica, podemos afirmar que el concepto de identificación por la voz, suele estar vinculado a dos grandes objetivos de análisis. En unos casos, estará relacionado con la tarea de verificar si un determinado hablante es realmente quien dice ser, y en otros, con la de establecer una asociación de identidad entre un número determinado de locutores y una muestra de habla anónima.

Existe un enfoque de estudio con un carácter más específico, donde confluyen multitud de factores no favorables a la propia tarea de identificación del habla: fuerte degradación de la señal, no cooperación por parte del sujeto emisor, insuficiencias cuantitativas del discurso objeto de análisis, variabilidad de los planos de expresión, no contemporaneidad de las muestras, distintas situaciones psíquico-emocionales de cada acto de habla, patologías relacionadas con el aparato fonador,etc. Nos estamos refiriendo a la **identificación de locutores en el ámbito forense.**

Pero, ¿ por qué elegir el entorno forense para hablar de identificación de voz ?. Pues sencillamente, porque se presenta como un campo de investigación idóneo, donde se conjugan el abanico de factores reales que caracterizan a los registros de habla no obtenidos en laboratorio -algunos anteriormente citados- y la necesidad del máximo rigor a la hora de emitir conclusiones en la aplicación de la técnica. Nos moveremos en un espacio donde no podremos permitirnos evaluaciones o resoluciones que puedan lesionar en forma alguna, uno de los derechos fundamentales del Hombre: su libertad. Un erróneo dictamen pericial basado en la utilización de esta técnica de identificación podría determinar un incorrecto veredicto judicial de nefastas consecuencias.

Mas dejemos por un momento las particularidades concernientes al ámbito forense. Estaríamos ante una grave omisión si olvidásemos el principal inconveniente con el que se encuentra, tarde o temprano, todo aquel que busca la fórmula definitiva que relacione de forma concluyente un acto de habla con su sujeto emisor. Queremos referirnos a la naturaleza variable de nuestro objeto de estudio.

Cada elocución es distinta a otra cualquiera, aunque sea producida por la misma persona y de la forma más parecida posible a otra realizada con anterioridad. Incluso el habla registrada puede llegar a contener matices diferenciadores al ser reproducida o transmitida con distintos medios, en distintos instantes temporales o en distintos espacios acústicos. Todo ello, sin entrar en las consideraciones de tipo perceptivo relacionadas con el sujeto receptor.

Aunque sin lugar a dudas, la variabilidad intrapersonal de las emisiones habladas será el problema más importante de cara a la individualización de las mismas, no será ésta la única dificultad a la que nos enfrentaremos en nuestro intento de establecer una metodología de identificación por la voz que proporcione un alto nivel de fiabilidad a esta técnica de investigación forense.

¿ Significa esto que en la actualidad no existe una metodología de identificación de locutores suficientemente fiable para su aplicación en el campo forense?

Taxativamente, no. La técnica en cuestión es utilizada y desarrollada por multitud de científicos de todo el mundo, si bien, existen distintos enfoques de análisis adecuados a las necesidades de los diferentes ámbitos legales de aplicación. Aunque el aspecto legal de la identificación forense de locutores pueda resultar más o menos determinante, sus aparentemente distintas perspectivas metodológicas, ponen de manifiesto una realidad muy a tener en cuenta, de cara a preservar la fiabilidad de esta técnica.

¿Cual es la causa de estos distintos planteamientos de análisis?. ¿Está motivada por las características intrínsecas a la naturaleza del habla?. ¿Debemos asociarla con el carácter multidisciplinar de los sistemas de análisis?. ¿Existen referentes comunes en las distintas metodologías, o por el contrario las diferencias resultan críticas e insalvables?. Las respuestas a estos interrogantes conformarán la columna vertebral del presente estudio.

Iniciaremos nuestra investigación introduciéndonos en diferentes fundamentos teóricos tales como la acústica, fonética, percepción, proceso de la producción vocal, etc. En un segundo paso tomaremos contacto directo con el ámbito forense de la identificación por la voz, sus antecedentes históricos y estado de la cuestión en el momento actual. Conoceremos y evaluaremos las distintas opciones, sistemas y herramientas de análisis, para concluir con la definición de un modelo metodológico de aplicación práctica.

OBJETIVOS, ESTRUCTURA Y METODOLOGÍA DE LA TESIS

Dado el carácter casi inédito que en nuestro país tiene la **Acústica Forense** -o lo que es lo mismo, el *conjunto de técnicas científicas de investigación judicial cuyo principal objeto de estudio son los registros sonoros y/o sus elementos afines* (soportes y medios de grabación, transmisión, reproducción, almacenamiento, etc)- llegamos a la conclusión, de que el presente trabajo podría representar una excelente oportunidad para conocer en profundidad la que es considerada “buque insignia”de dichas técnicas: la identificación de locutores. Efectivamente, la naturaleza multidisciplinar de sus posibles enfoques de estudio, confiere a esta tarea de individualización un alto nivel de complejidad en relación a otras técnicas de su mismo entorno: autentificación, procesados de adecuación , ruedas de reconocimiento de voz, pirateo de registros, etc.

Pero al margen de este planteamiento de carácter general, hemos establecido unos objetivos parciales que concurrirán en el que será nuestro *objetivo fundamental*: el diseño de un método forense de identificación de locutores de la más alta fiabilidad. Dicho método, estará

sustentado en unos sistemas, opciones, herramientas y procedimientos de análisis, cuya definición estará determinada por la consecución de los siguientes *objetivos parciales*:

- planteamiento de los distintos factores de problemática relacionados con la técnica, mediante el análisis de sus antecedentes históricos y del estado de la cuestión en el momento actual.

- clasificación general de los métodos forenses de identificación de locutores, señalando las ventajas, inconvenientes y eficacia de las distintas alternativas.

- definición y estructuración de los objetos de estudio .

- diseño de un procedimiento de evaluación de parámetros considerando los distintos factores y contextos en los que se desarrollarán las diferentes tareas de comparación.

- definición de los sistemas de análisis que integrarán nuestro modelo metodológico y ubicación de las referencias de estudio en cada uno de dichos sistemas.

- resolución de un caso práctico mediante la aplicación del método propuesto.

Cuatro capítulos generales estructurarán nuestras hipótesis de trabajo:

En el primero, situaremos nuestro objeto de estudio en sus fundamentos teóricos. Dada la diversidad de disciplinas en las que se referencia nuestra técnica, resultaría extensísimo analizar con total detalle los aspectos teóricos que de una u otra forma inciden en el conocimiento de su verdadera naturaleza. Por esta razón, procuraremos, por un lado no soslayar las informaciones imprescindibles y, por otro, no ir más allá de aquellas de carácter básico que guarden una relación directa con los diferentes factores teóricos que dimensionan la identificación forense de locutores.

Iniciaremos nuestro viaje adentrándonos en el conocimiento de lo que es la auténtica realidad física de una emisión de voz: el sonido. Analizaremos sus componentes fundamentales y sus distintos dominios de representación.

A continuación, abordaremos el proceso de producción acústica del habla conociendo la fisiología y función de los órganos de la fonación, las teorías de la mecánica fonatoria y las más comunes alteraciones del lenguaje y patologías relacionadas con el habla .

Estudiaremos cómo son percibidas las emisiones sonoras a nivel auditivo. Partiremos de un modelo clásico de comunicación por la voz, recordaremos las principales funciones del lenguaje y examinaremos la estructura fisiológica del oído y sus mecanismos de transmisión de estímulos. De la misma forma, evaluaremos diferentes teorías sobre los procesos de codificación de datos sonoros a nivel superior. Establecidas estas referencias, podremos proyectar nuestro análisis sobre el ámbito de estudio de las cualidades psicológicas del sonido: la psicoacústica.

básica para introducirnos en lo que va a ser nuestro entorno de trabajo: opciones y sistemas de

análisis; condiciones, factores, tipos de comparación y decisión, etc. De la misma forma, resultará imprescindible profundizar en el conocimiento de los antecedentes históricos relacionados con la técnica en sus distintas etapas y contextos, para lograr formar un criterio riguroso que nos permita evaluar en la mejor posición los fundamentos que sustentan nuestra propuesta.

Finalizaremos este segundo capítulo realizando una clasificación de las alternativas metodológicas que se han venido utilizando en las últimas décadas, matizando sus características y señalando que métodos y trabajos de futuro son propuestos por la comunidad científica internacional como los más idóneos.

Basándonos en la argumentación desarrollada en los dos capítulos precedentes, presentaremos el que será principal objetivo de nuestra tesis: un modelo de identificación forense de locutores fundamentado en la filosofía metodológica de los Amétodos combinados≅.

El carácter novedoso de nuestro modelo no sólo ha de asociarse a la combinación de los tres sistemas de análisis clásicos con una aplicación de reconocimiento automático concreta, sino también a la definición de un procedimiento de estructuración y evaluación de parámetros, y a la aportación de una nueva opción de análisis que pretende objetivar ciertas estimaciones perceptivas del experto.

En el capítulo final, analizaremos a través de nuestro modelo un caso típico en el que se simularán algunas de las habituales condiciones que caracterizan las muestras de voz en registros forenses. En este sentido, veremos cómo ha de efectuarse la evaluación cualitativa del material dubitado para la posterior obtención de unas muestras indubitadas en las mejores condiciones. Detallaremos tanto los meros aspectos técnicos como las estrategias y normas de procedimiento utilizados por distintas instituciones del mundo forense.

Tomando como referencia nuestro protocolo de evaluación, y situando los distintos parámetros en cada uno de sus correspondientes sistemas de análisis clásico, alcanzaremos unos resultados que serán complementados con las valoraciones obtenidas mediante la utilización de un prototipo de reconocimiento automático.

La metodología escogida para validar nuestra propuesta se sustenta en las siguientes premisas:

- la no existencia de un testimonio documental actualizado en el que se referencien las distintas alternativas metodológicas de identificación forense a nivel global, incluyendo sus antecedentes históricos, circunstancias científicas y los necesarios pormenores político-legales.

- la no existencia de una propuesta metodológica que claramente defina y estructure los objetos de estudio, así como sus procedimientos de evaluación de parámetros y las consiguientes reglas de decisión. En este sentido, han de reseñarse como excepción los estándares publicados por el subcomité de análisis acústicos e identificación de voz [VIAAS, 1991] de la International Association for Identification (I.A.I.), si bien, dichos estándares están referidos exclusivamente a la utilización del que más adelante denominaremos método auditivo-espectrográfico.

- la no existencia de una propuesta metodológica que conjugue los enfoques de análisis presentados en la nuestra.

Partiendo de estos supuestos, y teniendo en cuenta el planteamiento y contenido de nuestro modelo, podemos afirmar que, tanto nuestros objetivos parciales, como nuestra

hipótesis fundamental quedarán plenamente demostrados. Para ello, nuestro modelo combinado se fundamentará en un protocolo de evaluación que permitirá conocer con precisión el valor identificativo de unos elementos de estudio perfectamente definidos. Tales elementos, serán ubicados en unos sistemas y opciones de análisis concretos que posibilitarán su óptima apreciación.

Las principales aportaciones que caracterizarán nuestro modelo serán:

- una opción de análisis para la objetivación del enfoque de estudio de carácter perceptivo (APRES) y ,

- la incorporación a la estructura básica de análisis combinado (perceptivo- acústico-fonético) de un prototipo de reconocimiento automático basado en el modelado de clases fonéticas por mezcla de gaussianas (GMM).

La eficacia y funcionalidad de la primera de ellas se pondrá de manifiesto mediante su valoración a través de distintos presupuestos y leyes clásicas que regulan las tareas de codificación perceptiva.La agilidad y conveniencia de la aplicación automática será evidenciada mediante la valoración de estudios comparativos de sus prestaciones frente a otras alternativas de similar naturaleza.

Tanto la total constatación de las premisas arriba expresadas, como otro tipo de datos que serán vertidos en el desarrollo de esta tesis, sólo podrán ser justificadas a partir de informaciones de carácter reservado que, en cualquier caso, serán convenientemente referenciadas o puestas a disposición del tribunal evaluador. Por este motivo resulta oportuno comentar - llegado este momento - que el autor del presente trabajo, hace ahora diez años se iniciaba como especialista en identificación forense de locutores, en el laboratorio de acústica forense del Cuerpo Nacional de Policía. En el transcurso de esta andadura, inició su formación en el Instituto de Identificación de Voz de la Michigan State University, completando dos años más tarde (1992) su cualificación como experto internacional en Análisis Acústico e Identificación de Voz por la International Association for Identification. Desde entonces ha sido miembro del consejo de cualificación de expertos y del comité ejecutivo del VIASS de la IAI, siendo en la actualidad vicepresidente del mismo. En 1995, [Documento ENFOPOL 80] introdujo en la Unión Europea el proyecto de estandarización en acústica forense que en la actualidad se está desarrollando bajo su responsabilidad (área de identificación de locutores) dentro del entorno de expertos de la Red Europea de Institutos de Ciencias Forenses (ENFSI).

CAPÍTULO I

FUNDAMENTOS TEÓRICOS DE LA TÉCNICA FORENSE DE IDENTIFICACIÓN DE LOCUTORES

I.1.- ELEMENTOS BÁSICOS DE FÍSICA ACÚSTICA

I.1.0.- Introducción

Dentro del presente epígrafe realizaremos una presentación muy elemental de ciertos fundamentos teóricos relacionados con la naturaleza física del sonido del habla.Comenzaremos centrando el concepto de sonido dentro de los ejes dimensionales del proceso perceptivo humano. Analizaremos sus componentes y conceptos básicos, y efectuaremos una clasificación general de los sonidos para comprender con mejor criterio a que tipo de estructuras físicas nos enfrentamos en el caso de las emisiones de voz.

Una vez conocidos estos fundamentos, estaremos en condiciones de dar un paso adelante y entender el carácter trascendente y funcional del análisis de Fourier, o lo que es lo mismo, la posibilidad de descomponer los sonidos complejos - como es el caso del sonido del habla - en sus estructuras simples ; asimismo, veremos cómo las denominadas transformadas de Fourier permitirán la permutación de la representación dimensional de los sonidos entre diferentes dominios físicos de referencia .

Concluiremos, deteniéndonos brevemente en el proceso de conversión analógico/digital de la señal de audio y en el examen de los diferentes planos de representación gráfica del sonido digitalizado, haciendo especial énfasis en su específica funcionalidad dentro del análisis identificativo forense.

I.1.1.- La ciencia del sonido.

Acústica es la ciencia del sonido, incluyendo su producción, propagación, recepción, percepción y sus posibles efectos sobre la materia. La acústica forma parte de la física clásica, concretamente de la mecánica, y dentro de ésta, de la dinámica. Más adelante, nos adentraremos en un área concreta de la acústica dedicada al estudio de la percepción del sonido: la psicoacústica.

Existen multitud de posibles definiciones del sonido. En general, se puede decir que hablar de sonido es hablar de las vibraciones o movimientos recurrentes de un cuerpo - moléculas de aire por ejemplo - dentro de un medio de transmisión elástico. De esta definición, se deduce claramente que el sonido es un fenómeno mecánico, y como tal, requiere de masa y elasticidad para producirse. Dentro de este orden de ideas podemos definir el sonido como un conjunto de energía mecánica modulada.

Si efectuásemos una emisión hablada en la luna, donde el medio de transmisión Aaéreo≅ entre interlocutores es inexistente por la ausencia de atmósfera, no seríamos capaces de percibir dicha emisión.

También puede ocurrir que una emisión sonora no sea percibida, aun ante la existencia de un medio elástico de transmisión. Todos conocemos los silbatos que emiten sonidos para dar

órdenes a ciertos animales, que sin embargo no son perceptibles para el oído humano. En estos casos, donde sí se produce la transmisión de las vibraciones aunque éstas no llegan a ser percibidas como una sensación por nuestro cerebro, estaríamos ante los denominados ultrasonidos o infrasonidos; es decir, aquellas recurrencias de presión o frecuencias sonoras que están por encima o debajo del umbral de audición del oído humano.

Ante este fenómeno, cabe plantearse si resulta funcional la definición general de sonido en el particular terreno de las comunicaciones habladas. Probablemente no.

Al igual que un cuerpo necesita ser excitado por una clase de luz para absorber o reflejar determinadas longitudes de onda que a su vez nosotros percibimos como colores, nuestro oído requiere que las vibraciones mecánicas de las moléculas de aire, se produzcan dentro de unos límites de recurrencia y presión determinados, fuera de los cuales, los sonidos no pueden ser percibidos. Por este motivo, abordaremos la definición del sonido dentro del marco perceptivo de la audición humana.

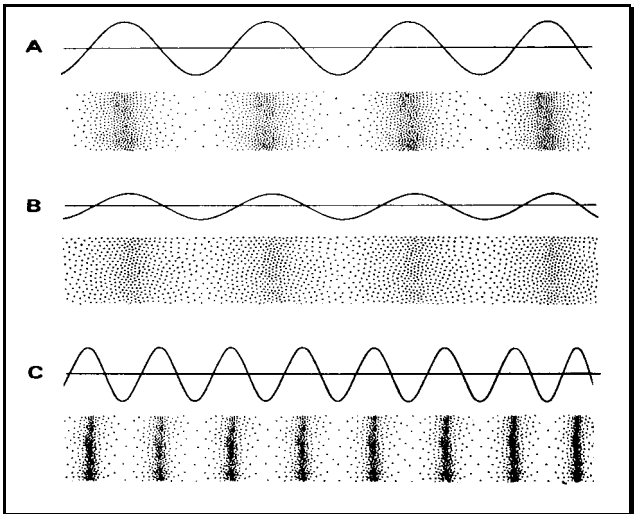
I.1.2.- El sonido y sus componentes fundamentales

Desde el enfoque perceptivo del ser humano, denominamos sonido a las vibraciones con determinada frecuencia -normalmente entre 20 Hz y 20 KHz- de las moléculas de un medio de transmisión elástico, siempre que la presión acústica generada por tal recurrencia sea igual o superior al mínimo nivel de intensidad necesario para percibir dicha frecuencia.

Esta definición de sonido introduce nuevos conceptos. Cuando hablamos de *frecuencia*, nos estamos refiriendo al número de recurrencias por unidad de tiempo. La duración de una recurrencia de la masa vibrante es denominada *período*. Una vibración, ciclo o recurrencia por segundo se corresponde con una frecuencia de 1 hertzio (1 Hz). Por tanto, es fácil deducir que el período y frecuencia de cualquier vibración estarán siempre en una relación inversamente proporcional. Cuanto mayor es la duración del período menor es la frecuencia y viceversa.

La *intensidad* o energía acústica de un sonido es proporcional al cuadrado del máximo desplazamiento de la vibración de una partícula en torno a su posición neutral de equilibrio dentro de un medio elástico. Normalmente es medida en decibelios (dB) , los cuales son un ratio logarítmico entre la intensidad (dB_{sil}) o presión (dB_{spl}) del sonido medido y una referencia dada, por lo general, el umbral de audición a 1 Khz. Esta intensidad o presión sonora es percibida como *sonoridad*, aunque las relaciones entre los distintos componentes físicos del sonido y sus correlatos perceptivos serán comentados más adelante en la parcela de la Psicoacústica.

En la ilustración n11 podemos observar cómo se asocian la frecuencia de vibración (el eje de referencia se corresponde con el nivel de presión estática) y sus correspondientes desplazamientos en las partículas de un medio elástico. En el caso de las ondas A y B la frecuencia de vibración es la misma, sin embargo, el desplazamiento (intensidad o presión) es superior en el caso A. La onda C presenta una intensidad similar a la A, pero una frecuencia de vibración superior.



Un oído humano en óptimas condiciones puede percibir como sonido las vibraciones comprendidas entre 20 y 20.000 ciclos por segundo o hertzios, siempre y cuando su intensidad o presión sonora esté por encima de un umbral determinado. En el caso de los sonidos del habla el rango de frecuencia abarcado comprendería aproximadamente una banda entre los 100 y los 7.000 Hz, aunque en la mayoría de los casos la información disponible en condiciones forenses la encontraremos en el rango 300/4.000 Hz.

Otro parámetro fundamental de la onda sonora es su longitud. Entendemos por *longitud de onda* la distancia recorrida durante un período, por una onda que se propaga a través de un medio desde una fuente emisora (un hablante por ejemplo) a otra receptora (un oyente). La velocidad de propagación es constante y depende fundamentalmente de la densidad del medio por el que transcurre la onda. Para ondas sonoras, en condiciones atmosféricas de presión y temperatura normales (201C) , la velocidad de propagación en el aire es aproximadamente de 344 m/s. Resulta evidente que cuanto más alta sea la frecuencia de un sonido, más corta será su longitud de onda.

I.1.3.- Sonidos simples y sonidos complejos

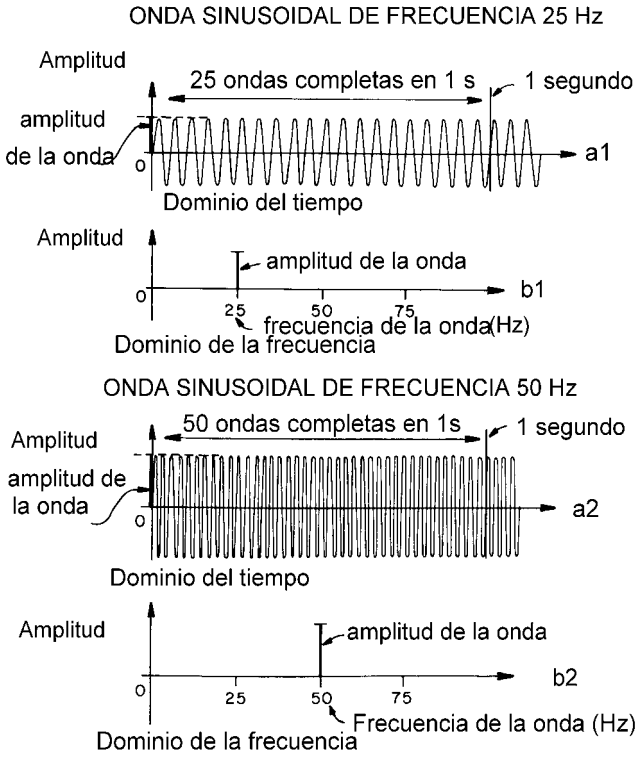
Desde un punto de vista puramente físico, podemos dividir los sonidos en dos grandes grupos: simples y complejos. Un sonido simple es aquel que es consecuencia de un tipo especial de movimiento recurrente de una partícula, denominado movimiento armónico simple (M.A.S.). La trayectoria de vibración de dicho movimiento es una línea recta. Esta vibración depende de fuerzas elásticas centrales, es decir, fuerzas de una magnitud proporcional a la elongación o desplazamiento de las masas vibrantes desde una posición de equilibrio. La distancia entre los puntos de máximo desplazamiento y el eje de posición neutral de la masa recurrente (en el caso del sonido, presión estática atmosférica) se denomina *amplitud* de onda. El M.A.S. tiene una representación gráfica sinusoidal en torno a un eje recto horizontal que referencia el tiempo. Sobre su eje vertical podremos trazar niveles de elongación, desplazamiento, intensidad u otros parámetros equivalentes, y puede proyectarse como un movimiento circular armónico (ilustración n14). De acuerdo a este gráfico la amplitud de onda en cualquier instante podrá obtenerse de la ecuación del MAS: $y = A \operatorname{sen} \theta = A \operatorname{sen} \omega t$.

A pesar de la representación sinusoidal de los sonidos con M.A.S., es conveniente

recordar que los desplazamientos de masas con este movimiento son siempre una línea recta, y los sonidos que representa son percibidos por el oído humano como un tono puro continuo.

En la ilustración n1 2 podemos observar las referencias amplitud en el dominio del tiempo (a_1 y a_2) de dos ondas sonoras sinusoidales de diferente frecuencia o recurrencia de vibración.

El mismo tipo de vibración, que hasta el momento hemos representado en el dominio del tiempo, puede ser representado en el dominio de la frecuencia. En este caso estamos ante otra forma de representación gráfica del sonido: el espectro (en la ilustración n12 apreciamos esta forma de representación en los gráficos b_1 y b_2).



En el espectro, el eje horizontal es la referencia de frecuencia, y dado que estamos ante un M.A.S. con una única frecuencia de vibración, su representación espectral en relación a la intensidad o amplitud estará definida por una sola línea.

Un sonido complejo es aquel que no es simple o sinusoidal. Todos los sonidos del habla son complejos ya que, necesariamente, toda producción acústica humana ha de atravesar la cavidad bucal viéndose sometida al efecto de resonancia que posteriormente comentaremos.

A su vez, todo sonido complejo puede ser periódico, aperiódico o quasi-periódico. En el caso de la onda compleja periódica cada ciclo se completa en el mismo tiempo o período. Sin embargo, en la onda aperiódica la duración de cada recurrencia es aleatoria. Las ondas complejas quasi-periódicas son aquellas en las que el período de cada ciclo no es constante pero la diferencia

es muy pequeña. Todos los sonidos "sonoros" del habla (en los que intervienen las cuerdas vocales) se corresponden con ondas complejas quasi-periódicas. Una excepción a este fenómeno - por su mayor carácter períodico - podría ser un fragmento estable de la realización cantada de un sonido vocálico sostenido, emitido por un cantante de ópera entrenado y registrado en su fase de radiación, o el mismo sonido registrado a nivel glotal.

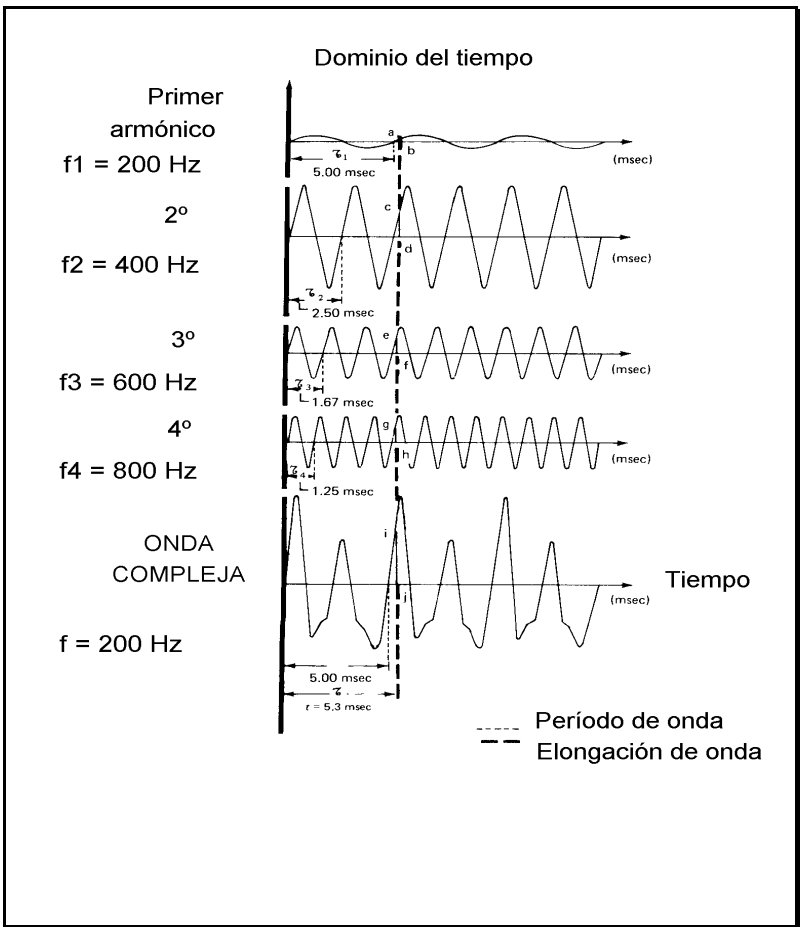
De la misma forma que ocurre con los sonidos simples, los sonidos complejos pueden ser representados de forma gráfica en el dominio del tiempo (*onda sonora u oscilograma*) y en el dominio de la frecuencia (*espectro*). Dado que las representaciones oscilográficas y espectrales de la señal sonora están íntimamente relacionadas con la casi totalidad de los sistemas de análisis utilizados en la identificación de locutores, parece inevitable efectuar un breve comentario en relación con los fundamentos básicos en los que se basa la descomposición de las ondas sonoras complejas.

I.1.4.- El análisis de Fourier

A finales del siglo XIX, el físico y matemático francés Fourier descubrió que cualquier sonido complejo, periódico y estable puede descomponerse en un grupo de ondas simples o sinusoidales de diferentes frecuencias o intensidades. Como ilustración a este importante principio, la figura 3 muestra una onda compleja descompuesta en cuatro sonidos simples, denominados *armónicos* del sonido complejo. La frecuencia del primer armónico o *frecuencia fundamental* (F_0), coincide con la frecuencia del sonido complejo. La frecuencia del resto de armónicos son múltiplos enteros del primer armónico. Por ejemplo, considerando que la frecuencia de una onda compleja es 200 Hz, la frecuencia de su primer armónico o frecuencia a nivel glotal sería también 200 Hz. La frecuencia del segundo armónico sería 400 Hz (dos veces el 11), la del tercero 600Hz, la del cuarto 800 Hz, etc. En términos generales puede expresarse que:

$$f_n = n\, f_1$$

donde f_n es la frecuencia del enésimo armónico y f_1 es la frecuencia del primer armónico o frecuencia fundamental (F_0).



Los cálculos y hallazgos de Fourier son de extraordinaria importancia para el análisis del sonido en general y del habla en particular. Parámetros mensurables de las emisiones habladas, como puede ser el tono o frecuencia fundamental de un determinado locutor, son fácilmente calculables sin necesidad de obtener una muestra de voz a nivel de las cuerdas vocales, puesto que la onda compleja que sale más allá de los labios, tras haber cruzado el tracto vocal y haberse sometido al efecto de resonancia, tendrá la misma frecuencia que su primer armónico o frecuencia fundamental generada a nivel de la glotis.

El número de armónicos que componen una onda compleja es teóricamente infinito, aunque en la práctica sólo unos cuantos -generalmente unos 20- pueden producir una aceptable aproximación de la onda compleja analizada. Si nos fijamos en la figura 3 podremos percibir que la elongación o desplazamiento de la onda compleja en un instante cualquiera, se corresponde con la suma algebraica de las distintas elongaciones en dicho instante, de los diferentes armónicos en los que se descompone la onda compleja.

El análisis de Fourier es la operación matemática que permite descomponer la onda compleja en sus diferentes armónicos simples. De forma inversa, la síntesis de Fourier integra diferentes armónicos de un sonido para la obtención de un tipo determinado de onda compleja. Este último procedimiento es uno de los utilizados en la producción de habla sintética a través de aplicaciones digitales específicas. También es utilizado en un método de investigación sobre las

características acústicas del habla, denominado " análisis por síntesis".

Una de las expresiones matemáticas del análisis de Fourier se corresponde con la siguiente igualdad:

$$v(t) = c_0 + \sum_{n=1}^{4} c_n \operatorname{sen} n\omega t + \nu_n$$

donde:

$v(t)$ = valores de las ordenadas verticales sucesivas de v (de amplitud, intensidad, presión etc.) en una onda períodica compleja, expresadas como una función de tiempo t.

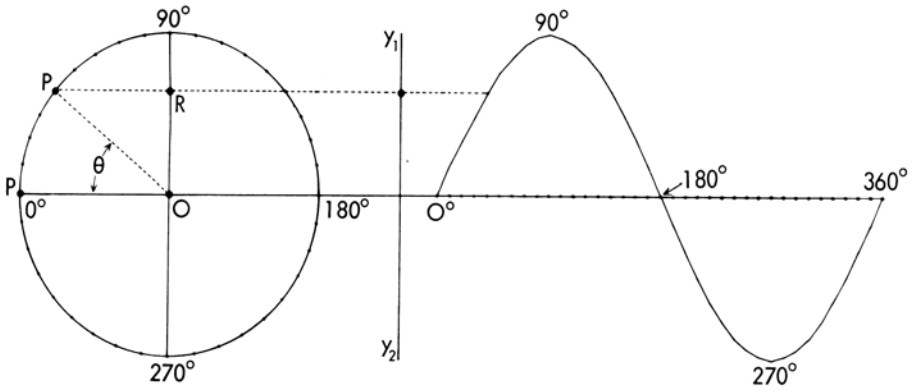
c_0 = un término constante del eje vertical (de amplitud, intensidad, presión etc.). Si ocurriera que la onda compleja es simétrica respecto del eje horizontal del tiempo, entonces el valor de $c_0 = 0$.

c_n = pico de ordenada (de amplitud, intensidad, presión etc.) del enésimo armónico en el Aseno≅ de la expansión de Fourier.

n = número de armónicos considerados.

ω = frecuencia angular del período de la onda compleja $\omega = 2\pi F_0$

ν_n= fase del enésimo armónico, o el tiempo o ángulo equivalente (período =3601) obtenidos por la variación de la onda sinusoidal en relación al eje horizontal para alcanzar una ordenada cero en el valor tiempo cero.



I.1.5.- Las Transformadas de Fourier

Dado que una onda estacionaria compleja puede descomponerse o expandirse como una suma de ondas sinusoidales simples, en determinados tipos de análisis puede resultar más interesante representar dichos armónicos en el dominio de la frecuencia que en el dominio del tiempo, o lo que es lo mismo, en su representación espectral. En la ilustración n15 vemos representados los espectros correspondientes a los mismos armónicos que aparecen como ondas simples -en el dominio del tiempo- en la n13. Aquí, cada línea es el espectro de cada uno de los componentes armónicos del sonido complejo de referencia. Por consiguiente el espectro de una onda compleja, estará compuesto de diversas líneas, cada una de las cuales representa un sonido simple o componente armónico.

La expresión matemática que permite la conversión al dominio de la frecuencia de los valores de representación de la onda compleja en el dominio del tiempo, fue desarrollada por Fourier y se denomina Transformada:

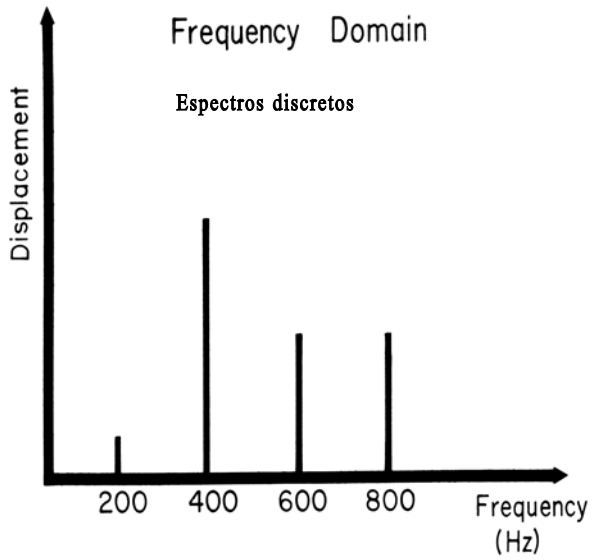
$$V(\omega) = \int_{-4}^{4} v(t)\, e^{-j\omega t}\, dt$$

donde:

$V(\omega)$ = los valores sucesivos de la ordenada V del espectro de una onda compleja como una función de la frecuencia angular ω.

$v\,(t)$ = los valores sucesivos de las ordenadas v expresadas en función del tiempo t.

La Transformada de Fourier proporciona el valor $V(\omega)$ de cada ordenada del espectro para cada frecuencia angular ω ($\omega = 2\pi f$) a lo largo del eje horizontal. De la misma manera, otra expresión matemática - otra transformada - permite llevar a cabo el proceso inverso. Dicha fórmula es conocida como la transformada inversa de Fourier.



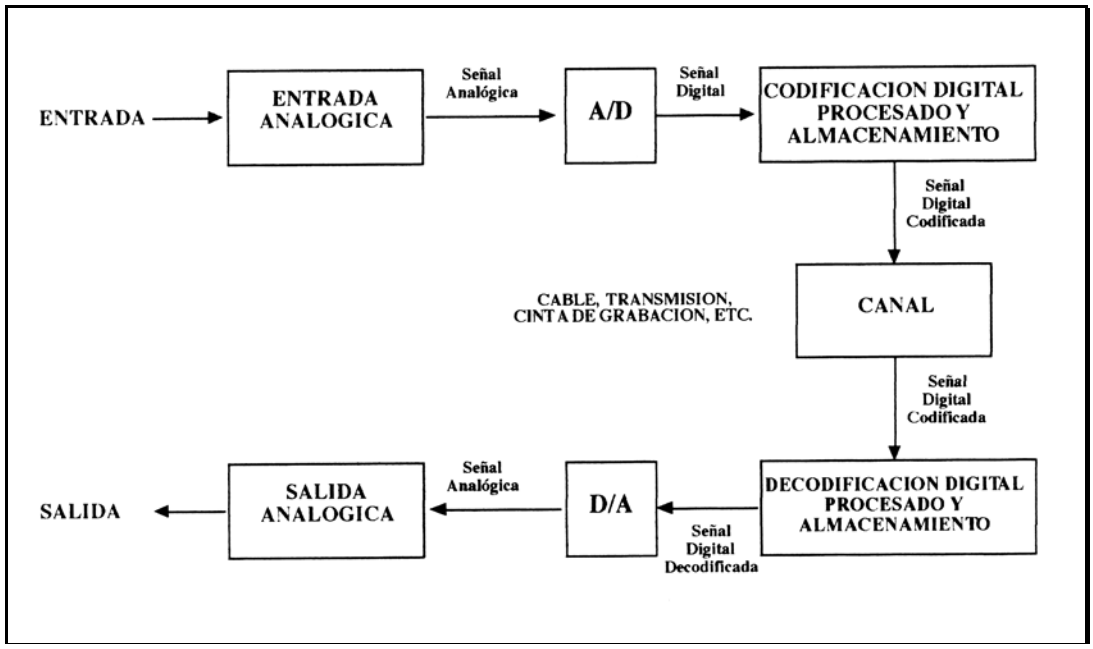
I.1.6.- Audio analógico, audio digital.

Las variaciones de presión generadas por las emisiones sonoras pueden ser procesadas y registradas en equipos y soportes de distinto tipo. Cuando las variaciones de energía mecánica o magnética son un reflejo fiel de las variaciones de presión sonora que fueron previamente traducidas a fluctuaciones de naturaleza eléctrica, nos hallamos ante un procesado analógico (similar) de la señal de audio. Sin embargo, si esas materializadas variaciones de corriente eléctrica, son transformadas mediante una codificación numérica binaria, nos encontraremos ante un procesado de tipo digital.

A primera vista, todo proceso de codificación digital implica una mínima degradación de la señal original. Si bien, es igualmente cierto que cualquier proceso de conversión digital convenientemente administrado (en sus fases de muestreo, retención, cuantificación, compresión/expansión, filtrado, modulación, etc.) no sufrirá en ningún caso los graves inconvenientes del efecto canal propios de la transferencia analógica (ruido, distorsiones lineales o no lineales, etc.). Por otra parte, todas las copias digitales de una señal de audio original (de primera o enésima generación) serán siempre prácticamente idénticas en calidad a la original, circunstancia que nunca se produce en el supuesto analógico.

Además de las citadas, existen otras muchas ventajas del procesado de audio digital respecto del analógico: funcionalidad y agilidad de los procedimientos de trabajo, almacenamiento, conservación y transmisión de la señal, etc.

La siguiente ilustración, muestra el esquema clásico de procesado completo de una señal de audio en su conversión analógico/digital (A/D) y digital/analógico (D/A).



En el proceso arriba descrito, las fases más críticas se asocian a las tareas de conversión del campo analógico al digital y viceversa. En síntesis, un convertidor A/D transforma la señal de entrada analógica (tensión o corriente) en una frecuencia o serie de impulsos cuyo tiempo se mide

para proporcionar una salida digital representativa y proporcional; también puede efectuarse una comparación de la señal de entrada con una referencia variable utilizando un convertidor interno

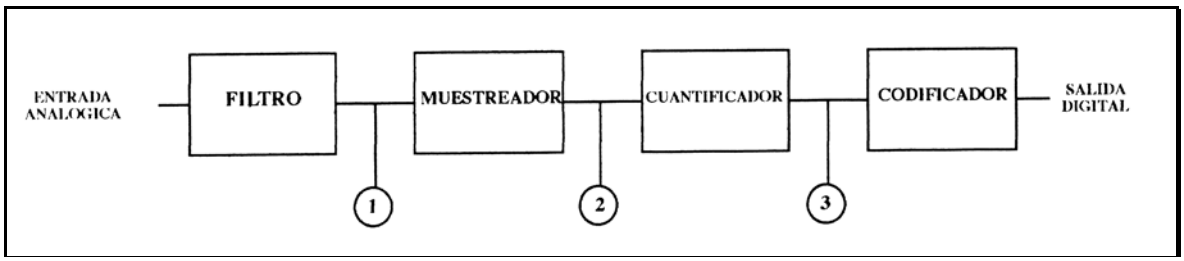
D/A para obtener una salida digital.

La conversión A/D se hace en varias etapas:

- *filtración:* limita la anchura de banda de la señal analógica.
- *muestreo:* convierte una señal de tiempo continuo en una señal de tiempo discreto.
- *cuantización:* convierte una señal de valor continuo en una de valor discreto.
- *codificación:* define el código de la señal digital según la aplicación que se ejecuta.

En el siguiente diagrama de bloques se muestra el proceso simplificado de un sistema de conversión A/D:

∈ Señal analógica filtrada, continua en el tiempo y en el valor.



∉ Señal muestreada, discreta en el tiempo y continua en el valor.

∠ Señal digital, discreta en el tiempo y en el valor.

Los convertidores D/A producen el efecto inverso al convertidor A/D. Estos circuitos, han de ser capaces - a partir de la lectura del correspondiente código digital - de generar una tensión o corriente continua en el tiempo y en el valor lo más similar posible a la señal analógica originalmente introducida en la fase inicial del proceso.

Otros dos importantes elementos en la metamorfosis de conversión A/D, son las premisas insoslayables relacionadas con el Teorema de Nyquist o Shannon y el correspondiente efecto Aalias≅ o aliasing. Según Nyquist, para obtener una señal correctamente muestreada, la frecuencia de muestreo ha de ser, como mínimo, el doble de la frecuencia máxima objeto del muestreo. Si

no ocurriera así, el espectro original solaparía a la parte modulada del espectro y estaríamos ante la presencia de un ruido extraño o Aalias≅. Por esta razón, los sistemas de conversión A/D llevan incorporados filtros de orden muy elevado (filtros anti-aliasing) a fin de eliminar estos efectos

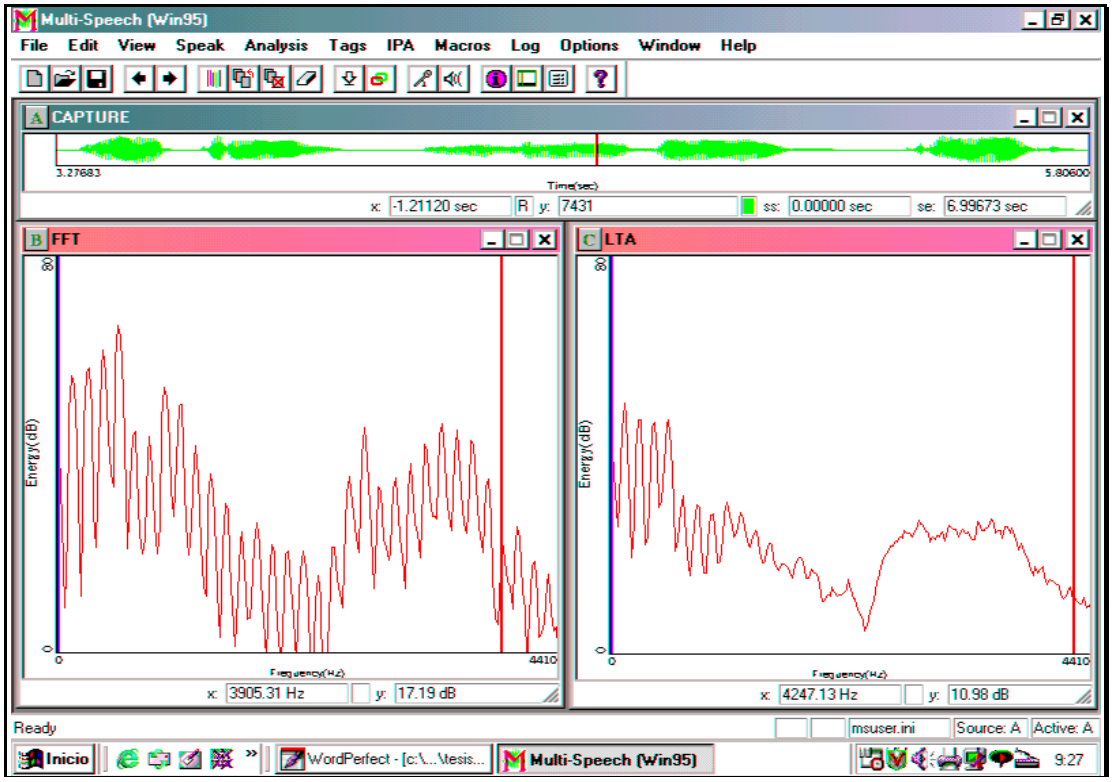
no deseados.

En la actualidad, cualquier ámbito de audio profesional, incluido el forense, se beneficia grandemente del procesado digital de la señal en muy diversas tareas: análisis, grabación, reproducción, transferencia, almacenamiento, display, edición, filtrado, ecualización, expansión, compresión, etc. El análisis del habla con fines identificativos no es una excepción pues, como veremos más adelante, las opciones digitales utilizadas en distintas fases de estudio poseerán una gran agilidad operativa que nunca podría ser alcanzada con herramientas de carácter analógico.

I.1.7.- Espectros FFT, LTA, Espectrogramas, Sonogramas.

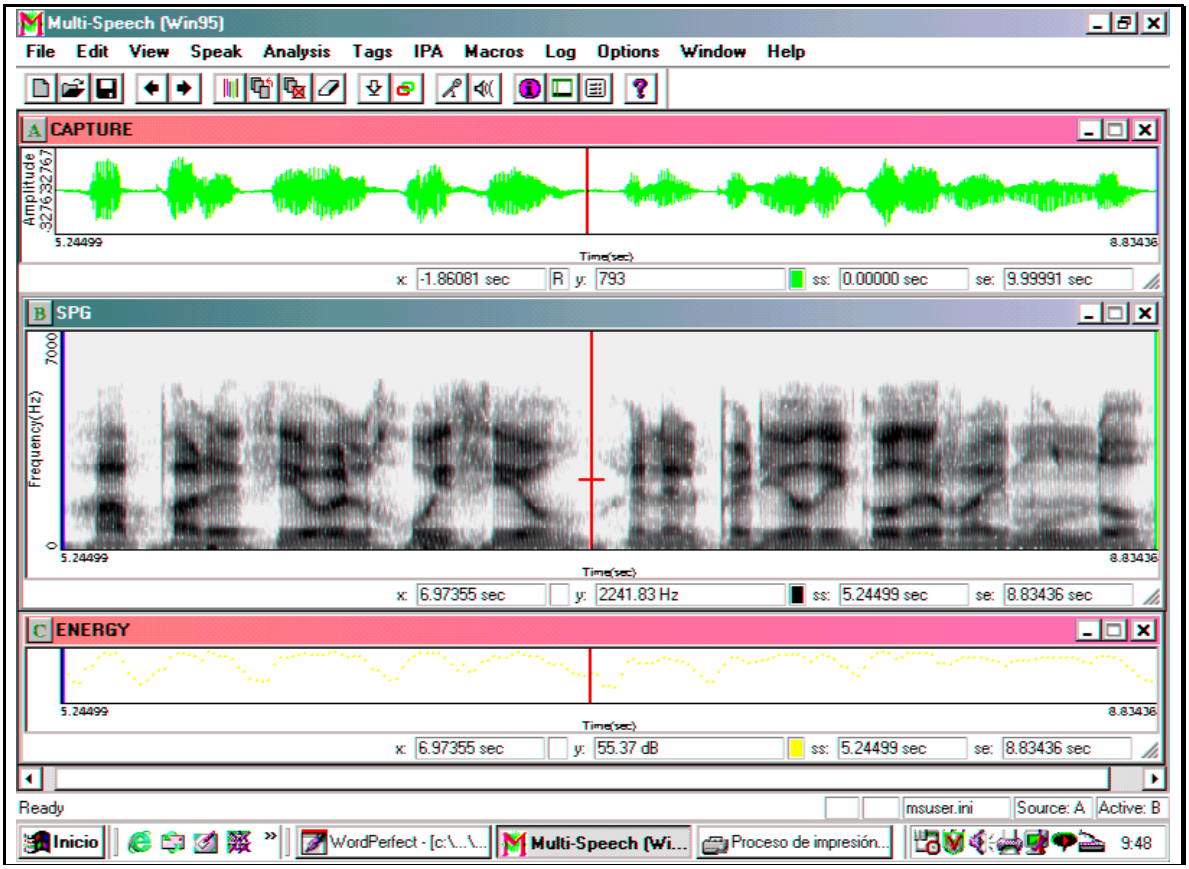
Sabemos que el sonido del habla puede representarse como un desplazamiento de valores de amplitud en el dominio del tiempo. También conocemos que dichos valores de amplitud, presión, intensidad, etc, pueden ser dimensionados en el dominio de la frecuencia, denominándose espectros. Pues bien, la señal digitalizada con la que habitualmente trabajan los profesionales del sonido presenta dos tipos de representaciones espectrales: una sucesión de espectros a corto plazo, *short-term spectra (FFT)* o un espectro a largo plazo promediado o *long term average spectrum (LTAS)*.

En el primer caso - espectro a corto plazo- nos referimos a un segmento temporal de señal muy pequeño (en torno a los 25 ms) en el que dicha señal puede considerarse quasi estacionaria. En el segundo, se efectúa un promediado de la intensidad de las distintas frecuencias en un intervalo temporal de mayor duración (segundos). En la siguiente ilustración (n18) podemos observar estas distintas formas de representación: ventana A, oscilograma; ventana B, espectro a corto plazo (FFT) correspondiente al instante temporal señalado por el cursor de la ventana A. Ventana C, espectro a largo plazo o LTA de los 2,5 s. representados en la ventana A.



Cuando queremos observar la continuidad espectral -sucesión de espectros a corto plazo- en relación al tiempo nos encontramos ante otra forma de representación gráfica del sonido: el

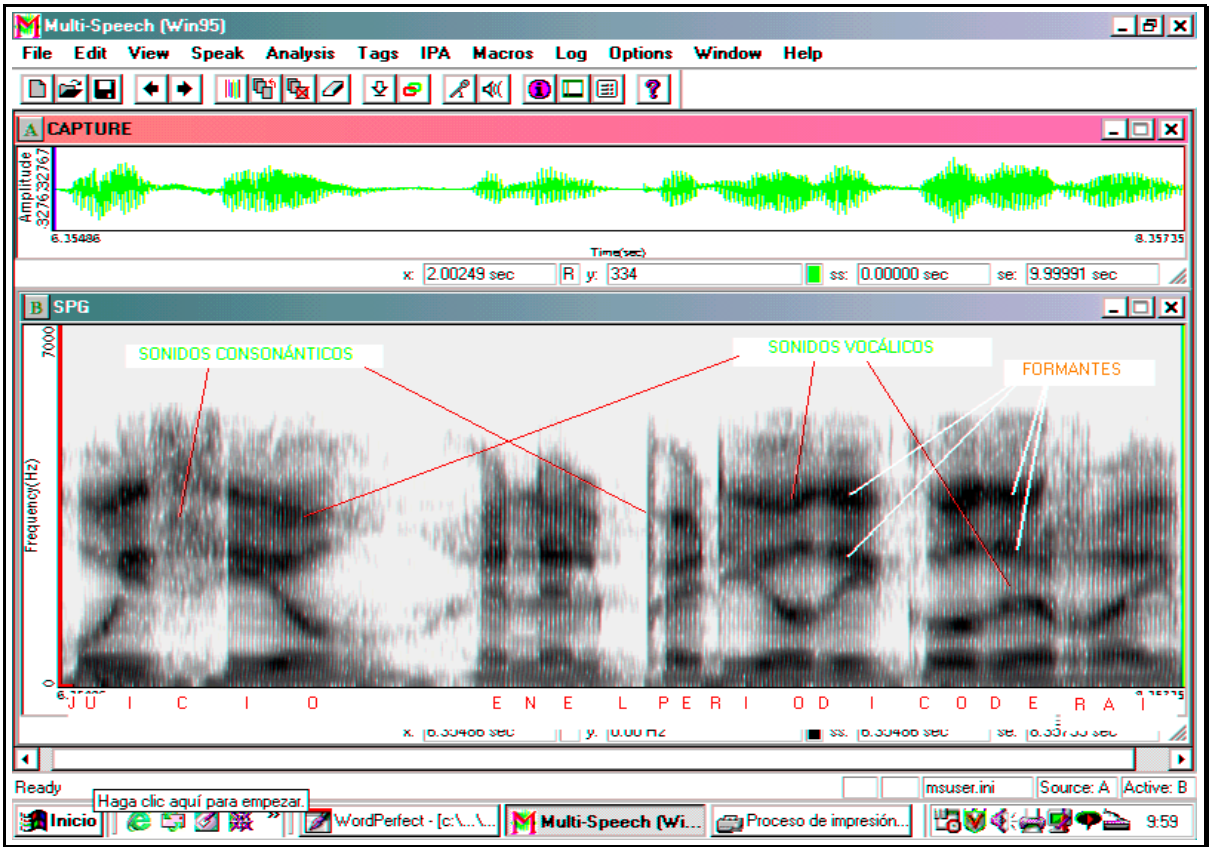
espectrograma o sonograma. En este caso, estamos representando los espectros de diferentes frecuencias con sus correspondientes niveles de intensidad o amplitud, en el dominio del tiempo.



En la ilustración n19, ventana B, observamos el sonograma del fragmento de voz representado en forma de onda en la ventana A. El eje de abscisas representa el tiempo (3,5 s.) mientras que el eje vertical o de ordenadas es la referencia en rango de frecuencia (7000 Hz). Las intensidades de energía se manifiestan de acuerdo a una escala de tonos de grises o, en otros casos, como una gradación de diferentes colores. En la ventana C, observamos el contorno de energía.

Existe cierta controversia en el ámbito del análisis acústico sobre si es más o menos acertado el uso de un término u otro - sonograma o espectrograma - para denominar este forma de representación gráfica de la señal. En nuestra opinión, resulta más conveniente y funcional emplear la palabra sonograma para evitar posibles confusiones entre los conceptos de espectro y espectrograma. No obstante, y concretamente en el ámbito forense, está bastante extendido el uso del término espectrograma para nombrar la representación de la frecuencia de sonido en el dominio del tiempo, no siendo a nuestro entender, un problema de nomenclatura sino más bien una cuestión pragmática.

Los primeros sonogramas o espectrogramas fueron obtenidos con un espectrógrafo

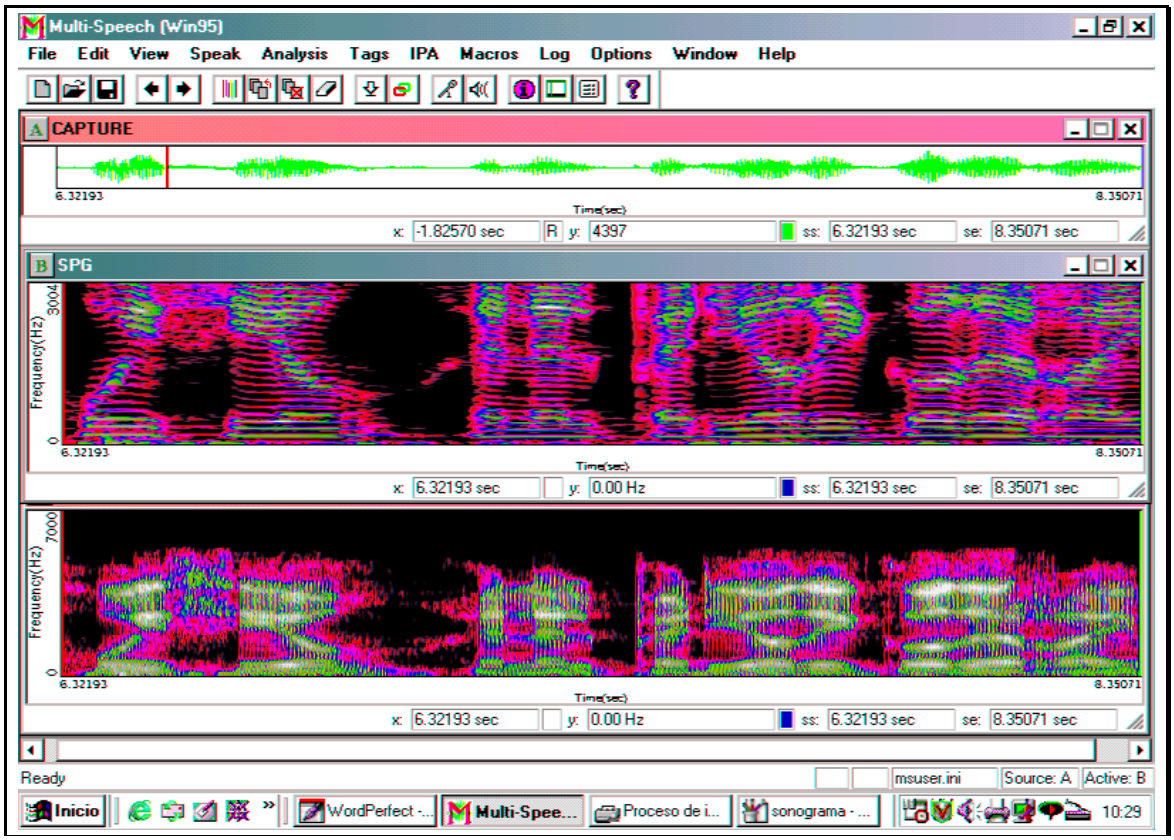


acústico analógico. En la actualidad, algunos ordenadores ejercen la función de sonógrafos - o espectrógrafos acústicos - permitiéndonos visualizar la señal sonora en sus diferentes formas de representación, en tiempo real y de forma simultánea. Además, dicha señal puede ser almacenada y recuperada, procesada, editada, etc.

El sonograma es, con toda seguridad, la forma de representación gráfica más funcional para la observación simultánea e inmediata de los componentes fundamentales que a nivel perceptivo referencian el sonido del habla; o lo que es lo mismo: el tono, el timbre, la intensidad acústica y la duración.

En la anterior ilustración, el tono estaría relacionado con la distancia existente entre las estriaciones verticales de los sonidos vocálicos. El timbre se identificaría con las manchas (formantes) en las que la energía se distribuye de forma uniforme y paralela al eje de abscisas o eje temporal (en la ilustración se corresponden con las grafías de los elementos vocálicos). La intensidad se asocia a la mayor o menor negrura de la energía: mayor intensidad fonatoria =

mayor oscuridad y, la duración, es directamente proporcional a la longitud de la energía respecto



del eje de abscisas.

Utilizando las distintas opciones de análisis que cada caso requerirá - frecuencia de muestreo, filtros de representación sonográfica, rangos dinámicos y de frecuencia, enventanados, intervalos temporales, etc. - obtendremos distintas clases de sonogramas que nos proporcionarán informaciones complementarias de gran importancia, en la búsqueda de las peculiaridades que caracterizarán las diferentes locuciones.

En la ilustración de arriba observamos dos sonogramas de la misma emisión de voz. En la ventana B, con filtro de banda estrecha (26,92 Hz), sobre un rango en frecuencia de 3.004 Hz y un intervalo temporal de 2 s. En la ventana C, con filtro de banda ancha (161,5 Hz), sobre un rango en frecuencia de 7000 Hz y un intervalo temporal idéntico. El primero de ellos puede aportarnos datos muy útiles sobre la estructura fina de la señal: distribución y número de armónicos, frecuencia fundamental, etc. El segundo, puede informarnos sobre el mayor o menor grado de apertura o cierre del resonador bucal, el adelantamiento o retraso de la lengua en las diferentes realizaciones, grados de tensión articulatoria, etc. En términos generales, los sonogramas de banda ancha siempre presentarán poca resolución a nivel espectral

y,consecuentemente, mucha resolución a nivel temporal; mientras que en los de banda estrecha se producirá exactamente el efecto contrario.

En una segunda fase, las informaciones sobre índices sonográficos podrán ser asociadas a unas causas fonoarticulatorias de carácter más específico contribuyendo, en definitiva, a la construcción del perfil característico de un locutor determinado.

Es decir, cada fonema o grupo fónico de una emisión hablada, presentará en su representación sonográfica una forma o "pattern" que nos pondrá de manifiesto informaciones instantáneas sobre las peculiaridades fonoarticulatorias de dicha emisión.

Como hemos podido comprobar existen diferentes formas de representación gráfica de la señal de habla. Aunque la más práctica de ellas sea el sonograma - de cara a efectuar una rápida observación de los diferentes índices acústicos - cada tipo de análisis sobre señales vocales requerirá una forma de representación determinada (oscilogramas u ondas sonoras, espectros, etc.).

Solamente el experto sabrá cual es la opción más idónea para cada caso, aunque no debemos olvidar que en cualquiera de las formas de representación gráfica citadas, están incluidas todas las referencias que dimensionan la señal sonora.

I.2.- EL PROCESO DE PRODUCCIÓN ACÚSTICA DEL HABLA: LA FONACIÓN.

I.2.0.- Introducción

Este segundo apartado de fundamentos teóricos lo dedicaremos a describir aquellos elementos, mecanismos y aspectos más relevantes relacionados con el acto de producción de los sonidos del habla: la fonación. Comenzaremos diferenciando de forma muy simple los tres niveles fisiológicos en los que pueden agruparse los órganos que intervienen en la fonación. Analizaremos las distintas teorías de la mecánica de producción vocal a nivel glótico y las fuentes de energía que generarán cada una de las realizaciones concretas de los diferentes elementos fónicos del habla. Haremos especial hincapié en el fenómeno de resonancia vocal, responsable directo de las estructuras del timbre, factor de vital importancia en el proceso de individualización de las emisiones de voz.

Conscientes de la existencia de excelentes manuales específicos sobre la materia [Prater y Swift ,1986], [Le Huche y Allali, 1993], [Perelló y Salvá, 1980], etc., hemos considerado oportuno no pasar de largo sin detenernos brevemente en algunas de las más comunes

alteraciones del lenguaje y ciertas patologías características relacionadas con los procesos fonatorios.

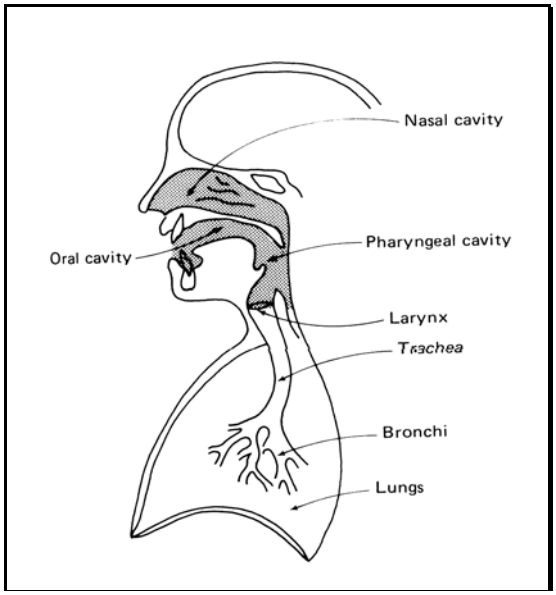
I.2.1.-Fisiología de la fonación.

El conjunto total de órganos que intervienen en la fonación, pueden ser clasificados en tres grupos bien diferenciados:

-Órgano respiratorio.
-Cavidad laríngea u órgano fonador.
-Cavidades resonadoras.

I.2.1.1.- Órgano respiratorio.

Está formado por los órganos propios de la respiración: pulmones, bronquios y tráquea; los pulmones son los proveedores del aire suficiente para la fonación; tienen dos movimientos, inspiración y espiración, siendo en este segundo en el que se puede producir el sonido articulado. El aire contenido en los pulmones pasa por los bronquios y desde éstos a la tráquea.



Los pulmones están divididos en lóbulos separados por cisuras y recubiertos por la *pleura* que está formada por dos hojas: la interna o visceral y la externa o parietal, quedando entre ambas

la cavidad pleural. Se encuentran alojados en la cavidad torácica y tienen la posibilidad de expandirse en distintos planos (transversal, antero posterior y longitudinal), gracias a la acción externa de fuerzas que provocan un cambio de volumen constante.

La renovación del aire se hace mediante la inspiración y la espiración, pudiendo ser apreciada dicha coordinación de movimientos, junto con los de tipo diafragmático a través del *neumógrafo*.

El diafragma es un músculo impar y no simétrico - el lado derecho es más alto que el izquierdo- tiene forma de cúpula y separa la cavidad torácica de la abdominal. En la respiración sus movimientos son automáticos, si bien dichos impulsos automáticos pueden ser alterados por la necesidad de oxígeno o por la realización de un esfuerzo voluntario; al producirse el estímulo, el diafragma se contrae hacia abajo con lo que el tórax amplía su extensión y viceversa.

En la fase inspiratoria, el adulto normal realiza alrededor de 16 insp./minuto tanto a nivel vegetativo como para la fonación, entrando en funcionamiento los siguientes músculos:

Escalenos ..... Elevadores del tórax
Intercostales ..... Expansores del tórax
Diafragma ..... Extensores de la cavidad abdominal
Esternocleidomastoideo ..... Respiración más forzada
Extensores de la C. Vertebral ..... Respiración muy forzada

En la fase espiratoria no intervienen otro tipo de músculos, gracias al carácter antagónico de los mencionados anteriormente. El tiempo empleado en este paso (también utilizado en la fonación) es un poco más dilatado que el de la fase anterior, siendo también mayores los volúmenes de aire que se utilizan en la fonación que los empleados en una respiración en reposo.

Existen diversas formas de clasificar los tipos de respiración . En general, las formas de respirar dependerán siempre de las circunstancias (fonación , hematosis) y momentos en que se desarrollen:

Clavicular.- Conlleva elevación y descenso del tórax con hundimiento del pulmón en su parte superior. Su uso frecuente causa problemas en la voz debido a la utilización de los músculos del cuello que provoca insuficiencia de aire y tensión laríngea. Suele producirse un ataque de glotis debido a la posición que ocupa la laringe. Esta posición es utilizada en cualquier emoción repentina, provocándonos elevamiento del tórax..

Torácica.- Asociada a una expansión y retracción del tórax adecuada para el habla conversacional.

Abdominal o diafragmática.- Es el más sencillo de realizar desde un punto de vista mecánico, permitiendo realizar un mayor intercambio de aire. En el caso de ser necesaria una respiración de emergencia ponemos en funcionamiento la masa muscular de la columna vertebral para dar mayor apoyo y fuerza al acto.

La intensidad o fuerza de la emisión hablada depende siempre de la presión subglótica y no del volumen pulmonar. He aquí algunos valores de referencia habitualmente citados:

Conversación Avoz baja≅..... 30 dB_{SPL}.
Conversación normal ..... 60 dB_{SPL}
Reuniones - party..... 70 dB_{SPL}

También resulta interesante comentar algunos términos y datos importantes relacionados con el control de una respiración normal o en el uso fonatorio:

Capacidad Vital.- Cantidad de aire que podemos espirar después de una inspiración máxima (se mide con un espirómetro):

4.800 ml. para los hombres
3.200 ml. para las mujeres

Volumen Periódico.- Cantidad de aire que se inspira y espira en una respiración normal:

500 ml. tanto para hombres como mujeres

Volumen Inspiratorio de Reserva.- Cantidad máxima de aire que se puede inspirar después de una inspiración usual:

3.200 ml. para los hombres
2.000 ml. para las mujeres

Volumen Espiratorio de Reserva.- Cantidad máxima de aire que se puede espirar después

de una espiración usual:

1.100 ml. para los hombres
700 ml. para las mujeres

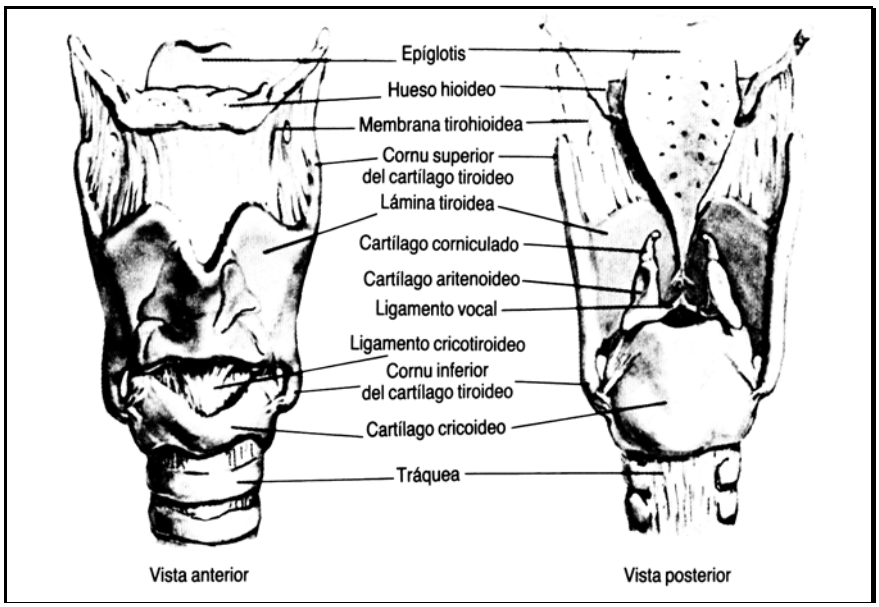
Tiempo Máximo de Fonación.- (con una vocal sostenida):

15 segundos para los hombres
14,3 segundos para mujeres

Entre los pulmones y la cavidad laríngea encontramos la *tráquea* que es un tubo grueso de unos 12 centímetros de longitud, formado por cartílagos (de 16 a 20); se dirige hacia abajo por detrás del esternón y delante del esófago hasta llegar a la mitad del pecho, donde se bifurca en dos *bronquios* que penetran respectivamente en cada uno de los pulmones. En el interior de los mismos se ramifican en ramas cada vez más finas denominadas *bronquiolos respiratorios* que derivan en los llamados *conductos alveolares* y éstos, a su vez, en *alvéolos* pulmonares rodeados de un gran número de capilares.

I.2.1.2.- Cavidad laríngea u órgano fonador.

La cavidad laríngea está situada inmediatamente por encima de la tráquea y comprende un complejo sistema de músculos, nervios, cartílagos, ligamentos, tendones, etc. Comúnmente es dividida en cuatro segmentos : subglotis, glotis, ventrículo y vestíbulo laríngeo. Fundamentalmente está constituida por una serie de cartílagos -cricoides, tiroides, epiglotis, aritenoides, corniculado, cuneiformes- que envuelven las llamadas cuerdas vocales; éstas son en realidad dos tendones cuyo reborde interior es algo más grueso y están situadas horizontalmente en dirección antero-posterior. Por su parte anterior se unen al cartílago tiroides (nuez o bocado de Adán) y, por la posterior, a los aritenoides. El espacio vacío que queda entre las dos cuerdas vocales recibe el nombre de glotis.



Las cuerdas vocales son responsables de la producción y clasificación del material fónico: si se aproximan y comienzan a vibrar se origina el sonido articulado sonoro; si por el contrario solamente se acercan pero no vibran, originarán el sonido articulado sordo. Para la formación del sonido vocálico sonoro, las cuerdas están más tensas y, en consecuencia, la frecuencia de vibración de las mismas es mayor; el grado de abertura de la glotis es mínimo y, por lo tanto, también lo es el gasto de aire. Para la formación del sonido consonántico sonoro, las cuerdas están menos tensas y, consecuentemente, la frecuencia de recurrencia es menor, siendo la abertura de la glotis mayor y, por ello, es también mayor el gasto de aire. De ahí que los

sonidos consonánticos tengan un ruido característico que se forma por el paso del aire a través de los pliegues vocales, de mayor intensidad que el de los sonidos vocálicos.

La función primordial de la laringe es facilitar la obturación de la tráquea en los actos de deglución, aunque para nosotros ésta será una función intrascendente en favor de la fonación. El hecho de que esté situada más baja que la de otros animales, implica un mayor volumen de resonancia de la faringe, pero impide la deglución y respiración conjuntamente, cosa que si puede realizar el bebé al tenerla en una posición más elevada.

Está suspendida mediante los músculos supra hioideos y los infrahioideos, interviniendo también los laríngeos intrínsecos y los extrínsecos cuando de forma sinérgica y antagónica hacen que suba y baje su posición, respectivamente, tanto en lo concerniente a los sonidos agudos y la

espiración como en los graves y la inspiración. Todos ellos participan en la fonación produciendo un equilibrio entre la fuerza realizada por los intrínsecos y la ejercida por la presión del aire. Si se rompe este equilibrio se produce una alteración en el tono, la intensidad y el timbre.

A nivel del cartílago tiroides se encuentran los repliegues vocales, elementos básicos de la fonación, que van progresivamente aumentado de tamaño a medida que el niño se hace adulto: 5 mm. en el niño; de 16 a 23 mm en el varón adulto y de 12 a 17 mm en la mujer.

En la parte de atrás de la lengua se encuentra el *additus laríngeo* que es lo que une la boca con la laringe; es la entrada de la laringe entre la epiglotis, los pliegues aritenoepiglóticos y la incisura interaritenoidea.

La laringe está constituida por un esqueleto cartilaginoso unido entre sí, y a los órganos vecinos, por medio de un complejo sistema de músculos-ligamentosos. Toda la estructura está recubierta por una mucosa e inervada por dos ramas del nervio vago. Se relaciona con la glándula tiroides en la parte delantera, con la arteria carótida y la vena yugular interna en la parte lateral, y con la laringo-faringe, boca del esófago y senos pinniformes en la parte trasera.

Tiene una capacidad reflexógena muy importante hasta tal punto que provoca, sobre todo en el niño, un conjunto de respuestas cardíacas (bradicardia, arritmia, cambios en la tensión arterial, cambios respiratorios, tos, disminución de la frecuencia de respiración) que pueden llegar incluso a producir una parada cardio-respiratoria y la muerte. De hecho, la actividad refleja inducida por estimulación laríngea ha sido frecuentemente relacionada con la muerte súbita en la infancia.

Esta compuesta por tres cartílagos impares (tiroides, cricoides y epiglotis) y tres pares (corniculados, cuneiformes y aritenoides), unidos por los siguientes ligamentos: membrana cricotraqueal, membrana cricotiroidea (dónde se realizan las traqueotomías), articulación cricotiroidea, articulación cricoaritenoidea y la membrana tirohioidea.

Se subdivide en cavidad *supraglótica*, que abarca desde el vestíbulo o additus laríngeo hasta los repliegues vocálicos, encontrándose en este espacio los pliegues ventriculares que no tienen función directa alguna en la fonación, cavidad *glótica* o espacio comprendido entre los repliegues o cuerdas vocales, y cavidad *infraglótica* que se extiende desde la anterior hasta el primer anillo traqueal. En los laterales glóticos se encuentra el elemento principal que da origen a las vibraciones de distintas clases de sonidos, incluido el lingüístico, nos estamos refiriendo a los *repliegues vocálicos,* cuyos músculos más importantes son:

Cricotiroideo .- Son músculos pares y en forma de abanico. Es el responsable de tensar los pliegues vocálicos, produciendo un balanceo del tiroides y provocando un acercamiento al cricoides.

Cricoaritenoideo posterior **.-** Son los únicos músculos que dilatan o abducen los pliegues vocales.

Cricoaritenoideo lateral **.-** Son músculos antagónicos a los cricoaritenoideos posteriores. Funcionan como aductores de los pliegues vocales.

Interaritenoideo.- Es el único músculo impar. Su función es aproximar los cartílagos aritenoides.

Aritenoepiglótico.- Desciende la epiglotis.

Tiroaritenoideo Superior.- Es el constrictor de la glotis

Tiroaritenoideo Inferior.- Es constrictor de la glotis y el que más directamente está involucrado en la voz. Se subdivide en tiromuscular y tirovocal (a medida que aumenta el tiempo de contacto entre ambos músculos tirovocales se enriquece el timbre).

Es necesaria una coordinación precisa de los músculos laríngeos para la correcta aducción de los pliegues vocálicos, provocando en casos de mal funcionamiento transtornos por hipo o hiperaducción.

La Inervación del sistema fonatorio comprende los siguientes nervios craneales:

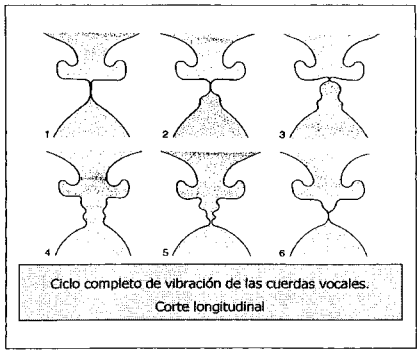
VAGO - ..... Par X (Inervación Intrínseca)
TRIGÉMINO - ..... Par V (Inervación Extrínseca)
FACIAL - ..... Par VII (Inervación Extrínseca)
HIPOGLOSO -..... Par XII (Inervación Extrínseca)

Es el nervio vago el que está más directamente relacionado con la laringe. Se subdivide en: laríngeo superior y laríngeo inferior o recurrente.

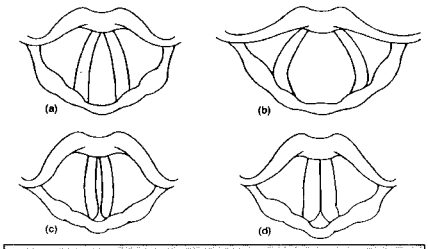
La irrigación de la laringe corre a cargo de la arteria laríngea superior, posterior e inferior. Todas ellas derivan de la arteria tiroidea.

El acto fonador que relaciona el proceso de apertura/cierre de las cuerdas vocales con el proceso respiratorio puede definirse en diez pasos:

1.- Inspiración de aire por los pulmones.
2.- Pulmones llenos de aire. Excitación nerviosa sobre las cuerdas vocales.
3.- Cierre cordal.
4.- Espiración. Aumento de la presión subglótica.
5.- Apertura y movilidad cordal.(VOZ)



Ciclo completo de vibración de las cuerdas vocales. Corte longitudinal



Vista transversal superior de la glotis: (a) Respiración normal. (b) Respiración profunda. (c) Sonido sonoro.

6.- Cierre cordal. Aumento de la presión subglótica.
7.- Apertura y movilidad cordal.
8.- Disminución del volumen pulmonar.
9.- Cese de excitación nerviosa.
10.- Nueva inspiración.

La vibración de las cuerdas vocales genera un armónico u onda sonora, que denominamos frecuencia fundamental. Esta onda adquiere inmediatamente un carácter de sonido complejo ya que crea -por efecto de la resonancia que acontece en las cavidades supraglóticas- una serie de armónicos añadidos que conformarán la estructura acústica de los sonidos vocálicos y semivocálicos. El proceso es sencillo: la onda compleja formada en la laringe pasa a las cavidades supraglóticas; éstas, actúan sobre aquellos armónicos que coinciden con las frecuencias de resonancia de dichas cavidades, potenciándolos. El conjunto formado por el tono fundamental y los armónicos filtrados (por las cavidades resonantes) constituyen la esencia de lo

que llamamos timbre.

El número de vibraciones de las cuerdas vocales en relación al tiempo depende de una serie de factores a su vez interdependientes: masa de la parte vibratoria de las cuerdas vocales, tensión de las mismas, área de la glotis durante el ciclo, valor de la presión infraglótica y amortiguación de las cuerdas vocales.

Los individuos con una frecuencia fundamental grave poseen unas cuerdas vocales más voluminosas que los que tienen frecuencia fundamental aguda; del mismo modo, cuando la frecuencia se eleva, el espesor de las cuerdas disminuye y, al mismo tiempo, se alargan. En general, la diferencia en la constitución fisiológica de las cuerdas vocales es, en gran parte, responsable de las diferencias de frecuencia del fundamental para las distintas edades y sexos de los individuos. Uno de los promedios dados para los valores más frecuentes de la frecuencia fundamental es de 350 Hz para los niños, 250 Hz para las mujeres y 125 Hz para los hombres [Jackson-Menaldi, 1.992].

La naturaleza de cada acto de habla tendrá una dependencia crítica de la configuración fisiológica que posean las cavidades resonantes del tracto vocal de cada individuo, tanto desde un punto de vista anatómico como articulatorio. Por ello, el timbre o cualidad de voz *(ver I.3.3.6*) se constituirá como un componente fundamental de la voz, y aportará informaciones clave en el proceso de identificación/eliminación.

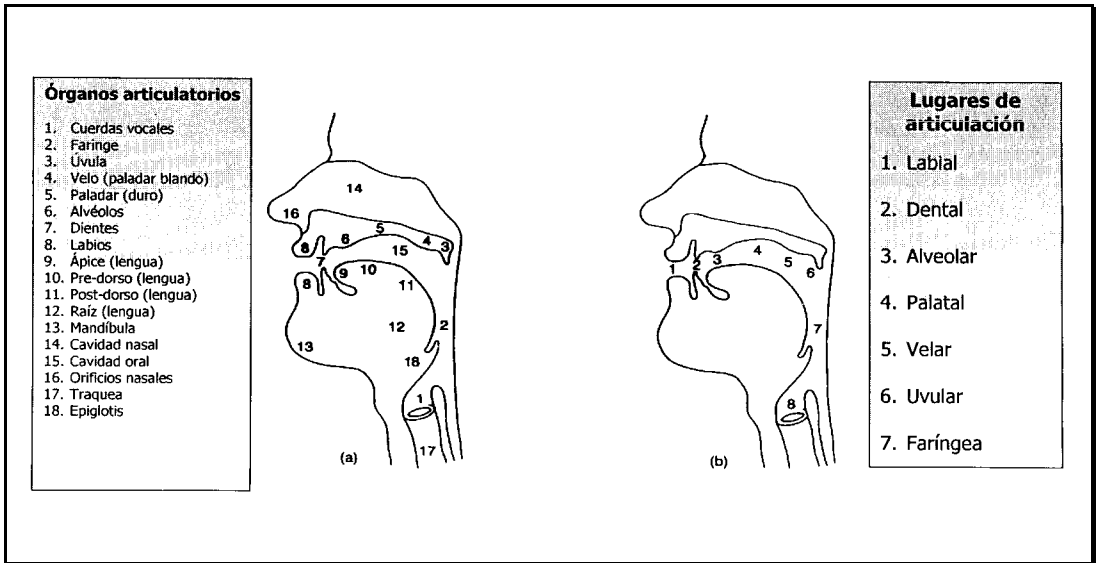
Pero volvamos a la laringe. Aunque de forma más indirecta, de ella también depende otra propiedad del sonido articulado: la intensidad de la voz. El aire contenido en las cavidades infraglóticas puede ser impulsado con mayor o menor energía hacia las cuerdas vocales y la presión del aire sobre ellas determina una mayor o menor amplitud vibratoria, que es la causa de la intensidad de sonido. Como vemos, de los cuatro elementos básicos del sonido del habla -tono, timbre, intensidad y duración- los tres primeros están muy relacionados con la laringe.

I.2.1.3.- Cavidades resonadoras

Una vez la corriente de aire abandona la laringe se adentra en la región laringofaríngea y, desde aquí, a la faringe oral, donde se produce otra división del material fónico. Si el velo del paladar está adherido a la pared faríngea, el aire fonador sale fundamentalmente a través de la cavidad bucal, dando origen a los sonidos articulados orales. Si el velo del paladar desciende, es decir, está separado de la pared faríngea, el aire fonador sale a través de la cavidad nasal -al

menos en parte- ya que también sale por la boca, produciendo los sonidos oronasales.

Cuando se trata del sonido oral, gracias a la movilidad de la lengua, al volumen y forma de la cavidad bucal, se produce la más variada gama de sonidos articulados.



La parte superior de la cavidad bucal está constituida por el paladar, dividido en dos zonas generales: la anterior, ósea, conocida con el nombre de paladar duro, y, la posterior, que recibe el nombre de paladar blando o velo del paladar. El paladar duro comienza inmediatamente por detrás de los alvéolos y se divide en prepaladar, mediopaladar y postpaladar; el paladar blando o velo del paladar está dividido en zona prevelar y postvelar.

La lengua es el órgano activo por excelencia y queda dividida, en su cara superior, en predorso, mediodorso y postdorso; el extremo anterior de la lengua recibe el nombre de ápice y tiene una importante función en la realización de diversas articulaciones.

En la parte anterior de la cavidad bucal están situados los incisivos superiores e inferiores; entre los incisivos superiores y el comienzo del paladar se sitúa la zona conocida por el nombre de alveolo (raíz de los incisivos superiores). Como órganos que cierran la cavidad bucal en su parte anterior están los labios superior e inferior, que por su movilidad cambian fácilmente la cavidad y, por consiguiente, modifican el timbre del sonido.

Los adjetivos correspondientes a las referidas zonas de articulación son : labial, dental, alveolar, palatal (prepalatal, mediopalatal y postpalatal) y velar (prevelar y postvelar). De la lengua: apical, dorsal (predorsal, mediodorsal y postdorsal). Podría añadirse, también, la zona radical (raíz de la lengua) y la uvular (úvula o campanilla). Ver ilustración n115.

Una vez descritos los principales órganos relacionados con la producción del habla, no debemos olvidar que no tienen en ésta su labor primordial (desde el punto de vista del soporte biológico) sino que dicha función, es más bien secundaria en favor de otras de carácter más vital: los pulmones sirven para la hematosis, la laringe como válvula que interviene en la deglución, respiración y oclusión de esfuerzo; el pabellón faringobucal para respirar y masticar, el velo para la deglución evitando la salida de líquidos y de alimentos por la nariz, etc.

I.2.2.- Teorías de la mecánica vocal

El habla o la mera emisión de un sonido vocálico, entraña una compleja orquestación de acciones mentales y físicas. La idea de producir un sonido se origina en la corteza cerebral -en el área del lenguaje. El movimiento de la laringe, controlado por este área, se transmite a través de diferentes nervios y, cumpliendo la orden cerebral, las cuerdas vocales vibran y generan un zumbido. La resonancia del sonido de vibración glotal por todo el área del tracto vocal epiglótico, donde se inscriben faringe, lengua, paladar, cavidad oral y la nariz, confiere al sonido las cualidades - timbre - percibidas por el oyente.

La relación entre el proceso de la fonación y el sistema nervioso central parece estar clara

en lo relativo a ciertos aspectos. Ya hemos comentado que el área del lenguaje en la corteza cerebral domina el acto consciente de la voz, estando regulada por el oído. La región bulbar parece estar relacionada con la regulación del estado tónico laríngeo. La modulación física de la voz, en cuanto a intensidad, tono y timbre, es responsabilidad del cerebelo; mientras que los nervios periféricos controlan la movilidad de la musculatura fonoarticulatoria.

La definición integral de este proceso mediante el cual distintos mecanismos interactúan y se coordinan en la laringe para la emisión de un sonido o vibración inicial, ha tratado de ser explicado por diversos autores a través de distintas teorías. En 1898, apoyando los estudios realizados por Johannes Müller, Ewald expuso la teoría *aerodinámica-mioelástica* caracterizada por dos conceptos:

11.- La vibración de las cuerdas o pliegues vocales es un acto pasivo.

21.- Las características del sonido generado a nivel glotal dependen exclusivamente de la presión infraglótica y de la tensión de los pliegues vocales.

Según esta teoría el proceso de la fonación sólo es posible cuando se establece un juego entre las fuerzas físicas de la aerodinámica - fundamentalmente a través del efecto Bernoulli - y la fuerza elástica del tejido de los músculos de la laringe.

La crítica a estas aseveraciones vino años más tarde de la mano de Husson quien mantenía la imposibilidad de ser explicado por medio de los postulados de Ewald el hecho de poder variar la intensidad de un sonido sin modificar al mismo tiempo su tono. En 1.950 Husson emite la teoría *neurocronáxica* afirmando que los pliegues vocales poseen una función activa inducidos por impulsos del nervio laríngeo recurrente; por tanto, la altura de los sonidos es independiente del mecanismo que regula la intensidad de los mismos (presión infraglótica).

Como reacción a la teoría neurocronáxica y apoyándose en experimentación fisiológica de la laringe, surgen a partir de 1.953 diversas teorías de oposición. En 1.960 Cornut y Lafón describen la teoría *impulsional*. Dos años más tarde, el español Perelló define la teoría *mucoondulatoria* y Vallancien la teoría *mioelástica perfeccionada*. En 1.968 Mac Leod y Sylvestre exponen la teoría *neurooscilatoria*.

En 1.974, Hirano distingue entre el cuerpo muscular del pliegue vocal y la mucosa que lo cubre. Estas consideraciones sustentan la teoría *osciloimpedancial* emitida por Dejonckère en 1.981.

Desde un punto de vista forense no tiene mayor relevancia el hecho de que las emisiones sonoras de la laringe sean efectuadas a través de unos mecanismos u otros, si bien en la actualidad - desde una perspectiva clínica - parecen ser más aceptadas las tesis con referencias mioelásticas y mucoondulatorias (los pliegues vocales parecen agitarse y ondular como una alfombra que se sacude con las manos).

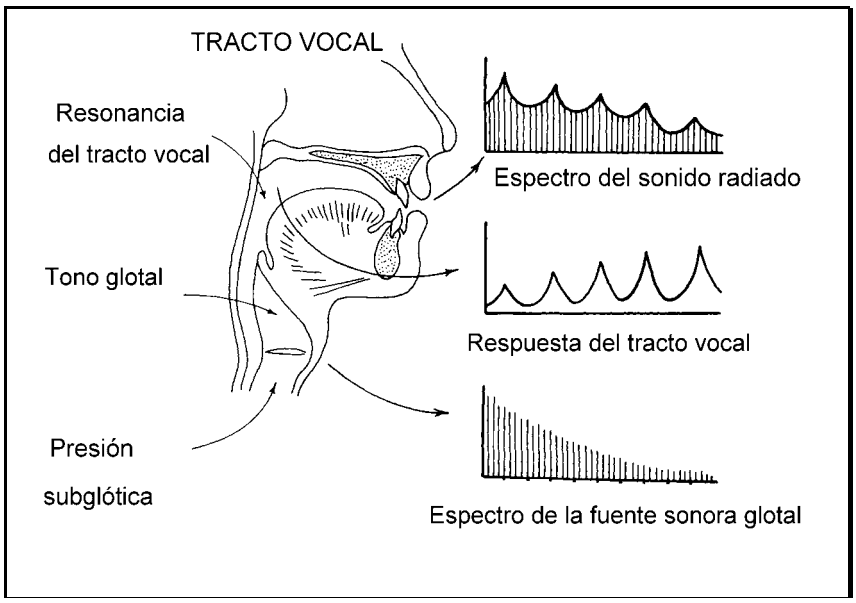
I.2.3.- Fuentes de energía acústica del habla. Resonancia.

La producción de los sonidos del habla se genera a través de tres fuentes: la fuente glotal, la fuente fricativa y la fuente explosiva.

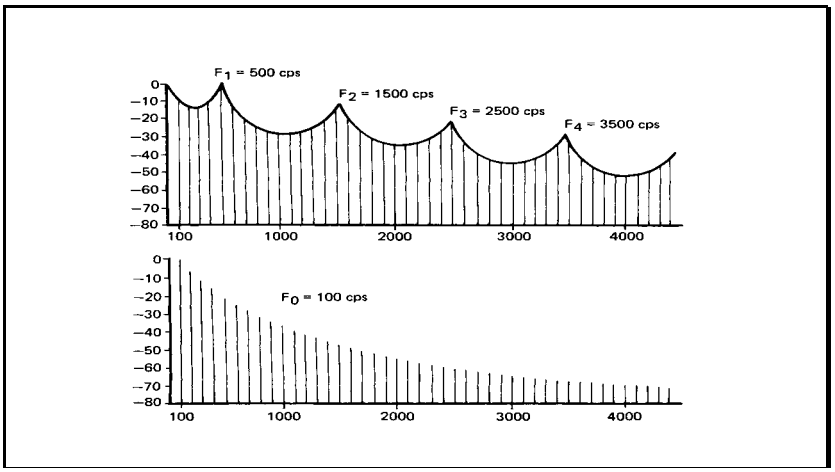
La *fuente glotal* radica en la modulación de una corriente de aire pulmonar al atravesar la glotis vibrante. Tal modulación genera una onda quasi periódica correspondiente a un sonido sonoro, armónico o periódico (sería el caso de un sonido vocálico). Dicha onda presenta como espectro discreto constituido por el tono fundamental de vibración de las cuerdas vocales y sus correspondientes armónicos. La intensidad de estos sucesivos armónicos decae a una razón de 12 dB por octava (Ilustración n1 17, debajo). El intervalo de una octava representa una relación del doble entre una frecuencia o tono y otra de referencia. La fuente glotal combinada con la fuente fricativa y la explosiva se utiliza para la producción de consonantes sonoras de dichos tipos.

La *fuente fricativa* está asociada a la turbulencia de la corriente de aire pulmonar que se genera por la constricción de la cavidad bucal por parte de los diferentes componentes del tracto vocal (lengua, labios, dientes...). Durante esta producción fricativa la glotis se mantiene abierta. Esta fuente fricativa de energía "sonora" es aperiódica o inarmónica ; el sonido generado es sordo. Estamos en el caso de los sonidos consonánticos fricativos : /s/, /□/, /x/...

La *fuente explosiva* es aquella relacionada con la expulsión instantánea a una presión superior que la atmosférica -explosión- de un caudal de aire retenido en la cavidad bucal -implosión- . Es el caso de los fonemas consonánticos sordos y sonoros oclusivos /p/, /k/, /b/, etc .



Los sonidos del habla tendrán su origen en alguna de las tres fuentes citadas, actuando por sí solas o en combinación. Con posterioridad, el sonido generado en dichas fuentes, atravesará el filtro variable del aparato fonador amplificando algunos de sus componentes armónicos o inarmónicos, para salir finalmente a través de la cavidad bucal y la nariz.



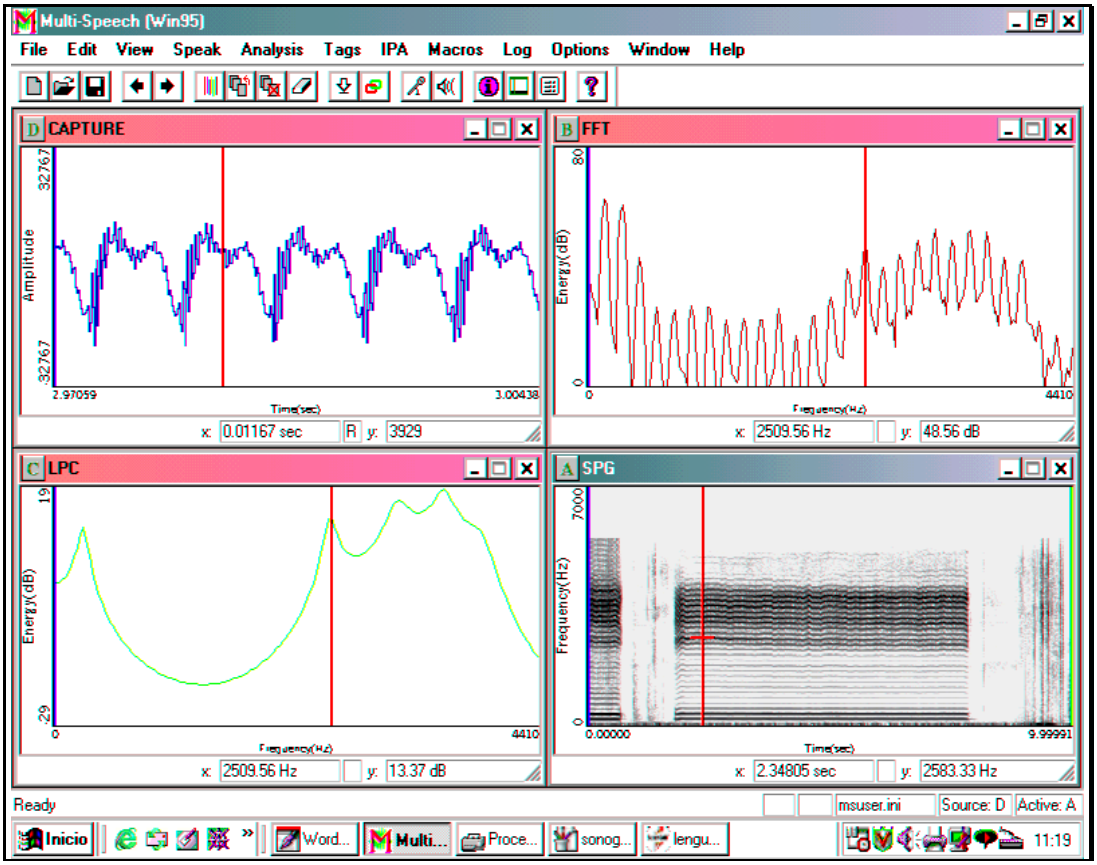
Por tanto, las distintas producciones acústicas del habla adquieren su forma definitiva con el fenómeno de la *resonancia*. Resonancia es la amplificación de un sonido de determinada frecuencia por la acción de un cuerpo pasivo denominado *resonador*, que es capaz de vibrar a dicha frecuencia . El resonador en función de sus características físicas -dimensión, forma, elasticidad, composición- poseerá una frecuencia natural de resonancia; será el máximo responsable en la formación del timbre característico de cada sonido. Sonidos de la misma intensidad y tono serán percibidos con timbres distintos en función de la caja de resonancia que los haya albergado. Un tono musical determinado presentará un timbre diferente al resonar en una cavidad metálica (por ejemplo una trompeta) o en una de madera (guitarra).

Si alguno de los componentes de frecuencia de un sonido coincide con la frecuencia natural de resonancia del resonador, será sensiblemente amplificado. Los componentes adyacentes a dicha frecuencia de resonancia también son amplificados.

La amplificación de una serie de frecuencias por parte de un resonador, nos introduce en un concepto que estará directamente relacionado con la representación sonográfica de las estructuras vocálicas; estamos hablando del *ancho de banda*, el cual, está definido por el conjunto de frecuencias contiguas a la frecuencia natural de resonancia, que son amplificadas dentro del intervalo de 3 dB, tomando como referencia la frecuencia de máxima amplificación del resonador o frecuencia natural de resonancia.

I.2.4.- Las estructuras formánticas y órganos resonadores.

Consecuencia directa del fenómeno de resonancia producido por las distintas conformaciones de la cavidad bucal (resonadores) , las representaciones espectrales de los sonidos vocálicos generados a nivel glotal, van a presentar diferentes grupos de armónicos con unos máximos relativos de intensidad. Estas agrupaciones de armónicos vocálicos con mayor amplitud se denominan *formantes*. En el espectro superior de la ilustración n1 17 observamos los picos correspondientes a los cuatro formantes de un sonido vocálico en su fase de radiación. En las ventanas B, C y D de la ilustración n1 18 podemos observar las estructuras formánticas del vocálico /i/ en su representación sonográfica de banda ancha, espectro FFT y contorno LPC.


Multi-Speech (Win95)
File Edit View Speak Analysis Tags IPA Macros Log Options Window Help
D CAPTURE
Amplitude
32767
-32767
2.97059
3.00438
Time(sec)
x: 0.01167 sec
R
y: 3929
B FFT
Energy(dB)
80
0
4410
Frequency(Hz)
x: 2509.56 Hz
y: 48.56 dB
C LPC
Energy(dB)
29
0
4410
Frequency(Hz)
x: 2509.56 Hz
y: 13.37 dB
A SPG
Frequency(Hz)
7000
0
0.00000
9.99991
Time(sec)
x: 2.34805 sec
y: 2583.33 Hz
Ready
msuser.ini
Source: D
Active: A
Inicio
Word...
Multi...
Proce...
sonog...
lengu...
11:19

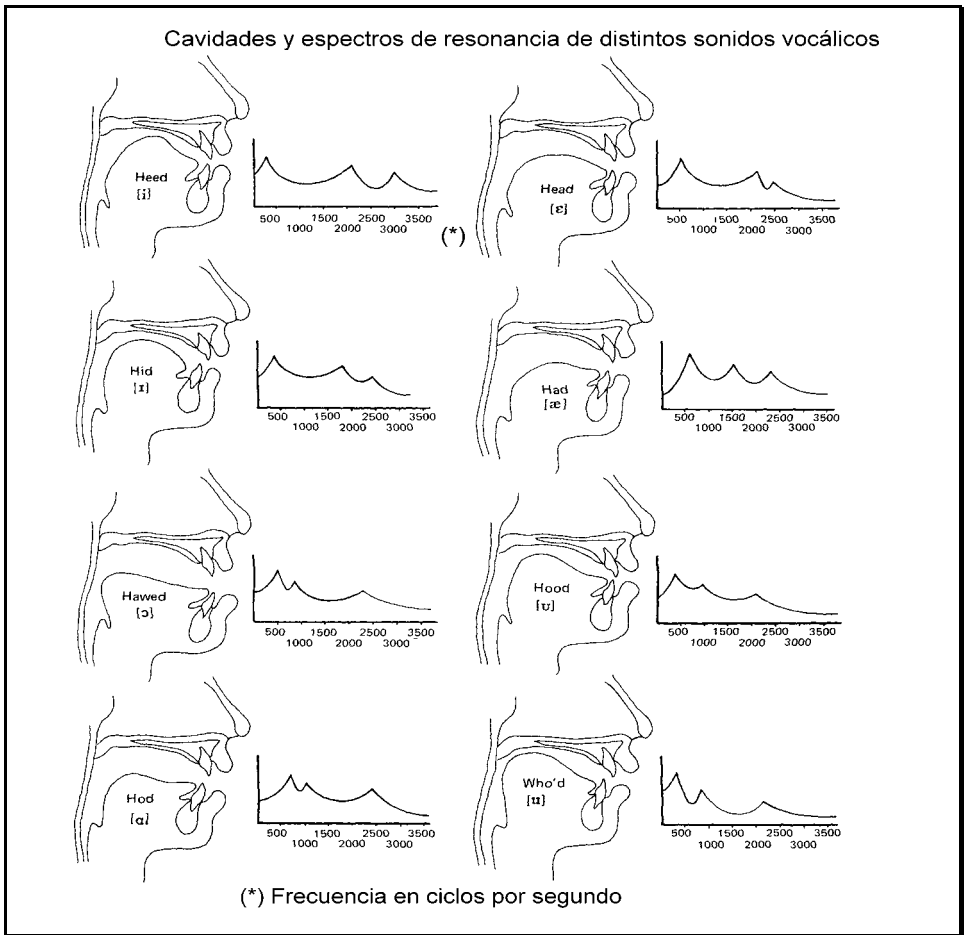
Los formantes serán numerados en razón a su posicionamiento en la escala de frecuencias, como F1, F2, F3, F4, etc.; El primer formante o F1 representará la agrupación de los armónicos amplificados de más baja frecuencia, el F2 se corresponderá con el siguiente grupo amplificado en un rango de frecuencia superior, y así sucesivamente. Según Sundberb [1.977] el F1 está relacionado con el grado de apertura/cierre del resonador, el F2 con la forma del cuerpo de la lengua y el F3 con la posición del ápice de la misma.

Dichos formantes, tendrán un ancho de banda y una frecuencia central de referencia, cuya ubicación en el rango de frecuencia dependerá exclusivamente de la acción del resonador. Es decir, la altura frecuencial de un formante está en relación directa a la configuración del tracto vocal en la articulación de un sonido vocálico. Teóricamente, un cambio en el tono fundamental - por la acción de las cuerdas vocales - si no es acompañado de una alteración del resonador, no debe implicar un cambio en la altura frecuencial media de un formante [Zemlin, 1981]. Decimos "teóricamente" porque para un emisor no entrenado resulta bastante complicado efectuar un cambio a nivel de frecuencia fundamental sin producir simultáneamente algún tipo de modificación en los órganos del resonador bucal.

Este último aspecto debe ser tenido muy en cuenta en un proceso de identificación de locutores. Como veremos más adelante el experto en identificación por la voz tendrá que efectuar comparaciones entre actos de habla emitidos en planos expresivos muy diferentes, en donde el papel de ciertos factores (tono fundamental, intensidad, etc.) puede influir de forma crítica en la modificación de las estructuras formánticas, produciendo en última instancia ciertos cambios de la cualidad tímbrica.

Del anterior razonamiento podemos deducir la existencia del correlato *RESONADOR-FORMANTES-TIMBRE* , fruto de la actividad del tracto vocal durante la fonación.

La ubicación y distribución en distintas áreas frecuenciales de las estructuras formánticas y energías de otros índices acústicos dependerán de la configuración fisiológica y articulación de los diferentes órganos de los resonadores vocales: faringe, mandíbulas, dientes, lengua, labios, fosas nasales, senos, etc. (Ilustración n1 19).


Cavidades y espectros de resonancia de distintos sonidos vocálicos
Heed [i]
Head [ɛ]
Hid [ɪ]
Had [æ]
Hawed [ɔ]
Hood [ʊ]
Hod [ɑ]
Who'd [u]
500 1000 1500 2000 2500 3000 3500
(*)
(*) Frecuencia en ciclos por segundo

Los *labios* poseen una movilidad enorme debido a los trece pares de músculos que intervienen en la sinergia, adecuación y antagonismo, siendo el más importante el *orbicular*, que va de una comisura hacia la otra. En reposo mantienen un contacto entre sí, que puede ser mucho mayor cuando el tono muscular se intensifica. Una vez separados, la cavidad bucal se pone en contacto con el exterior, produciéndose la radiación del habla.

Lógicamente juegan un papel trascendental en la realización de los llamados fonemas labializados, donde se produce un alargamiento de la cavidad anterior consecuencia de la protrusión de los mismos.

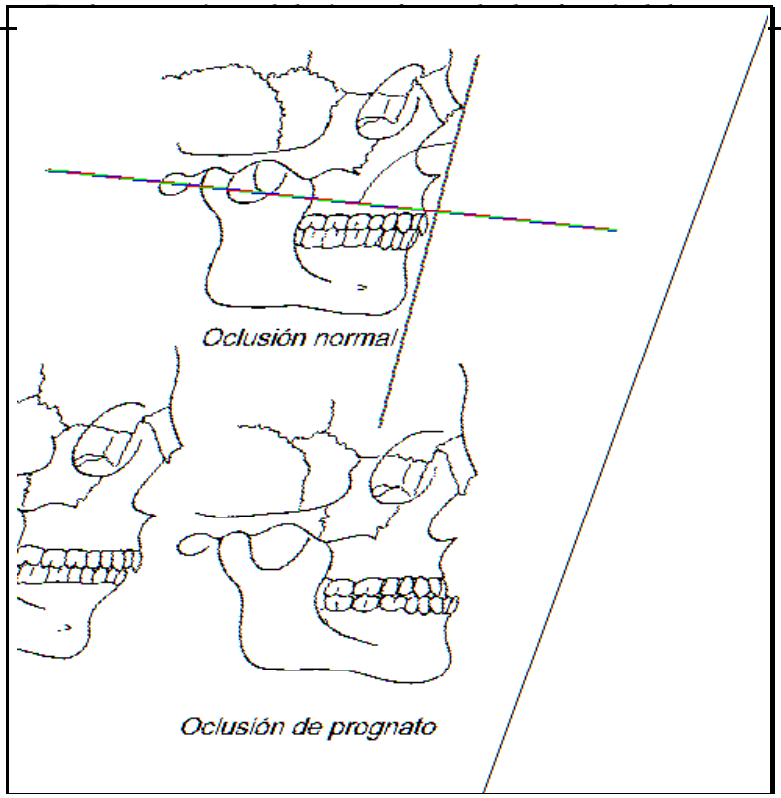
Las *mejillas* intervienen en la presión intrabucal, teniendo una relación muy directa en la pronunciación de los fonemas oclusivos bilabiales, sobre todo en posición inicial.

La *mandíbula* es un único hueso articulable a través del músculo temporal, y a nivel del cóndido de la misma. Su articulación permite realizar movimientos de descenso (no existe ningún músculo que intervenga en este tipo de movimiento) y elevación, propulsión, retropulsión y de diducción. Permite aumentar el volumen de la boca, mover las posiciones de la lengua y labio inferior.

A nivel del reborde alveolar se insertan los *dientes*, que son órganos fijos que también intervienen en la producción de sonidos. La dentición denominada caduca o temporal se completa alrededor de los dos años y medio, pasando por una etapa de transición dentaria desde los 6 a los 15 años aproximadamente, donde se completa la llamada dentición permanente.

En relación con la calavera, y para determinar el tipo de ángulo que forma, se utiliza el *ángulo facial de Camper* (modificado por Cuvier, Cloquet y otros) que está formado por dos líneas imaginarias : una que pasa por el punto *nasión* , en el arranque de la nariz, y por el punto *prostión,* en la raíz de los incisivos medios superiores*;* y otra, desde este mismo punto, al punto denominado *basión* en el borde anterior del agujero occipital. Aunque también se habla de hiperortognatos e hiperprognatos en los biotipos más extremos, la referencia de Camper determina la siguiente clasificación:

Ortognato - si el ángulo es superior a 851
Mesognato - si está entre 801 y 851



rognato - si es inferior a 801

Dicha clasificación suele corresponderse con los tipos más comunes de oclusiones mandibulares que reflejamos en el siguiente gráfico:

El dinamismo y los correspondientes elementos de transición que caracterizan la articulación de los sonidos del habla, son una consecuencia lógica de los cambios de posición de los diferentes órganos del tracto. De entre ellos, *la lengua* - que interviene en la realización de la

práctica totalidad de los fonemas- desempeña un papel protagonista debido a su versátil movilidad:

- Puede extenderse o estrecharse lateralmente, provocando el llamado bruxismo dental en el caso de exceso de presión,
- puede contactar los bordes con dientes o encías,
- incurvar el ápice hacia arriba o hacia abajo,
- elevar su dorso para contactar con el velo,
- mover su raíz,
- adelantar o arrastrar la epiglotis, etc.

Las *fosas nasales* se comunican con la cavidad oral mediante la nasofaringe , y en ella aparecen una serie de oquedades denominadas senos que se comunican a través de pequeños orificios llamados *ostium*, carentes de función en la fonación. Son dos cavidades que están situadas en el macizo facial superior, y constan de la nariz, que está formada por un tejido blando y cartilaginoso, y los conductos nasales, formados por dos conductos óseos. Se comunican al exterior por unos orificios denominados *narinas* y al interior por medio de las *coanas*.

Intervienen en la realización de los sonidos nasales y en menor medida de los orales, produciéndose unos trastornos de resonancia cuando se obstruyen total o parcialmente, o cuando existe una dilatación o estrechamiento acentuado en la fase productora del habla; dichos trastornos tendrán, en ocasiones, importantes consecuencias de cara al análisis forense de identificación por la voz.

En general, la nasalización implica una disminución del espectro sonoro en el rango comprendido aproximadamente entre los 1.150 Hz. y lo 1.800 Hz.. Asimismo, las fosas nasales absorben energía, de forma que un sonido cuya intensidad alcanzase los 125 dB a nivel laríngeo, puede perder en su fase de radiacción unos 10 dB, si el sonido se transmite por la cavidad oral y, entre 15 y 30 dB si la emisión es de carácter nasal [Prater y Swift ,1986].

Los *senos faciales* son cavidades neumáticas que prolongan las fosas nasales, con las que se comunican. Existen 4 pares:

- Maxilares
- Etmoidales
- Frontales
- Esfenoidales

La función de estas cavidades ha sido objeto de múltiples interpretaciones: sirven de aligeramiento de los huesos del cráneo, de aislantes térmicos, de protección en los traumatismos craneofaciales, etc.; no estando considerados como cavidades de resonancia, si bien, ciertos profesionales del canto dicen sentir este tipo de resonancia al efectuar una emisión cantada.

Actualmente, su función más probable se asocia al aislamiento de las vibraciones producidas a nivel óseo, para evitar así la transmisión de las mismas hacia la cóclea.

La *faringe* es un canal de músculo membranoso de unos 14 cm de largo y 4 cm de diámetro en su parte superior y 1,5 cm en su parte inferior; está situada en la parte anterior de las vértebras y se comunica con las fosas nasales, boca , laringe, base del cráneo y oído medio (este último a través de la trompa de Eustaquio). Es capaz de estrecharse lateralmente y de atrás hacia adelante diametralmente (debido a la acción flexora y de rotación del esternocleidomastoideo), variar su volumen verticalmente (debido a la elevación y descenso de la laringe) y de aumentar su volumen (al separar la mandíbula y aplicar tensión a los músculos supra e infrahioideos, como ocurre en el bostezo). Se divide tradicionalmente en:

- LARINGOFARINGE (Hipofaringe)
- OROFARINGE (Mesofaringe)
- NASOFARINGE (Epifaringe)

Interviene muy directamente en la mayoría de las resonancias de los sonidos lingüísticos y sobre todo en los denominados Avelares≅.

La bóveda del *paladar* se asienta en sus dos terceras partes sobre una placa de carácter óseo. La parte restante, es una membrana blanda que denominamos Avelo del paladar≅, que en realidad es un esfinter que se eleva en la realización de los sonidos orales, o desciende en el caso de los sonidos nasales. Por este motivo, el mecanismo velofaríngeo intervendrá muy activamente en la caracterización distintiva de diversos sonidos del habla.

Brevemente, y a título ilustrativo, citaremos los músculos que participan en la ejecución del mencionado mecanismo:

Periestafilino externo.- Es el tensor del velo del paladar y el dilatador de la trompa auditiva. Debido a su composición neuromuscular, participa en la propiocepción subconsciente de la tensión del velo.

Periestafilino interno.- Es el elevador del velo del paladar y el dilatador de la trompa auditiva. Es muy activo en la realización del habla.

Faringoestafilino.- Es el constrictor del istmo de las fauces, el que desciende el velo del paladar y el que eleva la faringe y la laringe.

Uvular Palatoestafilino.- Es el retractor de la úvula. La acción durante el habla se desconoce, aunque determinadas personas con voz denominada Agangosa≅ tienen la úvula demasiado grande.

Glosoestafilino (Palatogloso).- Es el constrictor del istmo de las fauces, desciende el velo del paladar y eleva la base de la lengua.

I.2.5.- Alteraciones del lenguaje

Desde un punto de vista identificativo, resulta muy interesante formar un criterio en relación a la nomenclatura utilizada para definir y clasificar con propiedad las distintas familias en que se pueden agrupar las alteraciones del lenguaje. A este respecto y, como es lógico, existen diversos manuales en los que se denominan y definen tales alteraciones. Nosotros hemos tomado como referencia básica una clasificación efectuada por la *American Speech Correction Association* o lo que es lo mismo, la Asociación Americana de Corrección del Habla :

1.- **DISARTRIA.-** Defectos de la articulación originados por lesiones de transmisión en el neuroeje. Las disartrias se clasifican en:

A) Anartria.- Falta total de articulación oral.

B) Bradiartria.- Articulación lenta y laboriosa, trastorno que suele presentarse en la personas que padecen parálisis.

C) Mogiartria.- Defecto en la articulación debida a la incapacidad de controlar los movimientos musculares en forma voluntaria. Este trastorno es muy frecuente en los paralíticos cerebrales.

Existen otros calificativos asociados a las disartrias:Flácida, Espástica, Atáxica, Hipocinética, Hipercinética, Mixta, etc.

2.- DISLALIA..- Defecto en la articulación de origen extra-neúrico. Puede ser debido a causas orgánicas, funcionales o psicosomáticas. Otros autores [Perelló, 1977] excluyen la lesión orgánica del aparato fonador. En este grupo se ubican todos los defectos articulatorios y fonéticos de tipo periférico. Se clasifican a su vez en:

A) Alalia, mutismo o ausencia de lenguaje, que comprende los siguientes trastornos:

a) Alalia cofótica o sordomudez.

b) Alalia orgánica, debida a daños anatómicos en el mecanismo periférico del lenguaje.

c) Alalia fisiológica, debida a defecto funcional.

d) Alalia prolongada, lenguaje retardado, que puede ser debido a mudez auditoria, mutismo auditivo y mutismo prolongado.

B) Barbarolalia. Articulación con acento extranjero o con cierto provincialismo.

C) Barilalia. Desorden sintáctico.

D) Idiolalia. Lenguaje inventado.

E) Paralalia. Sustitución fonética. (Ceceo, lambdacismo, etc.)

F) Pedolalia. Perseveración infantil del lenguaje.

G) Rinolalia. Defectos articulatorios con voz nasal que pueden tener un origen diverso:

a) Rinolalia megauvúlica, debido a una prolongación de la úvula.

b) Rinolalia microuránica, debido al paladar corto.

c) Rinolalia uranochismática, debido a las fisuras palatinas.

d) Rinolalia uranotraumática. debido a un trauma palatino.

e) Rinolalia abierta, debido a alteraciones patológicas de las aberturas nasales posteriores.

f) Rinolalia clausa o cerrada, consistente en la falta de resonancia nasal debido a la obstrucción parcial o total de las vías nasales.

Existen una serie de *dislalias disfuncionales* que resultan de alto interés en el análisis foarticulatorio identificativo: lambdacismo, yeismo, rotacismo, pararrotacismo, sigmatismo, jotacismo, betacismo, deltacismo, kappacismo, mitacismo, gammacismo, etc.)

3.- DISLOGIAS.- Defecto en la sintaxis y en la calidad de la expresión verbal debido a psicosis. Se clasifica en:

A) Agramalogía. Lenguaje incoherente.

B) Alogia. Ausencia de ideas.

C) Bradilogia. Lenguaje indolente, perezoso.

D)Catalogia o verbigeración. Perseveración de un sonido o palabras; estereotipia lingüística. A su vez comprende: habla ecoica y estereotipia.

E) Paralogia. Lenguaje desatinado. Razonamiento falso.

F) Polilogia. Locuacidad excesiva.

G) Taquilogia. Agitología o rapidez mórbida del lenguaje.

4.- DISFASIA.- Debilitación o pérdida de formación de las asociaciones verbales por disminución de la integración mental, debida a enfermedad, shock o trauma.

A) Afasia o pérdida del lenguaje oral o escrito. Se clasifica en:

a) Agrafia. Pérdida de la habilidad de escribir
b) Amusia. Pérdida de las asociaciones musicales. Cuando el paciente no puede identificar la música que escucha.
c) Amimia. Falta de habilidad en el lenguaje mímico.
d) Logofasia. Imposibilidad de expresar ideas por medio del lenguaje.
e) Alexia. Ceguera de palabras. El paciente ve lo que está escrito pero no puede leer porque no reconoce los símbolos gráficos del lenguaje.

B) Afasia Sensorial o pérdida de las asociaciones auditivo-verbales. Se clasifica en:

a) Afasia auditiva o sordera verbal. Incapacidad para entender el sentido de las palabras habladas.
b) Sordera psíquica. El paciente escucha la palabra y puede repetirla pero no la entiende.
c) Amusia Sensorial o Sordera a los tonos musicales.
*d) Afasia visual.*Alteraciones en el funcionamiento intelectual del lenguaje debido a falta de coordinación entre la imagen verbal y objetiva. A su vez comprende: la ceguera intelectual, la ceguera mental y la ceguera psíquica.
e) Agnosia. Pérdida de la capacidad para reconocer personas y cosas.
f) Alexia. Perdida de la capacidad para leer.

C) Afasia mixta:

a) Agramafasia. Ensalada de palabras, afasia sintáctica.
b) Hipofasia. Lentitud y monotonía del lenguaje.
c) Bradifasia. Habla titubeante.
d) Catafasia. Repetición constante de la misma expresión.
e) Parafasia. Sustitución de palabras.

D) Afasia total o afasia universal.

5.- DISFEMIAS.- Desorden del ritmo del lenguaje y tics debidos a psiconeurosis, (sinónimo de tartamudez). Se clasifican en:

A) Agilofemia. Habla agitada y nerviosa.

B) Afemia o mutismo, que puede ser:

a) Afemia o mutismo histéricos.

b) Afemia pathemática, debido a espanto o pasión.

c) Afemia plástica o mutismo voluntario.

d) Afemia espasmódica.

C*)* Parafemia. Balbuceo neurótico.

D*)* Espasmofenia.Tartamudez, tartajeo. Se divide en:

a) Afonía espasmódica.

b) Ritmo interrumpido.

c) Vacilación convulsiva.

d) Disfonía espástica

e) *Afonía espástica.*

f) Espasmofemia clónica.

*g) Espasmofemi*a *críptica. o silenciosa*

h) Espasmofemia tónica..

6.- DISFONIAS. Defectos de la voz debidos a perturbaciones orgánicas o funcionales de las cuerdas vocales o a respiración defectuosa.

A) Afonía. Ausencia de voz, que puede ser:

a) Afonía apofática, debida a negativismo de la conducta.

b) Afonía histérica.

c) Afonía orgánica, debida a anomalías en la estructura de la laringe.

d) Afonia paralítica (también una disartria).

*e) Afonia paranoica (*también una dislogia).

f) Afonia patemática , debido a espanto o pasión (también una dislogia o disfemia).

g) Afonia espástica, por espasmo de los músculos fonadores.

h) Afonía traumática, por trauma laríngeo

B) Baritofonía. Voz gruesa.

C) Guturofonía. Voz gutural.

D) Hipofonía. Voz susurrante.

E) Idiofonía. Características individuales de la voz: voz aguda, tosca, plana, lóbrega, grave, gruesa, dura, áspera o agria, infantil, estrepitosa, monótona, pasiva, rasposa, ronca o crascitante, quebrada, sepulcral, penetrante, sombría, estridente, sumisa o baja, atonal o quejumbrosa, etc.

F) Megafonía. Voz anormalmente alta.

G) Metafonía. Voz metálica.

H) Microfonía. Voz débil.

I) Parafonía. Alteraciones mórbidas de la voz:

a) Parafonía amazónica. Voz hombruna en las mujeres. Voz de virago.

b) Parafonía atímica. Cambios de voz en las depresiones.

c) Parafonía copiaca. Cambio de voz por la fatiga.

d) Parafonía adenopática. Cambios en la voz debido a alteraciones glandulares.

e) Parafonía eunucoide. Voz de falsete.

f) Parafonía gerática. Voz resquebrajada. Senil.

g) Parafonía microischica. Cambio de voz resultante de una disminución de vitalidad.

h) Parafonía neurasténica.

i) Parafonía puberal. Voz irregular en la pubertad. Voz tirolesa.

J) Neumofonía. Defectos de la voz debidos a falta de coordinación neumofónica, como ocurre con la voz aspirada.

K) Rinofonía o voz nasal.

a) Nasalidad.

b) Gangosidad. Voz Fañosa.

c) Rinismo.

*d) Rinolalia clausa.*Falta de nasalidad en los fonemas nasales debido a obstrucciones nasales.

L) Traquifonía. Ronquera o voz áspera.

M) Trombofonía. Voz tremolante.

7.- DISRITMIA. Defectos del ritmo en los que no se incluye tartamudez. Pueden deberse a defectos respiratorios o alteraciones endocrinas.

A) Disritmia neumafrasia. Debido a defectos respiratorios.

B) Disritmia prosódica. Defectos de la acentuación durante la lectura o en el habla espontánea.

C) Disritmia tonia. Defectos de la inflexión vocal.

I.2.6.- Patologías de los órganos o mecanismos de producción de voz.

Tanto las anteriormente relacionadas alteraciones del lenguaje, como las patologías relacionadas con los órganos o mecanismos de producción de la voz poseen un carácter intrínsicamente atípico o excepcional. Precisamente, ese carácter extraordinario confiere a estas emisiones de voz afectadas, un elevado grado de especificidad o individualización y, por consiguiente, un altísimo interés desde el punto de vista identificativo.

Generalmente, la detección de una alteración patológica del habla por parte del científico forense -al margen de los correspondientes estudios instrumentales clínicos, anamnesis, etc. que practican los profesionales de la medicina- pasará inexcusablemente por el estudio acústico y perceptivo de la propia cualidad de voz así como de otras circunstancias accesorias relacionadas con el comportamiento vocal del locutor: esfuerzo, respiración, ritmo, prosodia, melodía, etc.

Hemos de señalar la existencia de diversos criterios y nomenclaturas en las distintas clasificaciones nosológicas relacionadas con la producción vocal. Examinaremos algunos de ellos y sus referencias, para posteriormente describir aquellas disfunciones que por su carácter más habitual pudieran resultar más pertinentes.

Probablemente, el término *disfonía* es el más utilizado con carácter general para indicar la existencia de una disfunción vocal. Si partimos de la definición anteriormente aportada (defecto

de la voz debido a perturbaciones orgánicas o funcionales de las cuerdas vocales o a respiración defectuosa) la disfonía queda asociada a alteraciones en la masa, elasticidad o tensión de las verdaderas cuerdas vocales (músculo tiroaritenoideo interno); sin embargo, para otros autores, la relación se vincula a la alteración de uno o varios componentes fundamentales de la voz: timbre, intensidad y tono. Por ejemplo, Perelló [1.977] define la disfonía como la A*pérdida del timbre normal de voz*≅, con lo que la definición se ampliaría a la alteración de otros órganos de la fonación (cavidades infraglóticas y supraglóticas).

Como ya sabemos, la configuración anatómica de los pliegues glotales determinará los límites del rango tonal en un individuo. Aunque ya hemos apuntado unos valores estándar en relación con la frecuencia fundamental, el ser humano abarca una rango teórico de entre 30 y 700 Hz, si bien, los valores cercanos a los extremos han de considerarse absolutamente excepcionales en relación al habla conversacional, quedando éstos reservados para el ámbito de las emisiones cantadas. Sirva como ejemplo una típica escala de *tesituras* :

Soprano..... 349 - 698 Hz.
Mezzosoprano..... 330 - 659 Hz.
Contralto..... 294 - 587 Hz.
Contratenor..... 294 - 587 Hz.
Tenor..... 175 - 349 Hz.
Barítono..... 165 - 330 Hz.
Bajo..... 35 - 294 Hz.

En un capítulo posterior, *(véase II.1.5)* comentaremos más detalladamente aquellas alteraciones que con más asiduidad afectan al tono, tanto desde su propia naturaleza (cambios hormonales, factor edad, etc.) como por la influencia de agentes externos (medicamentos, tabaco, etc.). Por ejemplo, dejando al margen sus cambios drásticos consecuencia de las mudas propias de la adolescencia, puede afirmarse -*grosso modo*- que el tono se hace más agudo con la edad en el caso del sexo masculino, y se hace más grave en el caso del femenino. Ello es debido al lógico deterioro a nivel cerebral y fisio-fonatorio: cambios en la laringe por mala calcificación, por falta de flexibilidad, cambios en el aparato respiratorio y en el resonador, etc.; estas mutaciones suelen ir acompañadas de factores como la ronquera, temblores, fatiga, etc.

Al igual que a veces ocurre con el timbre, existe también cierta relación entre la intensidad y el tono producido, hasta tal punto que en muchas ocasiones no es posible elevar la frecuencia tonal sin un consiguiente aumento de la intensidad. Suele manejarse la siguiente

relación [Prater y Swift ,1986] como representativa :

Sonido a 80 dB .....aumenta entre 13 a 17 Hz.
Sonido a 90 dB .....aumenta entre 30 a 40 Hz.
Sonido a 100 dB. .....aumenta en torno a los 75 Hz.

El timbre o cualidad de voz es una energía acústica básicamente ligada al número, conformación e intensidad de los armónicos generados por la resonancia del fundamental. El intento de establecer una clasificación de los posibles timbres vocales es una misión de gran complejidad. Los profesionales de la foniatría suelen utilizar distintas escalas subjetivas para definir un timbre determinado. Entre las más utilizadas cabe mencionar la denominada AGRBAS≅, propuesta por la sociedad japonesa de logopedas y foniatras. Dichas siglas, establecen cuatro campos dimensionales que se corresponden con las referencias siguientes:

G = Grado de disfonía
R = Áspera o Rasposa (Inestabilidad en la F.Fundamental)
B = Soplada (escape de aire)
A = Asténica o Débil
S = Constreñida (F.Fundamental alta y ruidos en agudos)

Una de las nomenclaturas más comunes en el ámbito patológico para calificar el timbre incluye estas nueve opciones: ÁSPERA, RONCA, DICRÓTICA, FALSETE, VELADA, CUBIERTA, DIPLOFÓNICA, GUTURAL y BLANCA.

Una prueba evidente de las dificultades relacionadas con el intento de clasificar las diferentes patologías de voz, es la existencia de diferentes teorías sobre las causas matrices en las que originalmente radican. Citaremos las teorías más manejadas, aunque desde un enfoque forense esta cuestión no presenta mayor relevancia:

Los *Organicistas* observan como causa principal la desproporción existente entre los distintos órganos bucofonatorios.

La teoría de la *alteración endocrina* argumenta que existen determinadas enfermedades que implícitamente llevan asociado un trastorno específico de la voz como son eunuquismo, el mixedema,... y que por lo tanto debe generalizarse a las distintas patologías.

Según la teoría de la *alteración neurológica* los trastornos vocales son consecuencia de la alteración de los circuitos neurológicos que intervienen en el habla. Aunque este axioma es

cierto en algunos casos, no se puede generalizar.

Desde otros puntos de vista, la causa del trastorno vocal se asocia directamente a una lesión en el órgano vocal, como por ejemplo la Alaringitis≅ que conlleva alteración del timbre; sin embargo, esta teoría no puede argumentar otros capítulos como la fonastenia.

La teoría de las *etiologías psicológicas* encuentra como causa primigenia de las disfunciones de voz los factores de tipo psicológico ...

Las disfonías más habituales son las relacionadas con el mal uso o abuso vocal. La causa más frecuente de un mal uso vocal es la hiperaducción de la musculatura laríngea (intrínseca y extrínseca), que generalmente conlleva una vibración de la cuerda vocal de forma violenta, provocando alteración en su masa, elasticidad y tensión, siendo la resultante de todo ello una voz con ronquera, soplo y tono grave.

Una deficiente higiene vocal puede implicar una vibración traumática de los pliegues vocálicos, con una incorrecta producción del tono y de la intensidad (ya hemos apuntado que generalmente la elevación del tono conlleva la de la intensidad).

Las vocalizaciones o *habla forzada* producidas generalmente en las Aonomatopeyas no vocales≅(ruidos de avión, de coches, ametralladoras,...) o cuando realizamos un esfuerzo físico (transportando objetos pesados) conlleva una mayor intensidad en los tonos agudos y por lo tanto unas cuerdas hiperaduccidas con la consiguiente irritación que, a su vez, desencadenará una miopatía por sobresfuerzo (agujetas) e inflamación congestiva de las vías aéreas superiores por la gran presión subglótica ejercida. Este fenómeno es muy doloroso y cuando ocurre de forma sistemática provoca sensaciones de cuerpos extraños, opresión respiratoria, etc.

Consecuencias similares provocan los *chillidos o gritos*, que son producidos por una hiperaducción y vibración violenta del músculo cordal que generan distintos grados de irritación, llegando a producir patologías por traumatismos (hematoma, ingurgitación vascular) en los casos de práctica continuada.

Todo individuo tiene un límite fisiológico en el uso incesante y exagerado de la producción vocal *(habla excesiva*), sin embargo, dicho límite es distinto en cada persona. Aquellos que utilizan de forma profesional su voz : profesores, cantantes, locutores, políticos, etc, están más predispuestos a patologías laríngeas.

Si a la circunstancia del habla excesiva, se une un uso frecuente de ataque glótico duro

(aducción de pliegues vocales + elevación de la presión) provocaríamos una voz áspera y ronca, llegando a los *nódulos o pólipos* en los casos de práctica sistemática.

Determinadas costumbres o comportamientos no locutivos como la Atos, aclarar la garganta, consumo de alcohol, irritaciones aéreas como el humo de los cigarrillos, inhalación de gases o de polvo, causan traumatismos en los pliegues vocálicos debido a la necesidad acuciante de tener siempre la laringe lubricada, dado que tanto el alcohol como el humo actúan como secantes laríngeos, provocando una ronquera.

Asimismo, la elevación del Atono habitual≅ con ataque glótico duro e irritación laríngea debida a Aalergias≅ o hábitos poco saludables para la voz como hablar abusivamente en Acondiciones no óptimas≅ (bares nocturnos en los que es preciso elevar el tono por encima de ruidos o de la música, hablar mientras se viaja en coche Adisfonía del conductor≅) y durante períodos excesivos, produce aspereza o ronquera que puede llegar a tener un carácter quebradizo.

No debemos confundir el tono habitual -el utilizado por una persona en la conversación diaria- y el óptimo, en el que la voz se realiza con la menor tensión laríngea y la mayor comodidad de esfuerzo físico.

No es necesario insistir sobre la conveniencia de que el científico forense se encuentre familiarizado con estas sintomatologías de carácter general para tenerlas en consideración durante el desarrollo de determinadas fases de su estudio. Como veremos más adelante (*ver IV.1.3*), en la mayoría de los casos será necesario realizar una toma de voz indubitada para su comparación, y siempre que las circunstancias lo permitan, habrá de procederse a la anamnesis del candidato, como el médico lo hace con su paciente. Lógicamente, alcanzar un diagnóstico preciso en este ámbito corresponderá siempre a otro tipo de expertos (foniatras, otorrinolaringólogos, logopedas, etc.) en los que habremos de asesorarnos en cualquier caso.

De forma complementaria, y teniendo en consideración que siempre debemos tener la puerta abierta a la opinión de autoridad de los verdaderos expertos en cada uno de los entornos de especialización que integran nuestra técnica forense, definiremos brevemente las principales circunstancias que caracterizan a algunas de las más típicas disfonías :

Laringitis: inflamación duradera de la mucosa de los repliegues vocales causada, entre otros muchos factores, por fumar cigarrillos, contaminación industrial, alcohol, etc. Produce ronquera, tono vocal grave, fatiga, tos no productiva y cosquilleo. Dependiendo del tipo de laringitis que nos ocupe, nos enfrentaremos a grado mayor o menor de ronquera, llegando incluso a ser severa y persistente.

En ciertas ocasiones esta patología aparece acompañada de una rinitis y de un olor del aliento muy ofensivo debido a la sequedad interior de la laringe.

Nódulo vocal : engrosamiento de la mucosa de los pliegues, localizado en la unión del tercio anterior con el tercio medio, en el denominado Apunto nodular≅. Puede ser uni o bilateral y su incidencia es mayor en mujeres; siendo la edad de aparición la comprendida entre los 20 y 30 años. Su formación parece deberse a una hipotonía acompañada de una presión de aire excesiva. En general, se produce una bajada del nivel del tono habitual del individuo, hacia el grave, siendo la calidad de la voz áspera y con soplo.

Generalmente son personas con temperamento nervioso o tendente a la ansiedad, muy habladoras, socialmente agresivas y tensas, acostumbradas a utilizar la voz muy fuerte; con anterioridad utilizaban un tono agudo y un ataque glótico duro.

El *pólipo vocal* es una tumefacción benigna del epitelio que aparece en el borde libre de los pliegues vocales, a la misma altura que los nódulos.

Aparece aproximadamente entre los 30 y 50 años, y al igual que en el nódulo, suelen ser propio de personas nerviosas o tendentes a la ansiedad. Suele ser unilateral, siendo la Adiplofonía≅ (registros del fundamental en dos niveles) una de las características que acompañan a este tipo de fonación.

Dependiendo del tipo de pólipo, puede originarse una rotura súbita de la periodicidad de la vibración (pólipo Asesil≅), o producirse una cualidad de voz ronca y soplada en el caso del Apedúnculo≅.

El *Aedema de Reinke*≅ es una transformación edematosa del corion (membrana exterior) de la mucosa del repliegue vocal que deforma la cara superior y el borde libre del mismo. Es una patología directamente relacionada con el abuso de alcohol, que se observa con más frecuencia en los hombres (aproximadamente sobre los 50 años). El timbre afectado por esta disfunción se caracteriza por una pérdida en los agudos, agravación del tono, pérdida de intensidad, y fatiga cuando se habla mucho tiempo. Estamos ante un timbre Aronco, cascado, fatigado≅, acompañado de problemas respiratorios en algunas ocasiones.

El *quiste mucoso* es una tumefacción originada por la obstrucción del conducto excretor de una glándula que provoca una secreción mucoide. Se caracteriza por una alteración del timbre con voz diplofónica, disminución de la intensidad, fatiga y Abaches≅ en la voz.

Similares características presenta la *Ahemorragia submucosa*≅ que es una dilatación vascular con salida y acumulación de sangre por debajo de la mucosa de los repliegues, que suele incluir rotura del tiroaritenoideo, normalmente, a causa de traumatismos agudos o violentos esfuerzos vocales. Además, suele ir acompañada de dolor cervical y afonía completa.

La *úlcera de aritenoides o de contacto* se caracteriza por quedar al descubierto los aritenoides como consecuencia de una pérdida de masa. Se asocia a personalidades introvertidas y coléricas. También, a individuos con problemas digestivos (la úlcera aparece en el lado correspondiente a la mano que no es dominante). Frecuentemente utilizan el susurro para sustituir la fonación normal que les genera una mayor irritación.

Acústicamente se manifiesta con tono muy grave, ataque glótico duro, timbre algo ronco, grave y con soplo.

La *disfonía espasmódica*, tanto en su versión Aaductora≅ como en la Aabductora≅ se consideran trastornos muy graves. En la primera, se produce una hiperaducción tan tensa que la vibración de los pliegues vocálicos no se puede producir. Estamos ante una voz distorsionada, muy forzada, acompañada de dolor en el pecho, muecas, parpadeos sincinéticos, etc.

En el caso de la abductora -bastante menos frecuente- se aprecian momentos de afonía y soplo intermitentes que dificultan la distinción entre fonemas sordos y sonoros. Concretamente, el problema emerge al realizar la transición de una consonante sorda a una vocal acentuada, debido a que no se produce una apertura correcta sino espasmódica de la glotis.

Cuando las denominadas bandas ventriculares se aproximan y se utilizan para la fonación, quedando los verdaderos pliegues vocálicos en fase de abducción, nos encontramos ante una *disfonía ventricular*. En el caso de que bandas ventriculares y repliegues vocales actúen de forma simultánea se produciría la voz diplofónica. Una disfonía de este tipo se corresponde con una emisión vocal de carácter grave, extensión vocal reducida, poca intensidad y ronca. Todo ello es debido a la gran masa de las bandas ventriculares que impiden o dificultan la vibración

Se denomina *puberfonía* a la incapacidad (característica de la pubertad) de alcanzar un tono grave desprendiéndose de otro más agudo, debido al desarrollo fisiológico de las estructuras laríngeas. También se le suele llamar voz de falsete de muda o eunucoide.

Por una parte puede haber una elevación de la laringe con la consiguiente agudeza del tono, y por otra un descenso o inclinación hacia abajo que provoca laxitud en los pliegues vocales; al intentar la fonación, los aritenoides se aduccen muy tensamente consiguiendo solamente la vibración del borde glótico. Es normal que aparezcan los llamados *Agallos*≅, que se producen al pasar de un registro grave a otro agudo sin interrumpir la emisión vocal.

La puberfonía adquiere su verdadera naturaleza disfuncional cuando se desarrolla fuera del marco de la adolescencia. No obstante, se desconocen las causas reales que la originan existiendo distintas opiniones:

- hay quienes las relacionan a factores de carácter psicosocial: sentimientos exagerados de cariño, rechazo de responsabilidades de adultos, etc.

- otros, a factores etiológicos: trastorno endocrino, pérdida grave de oído, enfermedades neurológicas o que producen debilidad, etc.

Las denominadas *disfonías psicogénicas* (depresivas, espásticas, suspirosas, afonía o rinolalias histéricas, fonofobia, etc.) pueden afectar a la pérdida total o parcial de la capacidad fonatoria por psiconeurosis conocidas como la reacción de conversión, mutismo, etc. En estos casos, el conflicto psicológico se resuelve utilizando como autodefensa el habla para provocar distintos grados de ronquera. Generalmente afecta a las mujeres y suele actuar sobre los tres componentes fundamentales: tono, timbre e intensidad.

El *temblor de voz* se manifiesta en periodos temporales de entre 5 y 10 segundos, prolongándose en presencia de estrés. Se trata de un temblor benigno, hereditario, de origen desconocido y que suele darse en individuos normales de edades comprendidas entre los 35 y 50 años. Acústicamente presenta alteración rítmica del tono e intensidad de la voz -en el dominio del tiempo- generadas por cambios de altura de la laringe en relación al cuello.

Teniendo en consideración la importancia del timbre en la gran mayoría de los procedimientos forenses de identificación -dado que aglutina la totalidad de los componentes de la estructura acústica de voz en su fase de radiación- resulta muy conveniente no abandonar este capítulo sin analizar de forma breve aquellos trastornos de la resonancia vocal más característicos.

Aunque no los comentemos en detalle, tampoco deben olvidarse los traumatismos laríngeos que evidentemente tienen una repercusión directa en la estructura y función del resonador.

Desde un punto de vista fisiológico, la resonancia vocal está afectada principalmente por la medida, forma y consistencia de los órganos bucofonatorios, especialmente la *faringe*; en la que cualquier modificación de su forma, tamaño y consistencia desembocará en una alteración de la resonancia.

El funcionamiento inadecuado del mecanismo velofaringeo puede causar una *hipernasalidad o rinolalia abierta* : resonancia excesiva durante la producción de las vocales y consonantes orales. El grado de nasalidad desde el punto de vista perceptivo, varía de una persona a otra, de unas regiones a otras y de unas modas lingüísticas a otras, pudiendo considerarse como alto para una lengua concreta cierto grado de nasalidad que sería estimado normal en otra lengua o, incluso en la misma, ante situaciones diferentes. Por ejemplo, el grado de nasalidad que en general presenta la lengua inglesa hablada en los Estados Unidos es muy superior al inglés del Reino Unido. Ciertos modelos locutivos radiofónicos o televisivos españoles se manifiestan con un claro grado de nasalización superior la estándar coloquial

(Matías Prats Jr., Javier Sardá, etc.).

Fisiológicamente, la hipernasalidad involuntaria es el resultado de la incapacidad del esfínter velofaringeo para lograr y mantener un cierre adecuado en la articulación de los sonidos que normalmente son orales; puede ser debido a defectos estructurales como: longitud corta del velo, lesiones, etc., o a defectos neurológicos como: parálisis palatofaringea, disartria espástica, etc.,

Existen también casos de hipernasalidad transitoria: mecanismos imitatorios, evitación del dolor (como por ejemplo en fases postoperatorias de amigdalitis), etc.

La hipernasalidad no sólo produce un aumento de la resonancia en las vocales y en las consonantes orales, sino que se acompaña de muecas en la cara y problemas laríngeos en cuanto a la intensidad. Este tipo de locutores producen errores de articulación -provocando una ininteligibilidad de mayor o menor grado- debidos a la imposibilidad de obtener una presión intraoral suficiente, generando elementos compensatorios como las omisiones y sustituciones al no poder dar la tensión adecuada en la realización de determinados fonemas que lo precisan (se aprecia principalmente en las fricativas, africadas y oclusivas).

Como fenómeno contrapuesto cabe citar la *hiponasalidad o rinolalia cerrada*, o lo que es lo mismo, la pérdida de nasalidad en las consonantes nasales y las vocales (falta de Abrillo≅, pero en menor grado).

En otras ocasiones hablamos de *nasalidad asimilativa* para referirnos a la resonancia nasal excesiva de una vocal en presencia de consonante nasal. Está asociada al fenómeno de *coarticulación,* y es provocada por la lentitud articulatoria del velo.

Las consonantes nasales tienen una influencia fuerte en la articulación de otros sonidos adyacentes, por ello, al producir una palabra en presencia de nasales y vocales, el velo debe comenzar a bajar cuando se articula la vocal anticipándose a la nasal siguiente, causando cierto grado de nasalización debido a la lentitud dinámica del velo y a que la lengua se mueve a la zona de los alveolos para realizar el fonema consonántico nasal; posteriormente, el velo debe volver a descender anticipándose a la vocal siguiente y causándole una ligera resonancia.

I.3.- PERCEPCIÓN AUDITIVA DE LAS EMISIONES HABLADAS

I.3.0.- Introducción.

Oír. Escuchar. Palabras sencillas, que albergan tras de sí uno de los más complejos y maravillosos procesos de codificación del cerebro humano: la percepción auditiva del habla. A lo largo de este capítulo intentaremos acercarnos a los diferentes mecanismos de dicho proceso, para conocer la verdadera entidad de la que será una herramienta de inestimable utilidad para el experto forense. En una etapa posterior, comprobaremos cómo un oído entrenado, realizando un trabajo sistemático y debidamente complementado, puede aportar informaciones de alto interés en la actual filosofía metodológica de identificación forense.

Pensemos por un momento en la dificultad que puede entrañar la percepción del más simple mensaje-estímulo hablado, no sólo en orden a las posibles variaciones de sus componentes físicos dimensionales - tono, timbre, intensidad y duración - sino también en función de los distintos planos expresivos, entonaciones, ritmos velocidades, etc. Eso, si solamente apreciamos el fenómeno perceptivo desde el punto de vista del sujeto emisor; porque si a ello, unimos los distintos factores que afectan a los medios y espacios de transmisión de la emisión sonora, al receptor, etc., el planteamiento se complica bastante más.

Por este motivo, y metiéndonos ya en materia, resulta imprescindible conocer todos aquellos factores que regulan o interfieren la actividad perceptiva.

La percepción, es obviamente algo más que la información que del mundo exterior nos proporcionan nuestros sentidos. Al entrar en juego la comprensión y el significado, hemos de considerar la percepción como un proceso bipolar, es decir, con dos aspectos fundamentales. Uno relativo a las características de los estímulos que inciden sobre los sentidos, y otro que comprende las características del perceptor - fisiológicas, psicológicas, experiencia, actitud, cultura ...

No debemos olvidar, especialmente en el ámbito del habla, otros factores tales como el marco de referencia - cada situación implica distintos valores y motivaciones en el receptor - el medio de transmisión, la calidad del registro en su caso (componentes de transducción, reproducción y recepción, acústica de los locales en que han sido efectuadas o reproducidas, los

condicionamientos ambientales, etc). Por tanto, en la regulación y tratamiento del proceso en el que se integra la percepción, intervendrán muy distintas disciplinas: Física, Biología, Sociología (sistemas de valores y constricciones sociales) y por supuesto la Psicología.

Como posteriormente podremos comprobar en un modelo cibernético de comunicación por la voz, la percepción auditiva se canaliza a través de los niveles acústico (oído externo y medio) y fisiológico (oído interno) del receptor, para materializarse en última instancia en el cerebro, en el denominado nivel psicolingüístico.

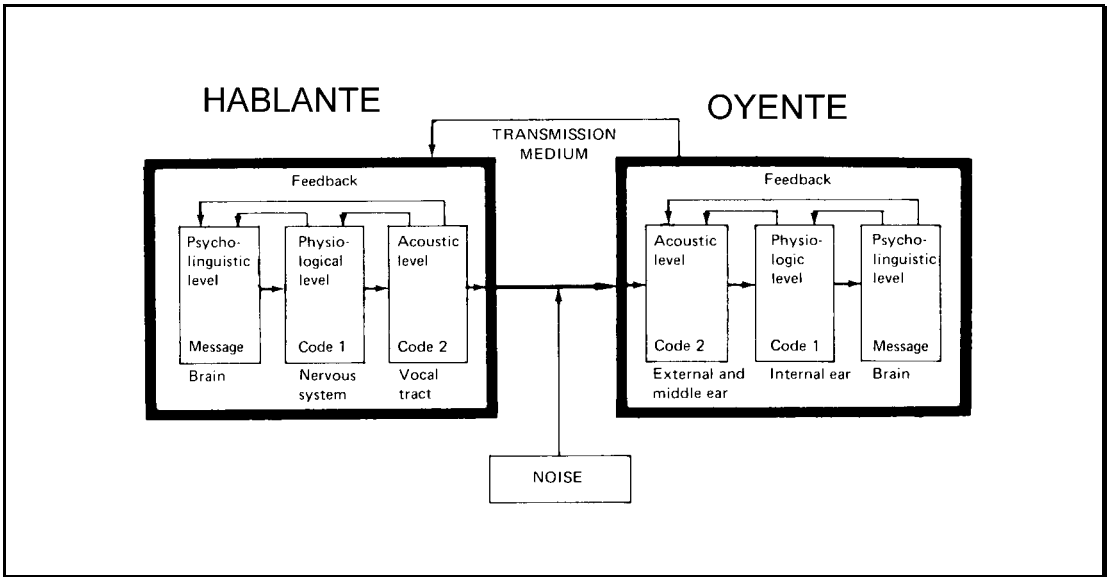
Hechas todas estas puntualizaciones, formulemos un planteamiento inicial:) Tenemos ante unos mismos estímulos, la misma percepción auditiva? Evidentemente, no. No necesariamente, a un mismo estímulo le corresponde el mismo mensaje de interpretación o de decodificación.)Cuáles son las causas?)Qué elementos intervienen en esa distinta percepción?

Las respuestas a estas preguntas las encontraremos contrastando los resultados fruto de la apreciación subjetiva a nivel perceptivo con la realidad del estímulo físico que la produce.

Por consiguiente, como punto de partida es insoslayable efectuar una serie de evaluaciones objetivas a través de mediciones físicas y estimaciones estadísticas. En el caso que nos ocupa, atenderemos a aquellos parámetros mensurables provenientes de actos de habla, para obtener una referencia empírica del objeto de estudio *(dimensión física y fonoarticulatoria del sonido*). Pero no debemos olvidar que las estimaciones realizadas a través del análisis físico, aunque de gran objetividad, llevan implícitos notables inconvenientes: variabilidad intrapersonal de los distintos actos de habla, componentes no deseados de la señal, etc.. Dado que no existen dos actos de habla idénticos (a excepción de los registrados) y la deficiente calidad de los registros forenses, la obtención de valores de físicos o matemáticos de alta precisión y muestra no garantiza en todos los casos unas referencias realmente representativas y, por tanto, válidas para su evaluación con fines identificativos.

I.3.1.- Proceso general de la comunicación hablada.

La percepción de mensajes hablados no tendría un sentido completo, en cuanto a los factores que en ella pueden incidir, si no estuviese referenciada en un marco o modelo de comunicación. Por ello, es conveniente definir un modelo cualquiera, que ponga de manifiesto el proceso de comunicación humano a través del habla.



Normalmente, tal como representa el anterior gráfico, en una situación de comunicación por la voz, están presentes tres elementos: el hablante, el medio de transmisión y el oyente.

El cuadro que representa al hablante se divide en tres compartimentos o niveles: uno psicolingüístico donde se genera el mensaje verbal o idea; un nivel fisiológico o neuroanatómico donde se producen series de impulsos de carácter neuromotor (eléctrico), codificados y ordenados en el mensaje psicolingüístico a través del Área de Broca de la corteza cerebral y otros centros neuronales, con la finalidad de activar las estructuras musculares del tracto vocal; y un nivel acústico (aparato fonador) donde se realizan, a través de la modulación de la corriente de aire que sale de los pulmones, las ondas del habla correlacionadas con las series de impulsos

motores, ya en forma de mensaje. Esta modulación se obtiene modificando a través de los músculos, las cavidades resonadoras del tracto vocal, por orden de los impulsos neuromotores.

Los mecanismos de recepción y traslación del mensaje a nivel auditivo presentan una clara e inversa correspondencia con el proceso de emisión de mensajes hablados anteriormente comentado. Dado que el proceso de percepción auditiva constituye el núcleo fundamental del presente estudio, más adelante conoceremos con detalle como se materializa dicha percepción a nivel fisiológico y neurológico.

Todas las traslaciones que se producen de un nivel al siguiente ocurren en un brevísimo lapso temporal. A su vez, se generan relaciones de retroalimentación (feedback) entre los distintos niveles, como se puede apreciar en el esquema gráfico del modelo.

La señal de habla (una onda acústica longitudinal compleja) se propaga a través de un medio elástico a una velocidad constante (la velocidad de propagación del sonido en ese medio) y en todas direcciones. La propagación de la señal de habla puede incluir distintos medios de transmisión, tales como el aire, líneas telefónicas, micrófonos, grabadores, altavoces, etc ; dichos medios, pueden incidir en mayor o menor medida en el deterioro de su cualidad original, produciendo ruido añadido, distorsión, solapamientos, etc.

El producto final del tracto vocal, una señal de habla codificada según las estructuras fonéticas y lingüísticas de cada lengua o dialecto, alcanza -en forma de variaciones de presión- las estructuras auditivas del oyente, donde el proceso descrito para el hablante se invierte para así poder decodificar el mensaje original

I.3.2.- Niveles del proceso de la comunicación y funciones del lenguaje.

En todo proceso de comunicación podemos distinguir los siguientes niveles:

- un nivel *social* en el que dos individuos se comunican (emisor y receptor).
- un nivel *extra lingüístico* al que corresponde el *referente*, es decir sobre lo que se habla.
- un nivel de *transmisión*: medium, o lo que es lo mismo, el vehículo o medio utilizado para el mensaje, ya sea el aire, el hilo telefónico, ondas hertzianas, etc.

En el caso de la comunicación hablada hay que añadir un cuarto nivel :

- un nivel *lingüístico*: el *signo lingüístico*, que tiene como función comunicar un mensaje mediante la utilización de un código.

El *signo* en general puede considerarse como aquella realidad que significa algo distinto de lo que es por sí misma, por ejemplo "una calavera"= peligro de muerte; "una bandera"= un país. La lengua es un sistema de signos; el signo lingüístico es la unión de un concepto y una imagen acústica, es decir, una entidad psíquica de dos caras como decía Saussure [1.955]: *significado y significante*; ejemplo, "mesa", el conjunto de los fonemas que la constituyen forman el significante, mientras que el concepto que tenemos de esa palabra es el significado. Algunos lingüistas han añadido un tercer factor: *la realidad extra lingüística*, pues una cosa es el concepto abstracto que tenemos en la mente, de las cosas, y otra la realidad concreta a que nos referimos en cada cosa; en el ejemplo anterior de "mesa" el concepto abstracto es "tablero con patas", mientras que la realidad concreta puede ser muy diversa en cada caso: mesa de despacho, de cocina, de operaciones, etc.

Los factores anteriormente reseñados determinan las distintas funciones del lenguaje y, que según Jakobson [1963], serían las siguientes:

10.- *Función referencial*, llamada también denotativa y cognitiva, que define las relaciones entre el mensaje y el objeto al que se refiere.

20.- *Función expresiva o emotiva*, que define las relaciones entre el mensaje y el emisor, es decir, la actitud del hablante frente a lo que expresa (alegría, tristeza, emoción, etc.). En este sentido hay que distinguir entre *denotación* y *connotación*: la primera está constituida por el significado concebido objetivamente, mientras que la connotación expresa valores añadidos al signo; ejemplo: la palabra "verano" denotativamente expresa una estación del tiempo, mientras que connotativamente para un agricultor puede significar época de recolección, y para un estudiante, vacaciones.

30.- *Función conativa o injuntiva*, define las relaciones entre el mensaje y el receptor considerando a éste como fin del mensaje, es decir, trata de influir en la voluntad del receptor; ejemplo: los mensajes publicitarios.

40.- *Función fática*, que tiene por objeto afirmar, mantener o detener la comunicación, es decir, comprobar que el canal de comunicación permanece abierto. Ejemplo el "sí" al descolgar el teléfono, "muletillas" o recursos retóricos utilizados en la conversación: "bien"; "vale"; "verdad"; "me entiendes?". Se trata de elementos vacíos de contenido semántico.

50.- *Función metalingüística*, que se da cuando hablamos del propio lenguaje; ejemplo: "hacia", se escribe con "h"; "El" es un artículo determinado, etc.

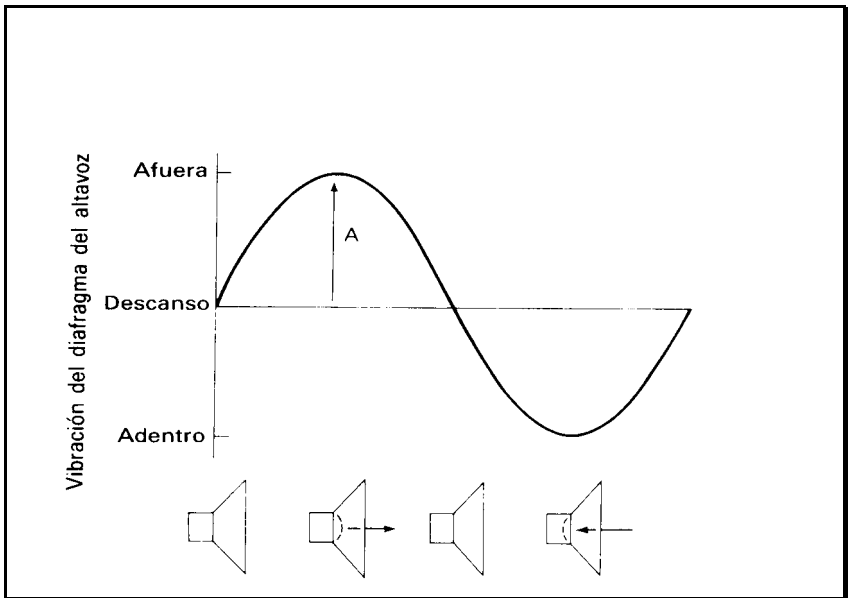
60.- *Función poética o estética*, es la que expresa la relación entre el mensaje y él mismo, es decir, lo que llamamos arte del lenguaje; ejemplo, la metáfora y demás figuras retóricas.

I.3.3.- El proceso de percepción auditiva

I.3.3.1.- El sonido como estímulo

El sonido, a efectos perceptivos, puede definirse como la sensación recogida a través del oído, producida por estímulos de cambio de presión de las moléculas del aire. Dichos estímulos se producen de acuerdo a la repetición de un ciclo de compresión/descompresión, cientos o miles de veces por segundo, creándose un patrón de presiones altas y bajas en las moléculas del aire que se desplazan - en el caso de que el medio de transmisión sea el aire - expandiéndose de forma muy similar a como se separan las pequeñas olas (ondas u oscilaciones) de donde se arrojó una piedra en un estanque tranquilo. Este patrón de cambios en la presión del aire recibe el nombre de onda sonora y este planteamiento pondrá de manifiesto que la naturaleza de los sonidos, especialmente sus tonalidades y sonoridades, está en función de las propiedades de sus ondas sonoras.

Partiremos de un tipo convencional y simple de sonido, al que se suele denominar tono puro. Los tonos puros no son frecuentes en nuestro entorno cotidiano, pero se emplean a menudo en los laboratorios de acústica como referencias patrón para el estudio de los mecanismos básicos de la escucha. Una manera de producir un tono puro es conectar un altavoz a un dispositivo electrónico denominado oscilador. El oscilador hace que la membrana del altavoz oscile hacia fuera y hacia dentro siguiendo un movimiento de onda sinusoidal (Ilustración n1 22). El oscilador nos permite realizar ajustes para lograr que la membrana vibre con la amplitud que deseemos - recordemos que *amplitud* es la distancia máxima de desplazamiento respecto a su posición de descanso, referenciada con AA≅ en la ilustración n122. También es posible ajustar el oscilador para que la vibración se ajuste a una *frecuencia* determinada - el número de veces por segundo que el diafragma se mueve siguiendo el ciclo ya descrito - ; es decir, hacia afuera, hacia dentro y de nuevo hacia afuera.



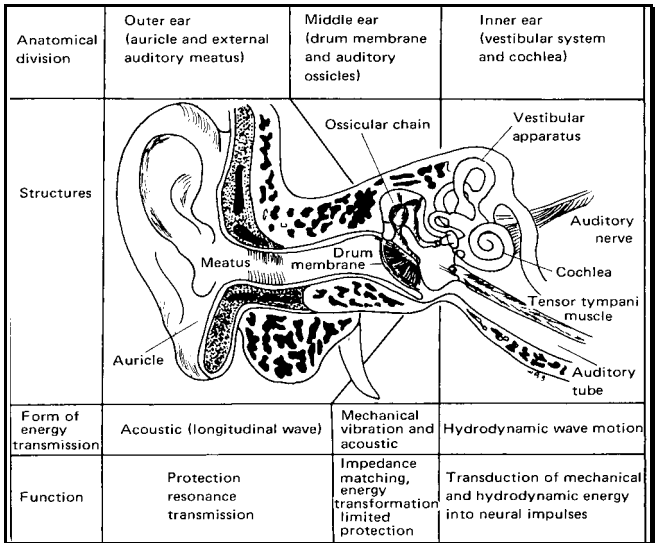
La vibración del diafragma se transmite al aire de sus proximidades y produce cambios en la presión de éste produciéndose un incremento respecto a la presión atmosférica y posteriormente una reducción respecto a la misma, pudiendo ser representado este incremento y decremento con la forma de una onda sinusoidal. El cambio sinusoidal perfecto en la presión del aire representaría un tono puro. Cada uno de estos estímulos sonoros suele describirse indicando su frecuencia de recurrencia en unidades Herzt (Hz), ciclos por segundo ó s^{-1}, en las que un Hertz representa un ciclo completo por cada segundo. Por tanto, un tono de 4.000 Hz es un tono puro cuyo ciclo se repite 4.000 veces en un segundo.

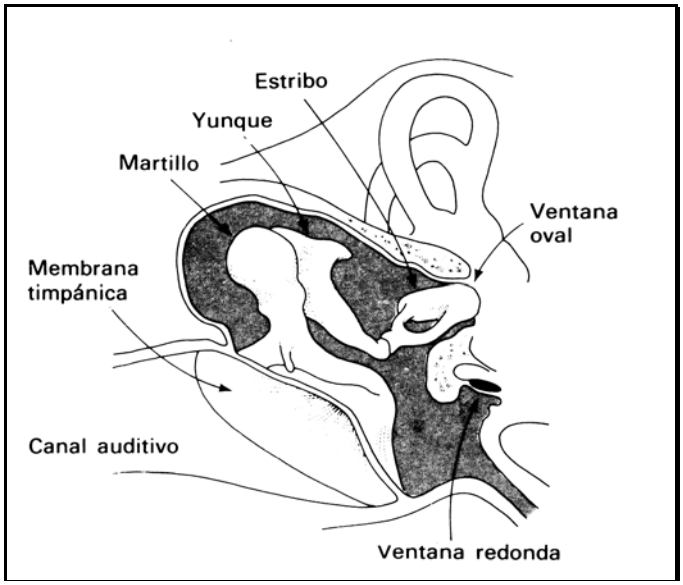
I.3.3.2.- Transducción del estímulo sonoro en el oído.

Ya conocemos que hablar de sonido es hablar de cambios en la presión del aire. Pero,)cómo estos cambios de presión llegan a transformarse en impulsos eléctricos para su trasmisión a través del sistema nervioso?. La estructura encargada de transformar las vibraciones del aire (energía mecánica) en energía eléctrica es el oído. En su parte más interna existe una subestructura con forma de caracol llamada cóclea, que está llena de líquido (endolinfa, perilinfa) y contiene unos receptores denominados células ciliares. Estas células,

Aunque los procesos de transducción más complejos se producen a nivel del oído interno, es necesario describir el camino que sigue la onda sonora desde su receptor de entrada: el pabellón auditivo u oreja. En la primera fase de su viaje hacia el interior del oído, las ondas sonoras deben pasar a través del *oído externo*, el cual está formado por la oreja y el canal auditivo. El conducto auditivo externo (o canal auditivo) es una estructura de resonancia en forma de tubo, de unos 3 centímetros de longitud, cuya función principal es proteger las delicadas estructuras del oído medio de potenciales peligros del mundo exterior. Los 3 centímetros de distanciamiento proporcionados por el canal auditivo, junto a los pelillos y la cera que contiene -que parece tener un efecto desagradable para los insectos- protegen la delicada membrana timpánica o tímpano, situada al final de este canal, al tiempo que permiten que ésta y las estructuras del oído medio se mantengan a una temperatura estable.

Además de su función protectora, el tubo del canal auditivo actúa como resonador para amplificar las intensidades de aquellos sonidos que se aproximan a su frecuencia de resonancia. Como ya hemos explicado, la frecuencia de resonancia del canal auditivo está determinada por su longitud y conformación, situándose alrededor de los 3.400 Hz. Las mediciones de las presiones sonoras en el interior del oído han mostrado que el canal tiene un efecto de amplificación moderada para las frecuencias comprendidas entre los 2.000 y 5.000 Hz. (casualmente las frecuencias medias donde se manifiestan las principales estructuras acústicas del habla).



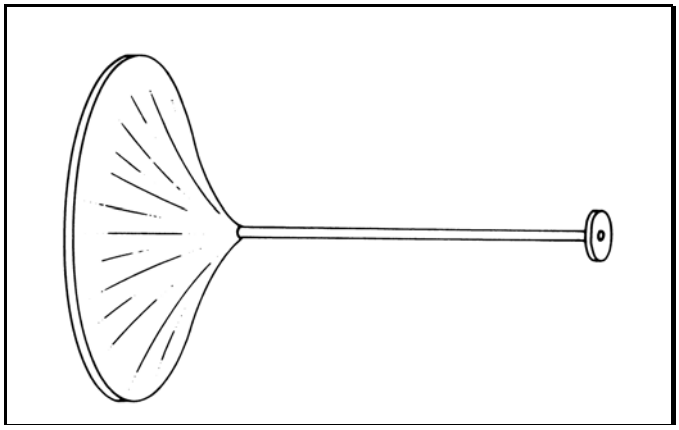


La membrana timpánica es considerada la frontera entre el oído externo y el oído medio. La ventana oval -donde se inserta el estribo en la cóclea- delimita el oído medio del oído interno. El *oído medio*, es una pequeña cavidad de unos dos centímetros cúbicos de volumen, en la que se hallan interconectados tres pequeños huesecillos: martillo, yunque y estribo. Al igual que el estribo conecta con la cóclea, el martillo lo hace con la membrana timpánica, produciéndose un

desajuste de impedancias -resistencia a la transmisión de la vibración- entre aquella existente a nivel timpánico (transmisión aérea) y la producida por los fluidos cocleares (transmisión hidrodinámica)

En el caso del oído, si las vibraciones de las moléculas del aire tuviesen que pasar directamente al líquido coclear, sólo serían transmitidas en un porcentaje aproximado del tres por ciento [Goldstein, 1.984]. Afortunadamente, dichas vibraciones pasan primero por el oído medio que tiene como principal misión resolver el problema creado por el desajuste de impedancias entre el aire y el líquido.

Cuando el tímpano, que es una membrana de forma cónica, comienza a vibrar, transmite la vibración al martillo, el primero de los huesecillos del oído medio. A su vez, el martillo transmite las vibraciones al yunque, y éste, por su parte, hace que vibre el último de los huesecillos, el estribo. Estos tres huesos, los más pequeños del cuerpo, resuelven el problema del desajuste de impedancias actuando en una doble vertiente.



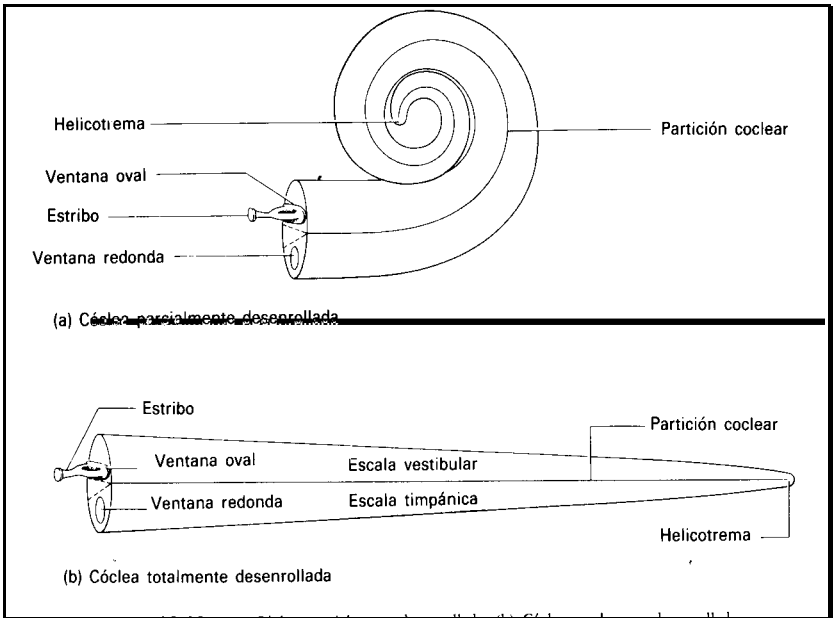
La primera solución al problema del desajuste de impedancias se pone de manifiesto al comparar las áreas de la membrana timpánica y la de la porción final del estribo (ver ilustración n1 25). La membrana timpánica tiene un área de unos 6 centímetros cuadrados, mientras que el final del estribo sólo la tiene de unos 0,032 -una razón de 17 a 1-. Este ratio proporciona un incremento considerable de la presión acústica por unidad de área que compensa en cierta medida la impedancia inherente a la transmisión por la cadena de huesecillos.

La segunda compensación al desajuste de impedancias está relacionada con la posición estructural de los huesecillos, los cuales se disponen de acuerdo al principio de la palanca con lo que logran multiplicar la intensidad de vibración por un factor aproximado de 1,3.

Según Goldstein [1.984] el incremento global de la intensidad de estímulo, con la combinación de las dos soluciones de compensación al desajuste de impedancias, alcanza un factor 22 (1,3 x 17), aunque otros autores sitúen el valor de dicho incremento en un factor multiplicativo tan alto como 100 [Schubert, 1980]. En cualquier caso, y al margen del valor del factor, queda patente la importancia de la cadena de huesecillos en la tarea de transmisión del estímulo auditivo sin menoscabo de la presión acústica del mismo.

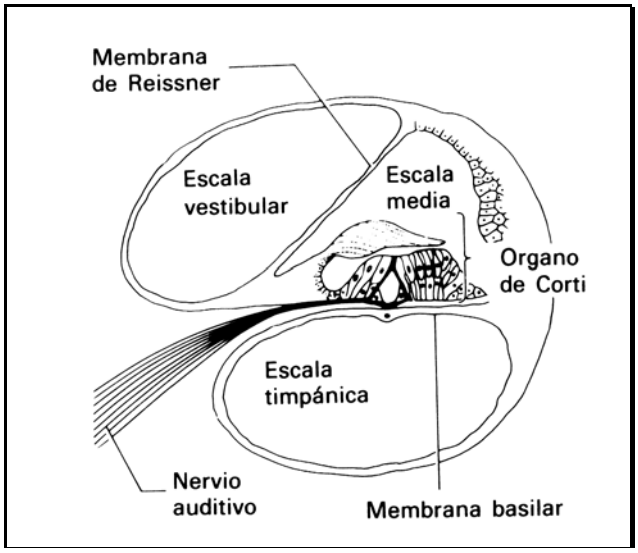
Las vibraciones de naturaleza aérea que llegaban al oído externo transformaron en vibraciones de naturaleza mecánica en la cadena de huesecillos; una vez abandonan el estribo a través de la ventana oval se produce una nueva transducción (mecánica-hidrodinámica) al entrar en vibración los líquidos de la cóclea. Nos encontramos en el *oído interno.*

La cóclea, que es una estructura ósea con forma de caracol, es difícil de visualizar porque está enrollada sobre sí misma en diversos giros. Su estructura interior es más sencilla de describir si la proyectamos desenrollada **:**

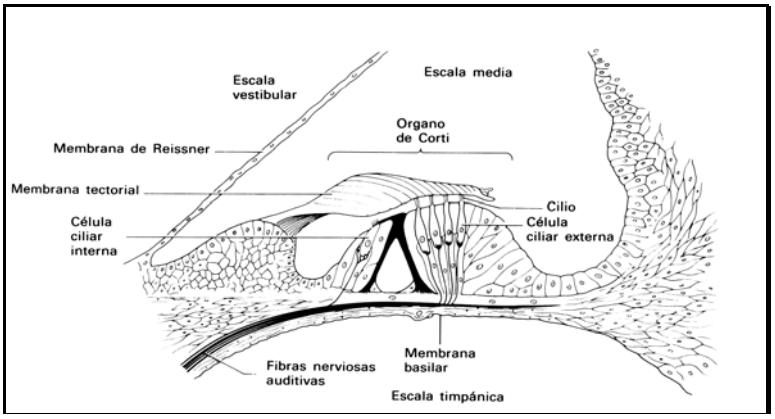


La denominada *partición coclear,* divide la estructura interna subcoclear en dos cámaras a lo largo de toda su longitud (salvo en la zona del *helicotrema o* extremo final); en dicha zona, se comunican la cavidad superior -*escala vestibular*- y la inferior -*escala timpánica*-. La cóclea desenrollada da lugar a una estructura cilíndrica - ahusada por la parte del helicotrema - de unos 2 mm de diámetro y 35 mm de longitud.

La estructura de la partición coclear se visualiza mejor a través de su sección transversal donde, además de las dos cavidades mencionadas, encontramos otro compartimento central -*escala media*- que se separa de ambas mediante dos membranas: la *membrana de Reissner* y la *membrana basilar* (Ilustración n1 27).



La escala media contiene el órgano de Corti, estructura de fundamental importancia en la generación de impulsos eléctricos hacia el cerebro:



Dentro de la escala media hemos de resaltar tres subestructuras fundamentales que ya inciden en un nivel cualitativo superior a lo que hasta este momento ha sido la mera transmisión del estímulo. Hablamos de *la membrana basilar, la membrana tectorial y las*

células ciliares. A partir de este punto, el mecanismo de trasmisión tamiza las vibraciones acústicas clasificándolas en relación a su frecuencia o recurrencia. Por vez primera, se generan datos (todavía de naturaleza hidro-dinámica) con un criterio más discriminativo, siendo las células ciliares las encargadas de su recepción y transmisión hacia el nervio auditivo en forma de impulsos eléctricos ; son de dos tipos : internas y externas. Como otras células nerviosas, las ciliares constan de un cuerpo celular y una fibra nerviosa, si bien, tienen un rasgo diferenciador: unos pequeños cilios o pelillos que sobresalen del cuerpo celular (Ilustración n129). Estos cilios, al moverse en la convulsión de los líquidos cocleares y las membranas basilar y tectorial provocan una nueva transducción generando señales eléctricas que se transmiten por la fibra nerviosa de la célula ciliar en dirección al cerebro.

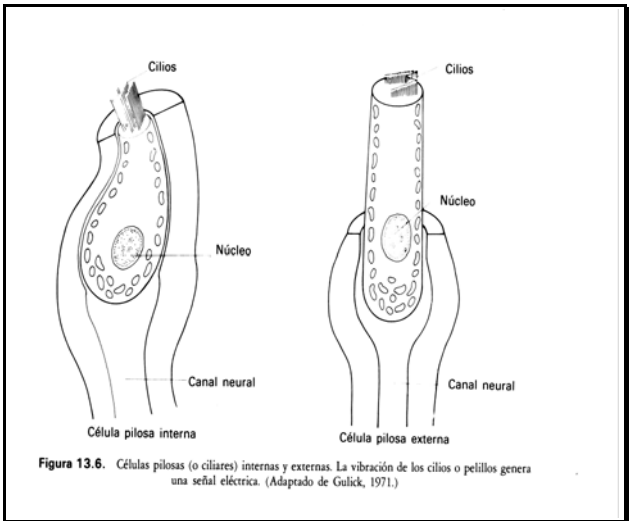


Figura 13.6. Células pilosas (o ciliares) internas y externas. La vibración de los cilios o pelillos genera una señal eléctrica. (Adaptado de Gulick, 1971.)

Una vez conocida la estructura interna de la cóclea, regresemos al mecanismo de propagación que habíamos abandonado en la ventana oval. Cuando el estribo transmite la vibración hacia el interior de la escala vestibular, parecería lógico pensar que dicha presión se trasladase hasta la escala timpánica a través del helicotrema. Pero en realidad esto no es así. El líquido de la escala vestibular desplaza su fuerza sobre la partición coclear. Este empuje, que mueve hacia abajo la elástica partición coclear, va seguido de una respuesta de la partición en la dirección contraria. Dicha respuesta, reduce la presión en la escala vestibular y es provocada por la acción del estribo al tirar de la ventana oval hacia atrás. Al final, se produce una sincronización perfecta entre la secuencia de empujes arriba/abajo de la partición coclear y la frecuencia de vibración en el área del estribo.

Analizados los procesos de transmisión del estímulo sonoro en las distintas áreas del oído, es momento de plantearnos una serie de cuestiones:)qué ocurre a partir de ahora con el estímulo?)existe un correlato del estímulo a nivel cerebral exactamente proporcional a sus originales cualidades físicas?)cómo se perciben a nivel neuronal los distintos componentes y estructuras del sonido del habla?

I.3.3.3.- Relación entre las cualidades físicas y psicológicas de los sonidos. Psicoacústica.

En la percepción de un estímulo acústico, existen dos aspectos de naturaleza diferente: uno es el *físico*, que se puede medir objetivamente en todos sus componentes; otro es el *psicológico*, que es el grado de sensación que ese estímulo produce en nosotros. Las relaciones de orden cuantitativo entre la sensación y el estímulo físico que la produce son el objeto de estudio de la *psicoacústica.*

)Significa esto que el oído es algo más que un simple receptor?. Sí. Puede actuar para filtrar los eventos neurales iniciales. La audición es un proceso de doble vía, donde además de transmitirse el mensaje al cerebro, la estructura neurológica induce al oído a la admisión, rechazo o modificación de ciertos aspectos de los estímulos auditivos.

Pues bien, la psicoacústica no sólo toma en consideración las posibles consecuencias de este proceso, sino también aquellas otras que puedan derivarse del propio acto perceptivo,

permitiéndonos mensurar las relaciones cuantitativas entre la sensación perceptiva (ψ) y el estímulo físico que la produce (ν).

Ya en el siglo XIX se establecieron las primeras ecuaciones para representar dichas relaciones. La más antigua conocida fue propuesta por Weber [1.834], quien suponía existía una relación de *linealidad* entre la sensación y el estímulo:

$$\Delta\nu/\nu = c$$

donde $\Delta\nu$ es la menor diferencia perceptible de estímulo, ν es su valor y "c" una constante que depende del tipo de sensación que se trate. Esta ley presenta notables errores, especialmente cuando se aplica a valores extremos de ν .

Fechner [1.860] trató de evitar tales errores proponiendo su ley *logarítmica* :

$$\psi = k \log \nu$$

donde "k" es una constante cuyo valor depende del tipo de sensación y de las unidades de medición de ν .

Considerando nueva información experimental obtenida, Stevens [1.936] demostró la falsedad del modelo anterior y propuso su ley de tipo *potencial*:

$$\psi = k \nu^n$$

donde"k" es una constante que depende de las unidades de medida adoptadas y "n" es un exponente que depende del tipo de sensación. Por ejemplo, para volumen sonoro n = 0,67.

Un estímulo acústico cualquiera comprende cuatro elementos físicos constitutivos que se complementan en un patrón complejo de dimensiones psicológicas:

Dimensión física		**Dimensión psicológica**
Tiempo	.....	Duración de la persistencia del sonido.
Intensidad	.....	Sonía, sonoridad o intensidad subjetiva. (Loudness)
Frec. Glotal	.....	Tonía, tonalidad o percepción de la altura tonal de un sonido. (Pitch)

Espectro de resonancia ..... Timbre o cualidad del sonido.

En realidad, los parámetros acústicos básicos son los tres primeros, ya que el timbre es el resultado del paso de la energía acústica producida por la fuente sonora, a través de una caja de resonancia. De cualquier forma, y en el caso de la voz, podremos referirnos a los ocho componentes fundamentales, puesto que desde un punto de vista identificativo, tanto la estructura espectral como su correlato psicológico tendrán un papel protagonista en la aportación de características individualizadoras. Por otra parte, resulta obvio señalar, que todo experto forense deberá tener muy en cuenta estas dos dimensiones de referencia a la hora de diseñar y ejecutar cualquier procedimiento de análisis con carácter auditivo.

Partiendo de este planteamiento general, y tratando de responder a las cuestiones formuladas en el apartado anterior, nos adentraremos en un estudio más detallado que nos permita conocer de que forma son conceptualizados cada uno de los componentes fundamentales de la voz.

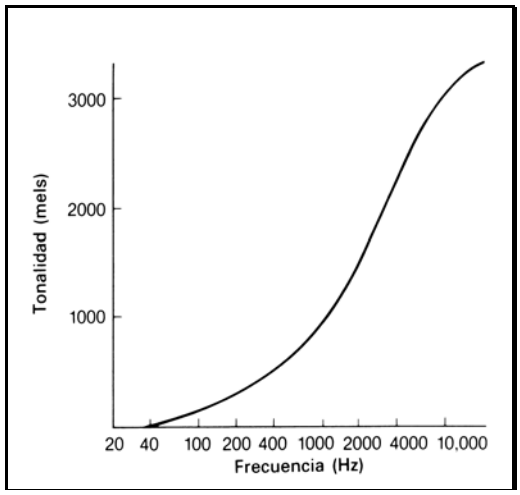
I.3.3.4.- Teorías sobre la percepción de la frecuencia

Se han propuesto diversas teorías para intentar explicar la percepción de la tonalidad, pero antes de comentar aquellas más significativas, es pertinente reflejar una circunstancia común a todas ellas: la percepción de la tonalidad está determinada, en gran medida, por las frecuencias del sonido percibido (recordemos que el término tonalidad se refiere a lo que escucha el sujeto y es, por tanto, de puro carácter psicológico). Sin embargo, al hablar de frecuencia, nos estamos refiriendo a una recurrencia de vibración de las moléculas de un medio elástico, que constituye el estímulo físico de la audición. La relación existente entre tonalidad y frecuencia no es exactamente lineal (Ilustración n130), aunque dentro de los umbrales de percepción para la frecuencia, puede decirse que existe una linealidad entre el incremento de la frecuencia y el incremento de la tonalidad, desde la baja frecuencia hasta los 1KHz . Por encima de dicho valor, la relación frecuencia/pitch se transforma en una relación logarítmica. A este respecto, existen otras fórmulas similares como la desarrollada por Fant:

$$F_{mel} = \square 1000/\log 2) \log [1+F_{Hz} /1000] = 1000 \log_2 [1+F_{Hz} /1000]$$

Como ya hemos comentado la unidad de medición de la frecuencia es el Hertz, ciclo por segundo, o el s^{-1} . Además, en función del ámbito de trabajo acústico de que se trate, pueden utilizarse otro tipo de unidades. Por ejemplo es muy común entre músicos el uso de la *octava* para relacionar las distintas frecuencias sonoras. Se denomina octava al intervalo de altura entre dos sonidos cuyas frecuencias son de relación 2. Es decir una, es el doble o la mitad de frecuencia que la otra. Este ratio de una octava se ha acordado en dividirlo en 6 *tonos*, y a su vez, subdividir éstos en 2 *semitonos*. Es decir, en una octava existen 6 tonos o doce semitonos. Weber y Fechner cuando en su Ley General trataban la tonalidad, denominaron *savart* al intervalo de variación en frecuencia más pequeño, perceptible por un oído sano.

La percepción de la frecuencia o tonalidad tiene como unidad de valoración el *mel*. No tiene mayor relevancia desde un punto de vista forense el conocer la relación exacta entre los valores de la frecuencia de un tono en Hertzios y sus posibles correspondencias en mel porque en la actualidad existe la posibilidad de obtener con absoluta precisión y de forma instantánea diversos valores estadísticos referidos al tono o frecuencia glotal (F_0) que origina una emisión hablada. Por otra parte, la sensación de tonalidad entra en el Ainter-juego≅ global de percepción junto con los otros ejes dimensionales del sonido: la duración y la sonoridad. Por este motivo, analizaremos con más detenimiento las relaciones entre las



realidades físicas mensurables y sus correspondencias perceptivas en el entorno de la sonoridad/presión acústica, centrándonos más en el proceso de transmisión y codificación psicofisiológica, en el caso de la tonalidad.

La necesidad de una clara comprensión de la actividad perceptiva por estímulos auditivos, a través del conocimiento de las distintas teorías de la percepción de la tonalidad, es extrapolable al resto de componentes fundamentales del sonido, y por tanto resulta de extraordinaria importancia a la hora de enfrentarse a los análisis de percepción auditiva del habla con finalidades forenses. El experto en acústica forense que sea consciente de sus limitaciones perceptivas, aunque se trate de alguien expresamente adiestrado para la percepción de las diferentes informaciones que integran los sonidos, será un experto riguroso. Aquél que esté seguro de que su oído no le puede engañar en ocasión alguna, estará expuesto a la comisión de graves errores, con nefastas consecuencias.

TEORÍA DE HELMHOLTZ SOBRE LA RESONANCIA

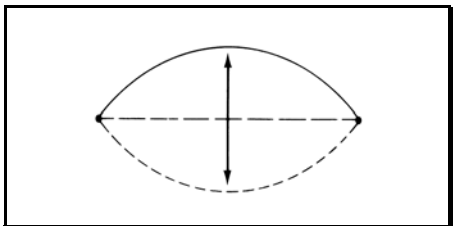
Hermann Helmholtz, pionero junto con Young de la teoría tricromática de la percepción del color, exploró también el ámbito perceptivo auditivo emitiendo su *teoría de la resonancia* [Helmholtz ,1863]; en ella, planteaba que la membrana basilar de la cóclea estaba compuesta por series de fibras transversales sintonizadas, para resonar, cada una de ellas, a una frecuencia determinada. Es decir, una frecuencia concreta sólo estimularía la fibra sensible que le correspondiese, provocando la activación de las fibras nerviosas que inervasen este área de la membrana basilar. Esta teoría fue refutada al demostrarse que las fibras de la membrana basilar están conectadas entre sí de forma tal que es imposible que una de ellas resuene independientemente de las otras en la forma propuesta por Helmholtz. En realidad, las áreas de la membrana basilar que vibran ante una frecuencia concreta suelen ser de gran extensión.

TEORÍA DE RUTHERFORD SOBRE LA FRECUENCIA

Años después, Rutherford [1.886] propuso que la membrana basilar vibraba como un todo, (de forma similar a la indicada en la ilustración n1 31) y que la frecuencia de tal vibración era idéntica a la recurrencia del estímulo sonoro que la producía.

De acuerdo a esta propuesta *(teoría de la frecuencia)*, debiera existir una relación directamente lineal entre la frecuencia del estímulo y el número de descargas de impulsos en las fibras nerviosas del nervio auditivo; por lo que estaríamos hablando de 8.000 impulsos por segundo ante la estimulación de un tono de 8 Khz.

Investigaciones posteriores demostraron que la membrana basilar no vibra exactamente en la forma propuesta por Rutherford, si bien es cierto que existe una coincidencia entre la frecuencia transmitida por el estímulo y la frecuencia de vibración en la membrana basilar. Es decir, un tono puro de 2 Khz implica una vibración de la membrana de 2000 veces por segundo. Entonces,)donde está el problema?. Las fibras nerviosas tienen un período refractario que limita su tasa máxima de disparo a unos mil impulsos por segundo. Por esta razón, aunque una frecuencia de 8000 Hz sí implicaría una respuesta de la membrana basilar de 8000 vibraciones por segundo, no podría sin embargo generar una similar respuesta de las fibras nerviosas.

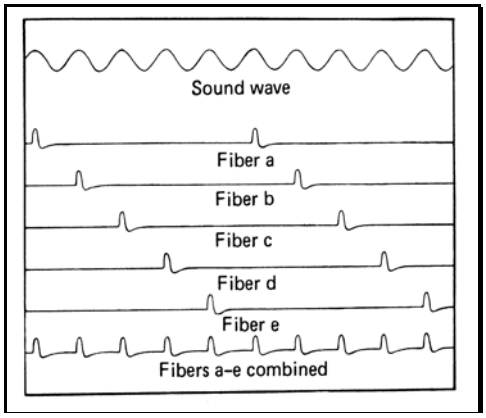


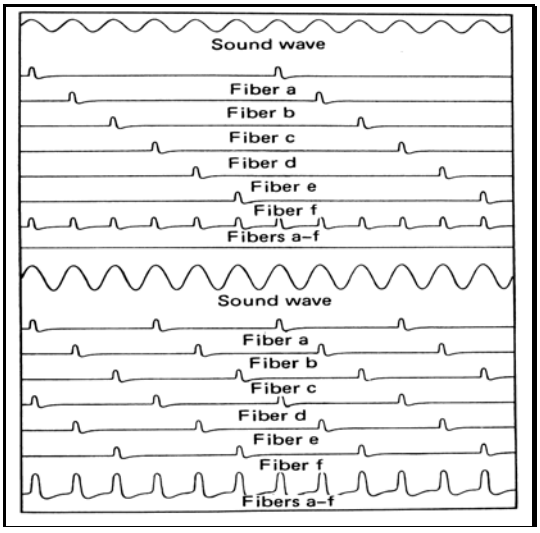
PRINCIPIO DE LA ANDANADA (O DESCARGA) DE WEVER

Intentando buscar una solución al principal inconveniente de la teoría de Rutherford, Ernest Wever [1949] propuso una modificación de la teoría de la frecuencia: *el principio de la andanada.*

Dicho principio, plantea una actividad de las fibras nerviosas que posibilita la obtención de altas tasas de descarga. Partamos de los gráficos siguientes:

En la ilustración n1 33 se muestran cinco fibras nerviosas funcionando en la forma conjugada propuesta por Wever, disparándose cada una de ellas tras cinco períodos de la onda sonora. La fibra "a" dispara en el primer período y, a continuación, se vuelve refractaria. Mientras "a" es refractaria, las fibras "b", "c", "d" y "e" descargan, respectivamente, durante los períodos segundo, tercero, cuarto y quinto. Pasados éstos, en el sexto período, ha dejado de ser refractaria y se descarga, con lo que se inicia un ciclo igual al descrito. Por tanto, mientras que la tasa de cada fibra individual está limitada por su período refractario, las agrupaciones de las 30.000 fibras nerviosas que parten del órgano de Corti, pueden conjugarse para producir altas tasas de descarga nerviosa. En la ilustración n1





32, observamos cómo el principio de la andanada trata de explicar la respuesta de las fibras nerviosas ante la diferencia de amplitud en tonos de idéntica frecuencia. A un incremento de la amplitud corresponde un acortamiento del intervalo refractario de cada fibra.

Jerzy Rose y sus colaboradores [1967] demostraron que las fibras del nervio auditivo descargaban en sincronía con el estímulo. Es decir, la relación existente entre las distintas porciones de un estímulo tonal puro y el disparo de distintas fibras nerviosas auditivas, provoca descargas irregulares en el tiempo de una regularidad temporal distinta a la propuesta por Wever. Según estos autores, la descarga sólo se efectúa cuando el estímulo tonal puro está en uno de sus máximos :

A la sincronización de la descarga de las fibras nerviosas con los máximos de un tono puro se le denomina *sintonización en fase*. El funcionamiento conjunto de grupos de fibras puede señalizar la tonalidad, tal y como propone el principio de andanada, aunque las fibras no trabajen conjuntamente de forma idéntica a la manera propuesta por Wever.

La mayoría de las evidencias que presentan la sintonización en fase como una posible explicación para señalizar la frecuencia del estímulo, procede de registros obtenidos a partir de las fibras nerviosas del nervio auditivo. Sin embargo, las evidencias de sintonización en fase se hacen más débiles en centros superiores del sistema auditivo y, cuanto más nos desplazamos en dirección al córtex, más complicado resulta encontrar neuronas con sintonización en fase para frecuencias superiores a 1.000 Hz. Incluso en el nervio auditivo, la sintonización de fase solo funciona bien hasta los 5.000 Hz. Esta serie de evidencias nos conduce al planteamiento de otra cuestión :)qué ocurre con la trasmisión de estímulos sonoros de alta frecuencia en los centros superiores?

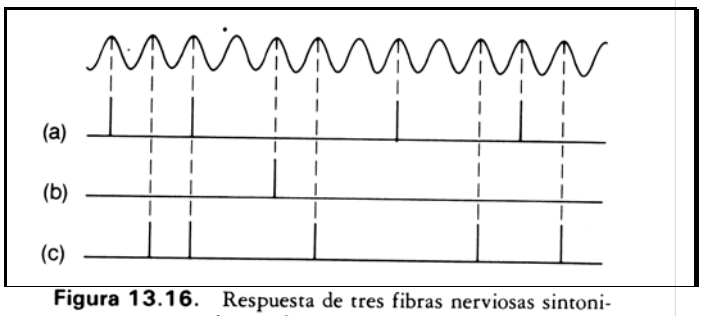
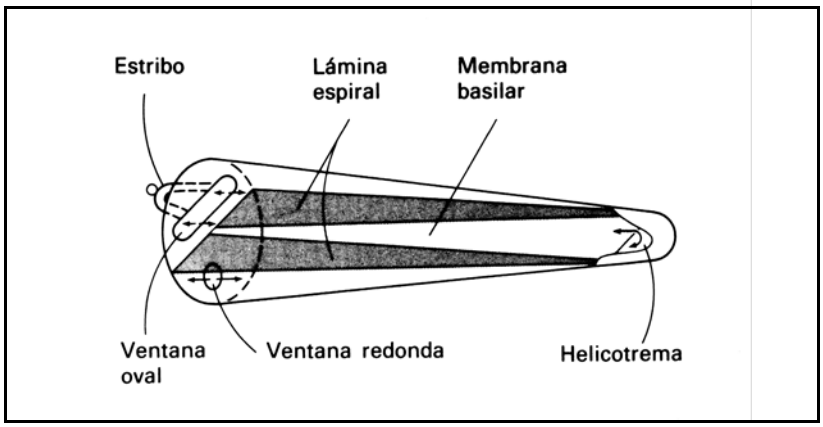


Figura 13.16. Respuesta de tres fibras nerviosas sintonizadas en fase con el estímulo.

La denominada *teoría del lugar* de Békésy formula una solución complementaria.

TEORÍA DEL LUGAR DE BÉKÉSY

El argumento principal de la teoría del lugar de Békésy [1960] plantea que cada frecuencia sonora obtiene una respuesta de las células ciliares en una zona concreta de la membrana basilar. En síntesis, las altas frecuencias excitarían especialmente el área próxima a la

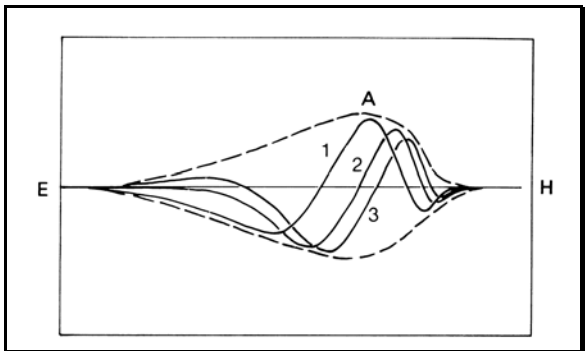


ventana oval, y las bajas frecuencias estimularían las células pilosas próximas al helicotrema. Es decir, teóricamente, cada uno de los posibles tonos puros estimularía con máxima intensidad un lugar de la partición coclear.

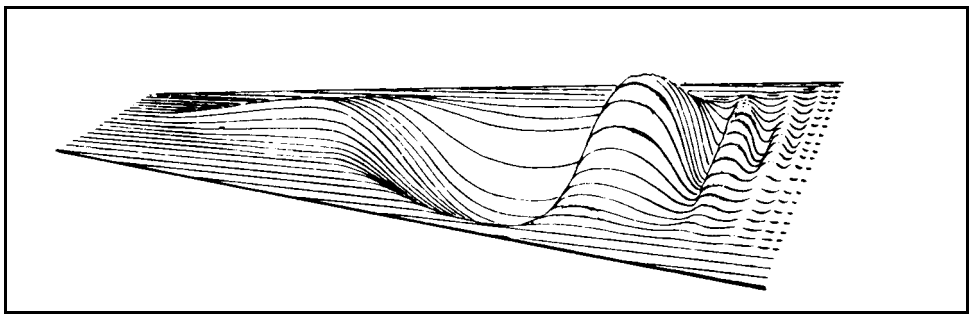
Este planteamiento es similar al expuesto en la teoría de la resonancia de Helmholtz, aunque difiere en el mecanismo. Mientras que dicha teoría habla de la estimulación de fibras resonadoras en sintonía a frecuencias concretas, Békésy piensa que las células o áreas excitadas, lo son, por el tipo de movimientos que los estímulos sonoros generan en la membrana basilar.

Békésy, estudió en profundidad las estructuras de la partición coclear concluyendo en que la vibración de la misma estaba absolutamente asociada al comportamiento de la membrana basilar. Por este motivo, focalizó todos sus esfuerzos en el estudio de dicha membrana, alcanzando dos conclusiones iniciales: la membrana basilar era tres o cuatro veces más ancha en zona del helicotrema que en la del estribo, y unas cien veces más dura en las proximidades del estribo que en el extremo del helicotrema:

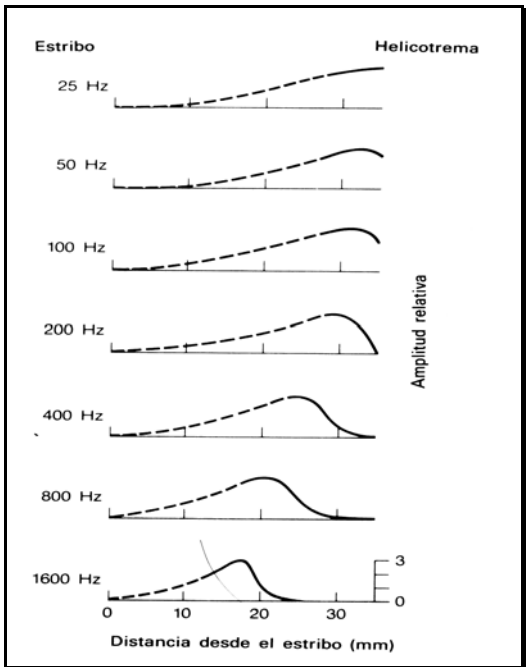
Partiendo de estos hechos, Békésy [1.960] introdujo el concepto de *onda desplazante,* que era la onda en que se configuraba la estructura de la membrana basilar ante los cambios de presión que afectaban al fluido coclear:



Para conocer el efecto que produce la vibración de la membrana basilar en las células ciliares - mayores tasas de descarga a mayor desplazamiento de la membrana en relación a su posición de reposo- es conveniente observar en detalle las denominadas *envolventes de la onda desplazante*. Estas envolventes, representan el movimiento de máximo desplazamiento de la membrana ante un estímulo tonal determinado (se indica mediante la línea discontinua de la ilustración n1 37).



Además la ilustración de arriba muestra la apariencia de una onda desplazante y los distintos puntos de excitación de la membrana en función de la frecuencia. La línea horizontal representa la membrana basilar en situación de reposo, siendo el final marcado con una AE≅ el área del estribo, y el marcado con una AH≅, la zona próxima al helicotrema. La curva 1 muestra la posición de la membrana basilar en un instante temporal de la vibración, mientras que las curvas 2 y 3 muestran la posición de la membrana en dos instantes posteriores.



Por tanto, la envolvente de la onda desplazante tiene dos importantes características:

- *presenta un desplazamiento máximo en un punto de la membrana basilar que coincide con el área de células ciliares que enviará más impulsos*. Por ejemplo, la envolvente de la ilustración n1 37 muestra que la membrana basilar sufre un desplazamiento máximo a causa de la onda desplazante en el punto A, lo que significa que las células pilosas próximas a AA≅ mandarán más señales que las próximas a otras partes de la membrana.

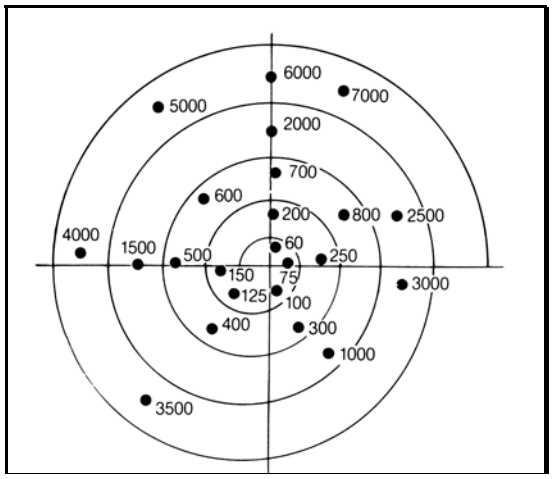
- *la posición de este máximo en la membrana basilar depende de la frecuencia del estímulo sonoro.* La ilustración n1 38 muestra las envolventes de la vibración producidas por estímulos cuyas frecuencias van desde 25 Hz (el más grave) hasta 1.600 Hz (el más agudo); como se puede observar, las bajas frecuencias producen vibraciones máximas en las proximidades del helicotrema, mientras que las frecuencias altas producen sus mayores vibraciones cerca del estribo.

Si observamos atentamente las envolventes correspondientes a las distintas frecuencias veremos cómo aquellas correspondientes a sonidos de baja frecuencia, además de producir una máxima respuesta en la zona del helicotrema, provocan un desplazamiento en buena parte de la membrana (a más baja frecuencia, mayor superficie de membrana afectada). Esta simple observación, explica un clásico fenómeno psicoacústico: el efecto Amasking≅ o de enmascaramiento, que se define como el número de decibelios que se eleva el umbral de audición de un sonido (enmascarado) ante la presencia de otro (enmascarante). Tal efecto perceptivo, se manifiesta en diversas situaciones de la vida real: si vamos escuchando una emisión radiofónica en nuestro vehículo con la ventanilla abierta, y entramos en un túnel, necesitamos aumentar el volumen para poder seguir escuchando con claridad; si nos encontramos viendo la televisión en nuestra casa y entra un funcionamiento un aparato de aire acondicionado o de ventilación, mínimamente ruidoso, muy probablemente efectuemos la misma operación. Los ruidos de ventilación, o el de rozamiento de las ruedas del coche contra el asfalto, llevan un elevado componente en bajas frecuencias que al estimular nuestros oídos de forma simultánea a otras frecuencias (habla, música) provocan su Aenmascaramiento≅ perceptivo. La explicación es sencilla: la onda desplazante de la membrana basilar, al responder a la baja frecuencia invade o incluso engulle la envolvente de otras frecuencias medias o altas (especialmente, si la intensidad del estímulo de las bajas frecuencias es elevado).

S. Stevens estudió la relación cuantitativa entre la percepción del tono o tonalidad y el nivel de intensidad, observando, que un aumento de la intensidad conlleva una elevación de la tonalidad de los agudos y disminuye sensiblemente la de los graves. Sin embargo, en el caso de las frecuencias medias (800-3000 Hz) la altura tonal quedaba prácticamente inalterada.

La asociación entre la frecuencia del estímulo y el lugar de la cóclea en el que las células ciliares son estimuladas al máximo podemos observarlo en la siguiente ilustración:

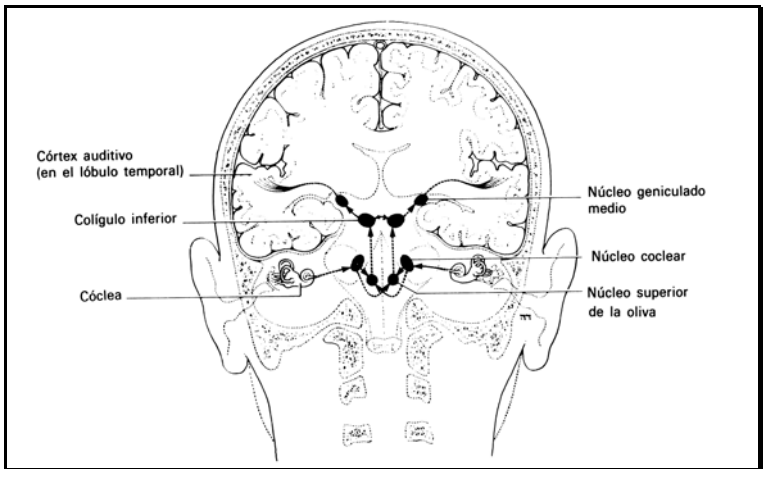
Este mapa de la cóclea se construyó colocando electrodos en localizaciones diferentes y determinando la frecuencia estimular que producía la mayor respuesta en cada localización [Culler et al.,1943]. Dicha asociación de correspondencia (área estimulada/estímulo) se ha constatado empíricamente mediante experimentos de Asordera por sobre estimulación≅, en los que se analizaron las estructuras dañadas del órgano de Corti de un animal, después de que éste hubiera sido expuesto a sonidos puros de gran intensidad. Los tonos de baja frecuencia producían graves daños en un área extensa próxima al helicotrema. Sin embargo, según se iba incrementando la frecuencia, el área lesionada se movía hacia la ventana oval y se hacía más localizada. De ello se deduce que la exposición dañina a sonidos de alta frecuencia es menos perjudicial -en cuanto a la dimensión de área afectada- que la exposición a estímulos de baja frecuencia. No obstante, los sonómetros -instrumentos utilizados para la medición de la nocividad del ruido a través de la valoración del nivel de presión sonora eficaz- suelen utilizar como escala de ponderación aquella que da un índice de la nocividad de los sonidos agudos pues son los considerados más agresivos para nuestro oído (escala A).



La experimentación desarrollada por Békésy -que dicho sea de paso le condujo al premio Nóbel en 1.961- demostró suficientemente la relación de correspondencia entre la frecuencia de un estímulo y el lugar de máxima estimulación en el oído interno. No obstante, quedaba por explicar como se codificaba en los niveles superiores la información obtenida nivel coclear. En la siguiente ilustración, observamos las distintas estructuras neurales por las que transcurre la señal auditiva hasta llegar al cortex:

Las fibras nerviosas procedentes de la cóclea realizan una primera sinapsis en el llamado *núcleo coclear*; a continuación, en el núcleo de la *oliva superior*; posteriormente, en el *colígulo inferior* del cerebro medio, y seguidamente llegan al *núcleo geniculado medio* del tálamo. Desde este punto, se dirigen a su destino definitivo: el *área primaria de recepción auditiva*, localizada en el lóbulo temporal del córtex. Pero al margen de esta ruta de transmisión directa, existen otras vías de interconexión más complejas entre los centros neurales de ambos hemisferios cerebrales. Por ejemplo, el núcleo coclear del oído derecho manda axones tanto al núcleo superior de la oliva como al colígulo inferior del oído izquierdo. Además, los colígulos inferiores izquierdo y derecho se comunican entre sí. De la misma forma, e insistiendo en el carácter activo del oído interno en el proceso de transmisión, existen canales nerviosos que envían señales descendentes desde el cerebro a la cóclea.

La experimentación con animales también ha permitido constatar la posibilidad de



establecer *mapas o representaciones tonotópicas* que representen la respuesta de los núcleos nerviosos ante los distintos estímulos de frecuencia sonora. Abeles y Goldstein [1970]

demostraron como ciertas neuronas o grupos de neuronas sitúan su umbral mínimo de sensibilidad en relación a una frecuencia concreta, que denominaron *frecuencia característica*.

Los pilares que sustentan la teoría del lugar parecen explicar satisfactoriamente los mecanismos de percepción para las frecuencias sonoras. Sin embargo, la teoría de Békésy no puede justificar la existencia de umbrales mínimos de discriminación entre tonos de baja frecuencia. Como ya hemos comentado, las ondas desplazantes por estímulos de baja frecuencia generan una envolvente de gran extensión, con un punto de máximo desplazamiento que se sitúa en el área del helicotrema. Partiendo de este principio, y conociendo que el umbral de discriminación perceptivo de frecuencia en el área de un tono de 100 Hz es de aproximadamente 3 Hz (lo que quiere decir que somos capaces de darnos cuenta de la diferencia entre un tono de 100 y otro de 97 Hz) resulta muy difícil explicar a partir de la teoría del lugar tal capacidad de discriminación. Es decir, ambas envolventes son muy amplias y tienen sus máximos demasiado cercanos.

Sin embargo, la explicación de este mínimo de discriminación en baja frecuencia, sí puede ser justificada mediante la *sintonización en fase* (que sin embargo no era capaz de explicar la respuesta de las fibras nerviosas para frecuencias por encima de los 4 Khz).

En resumen, podemos afirmar que si bien existen claras evidencias en favor de la existencia de los mecanismos de lugar y sintonización, ambos presentan ciertas limitaciones. Goldstein [1.984] apunta como explicación que la sincronización de los impulsos nerviosos es importante en las bajas frecuencias (por debajo de 500-1.000 Hz), que el lugar es importante en frecuencias altas (por encima de 4.500-5.000) y que ambos mecanismos pueden ser operativos para las frecuencias comprendidas entre estos límites.

Investigaciones más recientes han permitido explorar con más precisión las vibraciones de la membrana basilar. Uno de los métodos más utilizados es el de *Mössbauer,* que consiste en depositar un microcristal de cobalto 58 radiactivo sobre la membrana basilar. Cuando ésta se desplaza, la frecuencia de radiación emitida por el cristal comienza a variar siendo analizada por un ordenador a través de un filtro sintonizado, pudiéndose deducir así las referencias de movimiento de la membrana.

Esta nueva técnica ha aportado nuevos datos que permiten explicar la gran selectividad de frecuencias en la zona basal de la cóclea. Para un punto dado de la membrana basilar, la amplitud de las vibraciones disminuye muy rápidamente cuando la frecuencia se separa de la resonancia. La disminución es del orden de 24 dB por octava para las altas frecuencias y de 150 dB por octava para las bajas.

Como complemento a este análisis detallado de la estructura y funcionamiento de los componentes del oído interno relacionados con la percepción, resulta muy oportuno comentar la opinión de Zwicker [1.981] al respecto. Según tal autor, la membrana basilar se comporta como un analizador armónico con 24 filtros de banda (*bandas críticas o bark*) cuyo ancho de banda en la parte central del espectro audible es de un tercio de octava. Cada una de las bandas críticas (zonas de la membrana de 1,3 mm de longitud, divididas asimismo en 100 mel) están constituidas por una banda de frecuencias con una característica tal, que cualquiera que sea el ancho de una banda de ruido comprendido dentro de la banda crítica, la sonoridad es siempre la misma, si el nivel de intensidad de la banda ruidosa se mantiene constante.

Una vez analizados los diferentes mecanismos de transmisión y niveles intermedios de codificación, es momento de preguntarnos qué es lo que ocurre en la última instancia: el córtex cerebral. Durrant y Jovrinic [1.977], tras distintas experimentaciones con gatos a los que entrenaban en una tarea auditiva y eliminaban posteriormente sectores de su córtex con cirugía para volver a observar el nivel de ejecución del animal en dicha tarea, llegaron a la conclusión de que diversos trabajos auditivos pueden realizarse aun en ausencia del córtex :

1. Responder a la presentación de un sonido.
2. Responder a cambios en la intensidad de un sonido.
3. Responder a cambios en la frecuencia de un sonido.

Aparentemente, el córtex no es imprescindible para discriminar entre frecuencias distintas, lo que sugiere la posibilidad de que nuestro córtex auditivo no esté implicado en la percepción de la tonalidad. No obstante, experimentos como los desarrollados por Diamond y Neff [1975] demostraron que su papel es crucial en la realización de otros trascendentes trabajos auditivos, como la discriminación de patrones de sonidos o la duración y localización espacial de los mismos.

Durrant y Lovrinic [1.977] manifiestan que "cuanto mayor sea la complejidad de la estimulación sonora y la información en ella contenida, mayor será el grado esperable de implicación del córtex en su procesamiento≅. De igual forma, convienen en que A...indudablemente, el córtex auditivo está ampliamente implicado en la percepción del habla, así como las áreas corticales más complejas del córtex auditivo primario están implicadas en el manejo de las complejidades del significado y la memoria, componentes esenciales para que se produzca una adecuada percepción del habla≅.

I.3.3.5.- Percepción de la presión o intensidad sonora : sonoridad

Desde el punto de vista de la identificación forense a través de la voz, tanto las informaciones correspondientes a la amplitud - directamente relacionadas con la intensidad o presión sonora - como las relativas al tiempo que dura una emisión acústica, carecen de la relevancia individualizadora que aportan los datos provenientes del tono y el timbre.

No obstante, la percepción de la sonoridad - correlato psicológico de la amplitud de onda sonora- está inter relacionada con la frecuencia(s) del sonido(s) objeto(s) de tal percepción. Por ello, es necesario realizar una serie de consideraciones previas al respecto, ya que aunque no nos encontramos ante un componente fundamental a la hora de aportar rasgos característicos de un locutor, no debemos olvidar que podemos enfrentarnos a un relativo Amaquillador perceptivo≅ de las emisiones sonoras.

Como ya sabemos, el ser humano percibe el sonido a través de las vibraciones de la membrana timpánica, que son provocadas por las variaciones de la presión atmosférica que ejercen las moléculas del aire. Dichas moléculas se encuentran en un determinado lugar a una determinada presión atmosférica. Las variaciones de la temperatura ambiental provocan un movimiento de las mismas, generando simultáneamente una variación de la presión estática. Este movimiento se denomina movimiento Browniano molecular y se corresponde con una presión de 10^{-6} pascal. Afortunadamente, el oído humano necesitaría desplazar su umbral de sonoridad mínimo a -14 dB_{SPL} para poder escuchar el ruido térmico del medio ambiente. Decimos afortunadamente, por que de ser así estaríamos continuamente escuchando un soplido sonoro incluso en los lugares más recónditamente silenciosos.

Por tanto, cuando hablemos de la sonoridad en relación a una onda sonora, deberemos prestar atención a su amplitud o, lo que es lo mismo, la diferencia de presión existente entre la atmosférica (referencia estática) y la máxima alcanzada durante el ciclo de variaciones sinusoidales en la presión que produce dicha onda (referencia dinámica).

La simple experiencia nos muestra que la sensibilidad del oído a la intensidad o presión acústica no es la misma para todas las frecuencias. Las primeras evaluaciones experimentales relativas a sensaciones obtenidas a partir de los estímulos de intensidad sonora fueron realizadas por Weber y Fechner. Estos científicos se percataron de que al tomar como referencia un sonido de frecuencia 1 Khz, e incrementar de forma continua su amplitud desde su umbral mínimo de audición, de manera que apenas podía ser oído, hasta alcanzar la sonoridad que empezaba a dañar nuestros oídos, el intervalo de aumento de la amplitud se había incrementado en un factor aproximado de 10 millones.

Puesto que la utilización de números de tal dimensión resultaba poco funcional, los investigadores de la audición desarrollaron una escala cuya unidad era el *decibelio* y que tenía como principal virtud proporcionar factores numéricos mucho más manejables.

Los decibelios representan una relación logarítmica entre un factor de referencia y el nivel de señal del sonido medido. En términos generales, y desde nuestra perspectiva de estudio, dicha relación será relativa a valores de intensidad (dB_{sil}) o de presión sonora (dB_{spl}).

No siempre nos encontraremos con dB referidos a la intensidad o presión sonora. Sin ir más lejos, en muchos equipos de grabación/ reproducción cuando observamos los indicadores de niveles de entrada o reproducción de la señal (vumeter), vemos unos valores dB_v que miden diferencias de voltaje, siendo su nivel de referencia 0 dB_v = 1Voltio. En otros equipos de registro es muy común encontrarnos la medición de señal en dB_m, que son una versión similar a los dB_v pero que toman como valor de referencia 0dB_m = 0,775 Voltios.

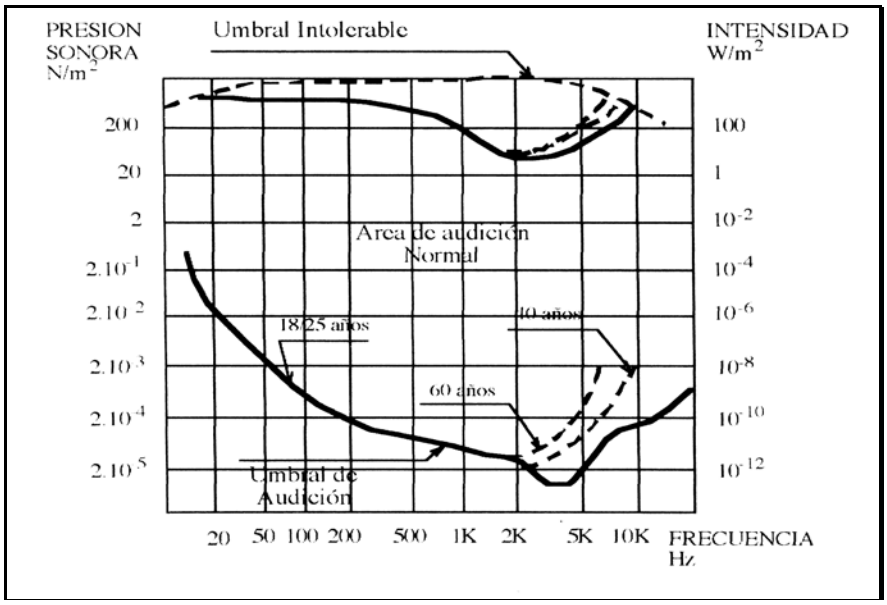
Cuando hablemos de mensurar simultáneamente distintos niveles sonoros *(presión sonora eficaz)* como ocurre en el caso de la medición del ruido ambiente, nos encontraremos con los dB_A , dB_B y dB_C que se corresponden respectivamente con tres escalas o curvas de ponderación para las bajas, medias y altas frecuencias que promedian el valor eficaz (RMS) de energía sonora de una frecuencia en el conjunto de las frecuencias que le acompañan en un mismo suceso sonoro.

En el ámbito de la acústica forense, suele utilizarse para definir la presión sonora a nivel perceptivo los dB_{SPL} , los cuales se definen a partir de la siguiente ecuación:

$$dB = 20 \log_{10} (P/P_0)$$

donde P es la amplitud de la onda sonora y P_0 es la presión de referencia elegida por el experimentador. La mayor parte de los investigadores utilizan como referencia la de 0.0002 dinas/cm^2, la presión del umbral de escucha en el rango de frecuencias a las que nuestro oído es más sensible (1.000-4.000 Hz). Igualmente pueden utilizarse unidades equivalentes como el microbar, micropascal, etc. . Para indicar que se ha escogido este nivel de presión como referencia se utiliza el término *nivel de presión sonora (SPL = "Sound Presure Level")*. Por tanto, cuando digamos que el SPL de un tono es de 50 dB, o simplemente hablemos de una presión sonora de 50 dB, sabremos que este valor se calculó a partir de una presión de referencia de *0,0002 dinas/cm^2*.

También es muy común el uso de los decibelios referidos a la intensidad o poder acústico del sonido. En este caso, hablaremos de los dB_{SIL} siglas correspondientes al *nivel de Intensidad sonora (SIL = ASound Intensity Level≅)* . La fórmula para el cálculo del ratio en este tipo de decibelios se expresa como :



$$Db = 10 \log_{10} (I/I_0)$$

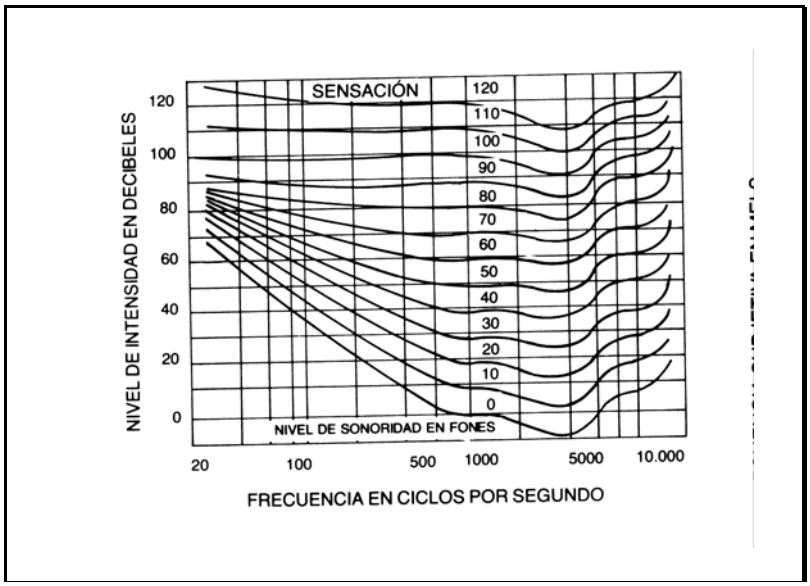
donde AI≅ es la intensidad sonora (en watios/m^2) , e AI_0A es la referencia estandarizada de intensidad para las frecuencias mencionadas en el caso del nivel de presión sonora(1 Khz-4KHz), usualmente 10^{-12} *watts* $/ m^2$.

La escala decibélica en el SPL permite convertir el rango de presión sonora que va desde 0,0002 *dinas/cm²* (la presión en el umbral de audición) hasta 2.000 *dinas/cm²* ó *N/m²* (presión del umbral del dolor) en fracciones de presiones sonoras mucho más manejables, que van de 0 a 140 dB_{SPL}. Esta escala es más funcional para trasladar los decibelios desde el ámbito de las matemáticas abstractas a referencias de la vida real. Así, podremos decir que un cuchicheo al oído estará en torno a los 20 dB, una conversación normal con un metro de distancia entre dos interlocutores se mantendrá en torno a los 60 dB y el ruido de un disparo alcanzará fácilmente los 120 dB; siempre por supuesto en la escala de presión sonora o SPL.

Acabamos de conocer cuales son las distintas unidades para medir la realidad física que representan la presión o intensidad sonora. Pero una cosa es medir esta magnitud real y otra algo distinta es conocer y mensurar su correlato perceptivo. Aunque el incremento en el nivel SPL suela provocar un aumento en la sonoridad, SPL y sonoridad no son la misma cosa.

Según Goldstein [1984] si utilizásemos el procedimiento de estimación de magnitud de Stevens para determinar la relación entre la sonoridad y la presión sonora, no nos encontraríamos con una relación uno-a-uno entre las dos. En intensidades moderadas, obtendríamos una ley potencial que relacionaría estas dos entidades mediante un exponente comprendido entre 0,4 y 0,7 (dependiendo de las condiciones de medida y de la frecuencia del sonido utilizado). Este resultado nos indicaría que en la audición, como en la visión se produce comprensión de respuesta en la relación existente entre la percepción y la medida física del estímulo. Esto es, partiendo de la existencia de unos umbrales físicos perceptivos (en el caso de la intensidad y en la zona central del espectro es de 0.2 dB) se observa que grandes incrementos en la presión sonora producen sólo incrementos moderados en sonoridad. Incrementar la presión sonora en un factor de diez (lo que es lo mismo que incrementar el SPL en 20 Db) sólo incrementará la sonoridad en un factor comprendido entre 2,5 - 4,0.

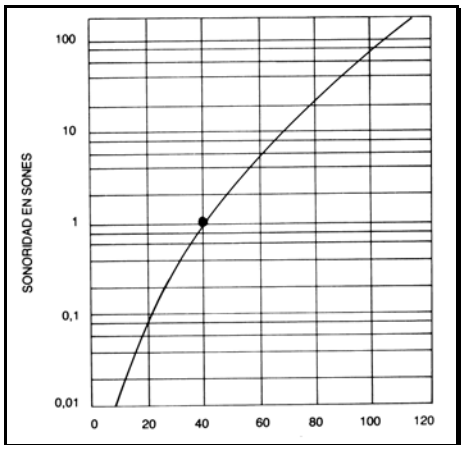
Tomando como referencia la escala SPL , Fletcher y Munson determinaron, dentro de los umbrales de percepción de la sonoridad y para el rango de frecuencias audibles por el hombre, una serie de curvas cuyo perfil se correspondía con aquellos valores de intensidad de tales sonidos, que proporcionaban la misma sensación de sonoridad que un tono puro de 1 Khz a diferentes valores de presión acústica:



Estas curvas normalizadas de igual sonoridad o *loudness* pusieron de manifiesto que a la altura de la frecuencia de 1000 Hz se producía una correspondencia a lo largo de todo el umbral de sonoridad , entre el incremento de ésta y el nivel de presión SPL. Teniendo en cuenta esta particularidad, se convino asociar el nivel de sonoridad de una señal acústica cualquiera con el valor numérico del nivel SPL de una frecuencia de 1 Khz que tuviese la misma sonoridad que tal señal. La unidad elegida se expresó como el *Phon*. Por tanto una señal acústica con la misma sonoridad que un tono de 1000 Hz cuyo nivel de presión sonora es de 80 dB_{SPL}, tiene un nivel de sonoridad de 80 Phon. La línea isofónica que se corresponde con el umbral mínimo de audición es la de 0 Phon, determinando la de 140 Phon el umbral del dolor.

No obstante, el Phon no expresa la verdadera sonoridad de un sonido, sino sólo un nivel de sonoridad. Hemos de tener en cuenta que ante un mismo sonido de una determinada sonoridad, cuando efectuamos una percepción binaural obtenemos una sensación dos veces superior a la que puede obtenerse utilizando un solo oído. Partiendo de la sensación de sonoridad percibida por un oído Stevens desarrolló una escala aritmética de sonoridades aceptando como unidad de medida el *son*, que es la sonoridad biauricular de una señal de 1000 Hz a 40 dB_{SPL} . A partir de aquí podemos establecer una relación entre las sonoridades (son) y los niveles de sonoridad (Phon). Es la llamada función de transferencia. Un nivel de sonoridad de 40 Phon es equivalente a una sonoridad de 1 son, duplicándose el número de son por cada incremento de 10 Phon.

Además de las ya citadas, existen diversas unidades para la medición a nivel perceptivo de la sonoridad ruidosa, es decir aquella no referida a valores tonales puros. Entre otras cabe citar el *laut* de Zwicker o el *Noy* definido por Kryter.



Los estudios sobre la sonoridad de las frecuencias introducen nuevos conceptos relacionados con las propiedades físicas y psicológicas del sonido. Hablaremos de *área de respuesta auditiva* cuando nos refiramos a la comprendida entre las curvas isofónicas superior (umbral del cosquilleo) e inferior (curva de audibilidad). Este área define las frecuencias y SPLs en las que nuestra audición es funcional. La curva de audibilidad o A Threshold of hearing≅ nos indica el valor del umbral mínimo de audición SPL ante las distintas frecuencias audibles; mientras que el umbral de cosquilleo o

AThreshold of feeling≅ indica las magnitudes que hacen que un sonido sea más sentido que oído. Es, entre estos dos extremos, donde experimentamos la percepción del sonido. Si el valor SPL se sitúa por debajo de la curva de audibilidad (0 phon), no podremos oír el sonido; si el valor SPL se sitúa justo por encima del umbral de cosquilleo, experimentaremos dolor y, de exponernos con cierta continuidad a esta estimulación, podríamos sufrir daños auditivos. El trauma ótico irreparable se produciría ante estímulos sonoros con valores de presión acústica más elevados (normalmente a partir de 140 dB).

I.3.3.6.- Percepción de la estructura acústica de resonancia: el timbre.

Desde un enfoque perceptivo, el timbre es el más complejo de los cuatro componentes psicoacústicos básicos de la voz. Como ya sabemos, la producción acústica del espectro de radiación - correlato físico del timbre- está basada en el fenómeno de resonancia, o lo que es lo mismo, la amplificación de unas determinadas frecuencias sonoras por la acción de un resonador (cuerpo que vibra a la misma frecuencia que otro que resuena, amplificando solamente una determinada gama de frecuencias). Desde el punto de vista identificativo resulta de gran relevancia, por tres razones fundamentales. En primer lugar, porque es el componente fundamental de carácter más invariable e individual al estar directamente relacionado con una estructura anatómico-fisiológica y una base de articulación concretas. En segundo lugar, porque integra todas las referencias físicas del sonido vocal, en su fase de radiación (fase en la que se encuentran siempre los objetos de estudio del audio forense). Y, en tercer lugar, porque el timbre -entendido en su dimensión más amplia- aporta informaciones relativas a tres dimensiones de interés identificativo: biológica, psicológica y socio-educacional.

En la dimensión biológica puede aportar información sobre las características anatómicas y fisiológicas del hablante, su posible edad, sexo, estados patológicos, hábitos fonatorios, etc.

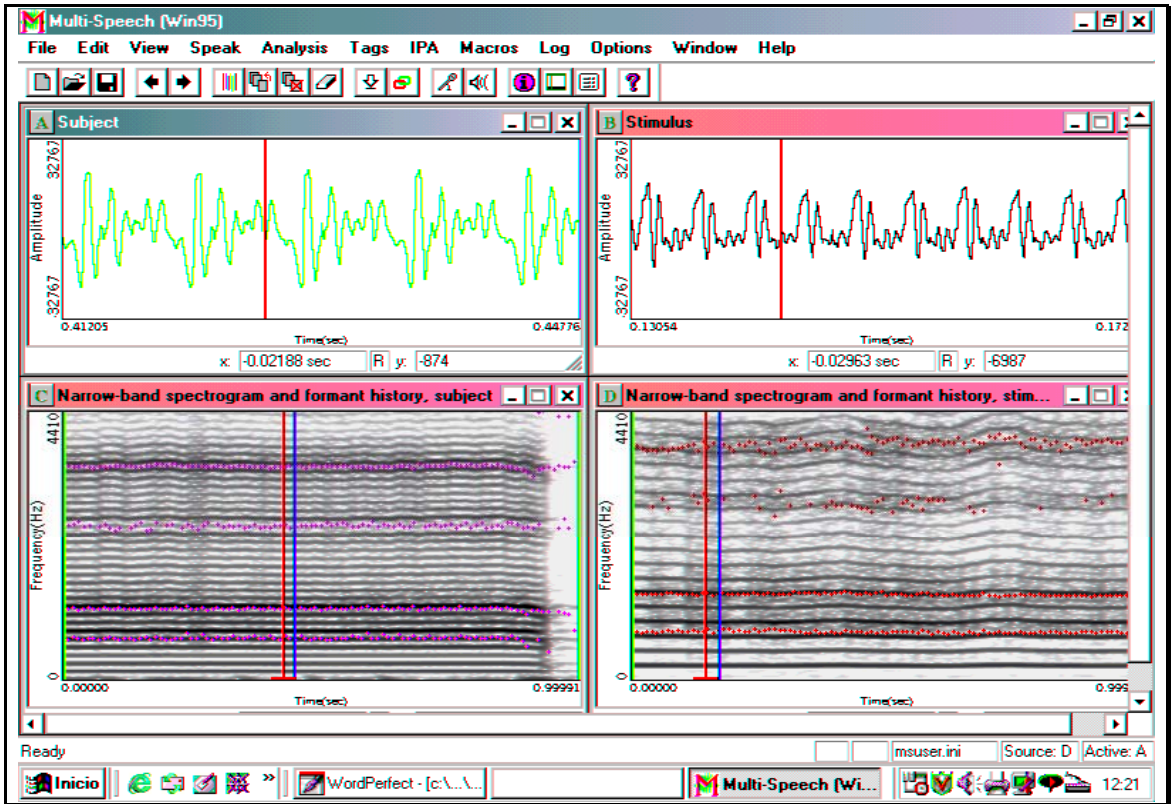
La dimensión psicológica puede proporcionar conocimientos sobre las características básicas de una personalidad o de un estado emocional concreto.

En el nivel socio-educacional encontraríamos aquellas informaciones derivadas de los hábitos de aprendizaje, e incluso aquellas otras de índole cultural relativas a factores diatópicos o etnográficos (normas de conducta comunicativa propias de una comunidad), etc.

Resulta muy complicado establecer una relación directa entre manifestaciones concretas de la cualidad de voz y sus dimensiones psicológica o socio-educacional, aunque también es difícil no admitir la existencia de tal vinculación. La denominada expresión fonoestésica

[Rodriguez, 1989] o lo que es lo mismo, la determinación de patrones de carácter del emisor a partir del análisis de sus emisiones, es un ejemplo en este sentido.

Pero dejando al margen este elemento de discusión, la naturaleza de un timbre vocal cualquiera estará siempre referenciada en base al *número, intensidad, distribución y conformación* de los armónicos que lo integran (ilustración n1 44). Igualmente entrarán en juego las cualidades transitorias de articulación: *Ataque, permanencia y extinción*.



Para el investigador forense resulta difícil definir el timbre. En términos elementales entendemos por timbre aquella cualidad que permite diferenciar dos sonidos del mismo tono y de la misma intensidad. Pero en el caso del habla, y cuando se trata de abarcar todos aquellos aspectos de una emisión individual que no pueden ser relacionados directamente con el tono o la intensidad vocal, no hay razón para soslayar cualquier aspecto o matiz que pueda enriquecer o precisar la definición de la cualidad vocal. Quizás es demasiado pretencioso atribuir al timbre componentes de tipo psicológico o educacional y debiéramos restringirnos a sus referencias más directas - fenómeno de resonancia y cualidades transitorias de articulación - pero en definitiva estamos ante un planteamiento de investigación donde lo más importante no es la ubicación de los objetos de análisis sino el no dejar de contemplar alguno de ellos.

Precisamente como consecuencia de la diversidad de factores que pueden integrar el concepto de timbre o cualidad de voz, la percepción auditiva del mismo se presenta como una de las tareas más complicadas en el análisis forense. Pero a esta dificultad hemos de contraponer la óptima posibilidad de analizar el timbre en un especio tridimensional gráfico. Los efectos sobre la vibración glotal de las cavidades resonantes -fundamentalmente faringe, boca y nariz- en consonancia con la posición de la lengua y los labios, pueden ser visualizados en tiempo real a través de la opción de representación gráfica digital que el experto estime más oportuna. El sonograma o espectrograma será en muchas ocasiones una herramienta de gran utilidad, permitiéndonos en algunos casos refrendar aquellos aspectos fonoarticulatorios constatados a nivel auditivo, y en otros, detectar la presencia de índices acústicos de interés identificativo.

El timbre no es algo cuantificable, ya sea interpretado como una realidad física sonora o como una constancia a nivel perceptivo. Por este motivo, no tiene unidades de medida como ocurre en el caso de la frecuencia o de la presión sonora. No obstante, a pesar de este inconveniente, las estructuras acústicas de resonancia serán para nosotros uno de los elementos más importantes a la hora de individualizar un acto de habla. Si importantes resultarán para el análisis acústico sonográfico, no lo serán menos para las distintas opciones de análisis y modelado relacionadas con los entornos de reconocimiento de voz automáticos o semiautomáticos.

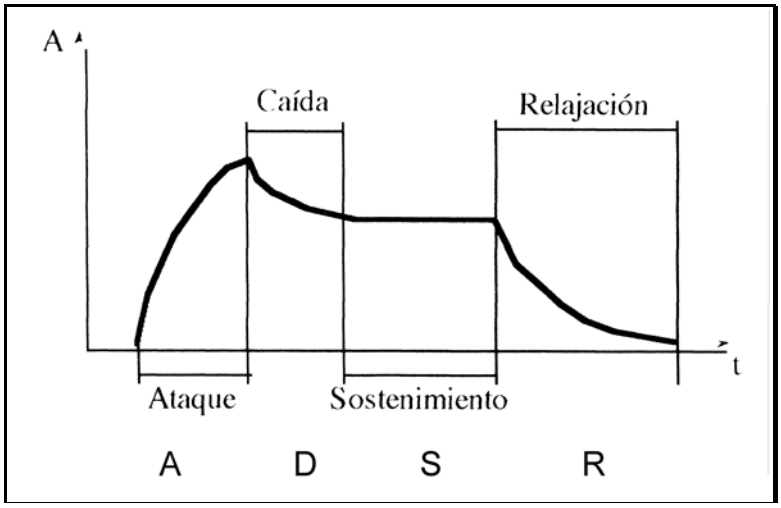
I.3.3.7.- Percepción del factor temporal: la duración.

La última dimensión de referencia perceptiva para cualquier evento sonoro es el tiempo. El tiempo físico u objetivo tiene su correlato perceptivo en el tiempo psicológico, también denominado *duración*. Velocidad de locución, ratios articulatorios, ritmo o fluctuación del tono, son algunos de los parámetros en los que interviene la duración, si bien la apreciación sistemática de los mismos no debe relativizarse a la subjetividad de los efectos perceptivos, pues existen las herramientas necesarias para una precisa estimación de los mismos.

No debemos olvidar sin embargo la existencia de una íntima relación entre las diferentes características subjetivas de los sonidos, por lo que una variación de tipo temporal puede afectar a la percepción del tono, a la sonoridad o al timbre. Realicemos un sencillo test: presentemos a distintos oyentes un tono simple de 1000 Hz en dos intervalos distintos de duración (un estímulo de 0.5 sg y otro de 2 sg). Administremos los estímulos de forma alternativa, dejando un intervalo de estimulación entre ambos de unos 15 sg. Si preguntamos a los sujetos receptores sobre la relación tonal de ambos sonidos comprobaremos, cómo sorprendentemente, algunos de ellos perciben los tonos en distintas alturas. Lógicamente, la sensación obtenida no sólo está

condicionada por el factor tiempo o las características perceptivas del receptor, sino también por la propia naturaleza del estímulo; de aquí, que idéntico test, aplicado sobre los mismos sujetos, pero utilizando diferentes estímulos frecuenciales, pueda deparar otro tipo de resultados.

Las conocidas siglas ASDR (Attack, Decay, Sustain, Release) representan la típica evolución temporal del clásico espectro vocálico: ataque, caída, sostenimiento y relajación.



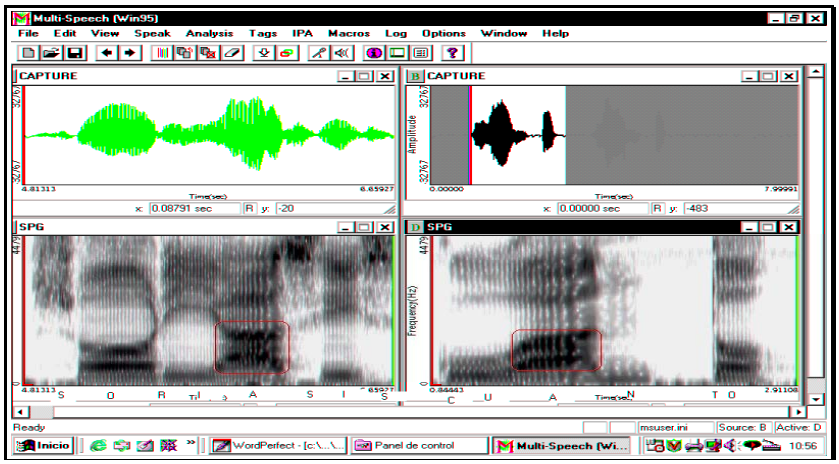
Durante el ataque y la caída inicial, el sonido experimenta una evolución temporal del espectro hasta alcanzar su fase estacionaria. En esta fase de mayor estabilidad, el sonido presenta su mayor periodicidad y su envolvente espectral está claramente definida. Sin embargo, en el intervalo de relajación el nivel de amplitud del espectro decrece progresivamente, amortiguándose la presencia de cada una de las componentes espectrales. Por su parte, la duración del intervalo de ataque puede conferir una seña acústica característica, pudiéndose alcanzar muy diferentes matices para un mismo sonido.

Por norma general, la referencia de duración de un suceso sonoro tendrá poca relevancia a la hora de aportar informaciones sobre las características individuales de una locución. Sin embargo, más adelante podremos comprobar que sí puede alcanzar cierto protagonismo en el diseño y ejecución de algunas tareas de discriminación o comparación perceptiva.

I.3.3.8.- La percepción del habla.

Existen claros indicios que parecen señalar la existencia de un código perceptual específico para los sonidos del habla. No obstante se aprecia un gran desacuerdo entre los distintos investigadores respecto de los mecanismos que permiten la percepción del habla.

En un principio, se manejó la hipótesis de que cada fonema estaba caracterizado por su propia clave acústica; hipótesis, más tarde refutada por el hecho evidente de que el patrón de frecuencias asociado a un determinado fonema siempre estará influenciado por aquellos otros que integran los fonemas adyacentes al mismo, ya sean los que le preceden o los que le siguen (efecto co-articulación). En la siguiente ilustración podemos observar la variación de la estructura acústica del primer y segundo formantes del fonema /a/ en función del contexto.



Libermann [1.967] trató de dar una explicación a la percepción del habla a través de su ya clásica *teoría motora* por la que establece un posible vínculo entre la percepción y la producción del habla. Libermann expone que cada fonema tiene un patrón de articulación exclusivo, determinado por el lugar, la forma de articulación y ciertas claves de identidad acústica, como es el caso en algunas consonantes de la duración del tiempo de iniciación vocálica (TIV). Experimentos relacionados con la denominada *percepción categorial* de estos efectos fueron desarrollados con niños observándose un comportamiento similar al de los adultos, lo que sugiere la posibilidad de que posean mecanismos para detectar los sonidos del habla.

Otros estudiosos del proceso de percepción del habla, defienden la postura de que en la propia señal acústica se halla la suficiente información para identificar los distintos fonemas, sin necesidad de la existencia de un de un decodificador fonético específico.

Eimas y Corbit [1.973] realizaron experimentos de *adaptación selectiva* para estímulos de habla. Los resultados revelaron la posible existencia de unos A*detectores de características*≅ que, en una etapa posterior, plantearon la controversia sobre si su carácter era de tipo lingüístico o acústico; es decir, si sintonizaban específicamente con estímulos del lenguaje o lo hacían con las señales acústicas en general.

En [1.976], McGurk y MacDonald descubrieron el denominado *efecto de percepción audiovisual del habla* que sustentaba la naturaleza acústica de los mencionados detectores del habla. Posteriormente, otros investigadores abogaron por la existencia de una combinación de detectores lingüísticos y acústicos.

Como podemos comprobar, no existe un criterio unánime entre los investigadores del proceso perceptivo del habla. Si además, a dicha circunstancia le unimos la del carácter activo de los centros superiores en el momento de procesar supra estructuras del habla a nivel conversacional (marcos semánticos o sintácticos), convendremos en afirmar que la gran complejidad que caracteriza al proceso de percepción del habla, no sólo se circunscribe a las meras informaciones acústicas o lingüísticas, sino también a la influencia del plano contextual.

En un capítulo posterior, citaremos de forma complementaria otra serie de circunstancias relacionadas con el proceso de percepción auditiva del habla, si bien, ya orientadas a su aplicación práctica como parte de un prototipo de comparación perceptiva en condiciones forenses.

I.3.3.9.- La influencia del paisaje sonoro en la percepción.

El término *paisaje sonoro* (soundscape) fue acuñado por el austriaco Murray Schafer [1979] para englobar en un mismo concepto el entorno acústico donde se producen los objetos sonoros que constituyen el estímulo y la relación perceptiva que en torno al mismo se establece por parte del sujeto receptor.

Hemos estimado oportuno introducir este concepto, pues el espacio auditivo además de constituir una realidad física de propagación sonora es, en sí mismo, un elemento más del proceso perceptivo de cada individuo. Y desde este punto de vista debe ser tenido en cuenta por el experto forense para alcanzar el máximo nivel de objetividad en la evaluación auditiva de los diferentes eventos sonoros. En este mismo sentido ha de considerarse lo que también Schafer denomina *"esquizofonia"* : *... a raíz de la experiencia electroacústica con la que el sonido puede ser registrado, puede producirse la esquizofonia o disociación espacio-temporal de un sonido grabado respecto de su referente original, al entrar en juego la arbitrariedad contextual*. Es decir, incluso en el caso de las emisiones grabadas, el hecho de que sean percibidas en un espacio o tiempo distintos a los originales puede conducir al receptor a distintos tipos de sensación.

Otro de los ejes a considerar a la hora de adquirir referencias sobre el espacio auditivo es la localización espacial del sonido, la cual, está directamente relacionada con la escucha binaural o utilización de ambos oídos. En los procesos perceptivos de reconocimiento de locutores no tiene una relevencia especial la consideración de esta referencia, puesto que el investigador conocerá y seleccionará la ubicación de la fuente sonora. Tan sólo habrá que prestar atención a las características dimensionales y acústicas de la sala donde se realice la tarea correspondiente para evitar posibles diferencias interaurales de tipo temporal o faltas de adecuación en la administración correcta de los estímulos. De producirse algún inconveniente de esta clase, deberá ser inmediatamente corregido.

Las teóricamente reducidas dimensiones y aislamiento acústico de las salas de trabajo en los laboratorios forenses, el uso de auriculares, etc, impedirán la presencia de sombras acústicas de intensidad (reverberaciones, ecos, etc) u otra clase de efectos no deseados para la correcta realización de un estudio de percepción auditiva. Otra serie de precauciones en este sentido, deberán ser tenidas en cuenta durante la ejecución de tareas complementarias al propio proceso de análisis identificativo: tomas de muestras indubitadas, grabaciones con micrófonos ocultos, etc. Más adelante, insistiremos en estos aspectos.

I.4.- LINGÜÍSTICA, FONOLOGÍA Y FONÉTICA

I.4.0.- Introducción

Lingüística, Fonología y Fonética deben ser consideradas como tres disciplinas fundamentales en el entorno actual de la identificación forense de locutores. Como veremos más adelante, la metodología "combinada" utilizará como referencia básica la denominada perspectiva "fonético-lingüística", la cual se revelará como uno de los sistemas de análisis imprescindibles tanto en las tareas selectivas de claves de identidad, como en las de servir de nexo asociativo Acausa/efecto≅ con las observaciones o resultados alcanzados a través de otras aproximaciones de estudio.

La lingüística general [Robins, 1.971] es el estudio científico del lenguaje como facultad universal del hombre. No se trata pues del estudio de una determinada lengua sino de los factores que rigen la comunicación humana por medio del lenguaje. La lingüística actual otorga mayor relevancia a la lengua hablada que a la escrita pues por un lado considera que ésta deriva de la hablada y, por otro, que la lengua hablada contiene muchos elementos enriquecedores - sobre todo a nivel suprasegmental o emocional - que no pueden ser apreciados a través de la lengua escrita.

Al margen de las diferentes definiciones básicas de "lenguaje" [Sapir, 1954] [Crystal, 1.968] podemos aceptar convencionalmente la concepción semiológica de Saussure [1955] o la de Hjelmslev [1.966], en la que el lenguaje es entendido como la facultad que tiene el hombre de comunicarse con los demás mediante la utilización de *signos lingüísticos*.

Para Saussure el lenguaje comprende dos aspectos: la lengua y el habla. La lengua es un producto social de la facultad del lenguaje y un conjunto de convenciones adoptadas por una comunidad para hacer posible la comunicación, es decir, un código en el que se produce una correspondencia entre imágenes acústicas y conceptos. El habla es la utilización de esa lengua por parte de los individuos, esto es, la realización concreta de la lengua en un lugar y en un momento determinados, por cada uno de los miembros de la colectividad lingüística. De esto se desprende que la lengua es un fenómeno social y abstracto, mientras que el habla es algo individual y concreto.

Más adelante podremos comprobar cómo este planteamiento dicotómico del lenguaje nos permitirá establecer una asociación entre lo que consideraremos referencias normativas de la lengua y las realizaciones concretas de la emisión de un hablante. De esta manera, estaremos en disposición de otorgar un mayor o menor peso individualizador y, por tanto identificativo, a las diferentes locuciones objeto de análisis.

I.4.1.- La dicotomía lengua/habla

La configuración lingüística de hace unos años definía como ciencia de los sonidos de la lengua (fonemas) la fonología, y como ciencia de los sonidos del habla (sonidos propiamente dichos o alófonos), la fonética. Hoy en día se vuelven a considerar materias estrechamente unidas, como formando dos aspectos de una misma disciplina.

Fonética y la fonología son las disciplinas de la lingüística que estudian los sonidos del lenguaje humano en la comunicación social. Todo acto de comunicación verbal supone la existencia de un hablante que emite sonidos a través de sus órganos de fonación en el momento de hablar y la de un oyente que los percibe por medio de su oído. Ahora bien, tanto la producción vocal de los sonidos como su interpretación están profundamente relacionadas con la actividad psíquica del ser humano. No debemos olvidar, que si bien el aparato fonador nos permite emitir sonidos generando los elementos fundamentales constitutivos del habla - tono, timbre e intensidad - realmente hablamos con nuestro cerebro, y la interpretación de los sonidos del lenguaje que llegan a nuestro oído depende también de los distintos centros superiores de codificación.

Las unidades básicas de codificación o *signos linguísticos* presentan dos aspectos fundamentales: el *significante* o expresión y el *significado* o contenido. También sabemos que el significante en el plano de la *lengua*, por ejemplo la palabra Arealidad≅, en el plano del *habla*, según el nivel sociocultural o zona de habla, puede pronunciarse de distintas maneras:

Ejemplo:

LENGUA	HABLA
/realidád/	[reali_a_] [reali_a] [reali_a□]

	[rali_á□] [,rali_á] [rjali_á]

El significante en el plano de la lengua es, como sabemos, de naturaleza psíquica, es el modelo que tenemos en nuestra mente. Cuando lo trasladamos al plano del habla, cuando pronunciamos ese modelo al hablar, cada individuo lo hace de una manera particular según sus características personales (emocionales, fisiológicas, articulatorias, etc.). Para distinguir los dos niveles del significante, en el plano de la lengua lo representamos entre barras: /b/, y en el plano del habla lo hacemos entre corchetes: [b].

I.4.2.- Fonética y fonología

La *fonética* tiene por objeto el análisis de los sonidos del lenguaje en el plano del habla. Se interesa por los sonidos emitidos por los órganos fonadores del hombre en sus realizaciones concretas, incidiendo en la manera de pronunciar un individuo o un grupo de individuos, para así poder describir las características de su pronunciación.

Los sonidos no son realizados de igual manera por todos los individuos de una misma comunidad lingüística. Por ejemplo, la pronunciación del fonema /s/ (letra As≅) de la mayoría de los andaluces e hispanoamericanos es ligeramente diferente de las personas de otras regiones de España. Los andaluces e hispanoamericanos en general pronuncian este sonido acercando la punta de la lengua a los dientes incisivos, por esto notamos que pronuncian una [s,_] llamada Adental≅; mientras que en zonas como Castilla o en otras regiones del Norte de España la lengua no se dispone tan adelantada y para la pronunciación de la /s/ se acerca más a los alveolos, por lo que se dice que es una [s] alveolar.

También es frecuente en Andalucía y otras regiones de España pronunciar el fonema /s/ como una aspiración que representamos por [h] . Así, por ejemplo, cuando pronuncian la palabra Amosca≅ se oye [móhka].

Lo mismo ocurre con la pronunciación del fonema /□/ (letra All≅).Muchas personas la pronuncian como [j] en [kaβájo] , otras incluso como un sonido vocálico [i] en [kaβáio] , los argentinos la pronuncian de una manera especial [kaβá∞o] , etc.

Otro ejemplo mucho más claro y que podemos observar en nosotros mismos se produce cuando pronunciamos la [n]. Si nos fijamos bien iremos observando cómo en los distintos ejemplos que siguen a continuación tocamos con la lengua en diferentes áreas del techo bucal.

En la palabra AAna≅ /ána/, tocamos con la punta de la lengua en los alveolos. En Aanda≅ [an,_da], cuando estamos pronunciando la An≅ ya no estamos tocando en los alveolos, sino en los dientes.

En Aanca≅ [á‿ka],estamos tocando para pronunciar la An≅ con la parte posterior de la lengua contra el velo del paladar.

Incluso cuando decimos Aánfora≅[á□fora] , normalmente no tocamos con la lengua en ningún lugar determinado y solamente el labio inferior se aproxima a los incisivos superiores.

Como podemos ver la fonética se ocupa de ir describiendo detalladamente las diferentes maneras de pronunciar los sonidos del significante en el plano del habla. Sin embargo, aunque la *fonología* se ocupa también de los sonidos del significante, lo hace en lo referente a su valor funcional dentro de la lengua.

Por ejemplo, habíamos visto cómo la fonética describe detalladamente las maneras más comunes de pronunciar la [s] en español, los distintos tipos de sonidos que tiene este elemento fónico del significante. Sin embargo,la fonología no entra en ese tipo de detalles, lo único de lo que se preocupa es que en la lengua española hay una unidad /s/ que es diferente de otras unidades de la lengua, como son la /n/ , /p/ , etc. Su objeto de interés es saber que hay una consonante tipo, un modelo /s/ en la lengua y no las diferentes pronunciaciones de ese modelo en cada una de las personas que lo utilizan a la hora de hablar.

Esas diferencias de pronunciación del significante en el habla no conllevan un cambio de significado en la palabra. En una palabra como Acasa≅, cuyo significante en la lengua sería /kása/ y su realización en el habla puede tener diferentes soluciones :[kása] , [kás,_a], etc , las diferentes maneras de pronunciar la [s] no influyen en absoluto para que la persona que oye siga entendiendo la misma idea o significado del significante /kása/ ; pero si en lugar de pronunciar una [s] pronunciamos una [p] tendríamos otro significante /kápa/ que tendría un significado muy diferente. Y lo mismo pasaría si pronunciásemos una [n] en lugar de una [s] o [p] , tendríamos /kána/.

Como podemos apreciar, la fonética y fonología estudian las unidades del significante desde puntos de vista diferentes: la fonética estudia las variantes de las unidades en el habla que

no tienen carácter funcional y que no cambian el valor significativo del significante. A estas variantes de las unidades fónicas se les llama *sonidos o alófonos*. La fonología estudia esas unidades en la lengua en cuanto tienen un valor funcional y significativo. Las unidades, o modelos tipo con los que opera la fonología se llaman *fonemas*.

La metodología que debemos aplicar en cualquier estudio actual de fonética - incluída nuestra aplicación específica forense - ha de considerar en primer lugar lo fonológico y en segundo lugar lo fonético, con su doble perspectiva fisiológica y acústica.

I.4.3.- Los rasgos acústicos

El método empleado por la fonología para proceder a la identificación de los fonemas de una lengua es el de crear contrastes u oposiciones significativas por medio de la conmutación o sustitución de cada una de las unidades fónicas de una palabra por otras.

Cada fonema de una lengua posee una serie de rasgos, algunos de los cuales coinciden con los de otros fonemas *(rasgos redundantes*) y, otros que sin embargo crean una relación de contraste con ellos. Los rasgos que determinan la *oposición* entre dos fonemas se llaman *rasgos pertinentes* (también se les llama distintivos o funcionales).

Desde un punto de vista fisiológico, cada sonido del español suele caracterizarse en razón a cuatro referencias: *lugar de articulación, modo de articulación, contraste sonoro-sordo, y contraste oral-nasal*. En nuestra lengua sólo tenemos tres fonemas nasales, el resto de sonidos son normalmente orales porque durante su pronunciación el aire emitido sale exclusivamente por la cavidad bucal. Por ello, para acotar los distintos fonemas suele obviarse la ubicación nasal/oral.

Por ejemplo, para reconocer si en español /p/ y /t/ son fonemas diferentes nos basta con oponer: /poko/ Apoco≅ y /toko/ Atoco≅. Estos dos significantes se distinguen solamente por el contraste existente entre /p/ y /t/.

- /p/: es una consonante:

1) *oclusiva*, porque se establece un contacto completo de dos órganos articulatorios que impiden la salida del aire fonador.

2) es *bilabial*, porque para pronunciarla juntamos ambos labios.

3) es *sorda*, porque no vibran las cuerdas vocales durante su emisión.

- /t/: es una consonante:

1) *oclusiva*, como la anterior.

2) es *linguodental*, porque para pronunciarla el ápice de la lengua toca con los dientes incisivos superiores.

3) es *sorda*, como la anterior.

Es decir /p/ y /t/ coinciden entre ellos en que tienen dos rasgos comunes o redundantes: 11 y 31, y solamente se diferencian por el 21 rasgo, ya que /p/ es bilabial y /t/ es linguodental, este rasgo es el que crea la oposición, y por tanto es su rasgo distintivo. Lo mismo pasa con otros dos fonemas /r/ y /r,_/:

/r,_/ es linguodental, vibrante, *múltiple* (varias vibraciones)

/r/ es: linguodental, vibrante, *simple* (una sola vibración).

El contraste entre estos dos fonemas lo encontramos en que el primero es múltiple y el segundo es simple. Ejs.:

/kar,_o/	Acarro≅
/karo/	Acaro≅

Pero ocurre, que en determinadas posiciones dos fonemas de la lengua pierden su función distintiva. Decimos entonces que se neutralizan. Hemos visto como, en español /r, / y /r,_/ son dos fonemas y crean un contraste significativo en posición intervocálica. Pero en posición inicial de palabra o final de sílaba el hecho de pronunciar [r,] o [r,_] no crea un contraste de significación en la palabra. Por ejemplo pronunciar [amór] o pronunciar [amór] no suponen significados distintos.

Otro ejemplo, en español existen tres fonemas nasales: /m/, /n/ y /□/ que crean contrastes significativos en posición inicial de palabra o en posición intervocálica:

/Kama/	Acama≅
/kána/	Acana≅
/ká□a/	Acaña≅

Pero en posición final de sílaba estos fonemas pierden sus rasgos distintivos y se neutralizan.

En la transición fonológica, cuando dos fonemas se neutralizan se representan por un *archifonema,* el cual se representa por medio de un signo diacrítico que es una letra mayúscula que refleja el rasgo común a los fonemas neutralizados. Así, por ejemplo:

/amáR/	Aamar≅
/rodáR/	Arodar≅
/áNdaN/	Aandan≅

El estudio de los rasgos distintivos desde el esquema binarista de Jakobson [Jakobson et al., 1952] resulta de extraordinaria importancia para cualquier estudio fonológico o fonético, incluida la perspectiva forense. En síntesis, el principio binario establece comparaciones entre unidades fónicas basándose en la presencia o ausencia de un determinado rasgo distintivo, para otorgar una determinada cualidad a las mismas.

Según el planteamiento de Jakobson y sus colaboradores [1952] los rasgos distintivos pueden ser de dos clases: *rasgos prosódicos* y *rasgos intrínsecos o inherentes*.

Los rasgos prosódicos o suprasegmentales sólo los presentan aquellos fonemas que constituyen el núcleo silábico, y se formulan a través de referencias intersilábicas e intrasilábicas. Los rasgos prosódicos pueden ser de tres tipos:

- Rasgos prosódicos *de tono*: desde el enfoque intersilábico (entre distintos núcleos silábicos de una misma secuencia) existen dos rasgos generales de tono: *alto o bajo*. Desde el punto de vista intrasilábico, a través del rasgo de *modulación* se efectúa el contraste entre el tono bajo de una parte de un fonema y el tono alto de la parte o fonema siguiente, o viceversa.

- Rasgos prosódicos de *fuerza*: a nivel intersilábico el *acento dinámico* efectúa el contraste de intensidad entre núcleos silábicos. Desde el enfoque intrasilábico se comparan dos partes contiguas del fonema acentuado. Sólo se presenta en la lengua danesa y el rasgo se denomina A*stosson*≅.

- Rasgos prosódicos de *cantidad*: desde la perspectiva intersilábica se contrasta un fonema de duración breve con otros fonemas de mayor duración. Intrasilábicamente, el rasgo denominado de *unión* coteja las diferentes duraciones de una vocal y la consonante siguiente.

Los rasgos distintivos intrínsecos o inherentes se denominan así porque inciden directamente en la propia naturaleza del fonema. A diferencia de lo que ocurre con los

prosódicos, su ausencia o presencia pueden determinar la propia cualidad de dicho fonema.

Los rasgos intrínsecos pueden ser de dos clases: de *sonoridad* y de *tonalidad*.

Los de *sonoridad* están asociados a los prosódicos de cantidad e intensidad. Se clasifican en :

- *vocálico/no vocálico:* el rasgo vocálico se corresponde con la presencia de una energía armónica que es generada a nivel glotal y potenciada en el resonador del tracto en algunos de sus componentes. A nivel acústico se manifiesta en las estructuras formánticas. Articulatoriamente se caracteriza por una ausencia de obstáculos en las cavidades supraglóticas al paso del aire de la fonación.

- *Consonántico/no consonántico*: el rasgo consonántico, en contraposición al vocálico, se corresponde con la ausencia de la energía periódica producida en los pliegues glotales y de las consiguientes resonancias. A nivel articulatorio, este rasgo se caracteriza por la presencia de algún obstáculo en las cavidades supraglóticas. Lógicamente el rasgo vocálico lo poseen las vocales y el consonántico las consonantes. Las consonantes líquidas poseen tanto rasgos vocálicos como consonánticos.

- *Compacto/difuso :* la caracterización de estos rasgos es más significativa desde el punto de vista acústico. En el caso de un rasgo compacto encontraremos una concentración de más elevada energía en una zona central de su área espectral, así como un aumento ponderado de la energía global y de su expansión en el dominio temporal. Sin embargo, en el caso del rasgo difuso tanto la concentración relativa de energía como la global y su expansión temporal son bastante más reducidas.

- *Tenso/laxo*: el rasgo tenso a nivel articulatorio implica una necesaria tensión general de los distintos elementos del tracto vocal en relación a lo que es su posición de reposo. Por este motivo, toda emisión producida con pereza articulatoria nunca comprenderá elementos fónicos tensos, resintiéndose en su propia naturaleza aquellos que más requieran de dicha tensión para realizarse como tales :oclusivos sordos, vibrante múltiple, ciertos fricativos, etc A nivel acústico, y en términos generales, el rasgo tenso conlleva además de un incremento en el nivel global de energía y en su extensión temporal, una mayor definición de los índices acústicos ya sean éstos resonancias armónicas, barras de explosión, estridencias, etc.

- *Sonoro/sordo*: el rasgo de sonoridad está en relación directa a la vibración de los pliegues glotales durante la emisión hablada. Si no hay vibración no hay rasgo de sonoridad. Acústicamente, el rasgo sonoro se relaciona con la presencia de una estructura formántica de bajo nivel asociada al fundamental. Esta estructura armónica llamada barra de sonoridad, en

algunas ocasiones tiende a confundirse con otra estructura -*hum*- ubicada en similar rango de frecuencia producida por el armónico de corriente alterna (50 ó 60 Hz.) y sus correspondientes armónicos y subarmónicos de resonancia.

- *Nasal/oral* : el rasgo de nasalidad aparece al ampliarse la cavidad de resonancia supra glótica en su zona posterior, debido a un descenso del velo del paladar. A nivel acústico, se manifiesta con una reducción de la intensidad de los primeros formantes vocálicos y la aparición de áreas formánticas a determinado rango de frecuencia en el caso de los sonidos consonánticos. Justamente se produce el efecto contrario cuando acontece el rasgo de oralidad.

- *Interrupto/continuo* : el rasgo interrupto se caracteriza a nivel acústico por un intervalo de ausencia de información a nivel frecuencial en el dominio temporal (silencio generalmente no superior a los 30 ms). Dicha ausencia ha de producirse en el continuum de la emisión ya iniciada y no concluida, y ha de estar flanqueada por relevantes distribuciones de energía en un amplio rango frecuencial. Lógicamente, el rasgo continuo se caracteriza por no poseer las citadas propiedades.

- *Estridente/mate* : estos rasgos conciernen exclusivamente a las consonantes. Desde un punto de vista acústico, el oscilograma y sonograma de un rasgo estridente se caracterizan respectivamente por su alto grado de aperiodicidad y por tanto una no uniformidad en la distribución de la energía. Sin embargo, el rasgo mate implica cierta uniformidad a nivel sonográfico (en algún caso aparecen distribuciones pseudoformánticas) y cierta recurrencia aleatoria de determinados períodos en el oscilograma. Articulatoriamente, el rasgo estridente es asociado a un mayor grado de energía y constricción durante la emisión [Quilis, 1981].

- *Bloqueado/no bloqueado* : dicotomía conocida también como glotalizado/no glotalizado. Desde un enfoque acústico, el rasgo glotalizado implica una significativa descarga de energía en un intervalo temporal reducido. A nivel articulatorio dicho rasgo se produce por una oclusión rápida y momentánea de la válvula glotal.

Los *rasgos de tonalidad* están asociados a los rasgos prosódicos basados en el tono. Son los siguientes:

- *grave/agudo* : desde una perspectiva acústica, el rasgo grave está asociado a una preponderancia de las frecuencias bajas del espectro, mientras que el agudo lo está a un predominio de las altas. Articulatoriamente, el rasgo grave se manifestará por la existencia de cavidades de resonancia amplias y no divididas en la realización del fonema. Los fonemas con el rasgo agudo se originan a partir de cavidades de resonancia pequeñas y divididas.

- *Bemolizado/no bemolizado* : articulatoriamente, el rasgo bemolizado conlleva una reducción del orificio anterior o posterior del resonador bucal acompañada de una velarización del mismo, que a su vez lo dilata. La consecuencia acústica puede ser un descenso frecuencial de todos o algunos de los campos de dispersión de las estructuras formánticas. Según Quilis, [1981] el rasgo de no bemolización se manifiesta por el efecto contrario. Jakobson [1957] y Halle [1957] han incluido dentro de estos rasgos los de retroflexión, velarización, faringalización, labialización y redondeamiento.

-*Sostenido/no sostenido :* Cuando se dilata el orificio faríngeo y simultáneamente se produce una palatalización que reduce y divide la cavidad central del resonador bucal, nos encontramos ante la realización del rasgo sostenido. La consecuencia acústica de dicho rasgo se traduce en una elevación de los F2 y una potenciación de ciertos armónicos del habla en alta frecuencia.

I.4.4.- Los alfabetos ortográfico, fonético y fonológico

Como todos conocemos, el *alfabeto ortográfico* de una lengua es el que utiliza letras que son la representación escrita o la imagen visual del significante en el lenguaje escrito. La ortografía de una lengua ha impuesto desde hace mucho tiempo unas reglas aceptadas por la comunidad de hablantes para escribir las palabras de una determinada manera. Pero en el español que escribimos hoy existen muchas letras, como Ah≅, que en siglos anteriores tenían un sonido determinado y correspondían a un fonema. Actualmente esta letra ni se pronuncia ni equivale a fonema alguno, porque con el transcurso del tiempo ha dejado de tener valor en el lenguaje oral. Si nosotros escribimos Ahecho≅ sabemos que corresponde al verbo hacer, y si escribimos Aecho≅ es del verbo echar. Pero cuando pronunciamos cualquiera de las dos, en realidad pronunciamos lo mismo aunque se escriban de manera diferente.

El *alfabeto fonético* ha surgido de la necesidad por parte de los lingüistas de estudiar las diferentes maneras de pronunciar que desarrollan las personas cuando hablan. Naturalmente hemos visto que hay muchas maneras de pronunciar y que nunca se producen dos sonidos exactamente iguales. Teóricamente existen tantas maneras de hablar como personas se expresan en una lengua e incluso el número de alófonos que puede emitir una misma persona es infinito. Precisamente, la imposibilidad de que un emisor produzca dos actos de habla idénticos es uno de los factores que confiere una especial dificultad a las tareas de identificación del locutor.

El alfabeto ortográfico de la lengua escrita es realmente muy reducido para poder reflejar

una serie de sonidos o alófonos - como por ejemplo los distintos tipos de [s]: [s], [s,,], etc., o los distintos tipos de [n]: [n], [n,,], [ɲ], etc. - que habitualmente son utilizados al hablar y que resultan muy importantes para el lingüista de cara a estudiar la pronunciación concreta de una persona, de un grupo social, de una región, etc. Para suplir esa necesidad se creó un alfabeto fonético con gran riqueza de signos que trata de reflejar las posibles variantes que se dan en la pronunciación de una lengua. La Asociación Fonética Internacional utiliza el denominado Alfabeto Fonético Internacional (IPA) -consensuado por lingüistas de distintos países - que es de gran funcionalidad para diversos fines, entre los cuales, podemos citar la materialización documental de ciertas tareas perceptivas en el análisis forense del habla.

También podemos reseñar el *alfabeto fonológico*, utilizado para representar las unidades abstractas con las que identificamos los sonidos en nuestra mente. Debemos recordar que esas unidades a las que llamamos *fonemas* en realidad no se pronuncian, sino que es el modelo o Aunidad ideal≅ con la que identificamos sonidos parecidos.

I.4.5.- Grafías. Fonemas. Alófonos.

En la lengua española tenemos cinco fonemas vocálicos /i/, /e/, /a/, /o/, /u/. Cada una de esas vocales las pronunciamos realizando sonidos y las escribimos por medio de letras o grafías.

Pero no debemos confundir los fonemas con los *sonidos* ni con las *letras*. Muchas veces un fonema está representado en la escritura por una letra: /f/ = Af≅; /a/ = Aa≅; pero otras veces un fonema está representado por dos letras diferentes /b/ = Ab≅ o Av≅ (Atuvo≅, del verbo tener, y Atubo≅, pieza hueca, se pronuncian igual). Lo mismo ocurre con el fonema /x/ =≅g≅ o Aj≅: podemos escribir Agaraje≅ o Agarage≅, y sin embargo la pronunciación de estas dos letras diferentes es la misma y equivale al mismo fonema. En otros casos, la misma letra equivale a dos fonemas diferentes Ag≅ = /g/ o /x/. En Agato≅ /gáto/ es totalmente diferente de Agesto≅ /xésto/.

También puede ocurrir que dos letras correspondan a un solo fonema: Aqueso≅ /késo/, porque Aq≅ + Au≅ = /k/.

En la palabra Ajorge≅ solamente hay cuatro fonemas y cinco letras diferentes, porque Ag≅ y Aj≅ equivalen al mismo fonema /xóRxe/. En la palabra Ahago≅ hay cuatro letras y solamente tres fonemas /ágo/, ya que la Ah≅ no tiene ningún valor en la comunicación oral.

En Acace≅ las letras diferentes son tres, pero los fonemas son cuatro, ya que la letra Ac≅

equivale a dos fonemas /k/ y /□/, por lo que fonológicamente representaríamos esta palabra /ka□é/.

En el siguiente cuadro vemos algunos ejemplos de los contrastes que existen entre los diferentes niveles.

FONEMAS (Lengua)	SONIDOS (Habla)	LETRAS (Escritura)
/i/	[i,˜] cerrada [i] normal [i,º] abierta, etc.	Ai≅: pipa Ay≅: ley
/b/	[b] oclusiva: [bár] [β] fricativa: [elβár]	Ab≅: tubo Av≅: tuvo
/□/	[□]: [□íne]	Ac≅ (ante e, i): cena, cine Az≅: (ante a,o,u): caza, zona azúcar.
/x/	[x] velar: [páxa] [h] aspirada: [páha]	Ag≅ (ante e, i): gente Aj≅ (ante i,e,a,o,u): jefe, jarra
/r/	[r]: [pára]	Ar≅ (entre vocales): para
/r,_/	[r,_] vibrante múltiple	Arr≅ (entre vocales): parra Ar≅ (inicial de palabra o precedida de l,n.s): rama, honra, alrededor, Israel.

El *fonema* puede definirse como la unidad mínima distintiva. Toda lengua tiene unas unidades mínimas convencionales que combinadas en virtud de unas determinadas reglas, son capaces de conformar cualquier mensaje dentro de esa lengua; así pues el fonema es abstracto, convencional y, por lo tanto, social.

Sin embargo, el *alófono* es la realización sonora del fonema y por lo tanto algo concreto e individual. Tal realización, debe ser desarrollada dentro de los límites convencionales que corresponden al fonema ya que de lo contrario se rompería el sistema; ejemplo: la realización del fonema /a/, puede ser más o menos cerrada pero siempre dentro de lo que llamamos su *campo de dispersión*, ya que si lo sobrepasamos pronunciaríamos una realización correspondiente al umbral acústico y articulatorio de otro fonema, y por lo tanto el interlocutor podría estar interpretando, por ejemplo, una Ae≅ en lugar de una Aa≅.

Como ha quedado reflejado, un fonema no se realiza siempre de la misma manera; sus diferentes realizaciones o alófonos pueden depender del plano de expresividad y/o del contorno fonético en el que están situados.

Estas diferentes variaciones pueden ser:

- *Combinatorias*, ejemplo el fonema /b/ tiene en español dos realizaciones [b] y [ß]: la primera se da después de pausa y de consonante nasal y la segunda en los demás casos.

- *Libres*, llamadas también estilísticas, ya que resultan de la elección más o menos consciente del hablante, ejemplo, el fonema /s/ frecuentemente en el español hablado en Madrid, y en posición postnuclear puede realizarse como [s], [h] o [X], es decir, ápico-alveolar, aspirada o velar (por ejemplo en la palabra Amosca≅).

- *Individuales*, aquellas que pueden dar indicaciones peculiares sobre un determinado hablante, pero que no son el resultado de una elección por parte del mismo, ejemplo: el yeismo más o menos africado y ensordecido de algunos hablantes madrileños.

I.4.6.- Fonemas y alófonos del Español.

Como todos sabemos, nuestra lengua tiene veinticuatro fonemas o segmentos mínimos distintivos con los que se elabora cualquier mensaje. De estos fonemas, cinco son vocálicos y los diecinueve restantes son consonánticos. Aunque posteriormente abordaremos todos ellos en detalle, haremos a continuación una breve descripción de los mismos desde una perspectiva de ubicación general.

FONEMAS VOCÁLICOS. Desde un punto de vista articulatorio, se definen fundamentalmente en función del lugar y modo de articulación. El lugar, es la zona de la cavidad

bucal donde actúan los órganos articulatorios: dientes, alveolos, labios, paladar, etc; y el modo es la manera o forma en que los órganos articulatorios se oponen a la corriente fónica, formando distintas cavidades de resonancia que caracterizan la articulación de cada uno de los sonidos.

En las vocales, su ubicación articulatoria se concreta en relación a la zona de la boca en que se articulan (el orden antero-posterior es: /i/, /e/, /a/, /o/ y /u/) y la abertura de la cavidad bucal (la más abierta es /a/, le siguen /e/, /o/, siendo las más cerradas /i/, /u/.

De acuerdo al lugar de articulación son denominadas: anteriores o palatales /i/, /e/; central /a/ y posteriores o velares /o/ y /u/. Por el modo de articulación son: altas /i/ /u/; medias /e/ /o/ y baja /a/.

FONEMAS CONSONÁNTICOS. Fundamentalmente son definidos en función del lugar de articulación, el modo y la acción de las cuerdas vocales.

Por el lugar:

Bilabiales:	/p/, /b/ y /m/.
Labiodentales:	/f/
Interdentales:	/□/
Dentales:	/t/, /d/
Alveolares:	/r/, /r,_/, /s/, /l/, /n/
Palatales:	/±/, /j, /, /□/, /□/
Velares:	/k/, /g/, /x/

Por el modo:

Oclusivas:	/p/, /t/, /k/
	/b/, /d/, /g/
Fricativas:	/f/, /□/, /x/, /s/, /j, /
Africadas:	/_,_/
Nasales:	/m/, /n/, /□/
Líquidas:	
- Laterales:	/l/, /□/

- Vibrantes: /r/, /r,_/

Por la acción de las cuerdas vocales:

Sordas (no vibran): /p/, /t/, /k/
Sonoras (vibran las cuerdas vocales): todas las demás

En el Español los fonemas se corresponden con las grafías o letras, salvo en los siguientes casos:

Grafía v corresponde al fonema /b/
Grafía b corresponde al fonema /b/
Grafía g + /a, o, u/ corresponde al fonema /g/
Grafía gu + /e, i/ corresponde al fonema /g/
Grafías c + /e, i/ corresponden al fonema /□/
Grafía z, en todos los casos, corresponde al fonema /□/
Grafía h, no se corresponde con ningún fonema
Grafías qu + /e, i/ corresponde al fonema /k/
Grafía c + /a, o, u/ corresponde al fonema /k/
Grafías g + /e, i/ corresponde al fonema /x/
Grafía j, en todos los casos corresponde a /x/
Grafía x corresponde al fonema /s/

ARCHIFONEMA.-Como ya hemos comentado, en algunas ocasiones - cuando van en posición postnuclear- ciertos fonemas se neutralizan constituyendo lo que llamamos archifonema.

Los fonemas /p/, /b/ en situación implosiva o postnuclear se neutralizan en el archifonema B. Ejemplo aptitud = /aBtitud/.

Los fonemas /t/, /d/, en situación postnuclear se neutralizan en el archifonema D. Ejemplo: atletismo = /aDletismo/.

Los fonemas /k/, /g/, en situación postnuclear, se neutralizan en el archifonema G.

Ejemplo: actitud = /aGtitud/.

Los fonemas /m/, /n/, en situación postnuclear, se neutralizan en el archifonema N. Ejemplo: empleado = /eNpleado/.

Y por último, los fonemas /r/, /r,_/, en situación postnuclear se neutralizan en el archifonema R. Ejemplo: cortar = /coRtaR/.

PRINCIPALES ALÓFONOS DE LOS FONEMAS DEL CASTELLANO:

El fonema /a/ tiene un sólo alófono [a]
El fonema /e/ tiene un sólo alófono [e]
El fonema /i/ tiene tres alófonos [i], [ψ], [i,º]
El fonema /o/ tiene un sólo alófono [o]
El fonema /u/ tiene tres alófonos [u], [w], [u,º]

El fonema /i/, se realiza como alófono semivocálico [i,º] en diptongo decreciente. Ejemplo: vaina = [bai,ºna] y como semiconsonántico en diptongos crecientes. Ejemplo: bien = [bψén].

El fonema /u/ se realiza como alófono semivocálico también en diptongo decreciente. Ejemplo: eutanasia = [eu,ºtanásia] y como consonántico en diptongos crecientes. Ejemplo: bueno = [bwéno].

Todos los fonemas vocálicos se realizan como alófonos oronasales [a,], [e,], [i,], [o,] [u,] cuando van precedidos de pausa y seguidos de consonante nasal, y cuando van entre consonantes nasales.

Hay que tener en cuenta que dentro de esos alófonos hay múltiples variantes en distribución libre, dependiendo únicamente de cada hablante y situación (grados de mayor o menor abertura o cierre, adelantamiento o retraso, etc.).

Los principales alófonos de los fonemas consonánticos son los siguientes:

El fonema /p/ tiene un sólo alófono: [p]
El fonema /t/ tiene un sólo alófono: [t]
El fonema /k/ tiene un sólo alófono: [k]
El fonema /b/ tiene dos alófonos: [b] y [β]
El fonema /d/ tiene dos alófonos: [d] y [_]
El fonema /g/ tiene dos alófonos: [g] y [π]
El fonema /f/ tiene un sólo alófono: [f]
El fonema /□/ tiene un sólo alófono: [□]
El fonema /x/ tiene un sólo alófono: [x]
El fonema /s/ tiene tres alófonos: [s], [s,,] y [s,_]
El fonema /j/ tiene dos alófonos: [j,], [→]
El fonema africado /_,_/ tiene un sólo alófono: [_,_]
El fonema nasal /m/ tiene un sólo alófono: [m]
El fonema nasal /n/ tiene los siguientes alófonos: [n], [m,,], [n,.], [n,,], [J], [n1]
El fonema nasal /□/ tiene un sólo alófono: [□]
El fonema lateral /l/ tiene los siguientes alófonos: [l], [l,.], [l,,], [l,,]
El fonema lateral /□/ tiene un sólo alófono: [□]
El fonema vibrante /r/ tiene un sólo alófono: [r]
El fonema vibrante múltiple /r,_/ tiene un sólo alófono: [r,_].

Como hemos visto, cada uno de los 24 fonemas del español puede manifestarse en la pronunciación con uno o varios sonidos en el lenguaje hablado, mientras que en el lenguaje escrito puede estar representado por una letra, dos, tres, o también una letra puede equivaler a uno o a varios fonemas e incluso a ninguno. Reseñemos algunos ejemplos:

Vocales

FONEMAS	Ejemplos	ALÓFONOS	Ejemplos	LETRAS	Ejemplo
/i/	/pipa/	[i]	[pipa]		pipa
	/pié/	[j]	[pjé]	i,y	pie
	/léi,º/	[i,º]	[léi,º]		ley
/e/	/péso/	[e]	[péso]	e	peso
/a/	/pása/	[a]	[pása]	a	pasa

/o/	/dós/	[o]	[dós]	o	dos
/u/	/dúro/ /kuátro] /káusa/	[u] [w] [u,º]	[dúro] [kwátro] [káu,ºsa]	u	duro cuatro causa

Consonantes

FONEMAS	Ejemplos	ALÓFONOS	Ejemplos	LETRAS	Ejemplos
/p/	/pípa/	[p]	[pipa]	p	pipa
/b/	/báka/ /túbo/	[b] [β]	[báka] [tuβo]	b,v b,v	baca,vaca tubo,tuvo
/t/	/páta/	[t]	[páta]	t	pata
/d/	/dós/ /lados/	[d] [_]	[dos] [la_os]	d	dos lados
/k/	/kása/ /késo/ /kiósko/	[k]	[kása] [késo] [kjósko]	c qu k,qu	casa queso kiosko, quiosco, kiosco
/g/	/gás/ /págas/ /pagé/	[g] [y]	[gás] [páyas] [payé]	g, gu	gas pagas pagué
/f/	/gáfas/	[f]	[gáfas]	f	gafas
/□/	/la□o/ /□íne/	[□]	[la□o] [□íne]	z,c	lazo cine
/s/	/kása/	[s] [s,,]	[kása] [kas,,]	s	casa

/j, /	/maj, o/	[j,]	[maj, o]	y	mayo
	/j, o/	[J,_]	[J,_o]		yo
/x/	/xéfe/	[x]	[xéfe]	g, j	jefe
	/xénio/		[xénjo]		genio
/c/	/kóce/	[c]	[kóce]	ch	coche
/m/	/Kama/	[m]	[Kama]	m	cama
/n/	/áNda/	[n,,]	[anda]	n	anda
	/kána/	[n]	[kána]		cana
	/áNka/	[ɲ]	[áɲka]		anca
/□/	/ká□a/	[□]	[ká□a]	ñ	caña
/l/	/ála/	[l]	[ála]	l	ala
/□/	/ká□e/	[□]	[ká□e]	ll	calle
/r/	/péra/	[r]	[péra]	r	pera
/r,_/	/pér, a/	[r,_]	[pér, a]	rr,r	perra
	/r, áma/		[r, áma]		rama

I.4.7.- Fonética acústica y fonética articulatoria.

Como ocurre en otras disciplinas, los estudios de fonética y fonología pueden agruparse y definirse en función de su objeto de estudio, de las perspectivas de enfoque desde las cuales se analiza dicho objeto de estudio, del método de análisis utilizado, etc.. Considerando que desde el punto de vista de la identificación de locutores, el interés prioritario se situará en torno a la localización de claves individualizadoras de los registros de habla, las aproximaciones de estudio que nos resultarán más funcionales serán, por una parte, la *fonética articulatoria o genética* - que estudia el modo de articularse o producirse los sonidos en el aparato fonador del hombre - y, por otra, la *fonética acústica*, que analiza las cualidades acústicas de los sonidos del habla. Mediante la combinación de ambas perspectivas, y el uso de las diversas aplicaciones de análisis

disponibles en la actualidad, podremos obtener estimaciones de altísima precisión que pondrán de manifiesto la auténtica dimensión de las diferentes características fonoarticulatorias en cada locutor.

Existen distintos manuales de fonética y fonología [Navarro Tomás 1961], [Canfield, 1962],[Malmberg, 1965], [Alarcos 1974], [Quilis 1963, 1975, 1981] donde son analizadas de forma exhaustiva la fonética articulatoria y acústica de las realizaciones de habla castellana. Por este motivo, nos limitaremos a extractar algunas de las características, rasgos y circunstancias descritas en ellos, especialmente, aquellas que nos ayudarán a detectar el carácter más o menos peculiar de una locución determinada. Merece especial mención el manual que consideramos buque insignia: la "Fonética acústica de la lengua española" de Antonio Quilis [1981], donde quedan recogidas las referencias acústicas más normativas de nuestra lengua, en perfecta asociación a su correspondiente correlato articulatorio.

La fonética acústica se ocupa del estudio de la onda sonora compleja del habla. Estudia su naturaleza y su trascendencia lingüística considerando estos fenómenos acústicos como el resultado de un proceso articulatorio de los órganos fonadores del emisor. En los últimos años, los trabajos sobre fonética y fonología parten esencialmente de criterios acústicos para establecer las descripciones de un sistema de lengua en el que los distintos sonidos se oponen o contrastan por la presencia o ausencia de determinados rasgos.

Como ya hemos referido, los grandes avances de la electroacústica moderna permiten estudiar con gran objetividad y precisión el sonido linguístico, dimensionando sus parámetros fundamentales a través de distintas formas de representación gráfica :oscilograma, sonograma, espectro, etc. Una aportación que puede considerarse fundamental en este sentido, viene representada por la incorporación del sonógrafo, que tiene como misión la descomposición automática de la onda sonora compleja en cada uno de sus componentes integrantes o armónicos permitiendo así un estudio pormenorizado de diferentes datos, desde distintas ópticas. Lógicamente, el sonógrafo -hoy en día digital- pude abarcar el análisis de toda la banda de frecuencias perceptibles de audio (es decir de 20 Hz hasta 20 KHz aproximadamente), quedando incluidos la totalidad de los sonidos lingüísticos.

Precisamente una de las primeras consecuencias de la incorporación de estas nuevas herramientas de análisis permitió a R. Jakobson, G. Fant y M. Halle (Preliminaries to Speech Analysis, 1952) establecer una clasificación de los sonidos del lenguaje, basándose en su estructura acústica.

Como nuestro principal objeto de estudio serán emisiones habladas en lengua castellana,

efectuaremos un amplio repaso de las características articulatorias y acústicas que definen cada una de las distintas estructuras de nuestra lengua para así poder detectar los rasgos de identidad que caracterizarán las diferentes locuciones analizadas, desde sus unidades más básicas, hasta los rasgos suprasegmentales.

I.4.7.1.- Los sonidos vocálicos (vocales).

Como afirma Quilis [1963] los sonidos vocálicos son aquellos que presentan la mayor abertura de los órganos articulatorios; el mayor número de vibraciones de las cuerdas vocales y, por tanto, el máximo de armónicos y mayor musicalidad. Además, en español, el sonido vocálico es el único capaz de constituir núcleo silábico (más adelante abordaremos la estructura de la sílaba).

Ya hemos comentado que los fonemas vocálicos se clasifican básicamente por el lugar y el modo de articulación.

En razón del lugar de articulación se clasifican en : *anteriores, centrales y posteriores*. Cuando la lengua ocupa una posición articulatoria en la región delantera de la cavidad bucal, se originan las vocales de la serie anterior o palatales: /i/, /e/.

Si es el postdorso de la lengua el que se acerca a la región posterior de la cavidad bucal, esto es, al velo del paladar o paladar blanco, se producen las vocales posteriores o velares: /o/, /u/.

Por último, cuando el dorso de la lengua se encuentra en una región cubierta por el medio paladar, se originan las vocales centrales, en el español /a/.

Por el modo de articulación se clasifican en: *altas, medias y bajas*. Si la lengua se aproxima hasta un máximo permisible para la articulación vocálica, bien al paladar duro o al paladar blando, se producen las vocales altas: /i/, /u/. Si la lengua se separa más de la bóveda de la cavidad bucal, se originan las llamadas vocales medias: /e/, /o/. Cuando la lengua se separa todavía más de la bóveda palatal y ocupa un límite máximo de alejamiento, se produce la vocal baja /a/.

Gráficamente los fonemas vocálicos se pueden representar mediante el llamado triángulo acústico vocálico, tal como se indica en el siguiente esquema:

Es necesario añadir otro rasgo para la clasificación de las vocales: la acción del velo del paladar, que nos ayuda a distinguir los sonidos orales de los oronasales. Cuando durante su emisión el velo del paladar está adosado a la pared faríngea, y por lo tanto la onda sonora sale únicamente por la cavidad bucal se producen los *sonidos orales*; si el velo del paladar está situado en una posición media entre la lengua y la pared faríngea, es decir, separado de ésta, la onda sonora sale al mismo tiempo por las cavidades bucal y nasal originando los sonidos llamados *oronasales*. En general la producción de estos sonidos se da en distribución complementaria: cuando van entre consonantes nasales y cuando aparecen entre pausa y consonante nasal.

Las vocales también pueden clasificarse por la acción de los labios en *labializadas o bemolizadas y deslabializadas o normales*; son labializadas /o/, /u/ y deslabializadas las demás. Por el grado de tensión articulatoria, se dividen en *relajadas* y *no relajadas*. En general se relajan /a/, /e/ y /o/ cuando están a final de palabra; también sufren relajación aunque en menor grado las átonas postónicas y en menos intensidad las pretónicas.

Los sonidos vocálicos oronasales se representan así: [a,-], [e,-], [i,-], [o,-], [u,-].

Las sonidos vocálicos relajados se representan: [Ξ], [χ], [_,_], [o],[n].

Por otra parte, es interesante comentar que en las vocales románicas las cuerdas vocales entran en vibración lentamente, por lo que no se produce golpe de glotis, dando lugar al llamado ataque vocálico suave. Por el contrario, en el caso de las vocales germánicas o de las sajonas el comienzo se realiza bruscamente, lo que se conoce con el nombre de ataque vocálico duro; se debe a que la entrada en vibración de las cuerdas vocales es muy rápida, lo que motiva el que se

perciba un ruido glotal al principio de la emisión. Sin embargo, la extinción en la pronunciación de las vocales románicas es brusca mientras que en el caso de las germánicas es suave, es decir se produce el fenómeno opuesto al de la iniciación.

Debido a su complejidad, el grupo de los sonidos vocálicos es el que ha despertado más interés en las investigaciones acústicas. Las cavidades de resonancia supralaringeas: oral, faríngea y nasal, actúan como resonadores. La cavidad oral es la más importante pues la lengua al adoptar diferentes posiciones modifica la capacidad del resonador oral, originando timbres vocálicos diferentes. La cavidad faríngea tiene un papel secundario en el timbre vocálico pues su volumen aumenta o disminuye poco y está en función de los movimientos de la lengua. Así, por ejemplo, para la emisión de /i/, /e/, la faringe se amplía considerablemente, mientras que ocurre lo contrario cuando la lengua retrocede para la emisión de las vocales /o/,/u/.

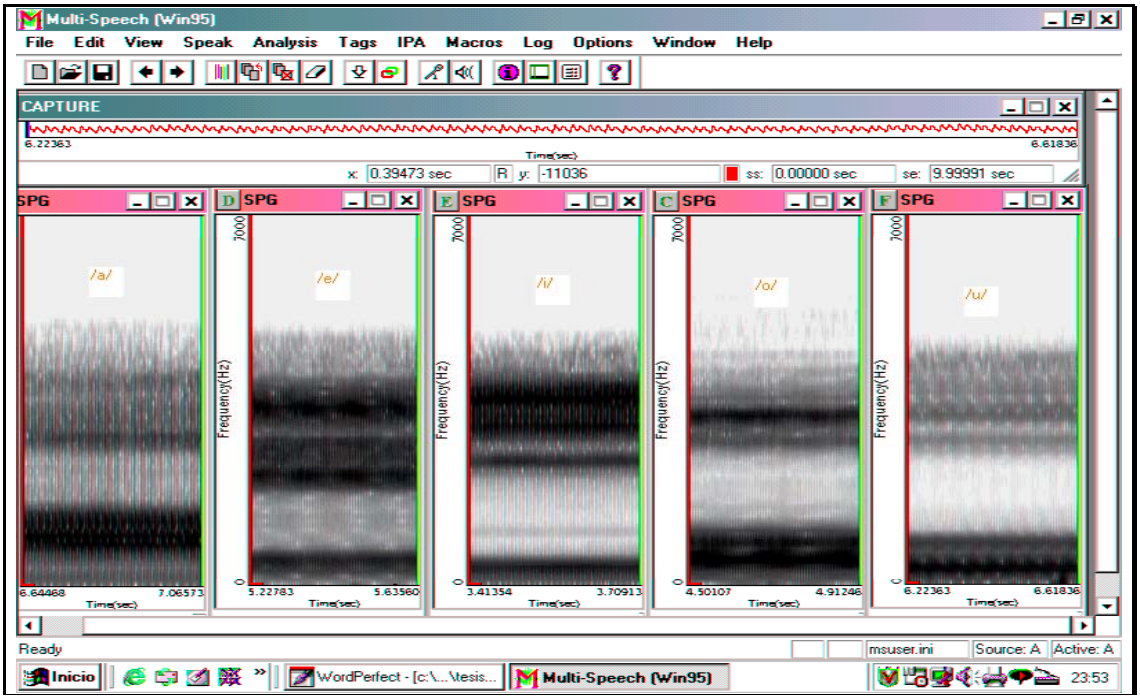
La cavidad nasal es invariable, limitándose su función a los sonidos oronasales tal y como se explica anteriormente.

Las resonancias que caracterizan el timbre de una vocal resultan de la filtración que por su paso a través de distintas cavidades sufre el tono glotal. La boca se comporta como un filtro que sólo deja pasar cierto tipo de vibraciones que se generan en la glotis. Dichas frecuencias son diferentes para cada vocal debido a que las cavidades de resonancia que las filtran cambian de forma y de dimensiones en cada realización.

Al ponerse en vibración las cuerdas vocales producen una onda compuesta periódica, o lo que es lo mismo, una onda constituida por un determinado número de ondas simples, cuyas frecuencias, como ya sabemos, se corresponderían con la relación: X, 2X, 3X, 4X , ... (considerando X como el valor de la frecuencia fundamental). Si mantuviésemos la misma frecuencia fundamental, cada uno de los sonidos vocálicos que emitiésemos tendría exactamente la misma configuración acústica a nivel glotal. Lo que en definitiva diferencia las estructuras acústicas de unas vocales y otras es su timbre o efecto de resonancia.

La onda simple de la glotis, pasa a las cavidades supraglóticas; éstas adoptan diferentes formas y volúmenes según la posición de la lengua. Cada uno de estos diferentes volúmenes actúa como una cavidad de resonancia: cuando la frecuencia de ciertas ondas simples coincide con la frecuencia de los resonadores bucales, estas ondas pasan con toda su plenitud y se potencian (formantes), mientras que las demás, cuya frecuencia no coincide, se atenúan. El resultado de estas filtraciones da lugar a las diferenciaciones entre las estructuras de unas vocales y otras, con sus consiguientes configuraciones formánticas :

Estos formantes o áreas de frecuencia potenciada son indispensables para la percepción acústica de cada vocal siendo directos responsables de la diferenciación inter vocálica los dos



primeros (F1 y F2). Además de estos dos formantes vocálicos indispensables y que podríamos denominar lingüísticos, existen los llamados formantes superiores o individuales, que en general aparecen por encima de los 3 KHz, y que varían en cada individuo según su configuración fisiológica laringo-bucal.

La altura frecuencial del primer formante guarda relación directa con la abertura del canal bucal: cuando la abertura es máxima, es decir, cuando la lengua está más separada del paladar, la frecuencia de dicho formante es más elevada; por el contrario, si la lengua se va acercando más al paladar, la abertura vocálica decrece y la frecuencia del formante disminuye. Por eso la frecuencia del F1 de Aa≅ es la más alta, siendo las de Ai≅ Au≅, las más bajas.

La altura del segundo formante F2, guarda una relación inversa con la longitud de la cavidad bucal de resonancia anterior; hay que tener en cuenta que el resonador bucal puede modificarse por dos causas principales: por la acción de los labios y por la posición de la lengua. Cuanto más anterior es la posición lingual, más pequeño es el resonador anterior, el sonido resultante es más agudo, y el F2 más alto; por el contrario, cuanto más posterior es la posición lingual, más grande es el resonador bucal anterior, el sonido resultante es más grave, y la frecuencia del F2 más baja. Por otra parte, cuanto más redondeados y abocinados se encuentren los labios, más baja es la frecuencia del F2 , ya que con este motivo queda más alargada la

cavidad anterior de resonancia.

Desde un punto de vista acústico todas las vocales presentan un rasgo común: su carácter *periódico, armónico o vocálico*. Esta circunstancia se proyecta por una estructuración formántica y una mayor concentración de la energía en las regiones comprendidas entre los 300 y los 800Hz aproximadamente.

En general, los formantes altos de las vocales están menos atenuados que los de otros sonidos pseudoperiódicos que posean una estructura formántica. Además, en términos generales, los sonidos vocálicos presentan una mayor intensidad acústica que los consonánticos.

Tomando como referencia los rasgos distintivos definidos por Jakobson**,** los fonemas vocálicos puede ser acotados de acuerdo a las siguientes oposiciones:

Vocales compactas o densas - vocales difusas.- El rasgo de compacidad/difusión de una vocal se manifiesta en la ubicación en rango de frecuencia del primer formante; cuanto más alto se encuentre y más próximo esté al segundo formante, más compacta será la vocal (con excepción de la u), por el contrario cuanto más bajo se encuentre y más separado del segundo formante, será más difusa.

La compacidad vocálica es directamente proporcional a la sección de paso que se establece entre los resonadores anterior y posterior; y por el contrario, la difusión vocálica es inversamente proporcional a la sección de paso entre los dos resonadores.

Vocales nasales y vocales orales.- Aunque no existe un criterio unánime sobre cuales son los índices acústicos que claramente caracterizan el matiz de nasalización, sí puede asegurarse que la vocal nasal se caracteriza por la atenuación de intensidad del primer formante F1 y un desplazamiento en mayor altura del tercero F3. Lógicamente, el rasgo de oralidad se revela en circunstancias opuestas al de nasalidad. En castellano el rasgo de nasalidad no es pertinente; este rasgo aparece sólo fonéticamente, cuando la vocal está situada entre dos consonantes nasales o en posición inicial absoluta.

Vocales graves - vocales agudas.- La diferencia entre vocal grave y aguda, se manifiesta a nivel sonográfico en el distinto rango frecuencial en el que aparece el segundo formante: cuanto más próximo se halle al primero, F1, la vocal será más grave, mientras que cuanto más cerca se encuentre del tercero, la vocal será más aguda.

El nivel de frecuencias del segundo formante es directamente proporcional al grado de

agudeza, e inversamente proporcional al de gravedad.

La cualidad de la vocal grave o aguda viene determinada por el volumen de la cavidad anterior de resonancia; cuando éste es pequeño el timbre resultante es agudo; por el contrario, cuando el volumen de la cavidad de resonancia anterior es grande, el timbre resultante es grave.

Vocales bemolizadas - vocales normales.- El factor de distinción entre vocales bemolizadas y normales viene determinado por la reducción del orificio labial, producido por un redondeamiento o alargamiento de los labios.

A estas posibles consecuencias acústicas relacionadas con los rasgos distintivos podemos añadir la realización *palatalizada* en la que se produce una considerable elevación del segundo formante y un ligero descenso del primero, o lo que es lo mismo, una mayor separación entre ambos; y la *velarizada*, que se caracteriza por un sensible descenso del segundo formante, es decir, una aproximación inter-formantica entre el F1 y el F2.

I.4.7.2.- Las secuencias vocálicas

Cuando dos vocales aparecen unidas en la cadena hablada pueden pertenecer o no a una misma sílaba. En el primer caso forman un *diptongo,* en el segundo, un *hiato*.

Diptongos.- De las vocales que integran un diptongo, una, constituye el núcleo silábico, y la otra, lo que se denomina margen silábico. La vocal que constituye el núcleo silábico es siempre aquella que reúne las condiciones fónicas óptimas de entre las vocales que están en la sílaba: mayor abertura, mayor tensión, mayor intensidad, mayor perceptibilidad, etc.

En el caso del hiato cada una de las vocales crea un núcleo silábico diferente.

Normativamente en castellano se considera diptongo a la unión en una misma sílaba de:
- 11: /i/ + /e, a, o/ ó /u/ + /e, a, o/ (diptongos crecientes).
- 21: /e, a, o/ + /i, u/ (diptongos decrecientes).
- 31: /i/ + /u/ ó /u/ + /i/ (Indistintamente crecientes o decrecientes).

Aunque no normativos, en las realizaciones habladas se producen otras series de diptongos; nos estamos refiriendo a aquellos formados entre las vocales medias y bajas e, a, o; en esta clase de diptongos cuando /a/ está presente, es el que forma el núcleo; en los casos de e, o, lo constituye la vocal de mayor intensidad al igual que sucede en el caso i, u.

En los diptongos crecientes la vocal que forma el núcleo silábico está situada en posición secundaria, y la vocal más cerrada ocupa una posición silábica prenuclear, recibiendo el nombre de *semiconsonante*. Estas semiconsonantes se transcriben fonéticamente como [j] para la i y como [w]para la u. Ejemplos: /pie/ [pje], /cuatro/ [kwatro].

En los denominados diptongos decrecientes la vocal más cerrada se halla en posición postnuclear, recibe el nombre de *semivocal* y se transcribe fonéticamente con los siguientes signos: i [i,º]; u [u,º]; ejemplos: /leí/ [lei,º] , /pausa/ [pau,ºsa].

Triptongos .- Cuando tres vocales se sitúan en la misma sílaba producen un triptongo. Como en el caso del diptongo, la vocal más abierta es la que forma el núcleo silábico. Las otras dos vocales serán semiconsonante o semivocal, según vayan situadas en posición prenuclear o postnuclear. Ejemplos: despreciáis; buey; averigüéis; que transcritos fonéticamente serían respectivamente, [despre□jai,ºs]; [bwei]; [aβeriπwéi,ºs].

Llegado este momento, resulta pertinente incluir ciertas consideraciones sobre las conjunciones Ay≅ , Au≅. La realización de la conjunción "y", varía en función del contexto fonético en el que se encuentre ubicada:

11) Cuando se halla entre dos consonantes, se realiza como la vocal anterior palatal [i].

21) Si está situada entre una consonante y una vocal, o entre dos vocales se realiza como la semiconsonante [j].

31) Cuando se encuentra precedida de una vocal y seguida de una consonante, se realiza como la semivocal [i,º].

La conjunción "u" al utilizarse sólo delante de palabras que empiezan por la vocal [o], se realiza siempre como la semiconsonante [w].

Hiatos.- Si dos vocales de una misma palabra concurren, y una de ellas es alta [i,u] y la otra media o baja [e,o,a] pueden no configurar un diptongo si cada una de ellas pertenece a una sílaba diferente. En este caso nos encontramos ante un hiato.

También puede producirse un hiato cuando en una misma palabra concurren dos vocales medias [eo, oe] o una media y otra baja o viceversa [ea,oa,ae,ao]; en tal caso, cada una de ellas constituye un núcleo silábico diferente, no formado por tanto diptongo; ejemplos: soez, beato, etc.

En ciertos actos de habla, estas vocales que normativamente forman sílabas distintas son pronunciadas en una sola. Por ejemplo, en lugar de "re-al" decimos "real"; este fenómeno se denomina *sinéresis*.

Desde un punto de vista acústico, la distinción entre diptongo e hiato viene determinada por la velocidad de transición entre las estructuras formánticas de las dos vocales; cuando la transición es lenta y su duración larga se trata de un diptongo; cuando la transición es rápida y, por tanto, su duración breve, se trata de un hiato.

En el ámbito del habla castellana no culta, aparecen en algunas ocasiones una serie de fenómenos que afectan a las secuencias vocálicas ya comentadas (diptongos, hiatos). Entre ellos, podemos citar los siguientes:

Metátesis.- Cambio de alguno de los elementos del diptongo, generalmente del margen silábico; ejemplo: naide (nadie).

Disimilaciones y asimilaciones: babiosos (babosos); pacencia (paciencia).

Pérdida de uno de los elementos del diptongo : Avinticinco≅

Tendencia a la igualación ei, ai.

Tendencia antihiática, o lo que es lo mismo, la conversión en diptongo de un hiato. La causa hay que asociarla con la tendencia a reforzar el límite silábico que entre dos vocales es muy débil, ejemplo: áhora, en lugar de ahora.

I.4.7.3.- Consonantes oclusivas o explosivas.

Desde el punto de vista articulatorio, consideramos consonantes oclusivas o explosivas aquellas que son producidas por un cierre del canal bucal. Teniendo esto en consideración, debieran incluirse bajo este epígrafe tanto las consideradas por Delattre [1958] como orales [p, t, k, b,d,g] como las nasales [m, n, □]. Desde una perspectiva acústica, tanto las explosivas orales como las nasales, comparten, la forma y dirección de las transiciones del segundo y tercer formantes.

Las que hemos denominado oclusivas orales, requieren para su emisión un cierre de la cavidad bucal y del conducto rinofaríngeo; se produce una retención del aire durante unos milisegundos hasta que nuevamente los órganos articulatorios abren el paso.

En castellano existen seis fonemas explosivos orales: /p/ sordo y bilabial; /b/ sonoro y bilabial; /t/ sordo y linguodental; /d/ sonoro y linguodental; /k/ sordo y linguovelar; y /g/ sonoro y linguovelar.

Estos fonemas funcionan plenamente en posición silábica prenuclear; ejemplos: taco, queso, etc. Sin embargo, cuando se encuentran en posición silábica postnuclear pierden su función distintiva neutralizándose; en estos casos la realización de tales fonemas es muy variada, dependiendo de ciertos hábitos socio o idiolectales del hablante. Ejemplos: actor > agtor> ator, etc.; si bien, lo más frecuente es que estos fonemas en esa situación postnuclear den lugar a unos alófonos sonorizados:

/-p/ > [β] < /-b/ ; /-t/ > [δ] < /-d/ ; /-k/ > [π] < /-g/

Por ello el resultado de la neutralización son los archifonemas / B, D,G /.

OCLUSIVAS ORALES

- BILABIALES:

Para su realización los dos labios se cierran momentáneamente, impidiendo la salida del aire a través de la cavidad bucal.

Oclusiva bilabial sorda.- Se representa fonéticamente por [p] y fonológicamente por /p/; ortográficamente responde siempre a la grafía p.

Oclusiva bilabial sonora.- Se representa fonológicamente por /b/ y fonéticamente por [b] cuando su realización es oclusiva, es decir, después de pausa y después de consonante nasal [m]; en los demás casos se fricatiza y se representa con [β]. Ejemplos: [ámbos] [aβía]. Ortográficamente responde a la grafía b ó v.

- DENTALES:

Se articulan con el ápice de la lengua contra los incisivos superiores.

Oclusiva dental sorda.- Se representa fonológicamente con /d/, y fonéticamente con [d] cuando su realización es oclusiva, esto es, cuando va detrás de pausa o bien de consonante nasal

ó l (ele); en los demás casos su realización es fricativa y se representa por [_]. Ortográficamente responde siempre a la grafía d.

- VELARES:

Se articulan con el postdorso de la lengua contra el paladar blando o velo del paladar.

Oclusiva velar sorda.- Se representa fonológicamente con /k/, fonéticamente con [k]. Ortográficamente responde a las grafías k, qu+e,i; c+a,o,u.

Oclusiva velar sonora.- Se representa fonológicamente con /g/ y fonéticamente con [g], cuando su realización es después de pausa o de consonante nasal; en los demás casos se fricatiza y su representación es [π]. Ortográficamente se representa por gu+e,i y por g+a,u.

Los fonemas oclusivos orales en las lenguas sajonas y germanas se realizan con un mayor grado de ensordecimiento las sonoras, y con una cierta aspiración las sordas.

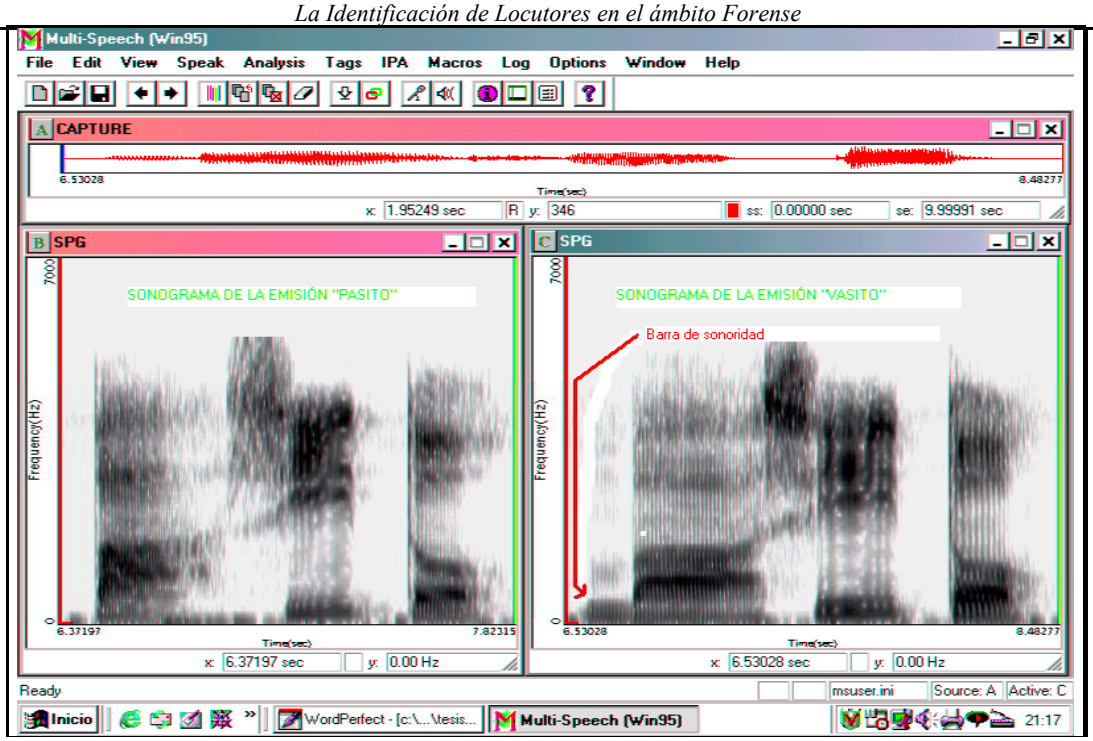
En su caracterización acústica los fonemas oclusivos presentan tres características fundamentales:

- Interrupción total del sonido articulado durante la fase de tensión.

- Explosión que sigue a dicha interrupción y que se traduce en un sonido turbulento impulsivo y de fuerte intensidad.

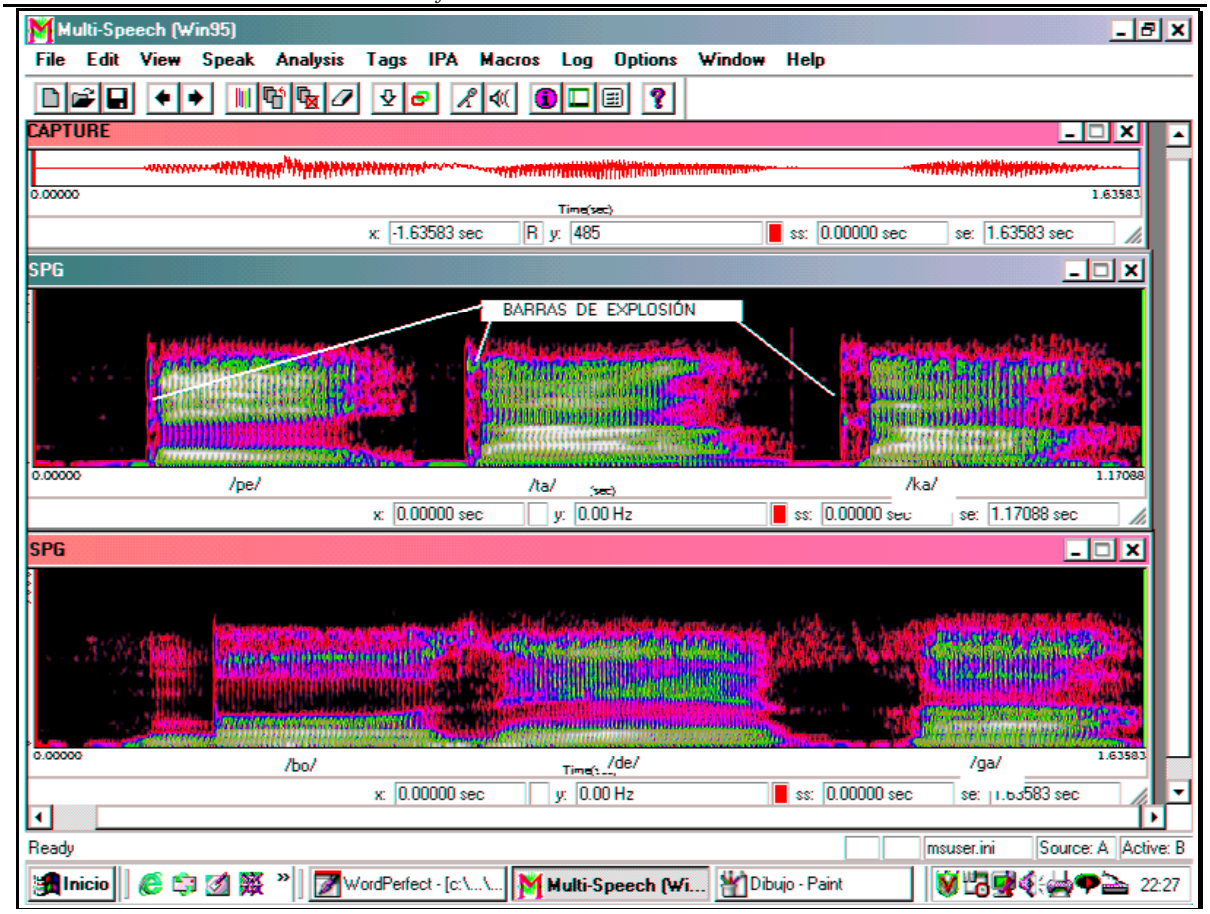
- Rapidez de las transiciones de los formantes de las vocales que les preceden o les siguen.

Como es lógico, las representaciones sonográficas de las oclusivas sordas se caracterizan por una ausencia de información armónica; en el caso de las sonoras, esta ausencia es también patente, pero una Abarra de sonoridad≅ de baja frecuencia las diferencia de las anteriores; esta barra de sonoridad se origina como consecuencia de la vibración a nivel glotal.


Multi-Speech (Win95)
File Edit View Speak Analysis Tags IPA Macros Log Options Window Help
CAPTURE
SPG
SONOGRAMA DE LA EMISIÓN "PASITO"
SPG
SONOGRAMA DE LA EMISIÓN "VASITO"
Barra de sonoridad
Frequency(Hz)
Time(sec)
Ready
Inicio
WordPerfect - [c:\...\tesis...
Multi-Speech (Win95)
21:17

A nivel espectrográfico es realmente complicado poder explicar por qué obtenemos distintas constancias perceptivas entre, por ejemplo, /p/ y /k/ ó /b/ y /g/. Inicialmente, algunos fonetistas intentaron buscar los índices acústicos que caracterizaban a estas consonantes, y pensaron que podrían estar asociados a dos factores : uno intrínseco -que es la explosión- y otro extrínseco, que son las transiciones. Estudios experimentales posteriores [Liberman, et al. 1954] pusieron de manifiesto la dificultad de establecer una relación directa entre la causa articulatoria y su efecto acústico. Precisamente, este es uno de los retos al que el identificador forense debe saber enfrentarse. Existen determinadas circunstancias que pueden ser detectadas a nivel auditivo y ser relacionadas a una causa articulatoria, no estando sin embargo clara su asociación a un índice acústico concreto. Lo mismo puede ocurrir a la inversa: en ocasiones, ciertas características acústicas son difícilmente adjudicables a un fenómeno articulatorio en particular. De ahí la conveniencia de utilizar una metodología en las que se apliquen distintos enfoques de estudio sobre el problema.

El momento de Aexplosión≅ de estas consonantes aparece en los sonogramas como una barra perpendicular al eje del dominio del tiempo, y se denomina *barra de explosión*.



Los índices de transición explosiva-vocal a nivel espectrográfico presentan unas orientaciones características en los formantes de las vocales contiguas. Según Delattre [1962] entre la tensión de una consonante y la tensión de la vocal siguiente, es decir, entre la fase cerrada y la fase abierta de una sílaba del tipo [ba], se produce un movimiento articulatorio hacia la abertura, combinado con un desplazamiento complejo de los órganos. Este movimiento o transición a nivel fisiológico se corresponde con un pattern espectrográfico de oscilaciones frecuenciales en los tres primeros formantes vocálicos. Estas transiciones se producen con todas las consonantes, presentando oscilaciones ascendentes o descendentes, que reciben el nombre de transiciones positivas y negativas, respectivamente.

En general la T_1 (Transición del primer formante) es positiva con los seis fonemas oclusivos, mientras que la T_2 (Transición de segundo formante) es positiva en el caso de las bilabiales y negativa con las dentales o velares. Todo ello referido a una transición con la vocal *Aa≅*, pues con el resto de las vocales se producen diversas variaciones (Ilustración 51).

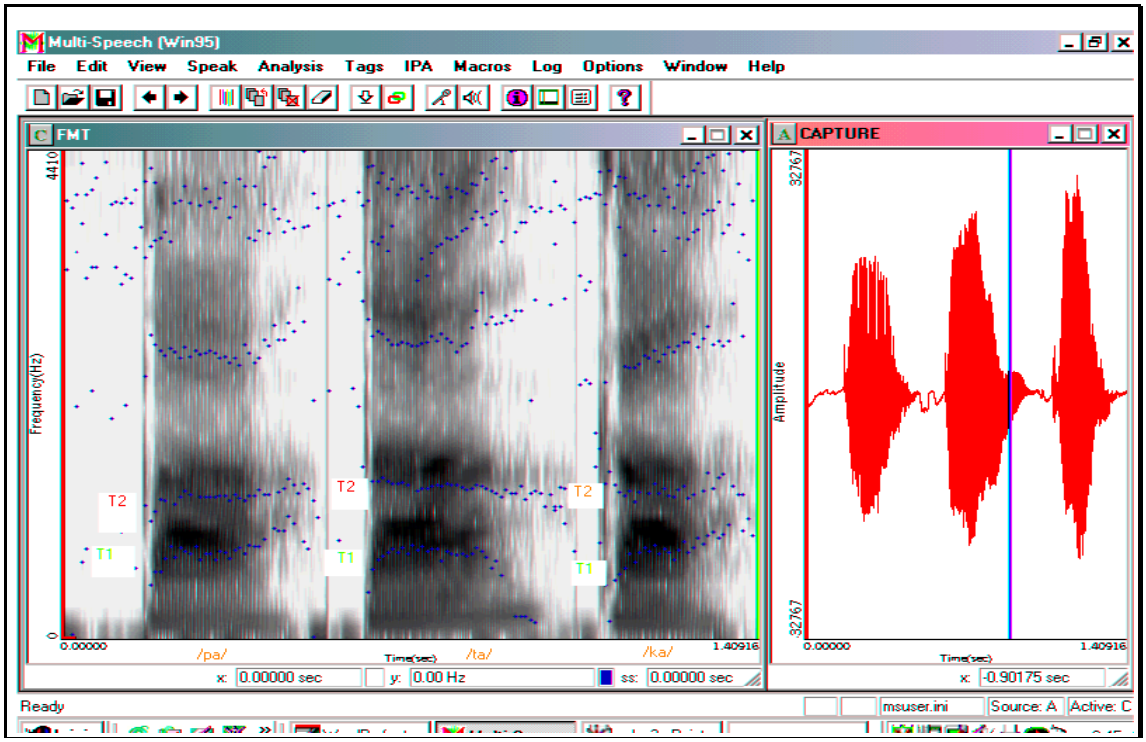
Desde un punto de vista acústico, y atendiendo a los rasgos distintivos que caracterizan los fonemas, los consonánticos oclusivos podrían clasificarse de la siguiente manera:

Densos o compactos-difusos. Son densos /k, g/ (velares) y difusos los cuatro restantes /p, b, t, d/. Es fácil comprender por qué /k, g/ son densos: la cavidad anterior de resonancia es mayor que la posterior, circunstancia que no se da en las difusas.

Graves-agudos.- Son graves los bilabiales y velares /p, b, k, g/ y agudos los dentales /t, d/. Se explica porque mientras las labiales y velares presentan un resonador indivisible, único, los dentales lo presentan dividido.

Estridentes-mates.- Los fonemas explosivos son todos ellos mate (en español sólo hay dos fonemas que presentan el rasgo estridente /s, c/).

Interruptos-continuos.- Los fonemas oclusivos son interruptos pero los sonoros /b,d,g/, cuando se fricatizan, son continuos.



OCLUSIVAS NASALES

Se consideran oclusivas nasales a aquellos sonidos consonánticos que se producen tras un cierre de los órganos articulatorios bucales y una apertura del pasaje rinofaringeo.

Desde un enfoque fonológico el español tiene tres fonemas nasales:

- Bilabial /m/
- Alveolar /n/
- Palatal /□/

Fonéticamente, presentan las siguientes realizaciones:

- El fonema /m/ bilabial presenta la realización [m].
- El fonema alveolar /n/ presenta:

  - Alveolar_____ [n]
  - Labiodental_____ [m,_]
  - Interdental_____ [n,_]
  - Dental_____ [n,_]
  - Velar_____ [□]
  - Palatal_____ [n,_]

- El fonema palatal /□/ tiene un sólo alófono [□].

- NASAL BILABIAL SONORA /m/

Su única realización se produce en situación silábica prenuclear. Cuando su posición es postnuclear, si no va delante de b, p (en este caso el alófono es [m]), su realización corresponde al alófono [n]; ejemplo: /album/ = [alβun]. Ortográficamente corresponde a la grafía m.

- NASAL ALVEOLAR SONORA /n/

La realización alveolar [n] se produce, o bien cuando se encuentra situado en posición silábica prenuclear o, cuando encontrándose en posición silábica postnuclear va seguido de consonante alveolar o de vocal.

La realización alofónica bilabial [m], se produce cuando le sigue uno de los fonemas

bilabiales Ap,b ó m≅. Ortográficamente corresponde a la grafía n. Ejemplo: un vaso = [úm báso].

La realización labiodental [m,_], se produce cuando el fonema /n/ está situado antes del fonema labiodental /f/. Ortográficamente coresponde a la grafía n. Ejemplo: /infame/ = [im,_ fame].

La realización interdental [n,_], se produce cuando el fonema /n/ va seguido del fonema fricativo interdental /□/. Ortográficamente se corresponde con la grafía n. Ejemplo: /en□erado/=[en□erado]. Este alófono no se pronuncia en las zonas donde existe Aseseo≅.

La realización dental [ℊ] se produce cuando el fonema nasal /n/ precede a una consonante dental, tanto sorda /t/ como sonora /d/. Ortográficamente se corresponde con la grafía n.

La realización velar [ɲ] aparece siempre que el fonema nasal precede a una consonante velar tanto sorda /k/ como sonora /g/. Ortográficamente se representa con la grafía n.

La realización palatal [n,_] se produce cuando el sonido nasal alveolar /n/ /m/ es seguido por una consonante palatal /c/ o /φ/. Ortográficamente corresponde a la grafía n.

- NASAL PALATAL SONORA. /□/

Tiene una única realización en cualquier contexto y su alófono se representa [□]. Ortográficamente se corresponde con la grafía ñ.

Desde una perspectiva acústica, las explosivas nasales comparten con las orales la forma y orientación de las transiciones del segundo y tercer formantes en las vocales contiguas. Sin embargo, su principal elemento de diferenciación lo constituye la existencia de ciertas estructuras formánticas en las nasales durante su intervalo de tensión; formantes que reemplazan la ausencia de información pseudoarmónica que acontece durante la fase de tensión de las explosivas orales, sonoras incluidas.

El pseudoformante F1 de las nasales está situado aproximadamente a una frecuencia de 250 hertzios y es de menor intensidad que los F1 de las vocales.

De todos los formantes que aparecen durante la fase de tensión en la emisión de las nasales, parece ser que el primero de ellos es el principal responsable de la percepción de la nasalidad; los superiores no dejan sentir apenas los efectos de la nasalidad y son de muy baja

intensidad. En un estudio experimental sobre síntesis del lenguaje [Liberman et al., 1954] se determinaron ciertos armónicos que pudieran ser los responsables de la diferenciación en la percepción de las nasales como clases de consonantes diferentes de las orales ; estas resonancias aparecían generalmente a unas frecuencias aproximadas de 240, 1.020 y 2.460 hertzios, siendo la de mayor intensidad la primera de ellas.

Acústicamente, y en relación con sus rasgos distintivos, los tres fonemas oclusivos nasales pueden clasificarse del siguiente modo:

/m/ = no vocálico, consonántico, nasal, sonoro, difuso y grave.
/n/ = no vocálico, consonántico, nasal, sonoro. difuso y agudo.
/□/ = no vocálico, consonántico, nasal, sonoro, denso y agudo.

En general la altura de frecuencias y la posibilidad de aparición de los formantes de las explosivas nasales se da como sigue:

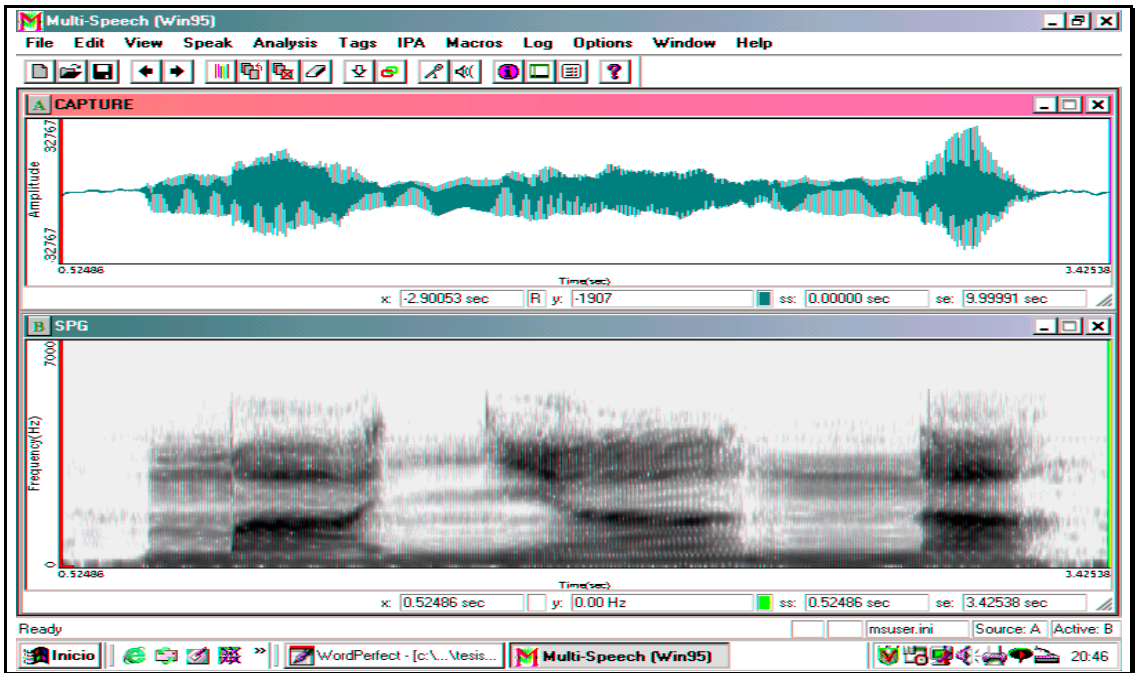
/m/.- Aparecen tres formantes:
F_1 a 250 Hz. (3mm)
F_2 a 1.020 Hz. (13mm)
F_3 a 2.000 Hz (25mm)

/n/.- Aparecen el primer y el tercer formante:
F_1 a 360 Hz (4,5 mm)
F_3 a 2.372 Hz (29 mm)

/□/ .- Suele aparecer sólo el formante primero a 300 Hz (4mm)

NEUTRALIZACIÓN DE LOS FONEMAS NASALES.

Los fonemas nasales cuando se encuentran en posición silábica implosiva, postnuclear, pierden sus caracteres distintivos. En esta situación los fonemas /m,n,□/, no se oponen, se neutralizan. Por lo tanto, desde una perspectiva fonológica , todos sus alófonos deberán sustituirse por el archifonema /N/.



I.4.7.4.- Consonantes fricativas

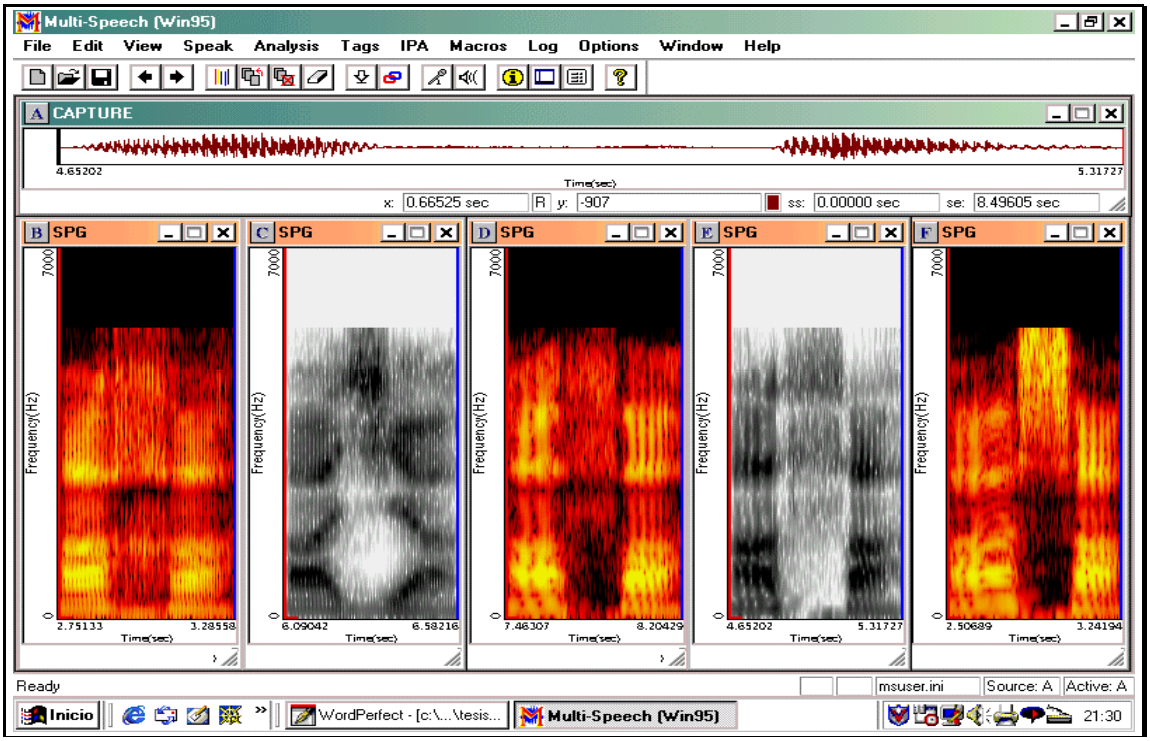
Denominamos *fricativas* a aquellas consonantes cuya principal característica acústica está relacionada con la fricción que produce el aire al atravesar distintas constricciones, ocasionadas por la función de dos órganos articulatorios del tracto vocal.

Fonológicamente son cinco los fonemas fricativos en español /f/ (labiodental sordo); /□/ (interdental sordo); /s/ (Alveolar sordo); /j,_/ (palatal sonoro) y /x/ (velar sordo). A ellos hay que añadir las realizaciones de los oclusivos /b,d,g/, cuando no van después de pausa o consonante nasal (y en el caso de /d/ también precedida de l), es decir cuando tales fonemas se fricatizan [ß] [π] [π].

El fonema labiodental sordo /f/ presenta una sola realización [f]. Ortográficamente responde a la grafía f.

Con relación al fonema /f/ conviene tener en cuenta que en ciertos hablantes o por peculiaridades de tipo regional aparece como bilabializado de ahí que su realización sea [ɤ].

El fonema interdental sordo /□/ presenta una sola realización [□]. Ortográficamente



corresponde a los grafemas c+e, i; z+a, e, i, o, u.

El fonema alveolar sordo /s/ presenta varias realizaciones, pero distinguiremos las tres fundamentales:

- *Apicoalveolar* .- Esta realización corresponde a la pronunciación normal castellana de /s/; se trata de una articulación cóncava pues el ápice de la lengua se eleva hasta los alveolos dejando así una amplia cavidad de resonancia. El comienzo de la energía de las estridencias a nivel sonográfico se produce en torno a los 3.500Hz.

- *Predorsoalveolar*.- Se articula situando el predorso de la lengua sobre los alveolos, por lo que la cavidad bucal de resonancia se reduce y las estridencias tienen menos intensidad. Se trata pues de una articulación convexa. Es la As≅ típica de Andalucía, Canarias y amplias zonas de

Hispanoamérica. Sus estridencias se inician en el espectrograma a una frecuencia más alta que la anterior, hacia los 4.000Hz o más.

- *Predorsointerdental*.- Se articula con el predorso de la lengua tocando en los incisivos por lo que la cavidad de resonancia se reduce aún más que en caso anterior. La posición de la lengua es convexa. Se corresponde con el fenómeno que denominamos Aceceo≅, aunque sus peculiaridades acústicas no coinciden con el sonido de /c/. Sus estridencias se inician hacia los 6.000 Hz.

Ortográficamente las realizaciones antes señaladas se corresponden con la grafía As≅ y la grafía Ax≅ en una pronunciación relajada.

El fonema alveolar sordo, cuando precede a una consonante sonora, suele sonorizarse y tal realización de representa [s,_]. Ejemplo: /los mismos/ = [los,_ mis,_mos].

La secuencia s+r,_ da lugar en la mayoría de los casos a la pérdida de la fricativa alveolar sorda, si bien en una pronunciación cuidada el resultado es la producción de una consonante fricativa sonora asibilada. Ejemplo: /Israél/ = [is,_rael].

El fonema fricativo velar sordo /x/ presenta en general un sólo alófono [x]. Ortográficamente se asocia a las grafías g+e,i; j+a,e,i,o u.

En Extremadura, Canarias, Andalucía, etc es muy frecuente la variante aspirada laríngea o faríngea. Ejemplo: caja = caha.

En Chile se dan variantes en distribución complementaría: Ante Aa, o, u≅ es velar como en castellano; ante Ae, i≅ es postpalatal y se representa foneticamente [ç]. Ejemplo: /xente/ = [çente]. Se percibe como A jiente≅.

El fonema fricativo palatal sonoro /j,_/. Presenta las siguientes realizaciones:

- Fricativo palatal sonoro, se representa fonéticamente [j,_], esta realización se produce cuando no va detrás de pausa, consonante nasal o l .

- Africado palatal sonoro, se representa fonéticamente [→], esta realización se produce

cuando va a continuación de pausa, consonante nasal o l. Se representa ortográficamente por la grafía Ay≅ así como hi+e.

En su caracterización acústica, los sonidos fricativos se dividen en dos grupos:

- De resonancias de baja frecuencia: [j,_], [→],[**δ]**, [ß], [ψ], [π] (corresponden a los fonemas sonoros).
- De resonancias de alta frecuencia: [f], [[κ], [□], [x], [s] (que corresponden a los fonemas sordos).

El único fonema fricativo que posee el rasgo estridente es el alveolar /s/, por la turbulencia que produce la corriente sonora contra los incisivos (tales turbulencias son concentraciones aperiódicas de energía). Se diferencia del fonema /□/ en que /s/ tiene una gran intensidad mientras que /□/ es más débil y las estridencias se inician a frecuencias más altas.

El fonema /f/ es muy similar a /□/ por lo que no hay distinción desde el punto de vista de las estridencias; es necesario recurrir a las transiciones, de forma parecida a lo que sucede con las oclusivas, es decir, teniendo en cuenta que en el fonema /f/ las transiciones son como en las bilabiales, mientras que en el fonema /□/ coinciden con las de las dentales.

El fonema /x/ por su condición de velar, aunque la hemos clasificado de fricativa de resonancias altas, en general también presenta ciertas estridencias en la parte baja del espectro. Por este motivo se distinguen fácilmente. Su espectro es una banda de estridencias con cierta progresión de intensidad de abajo hacia arriba.

La fricativa sonora /j,_/ se asemeja en gran medida a la i, es decir, que en espectrograma aparece barra de sonoridad + un pseudoformante de baja frecuencia y otro pseudoformante bastante alto.

Los sonidos oclusivos fricativos [β,δ,γ] presentan en el espectrograma concentraciones armónicas de energía (pseudoformantes), que normalmente aparecen a la altura del segundo formante de la vocal.

Los fonemas fricativos en relación a sus rasgos distintivos caracterizadores pueden definirse de la siguiente manera:

- /f/ = no vocálico; consonántico; difuso; grave; oral; continuo; mate; sordo.

- /θ/ = no vocálico; consonántico; difuso; agudo; oral; continuo; mate; sordo.
- /s/ = no vocálico; consonántico; difuso; agudo;oral;continuo;estridente;sordo.
- /j,_/ = no vocálico; consonántico; denso o compacto; agudo; oral; sonoro.

- /x/ = no vocálico; consonántico; denso o compacto; grave; oral; continuo; mate y sordo.

FENÓMENOS RELACIONADOS CON LAS FRICATIVAS.-

Es bastante habitual que cuando el fonema /s/ se encuentra en situación postnuclear o implosiva se realice aspirado o incluso que no se realice. Es decir, suele seguir la evolución aspiración-realización fonética 0.

s/ > [-h] > [cero]

La causa de este proceso es una disminución de la energía articulatoria (debido a la velocidad de elocución o pereza en la tensión articulatoria) que puede afectar tanto a una /s/ apicoalveolar como predosoalveolar. Este proceso de lenición es una manifestación lógica de la pérdida de tensión que caracteriza a toda articulación postnuclear [Chlumský, 1956].

De este fenómeno de aspiración se derivan dos consecuencias fundamentales:

11) Influencia sobre la duración de la vocal precedente (fenómeno compensatorio); cuanto más débil es la consonante que sigue, tanto más se reafirma la vocal precedente alargándose.

21) Influencia sobre la estructura acústica de las vocales precedentes. De este modo, las articulaciones del tipo /e/, /a/ se ven afectadas por esta posición lingual para la aspiración, abriendo y retrasando el lugar de articulación de dichas vocales.

Dos de los fenómenos más habituales y conocidos son el Aseseo≅ (uso de [s,_] estridente con timbre Aseseante≅) y el Aceceo≅ ([_,_] mate, de timbre Aceceante≅). En el caso del ceceo, la realización [_,_] andaluza es una articulación diferente de la [θ] castellana interdental; se realiza por medio de una constricción entre el predorso lingual y la cara interior de los incisivos superiores e inferiores.

I.4.7.5.- Consonantes africadas

Entendemos por africadas aquellas consonantes en cuya articulación intervienen un momento inicial oclusivo y otro, posterior, fricativo. Tanto la oclusión como la fricación se producen en el mismo lugar de articulación y en el momento de tensión.

En español sólo existe un fonema africado [±] que presenta dos realizaciones:

Alófono africado palatal sordo [±], que se realiza como tal en todos los contornos.

Alófono africado palatal sonoro [_,_], que se encuentra en distribución complementaria con la realización fricativa [j,], y se produce como africada, después de pausa o precedida de nasal o lateral.

El alófono africado palatal sordo tiene algunas variantes entre las que podemos destacar:

- Alvéolo-prepalatal [±] (corresponde a la Ache≅ castellana)

- Dento-alveolar [_,_] (tiene una articulación más anterior y una frecuencia más alta).

- Fricativo-prepalatal [□] (tiene una articulación similar a la Ach≅ francesa; se da en ciertas zonas de Andalucía y Extremadura).

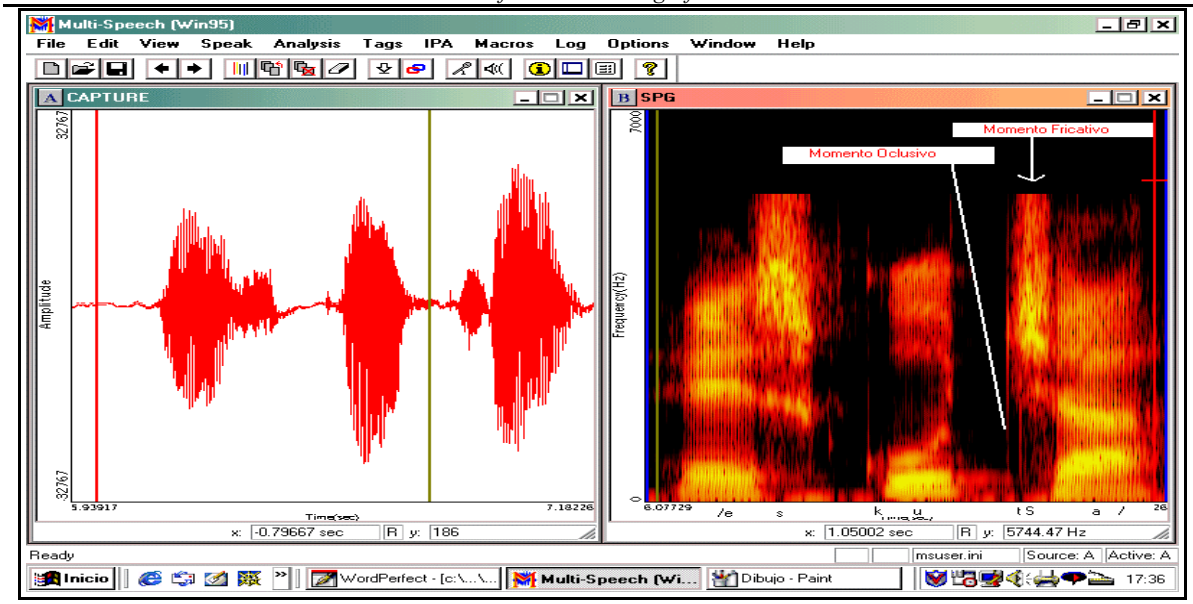
- Hay otra variante denominada adherente [Alvar y Quilis, 1966] que tiene un sonido de Ach≅ y Aye≅,; se da en Canarias, el español de América y también en Andalucía.

El alófono o sonido africado linguopalatal sordo ortográficamente se representa por la grafía ch.

El alófono africado palatal sonoro ortográficamente se representa por las grafías Ay≅ o hi+e en posición inicial.

El espectrograma de los sonidos africados presenta dos partes claramente diferenciadas: una primera parte en blanco (idéntica a la de las explosivas) y una segunda parte, con la turbulencia propia de las constrictivas o fricativas (Ilustración n1 54)

El intervalo de constricción de la africada es mayor que el que se produce en una



explosiva aspirada y menor que el de una fricativa.

Desde un punto de vista acústico, los sonidos africados en español se caracterizan por los siguientes rasgos distintivos: son densos o compactos, agudos, interruptos y estridentes.

I.4.7.6.- Consonantes líquidas.

Se incluyen bajo esta denominación las consonantes laterales y las vibrantes. Estos sonidos presentan ciertas características que les sitúan en una zona intermedia entre los sonidos vocálicos y los consonánticos. En general, las consonantes líquidas presentan estructuras formánticas similares a las de las vocales, si bien el tono fundamental y la intensidad media de aquellas son de un nivel más bajo..

Estudios clásicos sobre la caracterización acústica de las líquidas [O=Connor et al., 1957 y Delattre 1.958] pusieron de manifiesto diversas particularidades:

11) Presencia de un F1 en torno a los 400 Hz, que las diferencia del correspondiente formante de las nasales que no suele sobrepasar los 250 Hz.

21) En el momento de tensión aparecen formantes superiores al F1 que presentan una intensidad mayor que los de las nasales, pero menor que los de las vocales.

31) Las transiciones con las vocales se producen en trayectoria de continuidad a los formantes, mientras que en el caso de las nasales se apreciaban discontinuidades.

Se dividen en dos grupos:

- *Laterales*: /l/, /□/

- *Vibrantes*: /r/, /r,_/

Los fonemas laterales se comportan como tales en distribución silábica prenuclear, neutralizándose en posición postnuclear; los fonemas vibrantes funcionan como tales sólo en situación interior de palabras y en posición silábica prenuclear, neutralizándose en las demás posiciones.

- LATERALES.-

Son aquéllas en las que durante su emisión el aire de la fonación sale a través de un estrechamiento ocurrido en un lado o los dos de la lengua y el reborde o los rebordes homólogos de la región prepalatal o mediopalatal.

Desde el punto de vista fonológico es español sólo hay dos fonemas:

- Palatal...../l/
- Alveolar.... /□/

Desde una perspectiva fonética el fonema palatal, presenta sólo un alófono [□] y ortográficamente responde a la grafía All≅ (elle).

El fonema alveolar, presenta las siguientes realizaciones:

- Alveolar [l] que es la realización normal.
- Dental [l,_] cuando va seguida de fonemas /t/, /d/.
- Interdental [l,˜] cuando va seguida del fonema /□/.
- Palatal [l,_] cuando va seguido de fonemas palatales /_,_/, /□/, /j, /.

Existe además una variedad que se da generalmente en la zona de Cataluña y de Valencia que es la realización lateralvelarizada [≈].

Como ya hemos comentado, las líquidas laterales se caracterizan por su continuidad, lo que origina que su espectro presente ciertos formantes análogos a los vocálicos; concretamente el segundo y tercer formantes suelen variar y orientar sus frecuencias centrales en consonancia a las de los formantes vocálicos de su entorno.

La diferenciación acústica de los laterales /l/ y /□/, se puede concretar en los siguientes aspectos:

- El primer formante de /l/, aparece más elevado que el de /□/.
- El segundo formante del palatal /□/, aparece normalmente a una frecuencia más alta que el alveolar /l/ y es además más estable.
- El F_2 del fonema /l/ se contagia más de su entorno vocálico.
- Por otra parte, las transiciones de estos fonemas laterales con la vocal Aa≅ tienen un comportamiento diferente: Transición positiva con /□/ y algo negativa con /l/.

- YEISMO.-

Es un fenómeno consistente en la no diferenciación del fonema lateral palatal /□/ y el fricativo palatal /j,_/, efectuándose todas las realizaciones utilizando el fricativo. Cada vez está más extendido y asociado al entorno de las grandes ciudades. Las áreas geográficas españolas donde todavía permanece la distinción son Aragón, Navarra y gran parte de Castilla -León y Castilla La Mancha.

Igualmente en algunas zonas de Hispanoamérica, los dos fonemas /□/ y /j,_/, han perdido por un proceso de deslateralización su propiedad distintiva, llegando a confluir los dos en la fricativa central. Los alófonos de este nuevo sonido se extienden en abanico desde [j,_] hasta [→], pasando por una realización intermedia con la fricación propia de una rehilada [∞].

- VIBRANTES.-

Se denominan consonantes vibrantes a aquellas cuya característica principal está asociada a una o diversas interrupciones cortas de la corriente de aire fonador. Dichas interrupciones se

producen al contactar el ápice de la lengua en el área de los alveolos.

Fonológicamente, el español presenta dos fonemas vibrantes:

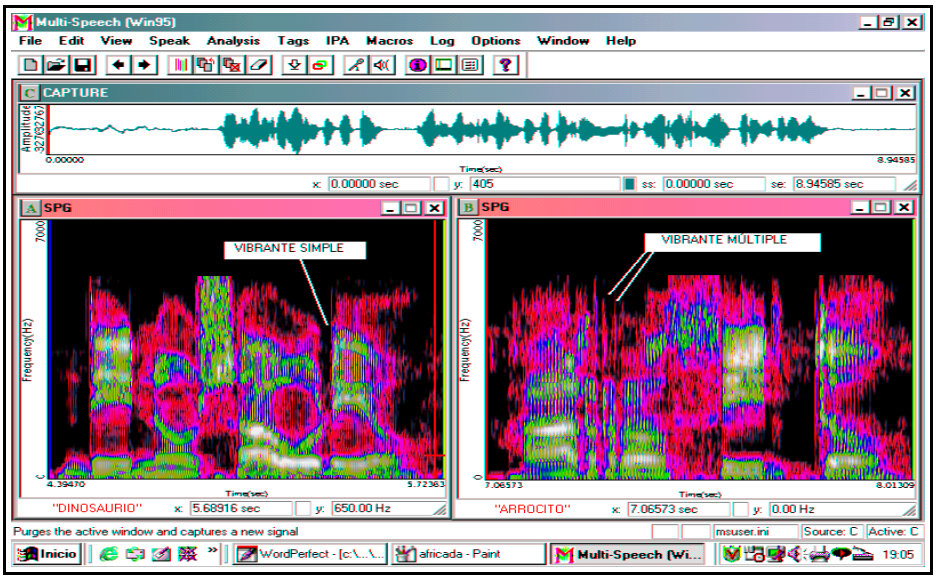
- Simple...../r/
- Múltiple...../r,_/

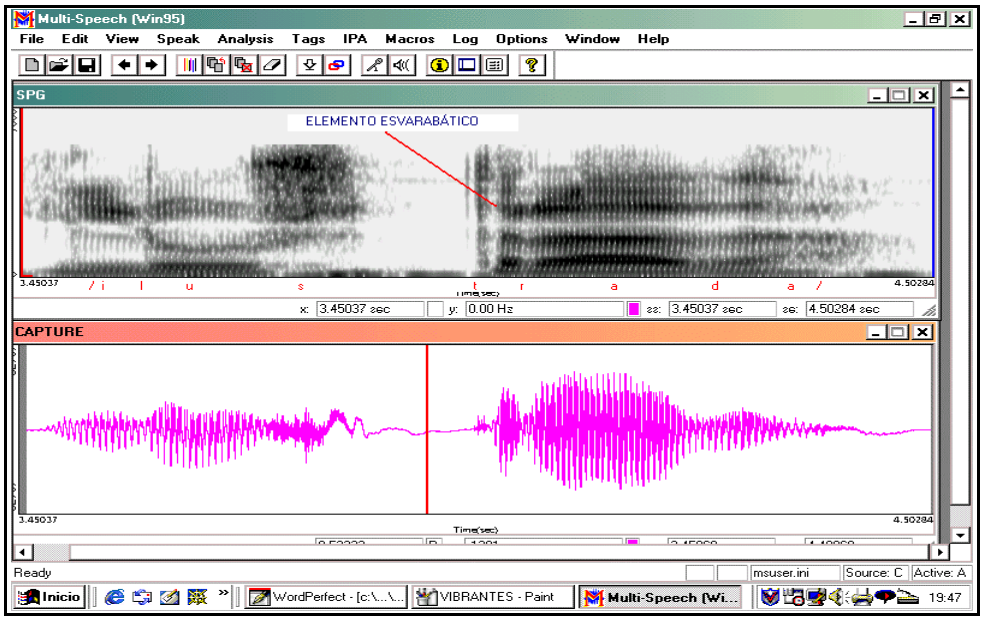
Fonéticamente, el español posee dos alófonos vibrantes:

- Simple .....[r]
- Múltiple....[r,_]

El sonido vibrante simple se produce por una breve oclusión del ápice de la lengua contra los alvéolos; se realiza de este modo en posición interior de palabra y en final. Si se producen dos o más oclusiones estamos ante el vibrante múltiple; el cual, acontece en principio de palabra, en posición medial o cuando se encuentra precedido de las consonantes n, l.

Las vibrantes se caracterizan por su carácter de interruptas. Si analizamos a nivel espectrográfico los dos sonidos vibrantes del siguiente gráfico observaremos que en el caso del múltiple puede verse después de /a/, una alternancia de espacios en blanco (en nuestro caso tres oclusiones que se corresponden con el momento en que el ápice de la lengua está tocando en la zona alveolar) con intervalos armónicos (de naturaleza similar a la vocal adyacente) producidos al separarse la lengua de los alveolos. Sin embargo, en Adinosaurio≅ sólo apreciamos una oclusión al ser un sonido vibrante simple.


Multi-Speech (Win95)
File Edit View Speak Analysis Tags IPA Macros Log Options Window Help
CAPTURE
Amplitude
32767
-32767
0.00000
8.94585
Time(sec)
x: 0.00000 sec
y: 405
ss: 0.00000 sec
se: 8.94585 sec
A SPG
B SPG
7000
0
Frequency(Hz)
VIBRANTE SIMPLE
VIBRANTE MÚLTIPLE
4.39470
5.72383
7.06573
8.01309
Time(sec)
"DINOSAURIO"
x: 5.68916 sec
y: 650.00 Hz
"ARROCITO"
x: 7.06573 sec
y: 0.00 Hz
Purges the active window and captures a new signal
msuser.ini
Source: C
Active: C
Inicio
WordPerfect - [c:\...\...
africada - Paint
Multi-Speech (Wi...
19:05



NEUTRALIZACIÓN DE LOS FONEMAS VIBRANTES.-

Cuando los fonemas vibrantes se encuentran en situación silábica implosiva, se realizan como un variante alofónico de cualquiera de los dos fonemas vibrantes en función del mayor o menor énfasis utilizado en su articulación. Al encontrarse en posición final y al perder su función distintiva las vibrantes se neutralizan resultando el archifonema /R/.

ELEMENTO ESVARABÁTICO.-

En español, los grupos formados por fonemas oclusivos + vibrante, o fricativo labiodental + vibrante: /pr-, br-, tr-, dr-, kr-, gr-, fr-/ son inseparables, y están situados en posición silábica prenuclear ; estos grupos consonánticos desarrollan en su realización un elemento vocálico añadido denominado *elemento esvarabático* [Lenz 1940, Navarro T.1918, Gili 1921, Malmberg, 1965, Quilis, 1970]. En algunas ocasiones también se produce este elemento cuando se pronuncia la *Ar*≅con sonoridad muy completa,como en las palabras: arte, cuerpo, trabajar, etc., donde entre el golpe de lengua de la *r* y las consonantes vecinas puede percibirse un perfecto sonido glótico.

La duración del elemento vocálico es muy variable. Su mayor o menor presencia tiene bastante relevancia a la hora de adjudicar un peso identificativo a dicho índice acústico. Sin embargo, la duración de la oclusión del sonido vibrante no aporta informaciones significativas de carácter individual.

Su estructura acústica es muy semejante a la de una vocal. El F1 del que Lenz denominaba elemento vocálico Aparásito≅ aparece a igual o algo menor frecuencia que los primeros formantes, en el caso de las vocales anteriores /i,e/. Con la vocal /a/ suele estar ubicado por debajo del comienzo de la transición del primer formante de /a/; con las vocales posteriores /o,u/ suele aparecer al nivel de su primer formante y a veces en el caso de /u/ a una frecuencia de mayor altura.

El F2 del elemento esvarabático suele vislumbrarse al nivel del segundo formante, en el caso de las vocales posteriores /u,o/ o a lo sumo en la misma frecuencia del inicio de la transición del segundo formante de /u,o/; con la vocal /a/ se observa al nivel frecuencial del comienzo de la transición de su F2. Con las vocales anteriores /i,e/ aparece situado en una altura frecuencial más baja que la correspondiente al comienzo del segundo formante de las mencionadas vocales.

Además del elemento esvarabático, podemos citar otros tres fenómenos menos comunes relacionados con las vibrantes:

1.- *La asibilación*.- En algunas áreas dialectales se produce una asibiliación de las vibrantes, siendo más común en el caso de la /r,_/ múltiple que pasa a /⊆/ . La realización de esta

última es fricativa y asibilada conservando su carácter de consonante líquida. En la parte superior de su espectro, aparecen las resonancias características de la fricación, y en su parte baja, dos estructuras formánticas que enlazan con las de las vocales contiguas.

La transformación de /r,_/ en /⊆/ es el resultado de un debilitamiento en la tensión ejercida con el ápice de la lengua; si el grado de debilitamiento afecta también a los bordes laterales con los que la lengua contacta con el paladar, puede suceder que /r,_/ pase a /l/.

2.- *Pérdida de Ar≅* o realización fonética cero en situación implosiva.

3.- *Articulación velar de Ar≅.*

Desde una nomenclatura acústica las líquidas pueden definirse de la siguiente manera:

El fonema /l/ posee el rasgo vocálico; consonántico; continuo.
El fonema /□/ posee el rasgo vocálico; consonántico; continuo.
El fonema /r/ posee el rasgo vocálico; consonántico; interrupto.
El fonema /r,_/ posee el rasgo vocálico; consonántico; interrupto.

I.4.8.- Los rasgos suprasegmentales.

I.4.8.1.- El acento.

Aunque desde un punto de vista identificativo tiene mucha más relevancia la entonación que el acento, merece la pena efectuar algunos comentarios en torno a este último.

Básicamente, el acento es el rasgo prosódico que permite poner de manifiesto una estructura lingüística superior al fonema, diferenciándolo de otras unidades similares. Por ejemplo, en la palabra caótica, la primera /o/ es fonema pero con categoría de sílaba.

Se suelen señalar cuatro funciones elementales del acento: *contrastiva, distintiva, demarcativa y culminativa*.

La contrastiva sirve para evidenciar las sílabas acentuadas frente a la inacentuadas. Se da en todas las lenguas, tanto en las de acento fijo (francés, turco, checo, etc) como en las de acento libre (español, inglés, etc).

La distintiva se manifiesta en el eje paradigmático pero sólo en las lenguas de acento libre. Su situación implica distintos contenidos semánticos. Ejemplo: Atérmino≅ es diferente a Aterminó≅.

La demarcativa en las lenguas de acento fijo, indica los límites entre las diversas unidades de una secuencia. Ejemplo: El final de una palabra en francés, el principio en checo o bien señalar una posición fija respecto al principio o al final como ocurre en polaco.

La culminativa señala en las lenguas de acento libre la presencia de una unidad acentual sin indicar exactamente los límites.

En español, toda palabra aisladamente lleva sólo una sílaba con carga acentual (a excepción de los adverbios acabados en Amente≅ o en caso de énfasis) ; no obstante, dentro de la cadena hablada algunas palabras pierden esa carga al formar grupo tónico con otra que la asume.

El acento es también considerado un rasgo funcional integrado por tres elementos: intensidad, tono y duración. En español, el parámetro que desempeña una función primordial en la realización y percepción del acento es la frecuencia del fundamental, sola o acompañada de la duración. En contadas ocasiones, cuando no actúan ni la frecuencia fundamental, ni la duración, es la intensidad la interviniente. También en nuestra lengua tiene una función diferenciadora que implica cambios semánticos muy significativos según donde se sitúe.

Por ejemplo, supongamos la sucesión de fonemas (a su vez integrados en sílabas) siguientes:

/ l + i - m + i - t + e /

Según donde nosotros hagamos recaer el acento tendremos un valor semántico u otro:

/límite/ -	/limíte/ -	/limité/
Alímite≅ -	Alimíte≅ -	Alimité≅

Existen numerosos ejemplos similares que ponen de manifiesto este rasgo funcional del acento en español. Comparemos los siguientes ejemplos en su representación ortográfica y

fonológica:

Atérmino≅	Atermino≅	Aterminó≅	/téRmino/	/teRmíno/	/teRminó/
Acántara≅	Acantara≅	Acantará≅	/káNtara/	/kaNtára/	/kaNtará/
Acélebre≅	Acelebre≅	Acelebré≅	/□élebre/	/□elebre/	/□elebré/
Adepósito≅	Adeposito≅		Adepositó≅	/depósito/	/deposito/ /depositó/
Aánimo≅	Aanimo≅	Aanimó≅	/ánimo/	/animo/	/animó/

I.4.8.2.- La entonación.

Según el punto de vista desde el que se pretenda efectuar un análisis de la entonación, pueden elaborarse múltiples definiciones de la misma.Unas, estarán referidas a lo que se considera pura sustancia (fluctuación del tono fundamental) : *A las variaciones en el tono de la voz del*

hablante≅ [Jones, 1909] , *Ala línea melódica del habla, la elevación y descenso del fundamental≅*[Bolinger, 1955] ; otras, serán relativas a la función lingüística de la entonación en la oración [Trager y Smith, 1935 ; Denes, 1959; Schooneveld, 1961; Artemov, 1965; Vasilyev, 1965; Lehiste, 1970...].

Para el investigador forense este rasgo suprasegmental tiene interés en ambas vertientes. Analizando la envolvente del tono fundamental, apreciada tanto en sus niveles de intensidad, sus diferentes duraciones, tonemas, etc. obtendrá una referencia que en algunos casos puede llegar a ser de especial utilidad. También deducirá datos interesantes sobre la componente emotivo-expresiva y sociolingüística del hablante cuando evalúe el carácter lingüístico de la entonación. Sin lugar a dudas, la entonación puede modificar el plano semántico de un contenido como puede reflejar el estado psíquico del individuo cuando se expresa (sus sentimientos, intenciones, estado anímico, etc.). Quizás - como afirma Quilis [1981] refiriéndose a la entonación - nos hayamos ante A*un prosodema que utiliza principalmente las variaciones de frecuencia del fundamental para desempeñar una función lingüística a nivel de la oración≅.*

Desde una perspectiva lingüística, en el estudio de la entonación deben considerarse tres niveles: *el lingüístico o denotativo, el sociolingüístico o connotativo y el expresivo*. Las funciones de la entonación en relación a éstos diferentes planos del acto del habla son las siguientes:

- *Función distintiva*: Permite distinguir las clases de oraciones. Ejemplo: una oración afirmativa de una interrogativa.

- *Función integradora*: Denominada así, en cuanto resulta indispensable para integrar las distintas palabras formando frases. Hay autores que piensan que sin entonación no hay lengua.

- *Función delimitadora*: La entonación delimita los enunciados y segmenta el continuum del discurso en unidades. En ocasiones tiene función semántica. Ejemplo: APepe come≅, APepe: come≅.

- *Función identificadora*: Asociada al nivel sociolingüístico. Cada lengua tiene su peculiar entonación, así todos los hablantes de una misma lengua usan secuencias tonales básicas. Naturalemente hay variaciones dialectales, sociolectales, incluso idiolectales. Por este motivo, en la identificación del hablante la entonación desempeña un papel importante.

- *Función impresiva*: Directamente relacionada con el plano expresivo. La entonación es un vehículo importante de la expresión afectiva, sola o combinada con otros factores.

Cabe preguntarse, si esta última función pertenece o no al dominio puramente linguístico. En general, puede decirse, que la entonación expresiva no interfiere la entonación comunicativa básica, sino que se superpone a ésta, es decir, que no afecta a la función linguística propiamente dicha, aunque es evidente que la expresividad acentúa ciertos significados.

Tampoco hay acuerdo unánime en asignar a la entonación -en un sentido general- la función expresiva, pues en algunas lenguas como el japonés existen muchas partículas expresivas que dentro de un plano puramente gramatical pueden desempeñar la expresividad [Abe, 1955]. En cualquier caso, no parece un disparate asociar expresividad y entonación. Según León [1970, 1972] en esa función expresiva intervienen varios factores:

- El *registro*: Un registro de tono alto evoca alegría, ligereza; un registro bajo: tristeza, gravedad.

- La *desviación* entre los puntos extremos del patrón melódico. A mayor separación de esos extremos, más acusada es la sensación expresada y al contrario.

- El *contorno del patrón melódico*, aunque es insuficiente, ya que un mismo contorno puede servir para muchas funciones.

- La *intensidad* : a mayor intensidad, mayor énfasis en el sentimiento expresado.

- La *duración* : relacionado con la evocación y connotaciones poéticas del sentimiento.

De alguna forma relacionados con el estudio de la entonación, conviene mencionar dos conceptos que regularmente utilizaremos para acotar nuestros posibles objetos de estudio. Nos estamos refiriendo a los términos: *Agrupo fónico≅ y Atonema≅.*

Grupo fónico:

El grupo fónico está constituido por la serie de unidades sonoras del discurso que se hallan comprendidas entre dos pausas.

Ejemplos: AEstoy triste≅ (un grupo fónico)
ADesde aquel día, estoy triste≅ (dos grupos fónicos)
ADesde aquel día, dijo el hombre, estoy triste (tres grupos fónicos)

Tonema:
Según Navarro Tomás [1948]: ASe entiende por tonema la altura musical correspondiente a la terminación de cada uno de los grupos fónicos en que se divide la frase≅.

En español, a nivel funcional, tenemos tres tipos de tonemas: *descendente, horizontal o suspensivo y ascendente*. No obstante, lo que tiene valor en la entonación desde el punto de vista significativo es el carácter del tonema:

Ejemplo:

Cantará
Cantará ... ___ si le pagan bien
)Cantará?

I.4.9.- La sílaba.

La unidad de agrupamiento inmediatamente superior a los fonemas es la sílaba, que a su vez, puede estar integrada por uno o varios fonemas. La estructura y límites de la sílaba han sido

estudiados detalladamente por distintos autores: [Malmberg, 1955; Rosetti, 1963; Hala, 1973...].

En español toda sílaba ha de contener al menos una vocal (que es el elemento principal de la sílaba). Nuestro interés se centrará en conocer las posibles consecuencias articulatorias y acústicas para los distintos fonemas, en función de su ubicación o distribución dentro de esta estructura básica. Para ello, nos limitaremos a describir aquellos fundamentos elementales relacionados con dicha estructura que nos permitirán referenciar con claridad las distintas peculiaridades fonoarticulatorias de las emisiones habladas.

En una sílaba como por ejemplo: [pa□] Apa≅. La consonante Ap≅ constituye el primer sonido de la sílaba y por lo tanto su *fase inicial*. La vocal Aá≅ es el *núcleo* o eje de la sílaba y es la que posee todas las cualidades articulatorias y acústicas en un grado máximo; es la fase culminante o central de la sílaba. Después viene el sonido [□] que constituye la *fase final* de la sílaba.

Partiendo de las teorías de Bohuslav Hála, la vocal que constituye el *núcleo* de la sílaba es el sonido que:

a) Posee la mayor intensidad
b) Es el más abierto.
c) Es el más perceptible
d) Es el que tiene una mayor tensión articulatoria
e) Es el que posee mayor grado de sonoridad

Partiendo de estos presupuestos, Hala considera que la vocal Aá≅ de [pá□] constituye el *núcleo* de esta sílaba y que los otros dos sonidos de la misma son los *márgenes silábicos*. El sonido [p] sería el *margen prenuclear* puesto que es el que encuentra antes del núcleo y el sonido [□] constituiría el *margen postnuclear* por encontrarse después del núcleo. Los márgenes silábicos tienen cualidades propias de los sonidos en un grado menor que el núcleo. Van aumentando gradualmente hacia el núcleo, donde se dan en un grado máximo, y a partir de él, van disminuyendo.

Los límites silábicos coinciden con la ausencia de las cualidades acústicas y articulatorias de los sonidos. Por lo que se manifiestan entre dos sonidos o grupos de sonidos.

Partiendo de estas teorías podemos considerar que en los grupos vocálicos (diptongos y

triptongos) las semiconsonantes [j], [w] son también *márgenes vocálicos prenucleares,* y las semivocales [º,i], [º,u] son *márgenes vocálicos postnucleares*.

Las sílabas pueden ser *tónicas* cuando el acento con mayor intensidad recae sobre el núcleo de una sílaba frente a las *átonas* o inacentuadas. Ejemplo:

[tér-mi-no]- [ter-mi-no] - [ter-mi-nó]

También suelen denominarse *abiertas* cuando el último elemento de la misma es una vocal y *cerradas* cuando es una consonante.

Ejemplos:

abiertas: [ká-sa]
cerradas: [ár-bol]

CAPÍTULO II

LA IDENTIFICACIÓN DE PERSONAS A TRAVÉS DE SUS EMISIONES HABLADAS

II.1.- LA VOZ , OBJETO DE INVESTIGACIÓN FORENSE

II.1.0.- Introducción.

Instruidos en el conocimiento básico de los aspectos teóricos que sustentarán nuestro principal argumento de trabajo -la identificación de voz- es momento de adentrarnos en el estudio de sus diferentes propuestas y entornos de investigación forense. Para ello, y en primer lugar, tomaremos conciencia sobre las especiales circunstancias que caracterizan la técnica de identificación de locutores, desde su concepción más general, a aquella otra más específica del campo forense. A continuación, nos familiarizaremos con una nomenclatura de términos fundamentales, y detallaremos las diferentes causas asociadas al que será principal factor de dificultad en las tareas de identificación: el carácter variable de los registros de habla.

Sin olvidarnos de comentar sus antecedentes históricos, finalizaremos este capitulo analizando las distintas alternativas metodológicas de la técnica, y sus repercusiones de valoración a nivel judicial.

II.1.1.- Introducción a la problemática de identificación de locutores.

A mediados del siglo XX, existía la convicción generalizada de que muchos de los jeroglíficos científicos del momento, estarían completamente esclarecidos en las primeras décadas del segundo milenio. Sin embargo, casi finalizada la andadura del mágico año 2000, asistimos impotentes a la no resolución de muchas de aquellas claves que se presumían accesibles. De la misma forma que graves enfermedades no han logrado ser erradicadas a pesar de los importantes avances tecnológicos, otros problemas científicos menores -como es el caso de la identificación de las personas a través de su voz- no han podido ser resueltos de forma definitiva.

Cualquiera que no sea buen conocedor de alguna de las disciplinas afines al entorno de las emisiones de voz, cuando oye hablar de métodos o tecnologías para el reconocimiento de locutores, tiende a relacionarlos con investigaciones de última generación. Probablemente, la filosofía multimedia que en la actualidad impera en todo tipo de aplicaciones informáticas, ha podido contribuir a la construcción de estas inexactas apreciaciones. Pero en realidad, como veremos más adelante, la utilización de medios de análisis electroacústicos con fines identificativos se iniciaba en el transcurso de la II guerra mundial, y los primeros pasos en reconocimiento automático de locutores se producían a primeros de los setenta.

Desde entonces, y durante casi quince años, las aplicaciones de carácter automático se han desarrollado sin deparar resultados espectaculares, si bien, fruto de unas necesidades de "mayor rentabilidad" (gestión bancaria, comunicaciones, seguridad...) que la que puede proporcionar el hecho de absolver a un inocente o inculpar a un criminal, durante la última década se viene desarrollando con auténtico frenesí, una carrera de alto nivel competitivo cuya meta es el diseño de los algoritmos definitivos que permitan a los ordenadores, sin margen de error, reconocer a una persona a través de un input de voz.

Oscar Tosi**,** uno de los padres de la identificación forense de la voz, murió en 1.994 con el convencimiento de la necesidad de fomentar el desarrollo de los que él denominaba métodos objetivos, o lo que es lo mismo, las aplicaciones informáticas de reconocimiento del habla. Aunque también es cierto, que siempre mantuvo la convicción de que no existía una solución

inmediata al problema de la identificación de locutores mediante el uso exclusivo de este tipo de sistemas. Así lo manifestó en [1973] en su informe a la L.E.A.A. (*Law Enforcement Association of America*) y posteriormente en 1.979 en su libro "Voice Identification". Veintisiete años más tarde, la situación en el campo forense puede decirse que es bastante similar. Las apreciaciones Asubjetivas≅ de los expertos continúan siendo determinantes.

Aunque más adelante abordaremos en detalle las particularidades, evolución y posibilidades de los sistemas automáticos en el reconocimiento de locutores, sí conviene adelantar que las especiales características que habitualmente presentan los registros de audio en el ámbito forense, dificultan gravemente la obtención sistemática de decisiones fiables por parte de los ordenadores.

Pero dejemos por el momento los sistemas automáticos y centrémonos en la técnica de identificación de hablantes. Las diferentes perspectivas de análisis que a lo largo del tiempo

se han proyectado sobre la misma, fruto de su intrínseco carácter multidisciplinar y de la naturaleza variable del objeto de estudio, han motivado algunas confusiones en relación a cierta nomenclatura y conceptos básicos. No obstante, éste es un problema menor que tiene una fácil solución con la elaboración y uso de estándares universales, aunque de este particular, ya hablaremos en un posterior apartado.

Lo que ya no parece tan sencillo, es la unificación metodológica, puesto que además de existir unos enfoques diferentes a nivel técnico, nos encontramos con problemas de aplicación en los distintos ámbitos legales de cada país. Si a todo esto le unimos los intereses particulares de la comunidad científica internacional, ya sea a nivel individual o colectivo, quizás entendamos porqué todavía en el año 2000 siguen latentes muchos de los problemas que surgieron ya hace casi treinta años.

Posteriormente contextualizaremos históricamente las discrepancias o reticencias que lógicamente surgieron en los primeros años de desarrollo, es decir, allá por los años 60. Si comparamos aquella situación con la actual, y teniendo en cuenta todas las particularidades ya citadas, resulta casi imposible comprender porqué no se han aunado los suficientes esfuerzos para la utilización de unas referencias comunes, que además de ser extraordinariamente útiles proporcionen el más alto nivel de fiabilidad a la técnica forense de identificación de locutores.

Esta importante cuestión de la falta de directrices generales no quedaría enteramente planteada si omitiésemos una de las causas más determinantes de tal situación. Nos estamos

refiriendo al distinto desarrollo que ha tenido la investigación forense de la voz en cada país, o más correctamente, en cada comunidad científica o pseudocientífica. En muchos casos, la insuficiente formación de los especialistas o una incompleta perspectiva del problema, ha desembocado en erróneas evaluaciones y conclusiones. Esta carencia del necesario rigor técnico y a veces ético, se ha traducido en unas consecuencias de indudable perjuicio para el resto de expertos que ejercitan correctamente su labor.

Nosotros, en el deber de ignorar posicionamientos interesados trataremos de mostrar una visión general y lo más objetiva posible de la evolución histórica de la técnica y el estado de la cuestión en el momento actual. Justificaremos razonadamente las posibles soluciones a los distintos interrogantes, introduciendo un modelo metodológico adecuado a las distintas opciones técnicas y tecnológicas conocidas en la actualidad.

II.1.2.- Nuestro entorno de investigación : el ámbito forense.

Como ya hemos comentado en el apartado anterior, durante las dos últimas décadas se ha venido observando un fuerte incremento del número de proyectos de investigación sobre sistemas automáticos de reconocimiento de locutores. Entidades bancarias, empresas de seguridad, redes informáticas, industrias militares, etc., apuestan y manifiestan un extraordinario interés por este tipo de proyectos. Pudiera ocurrir, que lo que otros científicos no han conseguido tras muchos años de trabajo al servicio de unos intereses más altruistas (caso de la aplicación forense) pudiera ser alcanzado a través de la investigación en estos nuevos caminos.

Mientras se produce el "milagro automático", los distintos entornos profesionales interesados, buscan las soluciones más oportunas a sus respectivas necesidades. Uno de esos entornos es el entorno forense.

Con carácter general, entendemos como *ámbito forense* aquel que guarda alguna relación con el esclarecimiento de una acción delictiva. Dicha acción, pudiera constituir una falta o un delito, tal cual son descritos en nuestra legislación penal. Por consiguiente, siempre que trabajemos con emisiones habladas que de alguna forma se relacionen con la investigación de un hecho delictivo, lo estaremos haciendo dentro de un entorno de análisis forense.

El análisis de la voz en las denominadas condiciones forenses, implica una serie de inconvenientes que incrementan considerablemente el nivel de dificultad en el que pueden desarrollarse otras experiencias de entornos afines. Las grabaciones anónimas o dubitadas, en

más del 90% de los casos son consecuencia de intervenciones telefónicas ordenadas por la Autoridad Judicial, lo que provoca un efecto inmediato de restricción de información en rango de frecuencia y otras alteraciones menores también relacionadas con las curvas de respuesta telefónicas (analógicas, digitales, móviles). En el resto de los casos, suele tratarse de registros recogidos con grabadores de pobre calidad (microcassette, walkmans, etc.), contestadores automáticos, centrales telefónicas de servicios públicos, etc.

En definitiva, nos hallamos ante una señal degradada por causas de índole cualitativo (curvas de respuesta, ruido, distorsiones, bajo nivel,etc.) e insuficiencias de tipo cuantitativo. A todo esto, hemos de añadir fuertes fluctuaciones de los planos expresivos entre las muestras anónimas e indubitadas debido al carácter habitualmente *Ano cooperativo*≅de la persona a quien se atribuyen los actos de habla objeto de análisis.

Este es el ámbito forense. Un entorno de trabajo dificultoso y atractivo. Sus especiales características, y la amplia diversidad de factores que acompañan a su casuística, serán los principales alicientes para el investigador, quien deberá de unir a una sólida, larga y continua formación, el más riguroso código de ética para el correcto desarrollo de su labor pericial.

II.1.3.-) Identificación, verificación, reconocimiento, autentificación?

Ha llegado la hora de introducirnos de lleno en una técnica que hasta este momento hemos venido denominando de diversas maneras: identificación de voz, reconocimiento de locutores, etc. Probablemente, la primera consecuencia de su carácter multidisciplinar -reiterada mente comentado- es la existencia de cierta confusión, o más bien de una falta de matización, sobre el propio nombre de la técnica.

No pretendiendo ser puristas, sino muy al contrario,tratando de utilizar una denominación idónea, universal y sencilla -aun conscientes de su posible inexactitud- debemos seleccionar un nombre adecuado al plano semántico que caracteriza el ámbito forense.

Los ingenieros de los laboratorios Bell, G.R.Doddington, Bishnu S. Atal y Aaron E. Rosenberg, considerados pioneros en el diseño de sistemas automáticos para el reconocimiento de locutores, tomando como referencia su campo de investigación, englobaron en el concepto de *"speaker recognition"* (reconocimiento de locutores) todas aquellas tareas relacionadas con la identificación, verificación, discriminación y autentificación de hablantes. Denominaron *identificación* al proceso en el que la máquina debe asociar una muestra de habla desconocida

con la más similar de entre un número indeterminado de muestras conocidas o indubitadas. Hablaban de *verificación*, cuando la muestra anónima es aportada por el sujeto emisor de forma "cooperativa", es decir, existe un interés por su parte en que se verifique una comparación con otro registro de voz para obtener un resultado positivo. A este primer requisito añadían un segundo: la necesidad de una única muestra patrón predeterminada para llevar a cabo el cotejo.

Sadaoki Furui, ingeniero de NTT Human Interface Laboratories, en Tokio , y una de las autoridades del momento actual en sistemas automáticos de reconocimiento de locutores, deja perfectamente definidos los tres anteriores conceptos:

- *Reconocimiento de locutores* : Todo proceso *automático* de reconocimiento de

hablantes basado en la información individual incluida en la señal de habla. Dicho proceso se divide en Identificación y Verificación de hablantes.

- *Identificación de hablantes*: Proceso por el que se determina a quien pertenece la muestra anónima aportada, de entre un número de muestras registradas pertenecientes a distintos hablantes (indubitados) .

- *Verificación de hablantes* : Proceso de aceptación o rechazo de identidad a través de la voz, solicitado por un hablante.

En relación con dichos conceptos, Furui [1994] puntualiza :

" ...la diferencia fundamental entre identificación y verificación es el número de decisiones alternativas. En identificación, el número de decisiones alternativas es igual al número de sujetos de la población que conforma la base de datos, mientras que en verificación sólo existen dos decisiones alternativas, aceptar o rechazar, con independencia de la talla de la población"

Esta nomenclatura utilizada en los entornos de sistemas automáticos debe ser tenida en consideración en el ámbito forense. No obstante, desde éste otro punto de vista, resulta necesario formular algunas aclaraciones.

En lo referente a los términos de "Reconocimiento" de hablantes y " Verificación o Autentificación" parece existir unanimidad de criterios entre los expertos forenses de diferentes escuelas, si bien , son dignas de comentario algunas de las matizaciones por ellos formuladas. Harry Hollien, de la Universidad de Florida, añade al concepto básico de "Verificación" suscrito por Furui y los ingenieros de los laboratorios Bell, la buena calidad que siempre presentan las

muestras objeto de comparación en ese proceso.

Tosi [1979], que denomina *"autentificación"* a la verificación, precisa, que además de esa buena calidad de la que habla Hollien, el proceso de autentificación incluye siempre una base de datos con un número limitado de muestras indubitadas , y los canales utilizados para la transmisión y grabación de las muestras habladas son siempre los mismos, con lo que los efectos derivados del uso de distintas curvas de respuesta quedan eliminados.

Analizadas estas definiciones, parecen quedar suficientemente claros los conceptos relativos a sistemas automáticos de reconocimiento de voz. En lo respecta a la nomenclatura del campo forense, el término acuñado con carácter general para este tipo de tareas es el de *"Identificación de locutores o Speaker Identification"*. No obstante, este concepto que puede ser perfectamente válido en nuestro ámbito, requiere de ciertas aclaraciones. Para ello, podemos partir de una clasificación metodológica ideada por Tosi.

Tosi, con buen criterio, adopta el término general de *" Métodos de Identificación y eliminación "* para albergar todas las posibles formas de identificar a una persona a través de su voz. Bajo este concepto general, establece una primera dualidad : *Métodos subjetivos y métodos objetivos*. Define los primeros, como aquellos en los que la decisión final es tomada por la mente humana. En el caso de los denominados objetivos, dicha decisión sería producida por medios automáticos. A continuación, divide los métodos subjetivos en *"auditivos" y "espectrográficos"*. Considera como "auditivos ", aquellos en los que se utiliza el oído para efectuar la comparación entre muestras. Dentro de éstos, contempla dos tipos de situaciones. Por una parte estarían aquellos casos en los que la muestra dubitada se encuentra en la memoria a *largo plazo* del sujeto, y por otra, aquellos en los que el cotejo perceptivo-auditivo se produce de forma casi instantánea, es decir, utilizando la memoria auditiva a *corto plazo* (reproducción de pasajes sonoros registrados). El método auditivo a largo plazo es, para Tosi, el que lleva implícito un mayor grado de subjetividad y lo refiere a situaciones en las que el sujeto no tuvo ocasión de visualizar a la persona que produjo la emisión (normalmente víctimas y testigos). En este capítulo debieran también recogerse las identificaciones por registros de memoria a largo plazo de *voces familiares*, o lo que es lo mismo, voces muy conocidas para el sujeto que intenta establecer una asociación de identificación.

El método "espectrográfico" es aquel en el que se comparan las representaciones gráficas del habla , por medio de sonogramas o espectrogramas (por lo general con filtros de

representación de banda ancha).

Dentro de sus métodos objetivos (en los que se introduce como elemento diferenciador el *Aordenador≅*) Tosi establece tres escalones de menor a mayor objetividad:

- *Semiautomáticos*.- *" Aquellos en los que es necesaria una fuerte y continuada interacción entre el analista y la aplicación de cálculo o análisis que utiliza"*. En estos sistemas, el operador desempeña un papel determinante tanto en la selección de elementos para la comparación, como en la interpretación de los resultados finales del proceso.

- *Automáticos* .- Aquellos en los que la interacción analista-máquina es muy limitada. Sería el caso - anteriormente relatado - de los sistemas de "identificación", tal y como son definidos por los expertos en reconocimiento automático de locutores. El ordenador es alimentado con las muestras, y de acuerdo a una aplicación específica, establece una comparación y toma una decisión. Pueden ser dependientes o independientes de texto. Más adelante, cuando abordemos la clasificación actual de los métodos de identificación de voz, trataremos más ampliamente el capítulo de los sistemas automáticos.

- *Autentificación de voz,* también conocida como verificación. Representaría el máximo nivel de objetividad dentro de la escala completa de métodos de identificación-eliminación descritos. Como ya hemos comentado, se trata de un proceso dependiente de texto en el que se usan los mismos canales y medios de registro para la introducción de ambas muestras, y se produce una reclamación voluntaria de identificación por el emisor de la muestra dubitada.

Como bien dice Tosi, en cualquier caso, la presencia del experto siempre es necesaria con independencia de la naturaleza del método empleado. La preparación de muestras, el diseño de las aplicaciones de análisis, cálculo o clasificación, la interpretación de resultados o la defensa de las conclusiones alcanzadas en una vista oral, son algunas de las tareas en las que dicha presencia se hace imprescindible.

Por todo lo anteriormente expuesto, y tratando de adecuarnos al plano semántico de otras técnicas de investigación forense, podemos tranquilamente adoptar el término *AIdentificación de locutores≅*para referenciar con carácter general las diferentes metodologías que integran esta técnica. Muchos expertos forenses, especialmente los partidarios de la aproximación espectrográfica denominan a esta técnica *"identificación de voz"*. Aunque ello no debe ser considerado ni mucho menos como una incorrección, sí parece más correcto hablar de identificación de hablantes o locutores ya que en la actualidad las emisiones de voz son abordadas como algo más que una pura referencia físico-acústica, es decir, como un elemento

integrado en una estructura de comunicación (habla), con sus correspondientes factores expresivos e incluso emocionales.

No estamos muy de acuerdo con la terminología utilizada por Tosi al hablar de Amétodos≅ de identificación/eliminación por la voz (VI/E), cuando en realidad debiera de hacerlo de "sistemas de análisis" de VI/E. Como más adelante podremos comprobar, la metodología de Tosi está basada en la utilización de distintos *sistemas de análisis* que lógicamente conforman un método de VI/E. Es muy importante hacer esta diferenciación pues,

como ya comentaremos, algunos analistas utilizan con carácter exclusivo alguno de los sistemas de análisis arriba descritos (véase II.4) . En estos casos, sí podemos entender como método el sistema de análisis utilizado. Por otra parte, la clasificación que realiza Tosi aunque es una buena referencia, no incluye todos los sistemas de análisis existentes. Por ejemplo, no hace mención a aquellos tan importantes como los que basan su investigación en las características fonoarticulatorias desde el punto de vista de la Fonética Acústica.

II.1.4.- Clases de tests en la identificación de locutores.

Dependiendo del tipo de discriminación a efectuar, las tareas forenses de identificación de hablantes, pueden ser clasificadas en dos grandes grupos: casos *abiertos* y casos *cerrados*.

En los casos cerrados, el investigador se enfrentará a una comparación entre una voz anónima y un número determinado de voces conocidas, alguna de las cuales pertenece al mismo sujeto emisor que produjo la voz anónima. En los denominados casos abiertos, la comparación debe realizarse entre un registro anónimo y un número determinado de voces conocidas, entre las cuales puede o no puede, encontrarse una emisión correspondiente al mismo sujeto que produjo la locución anónima.

En la gran mayoría de los casos forenses la tarea de discriminación es abierta, y más concretamente del tipo *uno vs uno* (una voz dubitada contra una indubitada) donde el experto ha de decidir si ambas voces pertenecen o no a la misma persona. Paradójicamente, esta tarea que al discernir de los profanos puede parecer sencilla, implica una altísima responsabilidad, pues en el caso de decisiones positivas supone implícitamente el reconocimiento de la inexistencia de otra persona que pueda haber realizado la emisión cuestionada, y la asunción opuesta en el caso de una decisión negativa.

II.1.5.- El carácter variable de las emisiones habladas.Variabilidad intra/interpersonal.

A los tres ejes físicos que dimensionan el sonido - frecuencia, intensidad y tiempo - en el caso del habla, se les une un cuarto factor que aportará elementos decisivos desde un punto de vista identificativo: el efecto de resonancia producido en la cavidad resonante del tracto vocal.

En condiciones psicofisiológicas normales, todo locutor dispondrá sus órganos de la fonación en función del tipo de emisión que desee generar, ejerciendo un absoluto y voluntario control sobre los mismos.

Esta posibilidad de modificar la caja de resonancia a voluntad del sujeto emisor provocará uno de los peores obstáculos a los que el examinador forense deberá enfrentarse : *la variabilidad intrapersonal de los actos de habla.*

El habla no posee el carácter inmutable de las huellas dactilares. Algunos pioneros en el estudio forense de la voz interpretaron lo que denominaron el "voiceprint" o huella de voz - representaciones sonográficas- como una forma gráfica de representación con un valor de identidad individual y exclusivo, paralelo al del "fingerprint" o huella dactilar. Por este motivo, tales investigadores pensaron que el adiestramiento en la técnica de identificación por la voz sería un proceso relativamente corto y sencillo al igual que ocurre en el caso de las impresiones dactilares. Craso error. La voz es un objeto de estudio con carácter variable que requiere del uso de distintos enfoques de investigación para evidenciar las peculiaridades articulatorias relacionadas con cada hablante en particular. Más adelante, describiremos cuales son esas distintas perspectivas y objetivos de análisis.

Por todo ello, la metodología de estudio forense para las emisiones habladas -al menos hasta el momento presente- no puede ser confeccionada de acuerdo a los mismos criterios que se siguen en la comparación de muestras con una estructura invariable.

Otro problema añadido al ya descrito, es el derivado de la relación variabilidad intra/interpersonal.Muchas son las circunstancias que confluyen en la imposibilidad de producir dos actos de habla idénticos. Salvo en caso de eventos hablados registrados, no es posible producir dos emisiones habladas iguales. Este es el principio que sustenta la existencia de una variabilidad a nivel intra personal, además de la que lógicamente se produce entre locutores distintos.

Estas evidentes deducciones no tendrían mayor relevancia si en todos los casos la variabilidad intrapersonal de las emisiones habladas fuera de menor rango que la interpersonal. Pero,)es esto siempre así?)pueden darse emisiones muy distintas en el mismo locutor?; o lo que es peor,)pueden existir locuciones de alta similitud entre emisores diferentes?.Las soluciones a estos interrogantes aportarán las claves para solventar nuestro planteamiento de investigación.

Una de los objetivos fundamentales en cualquier metodología de identificación de voz -siempre que ello sea posible- es conseguir reducir los umbrales de variabilidad existentes entre las grabaciones indubitadas o pertenecientes a un sujeto conocido y las correspondientes dubitadas. Esta no es una labor sencilla. Existen multitud de factores de variabilidad en los distintos actos de habla, muchos de ellos atribuibles al sujeto emisor y otros, al resto de circunstancias de emisión, transmisión y grabación en las que se desarrollan los distintos procedimientos de la técnica.

Con independencia de su carácter voluntario o involuntario, transitorio o permanente, podemos citar como *causas fundamentales de variabilidad en la señal de voz*, los siguientes capítulos:

A.- FACTORES DEPENDIENTES DE LA NATURALEZA DEL HABLA Y DEL SUJETO EMISOR (VARIABILIDAD INTRAPERSONAL)

- 11 Variaciones no relacionadas con el plano de expresividad :

- **Contemporaneidad o no contemporaneidad** de las muestras objeto de estudio. El lapso temporal existente entre dos emisiones -incluso de un mismo día- [Garrett y Healey, 1987] implica siempre un nivel de variabilidad. Resulta evidente, que a mayor duración de tal intervalo se produzca un mayor índice de variabilidad. Multitud de estudios así lo indican. Ya en 1.937, la doctora Frances Mc Gehee [1937 y 1944] llevó a cabo un experimento relacionado con los procesos de reconocimiento de voz a nivel perceptivo por memoria a largo plazo. Sus resultados ponían de manifiesto un decremento gradual en la precisión de correctas identificaciones, que transcurría del 83% un día después de haber escuchado la voz, al 13% tras cinco meses después. Tosi y sus colaboradores [1972]observaron que un lapso temporal de un mes incrementaba de forma considerable (10%) los errores de falsa eliminación y falsa eliminación en decisiones de identificación de hablantes. Estos resultados según el mismo autor no son extrapolables de forma directamente lineal al transcurrir del tiempo, como parece lógico por otra parte. Autores como Endress, Bambach y Flosser [1971] publicaron conclusiones sobre descensos en la frecuencia fundamental ó F_0 (número de vibraciones por segundo de las cuerdas vocales) de individuos tras lapsos temporales de 29 años. En la misma línea de trabajo existen experimentos asociando los

cambios de F_0 en función de la edad y del sexo [Helfrich 1,979;Hudson/Holbrook 1.981;Krook 1988; Endres et al. 1.971; De Pinto/Hollien 1.982]. Según Angelica Braun [1995] se puede considerar fuera de toda discusión en el caso de hablantes *varones* la afirmación suscrita por Böhme y Hecker en [1.979] sobre el alcance de

la madurez en F_0 en torno a los 15 años; si bien, la muda definitiva se suele iniciar antes en el varón (13-14 años) que en la mujer (14-15) [Jackson Menaldi, 1992]. Igualmente en el caso de los varones, parece observase un descenso gradual del valor dicho parámetro hasta la edad de 40 años, volviendo a incrementarse entre los 60 y 80 años [Hollien/Ship 1972; Mc Glone/Hollien 1.963]. En el caso de las mujeres, según Hollien/Paul [1.969] el cambio por decremento de F_0 parece ocurrir con anterioridad a los 15 años, y Duffy [1.970] puntualiza que en torno al 43% de este cambio tiene lugar con posterioridad a tal edad. Según este mismo autor resultan más relevantes (en cuanto al valor del tono fundamental) los cambios producidos por el advenimiento de la menstruación, que los debidos a la edad cronológica. También existen numerosos estudios poniendo de manifiesto el efecto virilizante (con descensos en F_0) de la menopausia [Stoicheff 1.981; Morris/Brown 1.988; Saxman/Burk 1967; De Pinto /Hollien 1.982; Krook 1.988]. La disminución en la producción de estrógenos parece ser la causa de una reducción del control fonatorio con la consiguiente alteración de la frecuencia fundamental. En opinión de Braun estos cambios no resultan críticos desde un punto de vista identificativo pues acontecen de forma paulatina y muy lentamente.

Salvo en el caso del radical cambio consecuencia de ciertas parafonías (pubertad, vejez, etc.), el factor edad no parece ser un elemento determinante de variabilidad intrapersonal, entendiendo esta apreciación, en lo referente a cambios drásticos y bruscos de los componentes fundamentales de la voz.

Una vez alcanzada la estabilidad en la voz (en torno a los veinte años en el hombre y dieciocho en la mujer) no suelen producirse alteraciones fundamentales hasta edades muy avanzadas como es el caso ciertas gerifonías (p.ej. agravamientos de F_0 por causa de la edad). No obstante, sí existe un proceso lento y paulatino de agudización y agravamiento del tono fundamental, respectivamente asociable al envejecimiento en el hombre y la mujer.

Pudiera darse el caso de producirse alteraciones más significativas -en los componentes fundamentales del habla- entre distintos momentos de un mismo día (primera hora de la mañana y última de la tarde por ejemplo) que entre emisiones correspondientes a lapsos temporales mucho mayores.

- **Circunstancias relacionadas con cambios en el proceso y órganos de la fonación.** Pueden ser agrupadas en tres entornos distintos :

- *Entorno anatómico*.- Cambios en la dentadura, dislalias protésicas, rinofonías, disglosias, disfonías orgánicas en general: nódulos, pólipos, tumoraciones, etc.

- *Entorno fisiológico*.- Dislalias funcionales y orgánicas, disfonías funcionales (fonoponosis, fonastenias, etc.), procesos inflamatorios, catarros, irritaciones, alteraciones por hipo o hiper funciones de las glándulas de secreción interna, menstruación, menopausia, etc.

- *Entorno psicológico/neurológico*.- Disartrias, temblor temporal, disfonías de origen psicogénico, falsete mutacional, cambios emocionales, Efecto Lombard, etc

- **La influencia de agentes químicos exógenos**, como el consumo de medicamentos, drogas, alcohol, tabaco, etc, que también pueden incidir de alguna forma en los dos últimos entornos del apartado anterior.

Las posibles consecuencias del uso de la píldora anticonceptiva y el estado premestrual son dos ejemplos válidos de los anteriores entornos de variación pues provocan un aumento de masa en los pliegues glotales debido al mayor volumen de progesterona segregado. El resultado, un tono áspero con descenso hacia los graves.

- ***21 Variaciones relacionadas con el plano de expresividad***:

- **Modificaciones de rangos fonatorios y articulatorios estándar** relacionadas con variaciones sensibles de componentes fundamentales como el tono o la intensidad, grados de tensión y relajación en la articulación, grados de nasalización y oralidad, sonoridad y ensordecimiento, apertura y oclusión, velarización y palatalización, fricatización, bemolización, etc.

- **Alteraciones elocutivas de elementos fonéticos simples** , en donde estarían incluidas las múltiples realizaciones alofónicas de cada fonema o grupo fónico y sus efectos asociados de ataque, extinción, transición y coarticulación.

-Alteraciones elocutivas relativas al Atempo≅ y carácter suprasegmental o melódico-expresivo. Reseñaríamos aquí todas aquellas variaciones relacionadas con la entonación (ascendente descendente suspensiva) , acentuación, velocidad (ratios silábicos o articulatorios de las emisiones), fluidez, ritmo (normal, bradilálico, taquilálico), pausas (articulatorias, respiratorias, dubitativas, de selección de unidades léxicas, de configuración sintáctica, de examen o preparación de la información comunicativa, de cesión de turno conversacional, etc.)

-Variaciones de construcción lingüística y de códigos de relación comunicativa : construcciones a niveles morfo-sintáctico, estilística, recursos retóricos, elementos paralingüísticos o extralingüísticos (chasquidos con la lengua, etc.), léxico, idiomas, dialectos, dialectos verticales o jergas, proxémica (estructuración cultural por micro espacios de conducta comunicativa), variaciones diafásicas (alternancia de códigos de expresión formal, por ejemplo la alternancia de plano coloquial y técnico).

- Alteraciones transitorias de la cualidad de voz no relacionadas con las ya citadas sobre márgenes fonatorios o articulatorios estándar o algún tipo de disfunción. Estarían contempladas tanto las de carácter involuntario (voz quebrada, áspera, ronca, cavernosa, fañosa, etc.) como las de carácter voluntario: voz cuchicheada, imitada, fingida o disimulada, etc.

- Variaciones de los componentes de construcción emocional o comunicativa del discurso relativos a los niveles de excitación, equilibrio, exclamación, tristeza, temor, amenaza, ansiedad, furia, alegría, persuasión, etc.

Un estudio orientativo del factor variabilidad en diferentes aspectos (técnico, fisiológico y psicológico) se recoge en [Braun, 1995]. Aunque está referido a un único parámetro (frecuencia fundamental) resulta bastante indicativo de lo que puede acontecer a un nivel más general con otros elementos de variabilidad intrapersonal.

La *variabilidad interpersonal* de los actos de habla comprendería la totalidad de vertientes de variabilidad intrapersonal anteriormente descritas, y aquellas otras relacionadas con las constituciones y referentes psicofisiológicos de los órganos y mecanismos de la fonación en cada locutor. También ha de incluirse como elemento diferenciador interpersonal, el papel que representan los distintos procesos de aprendizaje del lenguaje hablado, a nivel individual.

Supongamos el caso de dos hermanos gemelos que lógicamente presentarán una muy similar configuración anatómico-fisiológica de sus órganos y mecanismos fonatorios. Los rasgos diferenciadores entre sus locuciones pueden resultar más patentes en los códigos de construcción a nivel lingüístico que en los índices acústicos y fonoarticulatorios. De ahí la importancia de utilizar una metodología que incluya diversos enfoques de análisis.

B.- FACTORES AJENOS A LA NATURALEZA DEL HABLA Y AL SUJETO EMISOR

En el ámbito de la acústica forense existen otros factores de variabilidad no relacionados con la propia estructura de las emisiones habladas o con las circunstancias de los sujetos que las producen. Los grabaciones objeto de estudio, en la gran mayoría de los casos son obtenidas mediante la interceptación de líneas telefónicas y, por esta razón, es necesario añadir a los descritos una serie de elementos que pueden actuar alterando o distorsionando en alguna medida la señal de voz :

- Alteraciones relacionadas con los canales y procesos de transmisión o conversión.- Ruidos asociados a la línea por causa de interferencias: radiodifusión, inducciones electromagnéticas, Ahum≅ (ruido de baja frecuencia derivado de la recurrencia de corriente alterna -50 ó 60 c.p.s.- y sus correspondientes armónicos y subarmónicos), saturaciones o distorsiones de transductores microfónicos o altavoces, etc . Además de los posibles elementos ruidosos han de considerarse las consecuencias derivadas de las distintas dinámicas, curvas de respuesta telefónicas, microfónicas, de cajas acústicas,etc. Tanto los procesos de conversión de señal (analógico/digital, digital/analógico) como aquellos otros de codificación o compresión digital en distintos formatos (MPEG, PASC, ATRAC, etc,) no suelen representar alteraciones críticas sobre la señal fuente.

- Variaciones por causa de las emulsiones y soportes magnetofónicos. Las diferentes calidades o características de una emulsión o de un soporte, con las consiguientes distintas propiedades físicas (densidad, composición, homogeneidad,etc.) mecánicas (resistencia a la rotura o elongación, etc.) magnéticas (magnetización, coercitividad, remanencia,etc.) y electroacústicas (respuesta en frecuencia, MOL, etc) así como las condiciones de almacenamiento o tracción de dichos soportes, pueden producir cambios sobre la naturaleza original del sonido registrado: restricciones en rango de frecuencia, ruido, distorsiones, dropouts, etc.

- Alteraciones producidas por equipos de adquisición, grabación o reproducción de eventos sonoros . Son muy numerosas las posibles alteraciones que debieran recogerse en este epígrafe. La causa es fácil de comprender, existen multitud de equipos en el mercado de diferentes características técnicas y calidades que, por tanto, ofrecen distintas prestaciones. No sólo debemos pensar en los equipos de grabación o reproducción, sino también en contestadores telefónicos automáticos, centralitas telefónicas analógicas o digitales de grabación de sonido, mecanismos de adaptación para la adquisición de datos sonoros (pick up devices), etc.

Las diferentes clases de equipos de grabación/reproducción proporcionarán grabaciones de una mejor o peor calidad, o con una mayor o menor riqueza de información sonora en función de su curva de respuesta y calidades de sus componentes. Asimismo, en función de la calidad, uso y mantenimiento de los equipos se pueden producir desajustes (carga insuficiente de baterías, desalineamientos azimutales o cenitales, magnetización o lesiones de cabezales, holguras de poleas o correas de transmisión, etc.) que provocan distintos tipos de distorsión sobre la señal origen; en algunas ocasiones -caso de variaciones de la velocidad de playback- las consecuencias pueden resultar críticas.

- Eventos sonoros simultáneos a la señal analizada. Dentro de un mismo registro pueden producirse solapamientos entre la señal de habla objeto de estudio y otros sonidos. Ruidos aditivos de distinta naturaleza, registros musicales, tonos puros o multifrecuencia, influirán de forma más o menos determinante en la deformación de la señal hablada de cara a su posible análisis en un proceso de identificación de locutores.

Cuando el suceso acústico simultáneo incluye voz (efecto "cocktail party" por ejemplo) el asunto se complica aun más, pues nos encontramos ante emisiones sonoras de similar estructura.

- Las diferentes arquitecturas acústicas y las ubicaciones de las fuentes de registro relacionadas con la grabación de emisiones habladas, pueden proporcionar una causa añadida de variabilidad. Reverberaciones, absorciones acústicas, etc., pueden afectar de forma más o menos importante al timbre e intensidad real de ciertas elocuciones.

Una vez referida la larga lista de factores de variabilidad a los que pueden ser susceptibles las emisiones habladas, puede surgir el planteamiento de si a pesar de estos inconvenientes resulta posible establecer una metodología fiable para el análisis de voz con

fines identificativos. Por supuesto que sí. El simple hecho de mantener una conciencia permanente de ese carácter variable, y el de conocer cuales son las posibles causas asociadas a dicho carácter, es el primer paso para conferir a nuestras evaluaciones un alto nivel de

rigurosidad.

La propia variabilidad de nuestro objeto de estudio será la que nos dicte las primeras pautas a la hora de establecer la admisibilidad de las muestras para su estudio, o la que nos indique que herramientas u opciones de análisis serán las idóneas para abordar con garantías de éxito nuestras tareas de identificación.

Cualquier estudioso de nuestro entorno de investigación, tarde o temprano se encuentra ante la misma pregunta:)es siempre mayor la diferencia interpersonal de los actos de habla a la existente a nivel intrapersonal?. Muchos consideran que en la respuesta a esta pregunta puede encontrarse la clave donde sustentar una metodología de identificación de hablantes. Probablemente tengan razón, sobre todo, si contextualizamos el problema en la dinámica forense, donde las decisiones de identificación/eliminación inciden en los derechos fundamentales de las personas. Encontrar la respuesta a esta cuestión no es una labor fácil de alcanzar. Decimos Aalcanzar≅ porque solamente puede ser obtenida con la experiencia que proporciona el análisis de numerosos casos reales a lo largo de muchos años. En este sentido, podemos afirmar sin temor a equivocarnos que, en términos generales, en unas condiciones aceptables de calidad y cantidad de información, y utilizando los procedimientos adecuados para efectuar las correspondientes evaluaciones, la variabilidad interpersonal será siempre mayor que la intrapersonal.

No obstante, no hemos de olvidar que las suficiencias cualitativas o cuantitativas que acompañan a los experimentos de laboratorio, no suelen acontecer en los casos forenses reales. Puede, por tanto, darse la circunstancia -y de hecho así acontece- de que ante escasas cantidades de discurso o fuertes variaciones de los planos de expresividad entre un registro dubitado y otro indubitado, ya sea debido a causas fisiológicas o emocionales, voluntarias o involuntarias, exista una variabilidad a nivel intrapersonal mayor que otra interpersonal. Ante estas situaciones, sólo la experiencia y pericia del experto podrán auxiliarle en la determinación de si es posible o no seguir adelante. En cualquier caso, más adelante abordaremos y precisaremos cuales son los criterios que se deben seguir a la hora de aceptar o rechazar una muestra anónima para su análisis.

II.2.-ANTECEDENTES HISTÓRICOS DE LA IDENTIFICACIÓN FORENSE DE LOCUTORES POR

EXPERTOS.

II.2.1.- Introducción

Existen numerosas referencias anecdóticas que son citadas por distintos estudiosos de la técnica como posibles antecedentes históricos. Una de las más famosas, quizás por ser la más lejana en el tiempo, es la relatada en la Biblia cuando Isaac reconoció la voz de su hijo Jacob, quien ayudado por Rebeca, su madre, imitó la voz de su hermano Esaú para obtener los derechos de primogenitura.

Al igual que en el citado caso de la Biblia, los simples reconocimientos a nivel perceptivo por víctimas o testigos constituyen la primera referencia en relación con la admisión de la prueba de identificación de personas a través de su voz por parte de los tribunales de justicia. Así, en 1660 un tribunal inglés estimó válido un testimonio de este tipo en el caso de un tal William Hullet. En 1861, un tribunal de los Estados Unidos consideró admisible la identificación de un perro por su ladrido. El argumento utilizado para tal sentencia relataba que si " *...una persona puede ser identificada a través de su voz, un perro también puede serlo a través de su ladrido*".

En este mismo sentido, J.Thornwald [1965]comenta que durante el período 1754-1780, cuando John Fielding ocupaba el cargo de jefe de los ABow Street Runners≅, siendo ciego, consiguió identificar a numerosos delincuentes por su voz.

Una referencia emblemática y clásica del reconocimiento perceptivo, más cercana en el tiempo, fue el conocido caso del secuestro Lindberg (A*Los Estados Unidos contra Hauptmann*≅). En 1.935 Charles Lindberg, héroe nacional en Estados Unidos por ser la primera persona que sobrevoló en solitario el Océano Atlántico, sufrió el secuestro y asesinato de su hijo. Bruno Hauptmann fue arrestado como presunto culpable de tal acción. Durante el juicio, Lindberg reconoció la voz de Hauptmann como aquella del secuestrador que dos años antes había podido escuchar personalmente y a través del teléfono. Esta identificación fue considerada válida por el tribunal, y al parecer, tuvo un peso importante a la hora de argumentar su sentencia que, por cierto, fue de muerte. La gran resonancia del caso y el cuestionamiento que en relación a este tipo de reconocimiento perceptivo por memoria a largo

plazo formuló la doctora Frances Mc. Gehee, [1937] profesora de Psicología en la Universidad John Hopkins, hicieron del asunto Lindberg una referencia popular -aunque primitiva- de la técnica de identificación de voz.

No es necesario ser un experto en ingeniería o fonética acústica para identificar o reconocer a otras personas a través de la voz. Todos somos capaces de hacerlo. En muchas ocasiones respondemos a una llamada telefónica y reconocemos al otro interlocutor sin necesidad de que éste se identifique. Este ejercicio de reconocimiento perceptivo está basado en la memoria a medio o largo plazo. Es decir, comparamos un registro cerebral almacenado en estas parcelas de memoria con los estímulos auditivos que estamos recibiendo en un instante determinado. Pues bien, esta forma de reconocimiento fue la utilizada en las primeras experiencias forenses anteriormente comentadas.

El primer salto cualitativo en el desarrollo de nuevos sistemas de análisis tiene un antecedente en los trabajos de Alexander Melville Bell, que en 1867 ideó una forma de representación gráfica de las palabras en función de como eran pronunciadas; algo así como una transcripción jeroglífica que resultaba muy descriptiva para personas no expertas en este área de conocimiento. Este sistema, bautizado como "visible speech" (habla visible), fue empleado tanto por su creador como por su hijo -el famoso Alexander Graham Bell- para hacer más funcional el aprendizaje del habla en las personas sordas.

Los laboratorios Bell situados en Murray Hill, New Jersey, han sido una importantísima referencia para los estudiosos de la identificación de locutores en general. Una larga serie de ingenieros de estos laboratorios han contribuido de forma muy relevante al desarrollo de la técnica con diferentes aportaciones. Entre otros, pueden citarse a los Sres. Bell, Potter, Kopp, Green, Gray, Kersta, Atal, Rosenberg, Doddington, Presti, etc.

Precisamente en [1.947] los doctores Potter, Kopp y Green publicaban un libro titulado "Visible Speech" tomando prestado el nombre ya empleado por Alexander Melville Bell. En este libro se pretendía instruir sobre la interpretación lingüística de los sonidos del habla representados en forma de espectrogramas o sonogramas. A diferencia del Sr. Bell, ellos habían codificado el habla a formas gráficas utilizando una máquina de reciente invención: *el espectrógrafo analógico de sonido o sonógrafo***.** Este aparato, permitía la representación del sonido hablado en una referencia tridimensional (frecuencia/amplitud/tiempo) mediante la realización de sucesivos análisis de Fourier a corto plazo en una muestra de voz.

Ya a principios de nuestro siglo se hicieron los primeros progresos con espectrógrafos de naturaleza mecánica, como fue el caso del analizador de Henrici. En [1.937] Black obtuvo espectros tridimensionales de fonemas vocálicos pertenecientes a locutores distintos para analizar la variabilidad interpersonal del habla.

La segunda guerra mundial dio el impulso definitivo para la creación del sonógrafo, herramienta de indudable utilidad en los laboratorios de acústica forense de nuestros días. En 1941, los laboratorios Bell iniciaron su diseño en un proyecto conducido por el doctor Ralph Potter. La finalidad prioritaria de la máquina era la de ayudar al ejército de los Estados Unidos en la identificación de operadores de radio alemanes para poder detectar la ubicación y desplazamientos de las distintas unidades enemigas. El procedimiento era sencillo: en primer lugar los servicios de inteligencia asociaban las voces de cada operador de radio con una determinada división alemana (por ejemplo, de tanques) ; posteriormente, si era reconocida la voz del operador mediante el uso del sonógrafo, con ayuda de la goniometría calculaban la posición de dicho operador y consiguientemente, la de la unidad del ejército donde éste servía.

En [1.944], los doctores Gray y Koop se mostraban absolutamente entusiastas con la posibilidad de la utilización de los sonogramas con fines identificativos. Fue en este preciso momento cuando acuñaron el término *" voiceprint "* (huella de voz) en un intento de equiparar la representación gráfica del sonido hablado a otra técnica de identificación forense por entonces ya consolidada : la huella dactilar o "fingerprint". Seguramente, ni por lo más remoto pudieron imaginarse las nefastas consecuencias que posteriormente ocasionaría la utilización de tal concepto.

Dos circunstancias, la llegada del fin de la guerra y la dificultad de registrar en aquellos días grabaciones de voz, hicieron caer en el olvido el proyecto iniciado en los laboratorios Bell. Como contrapunto favorable, también a partir de ese momento la nueva máquina dejo de tener una aplicación exclusivamente militar y quedó a disposición de los científicos estudiosos del habla.

Como podemos ver, casi todas las citas o relatos conocidos relacionados con los primeros pasos de la identificación forense de locutores tienen su origen en los Estados Unidos, país que debe ser considerado pionero en ésta técnica por muchas razones. Es sabido, que en la antigua Unión Soviética y algunos de sus países satélites, se inició el desarrollo de la técnica poco después del final de la segunda guerra mundial. No se conocen muchos datos desde entonces, pero teniendo en cuenta el actual estado de la cuestión, que posteriormente

analizaremos, el planteamiento formulado en estos países viene a converger con las soluciones aportadas por otras instituciones y expertos occidentales.

II.2.2.- Lawrence Kersta,)héroe o villano?

Pero volvamos a los Estados Unidos. En torno a 1.960 una nueva moda delictiva surgió en Nueva York. El departamento de policía de la ciudad empezó a recibir multitud de llamadas telefónicas sobre amenazas de bomba a compañías aéreas. En aquellos días, la grabación magnetofónica de sucesos sonoros era algo tan viable como prácticamente lo es en la actualidad, salvando las lógicas diferencias entre los soportes y equipos de grabación/reproducción de entonces y los de ahora. Uno de los inconvenientes que quince años atrás paralizó el análisis de voz con fines identificativos, había desaparecido.

En estas circunstancias, la policía de Nueva York solicitó la ayuda de los laboratorios Bell para capturar a los individuos que perpetraban las llamadas amenazantes. Un físico que había participado en los experimentos iniciales del sonógrafo, Lawrence G. Kersta, fue el designado para desarrollar un método fiable de identificación por la voz. Kersta necesitó dos años para presentar su método, al cual otorgó una fiabilidad del 99,65% [Kersta, 1962].

En síntesis, el método de Kersta se basaba exclusivamente en la comparación de los "patterns" (figuras de representación gráfica de la frecuencia y la amplitud en el dominio del tiempo) que aparecen en los sonogramas; un proceso similar al del cotejo de las huellas dactilares motivo, por el que probablemente Kersta decidió volver a utilizar el término "voiceprint" para denominar el sonograma de voz. Es importante añadir, que el método de Kersta no contemplaba la utilización del análisis perceptivo/auditivo a corto plazo.

Hasta tal punto estaba Kersta convencido de la infalibilidad de su método, que en el transcurso de la presentación oficial del mismo ante la Acoustical Society of America (1.962), llegó a equiparar sus índices de fiabilidad con los de las huellas dactilares. En el período 1.962-1.966 Kersta colaboró exitosamente con distintos departamentos de Policía y agencias federales (United States Air Force, Civil Aeronautics Board, Federal Aviation Agency y otras de carácter confidencial).

En 1.966, Kersta abandonó los laboratorios Bell y creó su propia empresa: " Voiceprint Laboratories, Inc." . En esta nueva etapa, ofrecía distintos servicios de aplicación a casos

forenses: perito en identificación de voz para testificar de cara a los tribunales, procesado de señal, transcripciones sobre registros etc. En la introducción del catálogo de presentación comercial de sus productos, el "ex" de la Bell, dejaba patente su filosofía sobre la técnica de identificación de voz: *"...de la misma forma que la identificación dactilar se basa en las características individuales que aparecen en las impresiones de las huellas dactilares de las*

personas, la identificación por "voiceprint" tiene su fundamento en las características individuales que se ponen de manifiesto en las impresiones espectrográficas de las emisiones habladas de los sujetos. "

Aseveraciones de esta índole, conformarían el caldo de cultivo del que surgió un cisma metodológico en la identificación forense de locutores. Cisma, que iba a iniciarse en los años sesenta y que aun en nuestros días, en algunos casos, continua vigente.

Además de las mencionadas prestaciones, la empresa de Kersta ofrecía cursos de formación de expertos y producía sonógrafos con fines comerciales en competencia con la firma " Kay Elemetrics Co."

Kersta inició su primer curso de adiestramiento en 1.967. Asistieron al mismo dos miembros de la Policía Científica del estado de Michigan y Oscar Tosi como asesor de dicho departamento policial, en calidad de evaluador de los procedimientos utilizados por Kersta. En los meses sucesivos, aproximadamente veinte personas -la mayoría de ellos miembros de las fuerzas de seguridad- atendieron estos cursos cuya duración era de dos semanas. No todos los participantes concluyeron exitosamente este período, que por otra parte, debía ser complementado con otro, de dos años, en el que se efectuaba un entrenamiento continuado de comparación visual de espectrogramas bajo la supervisión de Kersta.

Como consecuencia de su carácter de prionero, Kersta, considerado primer perito que testificó ante un tribunal como experto en identificación de voz [*State v. Rispoli and Straehle, 1.966*] cometió diversos errores (alguno de ellos de dramáticas consecuencias para nuestra técnica).

En 1.973 Voiceprint Laboratories Inc. fue a la quiebra, siendo comprados sus derechos por William Hughes quien fundó "Voice Identification Inc." con el objetivo fundamental de continuar con la producción comercial del sonógrafo, aunque también de forma ocasional se ofrecían servicios periciales de identificación de voz, ya que parte del personal de la empresa de Kersta fue absorbido por la nueva compañía. Voice Identification Inc. sigue en la actualidad

trabajando en la técnica de identificación de voz, no conociéndose referencias comerciales suyas en lo relativo a la producción de sonógrafos, cuyo mercado está dominado por la firma Kay Elemetrics (probablemente por haber sabido anticiparse en su adaptación tecnológica al campo digital).

Kersta actuó como perito ante los tribunales en ocho ocasiones. En el caso [*People vs*

King (1.968)] (sobre incendio y pillaje en un barrio de Los Angeles) durante una entrevista televisiva alguien que no mostraba su cara a la cámara se hacía responsable de tales hechos delictivos. Kersta, que actuaba para el Fiscal, comparó este registro de habla con muestras de un individuo sospechoso al cual identificó. La defensa utilizó siete peritos (ingenieros y fonetistas) en contra de Kersta. Uno de ellos, el Dr.Peter Ladefoged de la Universidad de California, atacó exitosamente el método del voiceprint poniendo de manifiesto las importantes carencias del físico en el campo de la ciencias del habla. El acusado fue absuelto.

Si este caso representó para Kersta su tumba como perito y empresario en el ámbito de la identificación forense de hablantes, a nivel general supuso el desencadenamiento de una catástrofe. A partir de este momento, diversos científicos del campo de las ciencias del habla que hasta entonces estaban siguiendo de cerca el nacimiento de una técnica cercana a su conocimiento pero dominada por los ingenieros y físicos acústicos, se abalanzaron sobre una metodología que de forma involuntaria, pero injusta, les había ignorado.

Kersta cometió tres errores fundamentales:

11.- Situó en el mismo plano de infalibilidad la identificación dactilar y la identificación de voz, cuando los objetos de estudio de cada una de estas ciencias forenses presentan una naturaleza opuesta. La huella dactilar aunque diversiforme, es inmutable y perenne. La emisión hablada tiene siempre un carácter variable pues, salvo en el caso de las locuciones registradas, no resulta posible la emisión de dos actos de habla idénticos.

21.- Basado en insuficientes referencias experimentales, utilizó el análisis sonográfico como un método exclusivo, ignorando no sólo la perspectiva de estudio a nivel fonético, sino también la del análisis perceptivo/auditivo a corto plazo.

31.- Tenía la convicción de que cualquier miembro de las fuerzas de seguridad estaba capacitado para desarrollar la técnica de identificación de locutores tras un entrenamiento fundamentado en el único criterio de la comparación de "patterns" sonográficos.

Diversas causas se conjugan en el origen de estos críticos errores. La personalidad vanidosa de Kersta, un precipitado y ambicioso proyecto comercial forense, o las dificultades propias de los nuevos retos, podrían citarse como algunas de ellas. Aunque sin lugar a dudas, la columna vertebral de su fracaso fue la ausencia de un enfoque multidisciplinar del problema.

Esta pesada herencia que Kersta nos legó, aun en nuestros días utilizada por ciertos ignorantes voluntarios o involuntarios del progreso metodológico en la técnica, no debe

representar un impedimento a la hora de expresar el reconocimiento al primer científico que tomó la decisión de presentar ante los tribunales de justicia un nuevo instrumento de indudable utilidad para el esclarecimiento de muchas actividades delictivas.

II.2.3.- El surgimiento de una técnica. La referencia U.S.A.

La evolución histórica de cualquier técnica forense está en buena parte determinada por sus propias consecuencias a nivel judicial. En última instancia, las sucesivas sentencias emitidas por los órganos jurisdiccionales son las referencias válidas para evaluar su verdadero índice de fiabilidad. En este sentido, y volviendo a los orígenes judiciales de la técnica, las primeras valoraciones de admisibilidad recogidas en las sentencias de los tribunales de justicia de los Estados Unidos, provocaron los primeros posicionamientos metodológicos en relación con la práctica del método espectrográfico.

Los primeros testimonios de expertos ante los tribunales U.S.A. se producen en 1.966: [*ANueva York contra Rispoli y Straehle≅ (Kersta)*] y [*AU.S. Securities and Exchange Commission contra Klopp≅ (Kersta y Tosi)*] . No obstante, la primera vez que es emitida una decisión judicial para regular la admisibilidad del análisis forense de identificación de locutores realizado por expertos, acontece en 1.967. La sentencia es dictada por un tribunal militar en el caso [*"United States v. Wright"*]. El Juez introduce por vez primera la referencia de aceptación de evidencias científicas del caso *" Frye v. United States"* (1.923). El conocido " Frye test o Frye rule " fue también el standard de admisibilidad utilizado en el caso que provocó el declive como experto de Kersta (*"People v. King"*).

La referencia "Frye" fue dictada en [1.923] por un tribunal de apelación del distrito de Columbia para rechazar como evidencia admisible una especie de " test de la verdad " basado en un control de la presión sistólica de la sangre (un antecedente del actual "polígrafo"). En síntesis, el standard "Frye" señalaba que " *...cuando un nuevo principio o descubrimiento*

científico es utilizado ante los tribunales para demostrar alguna evidencia, éste, debe contar con la general aceptación de la comunidad científica de su entorno. "

También en 1.967 se produce la primera sentencia sobre la admisibilidad del método espectrográfico por parte de un tribunal no militar. Fue el Tribunal Supremo de New Jersey en resolución de apelación del caso [*" State v. Cary"*] . En esta ocasión, el dictamen no fue contrario a la fiabilidad de la nueva técnica forense aunque se estimaba era demasiado pronto para poder determinar si dicha técnica contaba con el requerido respaldo de aceptación científica

exigible a este tipo de evidencias.

La insuficiente evaluación del método espectrográfico en sus primeros pasos de aplicación práctica sobre casos reales y la ausencia de una referencia rigurosa de previa experimentación, fueron los principales argumentos esgrimidos en contra de su fiabilidad. En orden a paliar estas insuficiencias, el Departamento de Ciencias del habla y Audiología de la Universidad del Estado de Michigan (M.S.U.) subvencionado por el Departamento de Justicia de los Estados Unidos, desarrolló un largo experimento durante tres años [Tosi et al., 1972]. El padre y responsable de tal proyecto fue el Doctor en Ciencias Físicas Oscar Tosi. En dicho estudio -que más tarde abordaremos con detalle- se efectuaron 34.992 evaluaciones de identificación/eliminación espectrográfica de acuerdo a diferentes modelos de un diseño experimental. Aunque dicho experimento se llevó a cabo en un marco de laboratorio, ciertas condiciones forenses fueron consideradas en los distintos modelos: ruido, transmisión telefónica, no contemporaneidad de las muestras, tipos de tests, etc. . Concretamente , 11.664 del total de comparaciones, eran del tipo forense.

Simultáneamente al desarrollo de este experimento, la Policía del Estado de Michigan trabajó en casos reales bajo la supervisión y directrices de Tosi, si bien es cierto , que las conclusiones alcanzadas en relación con estos casos no constituyeron una evidencia legal durante este período de experimentación.

Como complemento a estos estudios, Tosi y Greenwald [1978] realizaron otro experimento contemplando la influencia de diversos factores en las tareas de identificación auditivo-espectrográficas: lapso temporal, sexo y entrenamiento del experto, etc.).

En 1.970 concluyen los estudios desarrollados por la M.S.U. y la Policía del mismo Estado con un balance altamente favorable en favor de la utilización del método espectrográfico con fines identificativos forenses. Por este motivo, el Departamento de la

Policía del Estado de Michigan decidió crear la que sería primera unidad policial de investigación en Identificación de Voz en los Estados Unidos (exceptuando el caso F.B.I.). Al mando de dicha unidad estaba el teniente Ernest Nash que fue la persona que había colaborado con Tosi realizando y coordinando los trabajos sobre casos forenses reales durante el período experimental.

Aunque Tosi había testificado en contra del uso del método espectrográfico de identificación de hablantes en casos concretos, tras el experimento de Michigan (1.970) su

propuesta metodológica se orientó hacia la utilización de la técnica auditivo-espectrográfica. Esta técnica combinada, fue utilizada con éxito ese mismo año en el caso [*AMinnesota v. Trimble≅*] donde el Tribunal Supremo de Minnesota reconoció como fiable la evidencia presentada por Tosi. Incluso el Dr. Ladefoged que actuaba en este caso como perito de la defensa -Tosi y Nash lo hacían para el Fiscal- estuvo de acuerdo en que en determinadas ocasiones era perfectamente factible la utilización del método auditivo-espectrográfico para identificar a una persona a través de su voz (recordemos que Ladefoged en otros casos se pronunció contrario al empleo de esta técnica).

Con la intención de institucionalizar diferentes aspectos y conceptos relacionados con la técnica, Kersta, Nash, Tosi y un asesor legal, fundan en 1.971 la I.A.V.I. (Asociación Internacional de Identificación de Voz). Inicialmente se proponen tres objetivos fundamentales: la formación y cualificación de expertos, el fomento de la investigación y el establecimiento de un código de ética para la práctica de la identificación de voz. Diversos especialistas fueron formados y participaron en numerosas vistas orales . En julio de 1980 la I.A.V.I. se integró en el VIAAS (Voice Identification & Acoustic Analysis Subcommitee) de la International Association for Identification (I.A.I.).

A principios de 1.978, tribunales estatales y federales de veintitrés estados U.S.A. , y otros de Canadá, Italia e Israel, habían admitido como evidencia la identificación de voz por examen auditivo-espectrográfico [Tosi, 1979]. Entre tanto, los detractores del método Kersta fueron ampliando sus críticas de una forma sistemática e injustamente generalizada a metodologías más desarrolladas que incluían en sus sistemas de análisis la comparación de patterns espectrográficos.

Consecuencia de este tira y afloja se produjeron distintos resultados sobre estudios de campo y veredictos judiciales en favor y en contra del análisis auditivo-espectrográfico, que provocaron cierta confusión sobre la fiabilidad de la técnica en cuanto a su aplicación legal.

La frecuentemente aludida regla Frye, que exigía la aceptación del método por la comunidad científica correspondiente, no determinaba sin embargo que comunidad científica era la competente para cada caso, y teniendo en cuenta el carácter multidisciplinar de la identificación de locutores, este último aspecto se tornó todavía más complejo.

Años atrás, a finales de los 50, el Federal Bureau of Investigation (F.B.I.) inició con carácter confidencial sus investigaciones sobre el análisis espectrográfico, obteniendo buenos resultados en la aplicación del mismo para la resolución de diversos asuntos de investigación interna. En 1.976, y ante la situación de confusión existente en los tribunales, el F.B.I. solicitó a

la National Academy of Sciencies un dictamen sobre la fiabilidad del método espectrográfico y su utilización como evidencia ante los tribunales de justicia.

Para dar cumplida respuesta a la solicitud efectuada por el F.B.I., la Academia Nacional de las Ciencias ordenó un estudio para evaluar la fiabilidad del método espectrográfico. Al frente de este proyecto estaba Douglas L.Hogan del Consejo Nacional de Investigación. Fue constituida una comisión de ocho expertos independientes (Cooper, Green, Hamlet, Hogan, Mc Knight, Picket, Tosi y Underwood) cuyas respectivas especialidades abarcaban tanto el ámbito legal como el científico (Derecho Penal, Leyes de la evidencia en general, Acústica, Electrónica, Ciencias del habla, Patologías, etc.) ; como chairman de dicha Comisión fue designado el Dr. R.H.Bolt.

En 1.979 la Comisión expresó sus resultados en un informe titulado *"On the Theory and Practice of Voice Identification"* [Bolt et al.,1979]. En su conclusión final no se pronunciaban ni a favor ni en contra del uso forense del método auditivo-espectrográfico, efectuando, eso sí, la recomendación de que si el método era utilizado de cara a los tribunales, deberían quedar claramente referidas ante el Juez o el Jurado tanto las limitaciones del mismo, como la cualificación y entrenamiento del experto que lo practicase: *AThe Committee takes no position for or against the forensic use of the aural visual method of Voice Identification, but recommends that if it is used in testimony, then the limitations of the method should be clearly and thoroughly explained to the fact finder, whether judge or jury≅.*

Tras conocer los resultados del informe de la Academia Nacional de las Ciencias, el F.B.I. continuó utilizando el método de identificación auditivo-espectrográfico sólamente de cara a su propia investigación, o como auxilio a cualquier otra fuerza de seguridad que lo requiriese para idénticos fines.

En 1.986 el Federal Bureau of Investigation publicó un estudio evaluando los resultados obtenidos con la utilización de su método en casos reales durante un período de quince años. El análisis comprendía 2000 comparaciones de identificación de voz realizadas por diez de sus expertos, todos los cuales, eran licenciados en Ciencias y habían completado como mínimo dos años de experiencia continuada en el desarrollo de dicha práctica. Los resultados ofrecían ratios de error inferiores al 1% . [Koenig, 1986]

El hecho de significar especialmente los resultados obtenidos por el F.B.I., cuando existen numerosos estudios de laboratorio evaluando la eficacia del método ante distintas

circunstancias [Kersta 1962; Young & Campbell 1967; Stevens et al., 1968; Tosi et al. 1972, 78 y 79; Bolt et al. 1970 y 1.973; Hennessy 1.970; Endrees et al. 1.971; Hazen 1.973; Black et al. 1.973; Smrkovski 1.975 y 1.976; Hall 1.975; Obrecht 1.975; Hollien & Mc Glone 1.976 y 77; Reich et al. 1.976 y 1.979; Rothman 1.977; Houlihan, 1.979; Greenwald 1.978 y 1.979, etc.] obedece a varias razones. Normalmente, en los estudios anteriormente referenciados la muestra es insuficiente, o están referidos a aspectos concretos del análisis (disimulo de la voz, influencia del lapso temporal, etc.) y por lo general no han sido realizados sobre casos reales o por expertos con una suficiente cualificación y trayectoria profesional en este tipo de casos.

Por otra parte, tampoco tiene mucho sentido entrar a analizar los distintos resultados y su carácter contradictorio o no contradictorio pues en la mayoría de los casos las referencias de análisis utilizadas no son comparables al existir severas descontextualizaciones de distinta índole: objeto de estudio, tipo de expertos que lo realizan, tamaño y clase de la muestra, número de comparaciones, reglas de decisión, etc.

A todos estos factores hay que sumar la evolución acontecida en el propio método espectrográfico desde sus comienzos con Kersta hasta el desarrollado por el F.B.I. y otros expertos policiales o privados en el momento del informe de la National Academy of Sciencies (1.979). Igualmente, habría que considerar las diferentes referencias y rigor de los análisis utilizados por unos y otros en el desarrollo de sus tareas.

En definitiva, debe de quedar muy claro que todo intento de agrupar en un mismo saco a todos los usuarios del análisis espectrográfico (expertos, especialistas, iniciados, charlatanes y perjuros), sin tener en cuenta las necesarias matizaciones de contextualización supondrá, en el mejor de los casos, un acto de manifiesta injusticia.

A mediados de los años setenta la identificación forense de locutores experimenta un cambio trascendental en su filosofía. Mientras unos perfeccionan el análisis espectrográfico y otros enfocan la cuestión desde otras perspectivas con un carácter de exclusividad (fonética acústica, reconocimiento automático o semiautomático, etc.) en la Universidad del Estado de Michigan el Dr. Tosi propicia el caldo de cultivo de lo que serán las nuevas tendencias metodológicas.

II.2.4.- Primeros pasos de la identificación de voz fuera de los EE.UU.

Hasta este momento, hemos analizado los inicios de la identificación de hablantes por expertos en el ámbito de los Estados Unidos. Hemos situado las primeras referencias a finales de

los años sesenta, pues precisamente en esta época comenzaron a producirse las primeras consecuencias de admisibilidad por parte de las instituciones de justicia de este país.

Los primeros datos fuera de los Estados Unidos son detectados en la antigua Unión Soviética poco después de la segunda guerra mundial [Solzhenitsyn, 1968]. No obstante, los planteamientos de aplicación práctica de la técnica no comienzan a desarrollarse hasta finales de los sesenta. Entre los primeros investigadores interesados en abordar el problema podemos mencionar a J. Ramisvili en la U.R.S.S. , S. Blasikievicz y Wojciech Majewski en Polonia, H. Habersbrunner en Alemania, Ion Anghelescu en Rumanía o los doctores Masao Onisi y Seiki Miyoshi en Japón.

En marzo de 1.963, un muchacho de cuatro años fue secuestrado en Tokio. La voz de su secuestrador fue grabada a través del teléfono en varias ocasiones y la policía solicitó la ayuda de fonetistas, lingüistas e ingenieros acústicos para analizar la voz del sospechoso. Esta fue la primera ocasión en que la identificación de locutores por expertos fue utilizada con una finalidad forense en Japón. Durante los siguientes años, las conclusiones de identificación de locutores tuvieron una aplicación exclusiva de apoyo a la investigación policial; posteriormente, fueron ya utilizadas de cara a los tribunales de justicia, siendo admitidas como evidencia a partir de 1.977.[I.R. n11]

A principios de los setenta, el Instituto Federal Físico-Técnico de Alemania confeccionó sus primeros trabajos basándose en el método auditivo-espectrográfico. Posteriormente, el Dr. Ernest Bunge al servicio del Kriminaltechnisches Institut del Bundeskriminalamt o B.K.A. (Policía Federal de Alemania), supervisó un método automático de identificación de voz

conocido como sistema "AUROS" (Automatic Recognition of Speakers) que aunque según su creador proporcionaba excelentes resultados en unas condiciones determinadas de laboratorio (del orden del 99,5 %) cuando se aplicó a la casuística real forense fue desestimado en favor de otra perspectiva de estudio fundamentada en el análisis auditivo-lingüístico. En base a este nuevo enfoque, la identificación de voz fue admitida como prueba ante los tribunales de justicia alemanes en 1.981 [I.R. n12].

Las referencias más antiguas de Europa en relación con la admisión de la evidencia de I/V se remontan al año 1.971 en la ya extinta Unión Soviética. En dicho año, el Laboratorio de Fonoscopia del Centro de Criminalística del Ministerio del Interior de la actual Rusia inició oficialmente sus actividades en el campo de la identificación forense de hablantes. Aunque pudiera suscitarse alguna reticencia de tipo político en relación a ello, es necesario aclarar que

especialmente desde principios de los ochenta, en Rusia se desarrolla una importante y extensa labor en relación con nuestra técnica, teniéndose constancia de la existencia de unos cincuenta laboratorios públicos en donde realizan sus trabajos de identificación de locutores ciento cincuenta expertos.[I.R. n13]

Al igual que en el caso de Alemania, la policía de Italia inició sus actividades de una forma sistemática a principios de los setenta. Con la década de los ochenta, otros laboratorios forenses policiales o de auxilio a la Justicia, iniciaron su andadura en el campo; es el caso del Gerechtelijk Laboratorium en Holanda, el Laboratorio de Acústica Forense del Cuerpo Nacional de Policía en España o el laboratorio de Investigación Acústica de la Academia de las Ciencias de Austria. A principios de los noventa se incorporaron nuevos laboratorios en el seno de instituciones públicas policiales de otros países europeos: Policía Técnica y Científica de Francia, Policía Judicial de Bélgica, Crime Laboratory del N.B.I en Finlandia, Laboratorio de Fonoscopia del M1 del Interior en Lituania, etc.[Documento ENFOPOL,144].

En el resto del mundo -exceptuando el caso de Canadá donde la técnica se comenzó a desarrollar por parte de la Policía Montada y algún laboratorio privado en 1.974- [I.R. n15] los nuevos laboratorios públicos y privados comienzan a practicar la técnica desde comienzos de los ochenta (algunos de ellos serán reseñados en el capítulo siguiente).

Como veremos a continuación, las soluciones al problema de la identificación forense de locutores, no fueron conducidas del mismo modo por los distintos expertos, pero sí es cierto, que casi todos ellos contribuyeron de alguna manera a la consolidación de esta técnica, compleja y apasionante.

II.3.- EL ESTADO DE LA CUESTIÓN EN LA ACTUALIDAD

II.3.1.- Estados Unidos.

Cuando hablábamos de los antecedentes históricos, dejamos la situación de la técnica en USA a finales de los años 70. Comentábamos que el Dr.Tosi con su nuevo enfoque metodológico sentó en cierta forma la filosofía en la que se sustentan las bases de las tendencias actuales.

En un momento dado, y debido a circunstancias de distinto tipo, Tosi se desmarca del ámbito de la I.A.V.I. e inicia una nueva andadura. De alguna forma intuyó que la utilización del

método auditivo-espectrográfico como método exclusivo era en ese momento insuficiente para abordar una tarea de identificación del locutor en toda su plenitud. Su cambio de enfoque coincide con la circunstancia de que a principios de los ochenta los ordenadores comienzan a ser una herramienta funcional y asequible para el desarrollo de distintas aplicaciones informáticas menores.

En síntesis, las aportaciones más relevantes hechas por Tosi pueden referirse a tres aspectos. En primer lugar, desarrolla e incorpora los sistemas de análisis automático TOSI I y TOSI III [Tosi et al., 1977] a su metodología de identificación. Asimismo, tomando como referencia los resultados de su experimento en la M.S.U., logra extrapolar mediante su curva P.S.S.(escala de probabilidad subjetiva) [Tosi, S.F.] los valores de similitud/disimilitud a valores de probabilidad, en un intento de objetivar en la medida de lo posible las valoraciones subjetivas de comparación efectuadas por el experto. Y por último, introduce unas reglas de decisión que permiten reducir notablemente el margen de error en las evaluaciones.

Aunque tanto los sistemas de reconocimiento automático TOSI I y TOSI III como la curva P.S.S. fueron considerados en su momento aportaciones muy novedosas y funcionales, lo realmente importante y reseñable es el hecho de la utilización de una metodología que combinaba distintas perspectivas de análisis e intentaba reducir el índice de subjetividad de las apreciaciones perceptivas de los expertos.

El "método Tosi" hasta 1.992 [I. C.n16] sistemáticamente había soslayado el enfoque de análisis complementario basado en el estudio fonético-lingüístico debido, entre otras cosas, a enconadas diferencias mantenidas ante posicionamientos radicales fundamentados en la exclusiva utilización de este tipo de análisis.

En 1.992, con ocasión de un curso de formación que impartió en la M.S.U. a miembros del laboratorio de Acústica Forense de la Policía Científica española, llegó al convencimiento de que el enfoque de análisis fonético-lingüístico -ampliamente desarrollado por estos alumnos en España- era de indudable utilidad dentro de cualquier metodología de identificación de hablantes; desde ese momento lo incluyó como un capítulo más de sus informes periciales.

El Dr. Óscar Tosi falleció el 4 de enero de 1.994 en East Lansing, Michigan, dejando un importante legado a aquellos que han sabido, o han tenido la oportunidad de interpretar correctamente sus criterios.

En la década de los ochenta, la identificación forense de locutores experimenta una importante actividad en los Estados Unidos. La técnica es desarrollada por los usuarios del método espectrográfico, aunque cuentan con la oposición de algunos fonetistas.Desde un punto

de vista institucional oficial, el F.B.I. es el organismo de referencia. Durante muchos años la investigación y desarrollo del método auditivo-espectrográfico en el F.B.I. estuvo a cargo del agente especial Bruce E. Koenig quien impulsó el uso del mismo durante su estancia en la División de Servicios Técnicos. Posteriormente, abandonó el F.B.I. para continuar trabajando en la empresa privada.

En la actualidad, la Sección de Tecnología y Vigilancia Electrónica del F.B.I. es la responsable del área de Identificación de Voz. Está comandada por el Dr. Hiro Nakasone, alumno de doctorado y discípulo de Tosi. Desde hace 40 años siguen trabajando con el método auditivo-espectrográfico, siendo norma del departamento no acudir a testificar a los tribunales y utilizar dicho método solamente como apoyo a la investigación de sus propios casos o de aquellos en los que son requeridos por otros organismos policiales o fuerzas de seguridad.[I.C.n15]

En el ámbito privado, diversos expertos trabajan en la técnica forense de identificación de hablantes en los Estados Unidos. La gran mayoría utiliza el método espectrográfico y están agrupados en el Subcomité de Análisis Acústicos e Identificación de Voz (VIAAS) de la International Association for Identification (I.A.I.).

En el aspecto legal, los estándares de admisibilidad para la evidencia de identificación del locutor en USA han incorporado una importante novedad. Antes de 1.993 se habían aplicado a este tipo de evidencia tres referencias de admisibilidad: el Frye test, las Reglas Federales para la Evidencia y las reglas para la evidencia existentes en algunos Estados. La regla Frye ha sido reiteradamente criticada por no ser considerada el test apropiado para evaluar el uso de la evidencia de identificación por la voz. Como ya sabemos, este estándar fue establecido y aplicado para la admisión de un tipo de evidencia muy diferente. En realidad, lo que se dilucidaba era la validez como prueba ante los tribunales de justicia de un procedimiento científico que determinaba si una persona decía o no la verdad. En este último caso la prueba incide directamente en el terreno de la investigación, mientras que en el caso de la voz estamos ante una evidencia pura de identificación, como es el caso de las huellas dactilares, balística, huella genética etc. Por otra parte, la regla Frye no determina cual es la comunidad científica competente para concluir en la aceptación o no del método científico y, en cualquier caso, como afirma Mc Cormick [1984]*A... la aceptación por parte de la generalidad de una comunidad científica es una referencia judicial sobre un hecho científico concreto, pero no un criterio para la admisibilidad o no de una evidencia científica "*

En 1.993 el Tribunal Supremo de los Estados Unidos cambió su norma de referencia para

la admisibilidad de las evidencias emitidas por expertos científicos, rechazando el Frye test como inconsecuente con las Reglas Federales para la Evidencia. Ocurrió en el caso [*Daubert vs. Merrell Dow Ph*.**,** 1993]. El tribunal determinó que las Reglas Federales para la Evidencia y no la referencia Frye eran el estándar para determinar la admisibilidad del testimonio de un experto científico. La "general aceptación" del test Frye fue sustituida por las Reglas Federales, concretamente la *Regla 702* es considerada como el estándar apropiado para evaluar la admisibilidad de la evidencia científica: *" Para poder cualificar un conocimiento científico, cualquier conclusión o afirmación emitidas deben deducirse de un método científico. El testimonio referido debe sustentarse en la correspondiente validación (por ejemplo una sólida formación en relación con el área de conocimiento sobre la que se opina). En definitiva, el requisito de que el testimonio de un experto pertenezca al conocimiento científico, establece por sí mismo un estándar de fiabilidad evidenciaria."*

Hasta el presente año, tribunales de treinta estados USA han admitido la identificación de locutores por el método auditivo-espectrográfico como evidencia; en siete estados la prueba ha sido en alguna ocasión desestimada, y en quince de ellos no se han emitido sentencias en uno u otro sentido.

II.3.2.- Europa y resto del mundo

En el momento actual, podemos afirmar sin temor a equivocarnos que la técnica de identificación de hablantes está plenamente consolidada en Europa. Incluso, puede hablarse de una situación de vanguardia en cuanto a investigación y desarrollo metodológico.

Las referencias válidas se sitúan fundamentalmente en los laboratorios policiales, si bien, existen algunas Instituciones universitarias y empresas o expertos privados que de forma eventual -en la mayoría de los casos- colaboran con estos organismos públicos o trabajan directamente para los tribunales de justicia.

En las dos últimas décadas, el desarrollo de la técnica en Europa ha surgido y evolucionado de forma distinta en cada uno de sus países, pero curiosamente, existe una orientación metodológica común en la gran mayoría de los laboratorios europeos a pesar de que las experiencias de iniciación en cada uno de ellos fue prácticamente independiente y en muchos casos autodidacta. Estamos hablando de la utilización de los que denominamos *métodos combinados*. Esta solución, a la que se ha llegado por diferentes vías de investigación y que

evidentemente no responde a una casualidad es, sin lugar a dudas, la mejor posible en el presente momento.

Los métodos combinados (*ver II.4.6*) son la consecuencia lógica de largos años de estudio invertidos en la búsqueda de la solución más idónea al problema de la identificación forense de locotures. En realidad, deben ser considerados como el correlato metodológico de una filosofía científica determinada por la naturaleza variable de su objeto de estudio.

Dadas las especiales características de nuestro entorno de investigación, resulta bastante dificultoso llegar a conocer todas las circunstancias y referencias concretas en las que se apoyan los criterios de análisis utilizados por cada laboratorio. Salvo en contadas ocasiones [Koval et al. 2000], [Koval y Krinov, 2000] no se encuentran publicaciones por parte de un organismo público o privado en las que se detallen con claridad y precisión cuales son tales criterios. En el mejor de los casos se pueden llegar a conocer los sistemas de análisis empleados, pero en último término nadie pone de manifiesto las claves exactas en las que se sustentan sus resultados o conclusiones. Otra excepción en este sentido, son los estándares de identificación espectrográfica establecidos por el Subcomité VIAAS de la I.A.I. en 1.991. Aunque ya en alguna medida obsoletos, constituyen una de las escasas referencias documentales metodológicas conocidas.

Además de las escasas divulgaciones científicas realizadas, la actividad desarrollada por los laboratorios policiales en los distintos ámbitos internacionales de acústica forense -ya se trate de organismos oficiales o asociaciones privadas- puede ser considerada como un índice bastante significativo de su posicionamiento metodológico. Resulta más pertinente hablar de "índices significativos" pues puede resultar desacertado el uso de términos más categóricos ante referencias del entorno policial (habitualmente acompañadas de un carácter más o menos reservado).

En este sentido, es imprescindible citar las dos asociaciones forenses que aglutinan buena parte de los expertos de nuestro campo: la I.A.I. (Subcomité de VIAAS) y la I.A.F.P. (International Association of Forensic Phonetics). El Subcomité de la I.A.I. es más antiguo (finales de los 70) y en términos generales puede decirse que ha venido representando la corriente metodológica de los "ingenieros" (su enfoque está basado en el análisis acústico de la señal) . La segunda (I.A.F.P., 1.991) tiene una perspectiva del problema fundamentada en la realización fonético/lingüística del habla y por tanto representa la corriente de los que denominamos "fonetistas".

La I.A.I. hasta la década de los 90 ha estado conducida por el F.B.I. y diversos expertos de laboratorios policiales y privados USA que basaban su metodología en la utilización del método auditivo-espectrográfico. A partir de 1.991 se fueron incorporando laboratorios policiales

de fuera de los Estados Unidos que aportaron distintos enfoques metodológicos. Entre los más relevantes cabe citar, el Instituto Nacional de Investigación de Ciencia Policial de Japón, el laboratorio de análisis y tratamiento de la señal de la Policía Técnica y Científica de Francia y el laboratorio de Acústica Forense de la Comisaría General de Policía Científica de España. Una característica muy a tener en cuenta en el caso de la I.A.I. es la pertenencia de la casi totalidad de sus miembros a laboratorios policiales ,o en su caso, a laboratorios forenses dedicados sistemáticamente a la identificación de hablantes.

En el caso de la I.A.F.P. , buena parte de sus miembros son expertos privados ubicados en centros de investigación universitarios, aunque también algunos de ellos, colaboran de forma permanente o eventual con organismos judiciales o policiales. Dentro de esta generalidad existen excepciones como es el caso del laboratorio de análisis acústicos del Instituto Técnico Criminal del Bundeskriminalamt dirigido desde 1.980 a 1.999 por el Dr. Hermann J. Künzel. Precisamente, Künzel y sus colaboradores, Harry Hollien y los suyos en la Universidad de Florida, y algunos fonetistas del Reino Unido, son considerados los expertos que marcan las pautas en dicha asociación y por tanto los más representativos.

Como hemos referido anteriormente, en estas dos asociaciones están incluidos casi todos los laboratorios, expertos o especialistas más relevantes de nuestra técnica. Además de los ya citados, en el caso de Europa es conveniente destacar los laboratorios policiales del Centro de Criminalística del M1 del Interior de Rusia que, en un número cercano a los 50, son los más antiguos del continente europeo.En este mismo país, desde 1.991 el Centro de Tecnología del Habla de San Petersburgo desarrolla una importante labor de investigación y formación en relación con nuestro entorno de investigación. A ellos pueden sumarse el laboratorio de la Policía Científica de Italia -segundo en antigüedad de Europa- el Laboratorio Central Forense de la Policía de Polonia , el de la Policía Nacional de la actual Chequia, el de la Policía Judicial de Bélgica, el del Instituto de Investigación Criminal de la Gendarmerie Nacional de Francia, el del Instituto de Investigación Forense de Lituania, el del National Bureau of Investigation de Finlandia o el Laboratorio Nacional de Ciencias Forenses de Suecia. En Alemania, además del laboratorio de la B.K.A. existen otros de carácter público, en los Landeskriminalamt o Policías de Estados Autonómicos (Munich, Brademburgo, Düsseldorf...). [I.C. n18]

Continuando en el ámbito europeo, no deben dejar de mencionarse otras Instituciones, laboratorios o expertos privados que desarrollan la técnica de cara a los tribunales de justicia; sería el caso de la Fundación Ugo Bordoni en Italia, diversos laboratorios en el Reino Unido, el Gerechtelijk Laboratorium de Holanda, el Laboratorio de Investigación Acústica de la Academia de las Ciencias de Austria y de otros muchos expertos de multitud de universidades europeas (Rusia, Francia, Suecia, Noruega, Suiza, Polonia, Hungría, Portugal, etc.).

Igualmente, son reseñables tres entornos de trabajo exclusivamente forenses: el *P.C.W.G.* (Police Co-operation Working Group) de la Unión Europea, los Simposiums de Ciencia Forense organizados por *INTERPOL* y la Red Europea de Institutos de Ciencias Forenses *E.N.F.S.I.*. Los dos primeros grupos tienen un carácter público y exclusivamente policial y el tercero es privado aunque conformado fundamentalmente por laboratorios policiales oficiales. En el seno del P.C.W.G. de la Unión Europea, desde 1.995 se están desarrollando tareas de estandarización en la técnica de Identificación de voz conducidas por la delegación española, la cual, está representada por el laboratorio de Acústica Forense de la Dirección General de la Policía. En nuestro país, éste es el único laboratorio policial que trabaja sistemáticamente la técnica de identificación de locutores para los tribunales de justicia. De forma muy eventual, algún laboratorio o especialistas privados relacionados con áreas afines a la técnica (ingenieros o técnicos de sonido, filólogos, foniatras, ...) emiten informes periciales para la Justicia, basando sus criterios de análisis en sus exclusivas y correspondientes perspectivas de estudio.

En lo que respecta al resto del mundo, se tiene constancia del desarrollo de la técnica en multitud de Universidades de los cinco continentes. De la misma forma, diferentes instituciones policiales incluyen en su estructura laboratorios forenses donde se llevan a cabo tareas de identificación de hablantes. Dentro de éstos, los probablemente más importantes se encuentran en Japón: Instituto Nacional de Investigación de Ciencia Policial de la Policía Nacional, Laboratorio de Investigación Científica del departamento de Policía Metropolitana de Tokyo y laboratorios de Investigación Científica de la Policía en las Prefecturas de Aichi, Osaka y Fukuoka. Igualmente, pueden citarse otros laboratorios policiales asiáticos: Policía Nacional de Israel en Jerusalén, Gendarmerie de Turquía, Ministerio de Justicia de la República China, Departamento de Ciencia Forense del M1 de Seguridad Pública de Arabia Saudí, Policía Nacional de Emiratos Árabes Unidos, Departamento de Investigación Científica del Crimen de Corea del Sur, Centro de Investigación Científica del Ministerio de Justicia de Taiwan y Real Policía de Hong Kong.

Como anteriormente hemos referido, la actividad fundamental de la identificación forense de locutores en lo que al continente americano concierne ha de situarse en los Estados Unidos. Dentro de su ámbito de influencia puede incluirse la situación de la técnica en Canadá, donde en el laboratorio de Análisis de Audio de la Real Policía Montada encontramos uno de sus exponentes más relevantes.

Dejando al margen el caso USA, el análisis de voz con fines forenses ha sido abordado por diversas universidades sudamericanas (Escuela Paulista de Medicina de Sao Paulo y el Departamento de Medicina Legal de la Universidad de Campinas en Brasil, Universidad de

Buenos Aires, Universidad de Perú, etc.) si bien es cierto, no existe constancia de un trabajo continuado por parte de departamento alguno. En lo que se refiere a laboratorios policiales, encontramos operativos tres laboratorios en Colombia (Fiscalía General de la Nación, Departamento Administrativo de Seguridad y Policía Nacional). No obstante,sí se ha detectado una eventual actividad de la técnica en Argentina, así como la intención de establecer unidades de identificación de locutores en las Policías de diversos países hispanoamericanos (México, Chile, Uruguay, Argentina, Perú, El Salvador, etc.)

Hasta el momento, las únicas noticias del continente africano provienen de Sudáfrica e Isla Mauricio, donde peritos extranjeros han testificado ante los respectivos tribunales de dichos países.(En el caso de Sudáfrica, también ha sido reseñada cierta actividad por parte de algún especialista nativo).

Por último, hacer una breve referencia a Australia, donde la práctica eventual de la técnica se circunscribe a la realizada por expertos universitarios (p.e.Monasch University) que apoyan la investigación de las agencias policiales al no disponer éstas de laboratorios específicos.

II.3.3.- Admisión de la evidencia de Identificación de Voz.

La finalidad última de las distintas técnicas forenses, es la aportación de evidencias o indicios para el esclarecimiento o resolución de los procesos de investigación judicial. Por este motivo, el éxito de nuestros trabajos estará íntimamente ligado a la apreciación de los mismos por parte de los diferentes estamentos judiciales. En el caso que nos ocupa, nos encontramos con dos aspectos que afectan directamente a las consecuencias judiciales relacionadas con la práctica y entorno jurídico de la técnica. Por una parte, existen ciertas diferencias metodológicas derivadas de su intrínseco carácter multidisciplinar y la ausencia de estándares de referencia común; por otra, la naturaleza variable del objeto de estudio -la voz no es una referencia biométrica inmutable como ocurre con el ADN o la huella dactilar- en algunos casos representa un serio obstáculo de cara a la aportación de niveles de conclusión de alta certeza.

Además, es necesario tener en consideración otra serie de importantes factores, comunes a la generalidad de las Ciencias Forenses e igualmente decisivos en la resolución del proceso. Nos referimos a la existencia de diferentes sistemas legales con diversos criterios de admisibilidad y valoración para la evidencia científica, a los códigos de ética y entidad técnica de los expertos, etc.

La exposición de estas consideraciones pretende llamar la atención sobre la necesidad de

contextualizar convenientemente las diferentes circunstancias que la técnica ha originado y origina en cada momento y en cada lugar, para así saber apreciar de forma correcta sus consecuencias a nivel judicial. Es decir, no tendría sentido argumentar la admisibilidad o no de nuestra evidencia sin tener en cuenta todos los factores anteriormente referidos: ámbito legal, político, trayectoria de la técnica en su dimensión espacial y temporal, etc.

Efectuadas estas salvedades, podemos hablar de la realidad que supone la admisión de conclusiones periciales de expertos en identificación forense de hablantes por parte de los órganos jurisdiccionales internacionales. No obstante, esta afirmación de carácter general debe ser analizada y matizada suficientemente para evitar posibles confusiones.

Las referencias de admisibilidad para la evidencia científica en los Estados Unidos -ampliamente tratadas en el capítulo anterior- pueden representar un claro índice de cómo evaluar con cierto rigor el peso de una evidencia dentro de un contexto determinado (ámbito legal, metodología utilizada, etc.). No obstante, resultaría extensísimo detallar todos los pormenores y antecedentes legales relacionados con la apreciación de la evidencia de identificación de locutores en cada uno de los tribunales de justicia en los que ha sido introducida. Un buen camino para constatar la validez de dicha prueba debiera pasar por examinar la jurisprudencia *ad hoc* en aquellos países en los que la técnica está siendo desarrollada al más alto nivel (en este sentido, ya hemos apuntado algunos laboratorios).

Pero con independencia del tipo de tribunal, experto o metodología, la prueba de identificación de voz por expertos ha sido admitida como evidencia o indicio en tribunales de justicia de Alemania, España, Francia, Rusia, Polonia, Lituania, Italia, Bélgica, Holanda, Reino Unido, Suecia, Noruega, Suiza, Austria, Finlandia, Chequia, Eslovaquia, etc. También es considerada evidencia en Estados Unidos, Japón, Canadá, Australia, Colombia, Sudáfrica y un largo etcétera en el que estarían incluidos numerosos países de todo el mundo.[I.R. n16]

En nuestro país, la primera referencia de admisión de la prueba de identificación forense de locutores por expertos se recoge en una sentencia de la Audiencia Provincial de Valencia en *febrero de* [1.991]. El informe pericial en cuestión fue emitido por expertos del laboratorio de Acústica Forense de la Comisaría General de Policía Científica.

Puede calificarse de misión imposible el pretender unificar las estructuras y códigos que rigen las normas judiciales de los distintos sistemas legales en cada país. Dicha labor, por otra parte, está fuera del marco de competencia de los expertos en Ciencias Forenses. No obstante, sí es bastante de su incumbencia el colaborar en la construcción de unas referencias metodológicas comunes, para erradicar definitivamente las discrepancias concernientes a criterios de selección sobre las distintas posibilidades u opciones de análisis.

En el entorno actual de nuestra técnica, la existencia de bases científicas metodológicas enfrentadas supone un crítico, aunque evitable error. La filosofía integradora de los métodos combinados tiene como pretensión prioritaria la contemplación de cualquier perspectiva de análisis que pueda aportar informaciones de utilidad sobre las características de una emisión de habla. Por este sencillo argumento, y asumiendo la idoneidad de esta metodología (constatada en las últimas reuniones de expertos a nivel internacional) [I.R. n17](ver II.4.6), en la actualidad no tiene ningún sentido que ante un tribunal de justicia, sean objetos de debate cuestiones de criterios metodológicos. En todo caso, las posibles disensiones habrían de referirse a posibles estimaciones contradictorias, eso sí, sobre la base de unos fundamentos científicos unánimemente aceptados por todos los expertos del campo. Es decir, dos expertos en identificación de locutores nunca tendrían que plantear sus desacuerdos en base a la fiabilidad del método de análisis utilizado; las matizaciones o diferencias de criterio tan sólo debieran surgir en torno a la obtención de unos u otros resultados o valoraciones, pero siempre, partiendo de la asunción de unos fundamentos metodológicos comunes.

II.3.4.- Futuras directrices de trabajo.

Desde la caída de Kersta tras las consecuencias del caso AKing≅, hasta prácticamente nuestros días, la identificación forense de locutores ha estado protagonizada por una dualidad metodológica todavía subyacente: *Aingenieros y fonetistas*≅. Afortunadamente, este posicionamiento maniqueo y obsoleto está llegando a su fin y, con toda probabilidad, dejará de existir en los próximos años.

Como más adelante comentaremos, en el momento actual se están desarrollando trabajos en el seno de la Unión Europea y la red ENFSI [I.R. n18] que pueden significar el despegue definitivo hacia un criterio metodológico normalizado. De hecho, la utilización conjunta o complementaria de diversas clases de expertos y sistemas de análisis es, desde hace tiempo, una visible realidad en la gran mayoría de laboratorios de audio forense.

La confluencia natural hacia el uso de métodos combinados, se ha visto favorecida por la gran eclosión de sistemas automáticos de reconocimiento de voz acontecida en la última década. Incluso, hay quien empieza a pensar que en un futuro no muy lejano la máquina podría sustituir al experto de forma definitiva. Desde luego, si ello aconteciese, estaríamos otorgando el máximo nivel de objetividad a nuestros criterios de decisión, aunque desgraciadamente, la probabilidad real de que esto ocurra en una fecha cercana, es una mera especulación.

Ciertamente, existen sistemas automáticos para la comparación de emisiones habladas que ofrecen buenos resultados de fiabilidad. Pero desgraciadamente, estos buenos comportamientos no lo son tanto cuando la muestra analizada se halla en las denominadas condiciones de registro forense (ruidos, curvas de respuesta telefónicas, distorsión, etc.).

Expertos de todo el mundo testean continuamente multitud de sistemas de este tipo, pero hasta el momento presente ningún laboratorio relevante ha divulgado oficialmente la utilización de un sistema automático fiable, como método exclusivo y único para la identificación forense de locutores.

En este caso, cuando hablamos de "laboratorios relevantes" nos estamos refiriendo a aquellos absolutamente rigurosos; aquellos, en los que se ha observado una dilatada trayectoria de profesionalidad no sólo relativa a una entidad técnica, sino también a un código continuado de responsabilidad ética. Eventualmente, saltan a la palestra supuestos "expertos" que olvidando las posibles graves consecuencias de una decisión pericial errónea, no tienen reparos a la hora de utilizar y proclamar las bondades de métodos automáticos de identificación, según ellos, "infalibles".

Pues bien, en tanto no contemos con un sistema automático realmente objetivo e incuestionable, nuestros esfuerzos deberán centrarse en dos fines prioritarios. Por un lado, trabajar en la solución metodológica de mayor idoneidad. Por el otro, elaborar unos estándares globales de referencia común, que confieran a nuestra técnica el mayor grado de fiabilidad de cara a las instituciones de justicia.

Los trabajos hacia la estandarización ya se han iniciado en el seno de la Unión Europea y en el grupo de expertos de la red ENFSI . Es una labor compleja y delicada en la que han de conjugarse criterios técnicos y objetivos políticos. Para empezar, hay algo en lo que todos están de acuerdo, y así ha sido reiteradamente manifestado: la necesidad imperiosa de contar con dichos estándares. Confiemos en que el beneficio común, prevalezca sobre los intereses particulares. Es el momento de poner punto y final a 30 años de estanqueidad e incomprensión.

II.4.- MÉTODOS FORENSES DE IDENTIFICACIÓN DE LOCUTORES.

II.4.1.- Introducción

En un capítulo anterior definimos algunos de los sistemas de análisis empleados en las técnicas de identificación o reconocimiento de locutores. Tomábamos como referencia una clasificación elaborada por Tosi en la que se nos hablaba de métodos objetivos y subjetivos. Ya entonces, apuntábamos cierta confusión en algunos conceptos fundamentales, como era el caso de la diferenciación entre sistema y método de identificación de hablantes. En general, un método de identificación forense de locutores vendrá conformado por distintos sistemas de análisis; solamente en aquellos casos, en los que el método se fundamente en un único sistema podremos hablar de coincidencia entre ambos conceptos.

De la misma forma, observamos una falta de nitidez por parte de los expertos, a la hora de definir con claridad las ubicaciones metodológicas de los procesos de análisis desarrollados en sus laboratorios forenses. En este sentido, resulta conveniente precisar que cuando nos dispongamos a referenciar una metodología forense de identificación de locutores, necesariamente hemos de definir y diferenciar cuales son las aproximaciones o sistemas de análisis utilizados, los objetos de estudio (su categorización, contexto de comparación, etc.), y cómo todas esas referencias son evaluadas y conducidas hacia unas reglas de decisión concretas.

En el contexto actual de expertos, la objetividad o subjetividad de un método de identificación del locutor, no se plantea como el elemento más crítico a la hora de evaluar su fiabilidad, si bien es indudable, que la mayor o menor fiabilidad del mismo, estará siempre en directa relación a su índice de objetividad. La explicación a este argumento, se fundamenta en dos razones. Por una parte, nos encontramos con que las herramientas de análisis de las que se dispone en la actualidad - ya sean de representación gráfica de la señal, cálculo, evaluación o comparación de parámetros del habla - posibilitan la obtención y manejo de datos absolutamente precisos. Por otra, nos hallamos ante la insoslayable interacción del experto en los procesos de selección e interpretación de tales datos. En esta coyuntura, siempre puede hablarse de cierta dosis de subjetividad, pues la presencia activa del analista seguirá siendo necesaria, en tanto no se disponga de un sistema de reconocimiento automático de total garantía. No obstante, esta dosis de subjetividad tiene que ser la mínima posible, y en este sentido, sí representan un papel trascendental los distintos criterios metodológicos.

Teniendo en cuenta las anteriores consideraciones, estableceremos la siguiente clasificación de métodos forenses de identificación de locutores por expertos :

A/ AUDITIVO - ESPECTROGRÁFICOS

B/ AUDITIVO - FONÉTICO/LINGÜÍSTICOS
C/ SEMIAUTOMÁTICOS O INTERACTIVOS POR ORDENADOR
D/ AUTOMÁTICOS POR ORDENADOR
E/ COMBINADOS

Aunque esta clasificación pudiera contemplar más opciones, debe ser apreciada como una clasificación de tipo general y representativa de la realidad actual de nuestro ámbito. No recogemos aquellos Amétodos≅ de identificación que no sean utilizados de forma sistemática por expertos, como es el caso de los basados en sucesos de percepción auditiva a largo plazo (normalmente relacionados con víctimas y testigos de actos criminales). Tampoco incluiremos los sistemas automáticos de *verificación* o *autentificación* con fines comerciales (banking o telefonía electrónicos, servicios de red informática, etc) pues no tienen una aplicación real en la identificación de locutores sistemática en condiciones forenses.

II.4.2.- Métodos Auditivo-Espectrográficos

Como ya hemos visto, los primeros pasos de la identificación de locutores por expertos estuvieron estrechamente ligados a la comparación visual de espectrogramas o sonogramas de banda ancha a los que, inadecuadamente, se denominó "voiceprint" o huellas de voz. Esta desafortunada conceptualización, unida a otras circunstancias ya referidas, propiciaron un rechazo sistemático y absurdo hacia este procedimiento de análisis por parte de ciertos expertos (especialmente a raíz del famoso caso judicial en los Estados Unidos : A*el Pueblo contra Edward Lee King*≅ en 1968).

De igual manera, los planteamientos metodológicos de Kersta - basados en la comparación sonográfica de banda ancha- presentaban graves lagunas y que en ningún caso podían ser asumidos como infalibles [Kersta, 1962]. No obstante, y partiendo de esta evidente consideración, debe entenderse como un grave error la circunstancia de ignorar las valiosas informaciones que pueden ser obtenidas a través de un estudio metodológico de los índices acústico-sonográficos.

Por tanto, y si en nuestra opinión son tan valiosas estas referencias gráficas)por qué ciertos expertos forenses mantienen una aparente actitud crítica ante su utilización?. La respuesta a esta cuestión es una de las claves en las que se ha sustentado la problemática de la identificación de hablantes en los últimos años. En realidad, estos expertos que aparentemente no se muestran partidarios del uso de esta aproximación de análisis están, en primer lugar, faltando a

la verdad y, en segundo lugar, generando un mayor o menor grado de confusión que siempre, en última instancia, incidirá negativamente sobre la fiabilidad de nuestra técnica forense.

)Pero, por qué hablamos de "faltar a la verdad"?. Existen diversas razones que argumentan esta aseveración. La más evidente de todas ellas, se relaciona con el hecho de que muchos de estos especialistas que reniegan de este sistema de análisis (fundamentalmente algunos Afonetistas forenses") en realidad lo están utilizando de forma sistemática , si bien, prefieren presentarlo como una aproximación secundaria del puro análisis perceptivo. Por otra parte, resulta indiscutible y fácilmente demostrable, el hecho de que diversos índices acústicos y distribuciones de energía sonora del habla, de la forma que mejor pueden ser apreciadas y mensuradas es a través de los sonogramas. Cualquier buen experto forense es consciente de la trascendencia de tales apreciaciones y, por consiguiente, del carácter insoslayable de este enfoque de análisis.

) Dónde radica entonces el problema ? Si realmente quienes claman la no relevancia de tal sistema de análisis lo están considerando,) donde encontramos la justificación a tal actitud? . La solución a este insólito planteamiento responde a la conjunción de diversas causas. La más profunda y en algunos casos inconsciente está fundamentada en la ya comentada dicotomía ingenieros/fonetistas. Desde los antiguos errores de Kersta, ciertos fonetistas forenses se han estado oponiendo de forma sistemática a su método y aún en la actualidad, algunos siguen obstinados en proseguir con dicha tarea. La mayoría de ellos ignora, o quiere ignorar, que muy pocos son ya los expertos que practican como método exclusivo de identificación forense el conocido como "auditivo/espectrográfico" ; y , que en cualquier caso, la aplicación de dicho método por tales expertos está mucho más evolucionada que la practicada por Kersta hace treinta años.

La mayor parte de los que inicialmente enfocaron el análisis forense de la voz agrupados en la perspectiva de la ingeniería acústica han evolucionado hacia la utilización de métodos "combinados". La misma tendencia ha sido detectada en otros expertos autodidactas y en los denominados "fonetistas" aunque éstos, aparentemente, y a pesar de utilizarlo - bajo la denominación de Afonética instrumental≅ - siguen empeñados en no ensalzar las bondades del análisis sonográfico. En el fondo de la cuestión, no parece encontrarse otra causa que el propio interés de presentar como más relevante el prisma de estudio de su especialización.

No obstante, y como suele ocurrir en otras discusiones de carácter científico, el punto de vista más idóneo para solventar una controversia suele estar referenciado en una posición de equilibrio. El caso de la identificación forense de la voz no constituye una excepción. Más adelante comprobaremos que ni desde los planteamientos radicales de ciertos "fonetistas" , ni

desde la utilización del análisis sonográfico como método exclusivo podrá formularse la mejor solución.

Ya conocemos que existen diversas posibilidades de dimensionar y representar gráficamente el sonido del habla. Recordemos que de todas ellas, la denominada sonográfica o espectrográfica es la que resulta más interesante para el investigador forense pues le sitúa en la mejor disposición para poder percibir de forma inmediata aquellos índices acústicos que caracterizan las distintas realizaciones vocales en relación a sus cuatro ejes de referencia : rango de frecuencia, nivel de presión sonora, duración de la emisión y factor de resonancia.

En síntesis, la comparación espectrográfica se basa en el cotejo de formas o "patterns" sonográficos, en orden a determinar el mayor o menor número de similitudes o diferencias existentes entre las muestras comparadas. Dicho así, este ejercicio comparativo de percepción visual no se presenta como una tarea complicada; de hecho, cualquier persona estaría en condiciones de emitir un juicio de este tipo, si bien, la apreciación de un experto implica consideraciones más complejas.

La principal diferencia entre el cotejo visual efectuado por un experto y el realizado por un profano estriba en la elemental circunstancia de que el segundo sólo puede emitir una opinión muy subjetiva sobre la similitud/disimilitud existente entre las formas que compara, mientras que el experto conoce, tanto la correspondencia de cada una de esas formas con una realización fono-articulatoria concreta, como la mayor o menor entidad de esas similitudes/disimilitudes desde un punto de vista identificativo e individualizador de las emisiones habladas objeto de análisis. Por consiguiente, para que un proceso de evaluación de características acústicas sea calificado como completo, resultará necesario que el analista esté en disposición de una sólida formación multidisciplinar.

Los seguidores del método auditivo-espectrográfico requieren de la necesidad de contar con el mismo contexto discursivo en las muestras objeto de comparación para llevar a cabo una evaluación de alta fiabilidad. El motivo no es otro que buscar una adecuación del efecto de co-articulación entre ambos actos de habla, dado que la realización de cada alófono o grupo fónico dentro de una cadena hablada está directamente influenciada por aquellas otras realizaciones sonoras que les preceden o les siguen.

Por tanto, el requerimiento del mismo contexto comparativo - premisa insoslayable para los partidarios del método auditivo-espectrográfico - ha de entenderse en su entorno como algo absolutamente lógico, pues al no utilizar otro criterio de análisis complementario (fonético, semiautomático, etc.) deberán ajustar en la medida de lo posible sus presupuestos de comparación.

Si tuviésemos que responder a la pregunta de si es posible alcanzar un alto grado de certeza en una decisión de identificación/eliminación mediante la utilización exclusiva del método auditivo-espectrográfico, la respuesta sólo podría ser una :"sí". Ahora bien, si de la misma forma debiéramos informar sobre si este método es en la actualidad la mejor opción para la identificación forense de hablantes, la respuesta sería : "no".

II.4.3.- Métodos Auditivo- Fonético/Lingüísticos

Bajo este epígrafe quedarán comprendidas aquellas aproximaciones de estudio cuya exclusiva base de referencia sea de carácter fonético/lingüístico. Además, estas opciones de análisis, deben estar alimentadas a través de inputs puramente perceptivo-auditivos, ya que si éstos se complementasen con evaluaciones deducidas de la observación de índices acústicos sonográficos, estaríamos introducidos en un enfoque metodológico "combinado".

El proceso general de un análisis de estas características comprende dos etapas. En la primera de ellas, se procede a una escucha de los registros objeto de estudio con el propósito de detectar las distintas realizaciones fono-articulatorias y lingüísticas emitidas por el locutor. En la segunda, se evalúan y asocian dichas realizaciones a nivel dialectal, sociolectal e idiolectal, con la finalidad de otorgar a las mismas un mayor o menor grado de relevancia, desde un punto de vista identificativo. Lógicamente, el rigor y entidad de esta sistemática de estudio estará en función de distintos factores: fiabilidad del proceso de adquisición de datos (normalmente materializado mediante transcripción fonética), número y tratamiento de los parámetros estimados, apreciación de los mismos con relación a referencias normativas experimentales de suficiente muestra representativa, etc.

El principal inconveniente de los métodos "auditivo-fonético/lingüísticos" lo constituye la naturaleza perceptiva que caracteriza a su fase de observación y captura de datos.A pesar de la funcionalidad que puede aportar el uso de la transcripción fonética, la dosis de subjetividad que acompaña a este tipo de evaluaciones, puede revelarse como un factor inquietante a la hora de considerar el grado de validez de las estimaciones obtenidas. Por esta causa, todo procedimiento metodológico de este tipo deberá fundamentarse en un protocolo de análisis que otorgue a sus estimaciones un aceptable rango de fiabilidad. Decimos "aceptable", porque difícilmente podrá ser de mayor entidad, ya que hoy en día no tiene sentido la práctica exclusiva de este método sin el refrendo añadido de otras aproximaciones de estudio.

Hasta el momento presente, no ha sido documentado formalmente protocolo alguno que

explícitamente describa una metodología de análisis forense "auditivo-fonético/lingüística" (descripción, categorización y evaluación de los parámetros considerados, etc.) si bien, de una u otra forma dicha técnica es desarrollada por diversos especialistas de esta disciplina científica. No obstante, hemos de admitir sus notables prestaciones cuando tal alternativa es apreciada como una opción complementaria y es desarrollada de una forma sistemática y de acuerdo al procedimiento adecuado. Ningún experto forense de nuestro ámbito, cuestiona las valiosísimas informaciones que un oído sano y adiestrado puede proporcionar al proceso de identificación forense de la voz.

Ahora bien, al igual que ocurría cuando hablábamos de los métodos "auditivo-espectrográficos", si alguien nos preguntara sobre la posibilidad de dar una opinión para identificar o descartar a un hablante a través de este único enfoque de estudio, la respuesta vendría formulada en idéntico sentido: sí, sería factible aunque no sería la mejor opción.

II.4.4.- Métodos Semiautomáticos o Interactivos por ordenador

Óscar Tosi [Tosi, 1979] denominaba "semiautomáticos" a los métodos de identificación de locutores en los que se produce una fuerte interacción entre el analista y la máquina, o lo que es lo mismo, entre el analista y la aplicación de análisis o cálculo por ordenador que se utiliza en el proceso. Según dicho autor, esta fuerte "interacción" concierne tanto a la discrecionalidad del operador para seleccionar los eventos objeto de comparación, como a la propia interpretación por el mismo de los parámetros que aporta el ordenador.

Aún teniendo en cuenta estas apreciaciones, resulta difícil precisar donde se encuentra la frontera que diferencia un entorno de reconocimiento de voz como automático o semiautomático. Métodos en su día considerados por Tosi como semiautomáticos, para otros en la actualidad podrían quedar enmarcados dentro de los sistemas de reconocimiento automáticos de primera generación, es decir, aquellos basados en esquemas de comparación de patrones (DTW, etc.).

La dinámica típica de un sistema semiautomático comprende una primera fase en la que el experto selecciona las referencias que estima más idóneas para cada proceso de identificación. Es decir, aquellos parámetros que por su propia naturaleza y en conexión a un contexto acústico determinado, le pueden aportar las informaciones más relevantes sobre el carácter individual de una emisión hablada concreta.

La selección de tales parámetros, forma parte de una sistemática más o menos variable, en razón al número de ocurrencias de los mismos en las emisiones objeto de análisis. Lógicamente, la mayor o menor suficiencia de estos elementos de estudio, a nivel cualitativo y

cuantitativo, incidirá directamente en la obtención y valoración de los resultados finales.

Una vez introducidos los datos para su muestreo y procesado en la aplicación de cálculo y análisis, el ordenador proporciona ciertos valores representativos de cada una de las referencias estudiadas. Este proceso, se efectúa de la misma forma sobre parámetros dubitados e indubitados. A continuación, el ordenador puede establecer, o no, una comparación de similitud entre las estimaciones correspondientes; y puede efectuar, o no, la consiguiente valoración de identificación/eliminación. Si este último presupuesto se produjese, probablemente estaríamos adentrándonos en un entorno de análisis automático, puesto que precisamente una de las premisas que caracterizan a un método como semiautomático es la interpretación final de los resultados procesados por parte del experto.

Como sistemas pioneros de análisis semiautomático, pueden citarse el publicado por Pruzanski en 1.966, [Pruzansky, 1966] , el desarrollado por Becker y el instituto de investigación de Stanford [Becker et al., 1972] que fue más tarde complementado por otro estudio de Hair y Requieta de Texas Instruments Inc [Hair y Requieta, 1972]. Posteriormente, en 1.977 debe reseñarse el SASIS (Sistema Automático de Identificación de locutores) desarrollado por la Aeroespace Corporation de El Segundo, California [1977] y el SAUSI desarrollado en la Universidad de Florida por H. Hollien y otros colaboradores [Hollien H. et al., 1977]

Tosi y sus colaboradores, los doctores Dubes y Anil, iniciaron un sistema semiautomático con el que pretendían establecer comparaciones ópticas por ordenador entre formas sonográficas mediante su previa digitalización [Tosi, 1979]. Algo similar a los sistemas automáticos actuales de comparación de impresiones dactilares (AFIS). La gran cantidad de factores o variables que debían ser consideradas por el sistema, lo hicieron inviable para su aplicación forense.

En el momento actual, distintos sistemas que pueden ser considerados como semiautomáticos (SIVE [Lipeika y Lipekiene, 1993, 1995, Lipeika, Lipekiene y Salna B. 1997], IDEM [Falcone y De Sario 1994], DIALECT [Popov et al. 1996], etc.) y otros no documentados, son utilizados en algunos laboratorios acústica forense con relativo éxito. No obstante, son considerados siempre como una herramienta de análisis complementaria o integradora de otras perspectivas de estudio.

El entorno forense se enfrenta a dos problemas en el caso de los sistemas semiautomáticos. Uno de ellos, viene representado por las malas condiciones que caracterizan a los registros de dicho ámbito. El otro, por el peligro de una inadecuada interacción por parte del operador del sistema. Por esta última razón resulta indispensable que tal operador sea un experto, pues sus actuaciones - ya sean de selección del material para el análisis o de interpretación de

resultados - siempre incidirán de forma directa y crítica en el correcto funcionamiento del sistema.

En nuestra opinión, y en tanto los sistemas automáticos de reconocimiento de locutores no deparen unos resultados de mayor fiabilidad, los métodos robustos semiautomáticos en manos de expertos cualificados pueden resultar un complemento perfectamente válido a otros sistemas de análisis más tradicionales.

II.4.5.- Métodos automáticos por ordenador

Sin obsesionarnos con la idea de establecer una frontera entre ámbitos de reconocimiento de voz semiautomáticos y automáticos, podemos convenir, con carácter general, que un sistema automático es aquel en el que la interacción del analista en el proceso es la imprescindible, quedando asumidas por la aplicación informática de análisis, las funciones de muestreo de parámetros, comparación y toma de decisión. Sería el caso - anteriormente relatado - de los sistemas de "identificación" y Averificación≅ tal y como son definidos por los expertos no forenses en reconocimiento automático de locutores.

Las primeras aproximaciones documentadas al reconocimiento automático de locutores se producen a principios de los años setenta. Los ingenieros de la Bell, Bishnu Atal [1.972, 1.974] y Aaron Rosenberg y Sanbur [1.975] publican sus primeros estudios utilizando ya, como base de extracción de datos, coeficientes cesptrum y coeficientes de predicción linear o LPC. En esta misma época, son testeados diversos parámetros acústicos del habla para su utilización en el diseño de sistemas de reconocimiento automático: en 1.972 Wolf analiza combinaciones de hasta veintisiete referencias extraídas de consonantes nasales, espectros de vocales, frecuencia fundamental, V.O.T de oclusivas, etc.[Wolf, 1972]. Su y Fu [1.973] utilizan como informaciones eficientes los espectros de consonantes nasales. Li y Hughes [1.974] toman como referencia matrices de correlación referidas a LTAS de fragmentos de habla continua.

La mayoría de estos primeros intentos estaban estrechamente ligados a comparaciones dependientes de texto, aunque igualmente se reportan estudios forenses de reconocimiento automático independientes de texto. Sería el caso del sistema AUROS [Bunge, 1977] el experimento de Hollien y Majewski [1.977], el de Tosi y sus asociados [TOSI III,1977] , el de Tosi y Nakasone [1989] o el proyecto CAVIS (Computer Assisted Voice Identification System) en [1.985].

No tendría mucho sentido entrar en detalle sobre las particularidades de cada uno de estos estudios, si bien es oportuno señalar que unos y otros se toparon tarde o temprano con la misma

dificultad: las condiciones que caracterizan la señal degradada de los registros forenses (restricción de información en rango de frecuencia, ruido, distorsiones, etc.) disminuían hasta tal punto la eficacia y fiabilidad del sistema que lo hacían inviable para su aplicación sobre casos forenses reales. No obstante, estas primeras experiencias aportaron datos valiosos para el desarrollo de futuros sistemas de reconocimiento automático más robustos. Entre otros, pueden señalarse la definición de ciertos componentes de alta estabilidad (vocales y nasales) o aquellos de menor permeabilidad al factor de influencia de la curva de respuesta telefónica (Cepstrum).

Desde entonces hasta nuestros días, la funcionalidad de diferentes generaciones de sistemas de reconocimiento automático de locutores (basados en DTW=s, V-Q, HMM, GMM, Redes Neuronales, etc.) han sido analizadas y testeadas en contextos de laboratorio [Doddington 1985, 2000], [Furui, 1.979, 1.981, 1.991, 1.994], [Klevans y Rodman, 1.997], [Matsui y Furui, 1.991, 1.992],[O=Shaugnessy, 1.986] ,[Reynolds, 1.994], [Rosengberg y Soong, 1.987, 1.991], [Tosi y Nakasone, 1989], etc., y frente a condiciones forenses [Gorban (CASVI), 1997], [Marescal, 1999], [Meuwly et al., 1998], [Suzuki, 1997], etc.. Diferentes alternativas de parametrización y modelación del habla, técnicas de normalización del canal, configuraciones de entrenamiento y comparación, etc., han sido combinadas y consideradas ; y aunque recientemente se están obteniendo resultados esperanzadores, persisten todavía diversos retos a resolver: variabilidad intra-personal, variabilidad por efecto del canal y condiciones de grabación, etc. [Furui, Reynolds, 1994]. Estos problemas se incrementan cuando nos enfrentamos a los habituales ingredientes de los registros de audio forense: ruido, distorsión, restricciones de información en rango de frecuencia, plano expresivo o emocional del discurso, voz disimulada o imitada, etc. Por todos estos inconvenientes, es por lo que la utilización en la actualidad de un sistema automático de reconocimiento de locutores (SARL) como opción exclusiva en el análisis forense no debe ser recomendada. Solamente podría ser considerada con tales propósitos en casos muy concretos, y siempre entendida como una herramienta complementaria a los análisis convencionales: perceptivo auditivo, acústico y fonético-lingüístico. En cualquier caso, la discrecionalidad para el uso complementario de este tipo de aplicaciones se revela como una tarea muy complicada por las lógicas cuestiones de compatibilidad inter-sistemas.

Al margen de los ya mencionados, otros prototipos de SARL basados en alineamiento temporal dinámico (DTW), modelado de clases fonéticas mediante cuantificación vectorial(V-Q) y modelos de mezclas de gausianas (GMM) han sido testeados en nuestro país ante casos forenses reales, tanto en tareas de identificación como de verificación [I.R. n1 9]. En términos generales los resultados suelen aparecer aceptables -aunque no totalmente satisfactorios- con muestras de habla de buena calidad. Sin embargo, cuando el test es ejecutado sobre señal degradada por factores de carácter forense, el nivel de éxito de los sistemas decrece de forma sensible. Además, en algunas ocasiones, los resultados de decisión del sistema automático ante

presupuestos de similares o idénticas características, se muestran contradictorios. Por este motivo, resulta extraordinariamente complejo determinar en tales casos, cual es la causa real ligada al origen de esta clase de problemas.

II.4.6.- Métodos Combinados de Identificación de locutores

Conocidos los principales argumentos que sustentan las perspectivas básicas de la identificación forense del habla (espectrográfica, fonético-lingüística, semiautomática y automática) vamos a adentrarnos en la que es considerada por nuestra comunidad forense de vanguardia, alternativa metodológica de mayor fiabilidad. Nos estamos refiriendo a los denominados "métodos combinados". Dicha denominación, se deriva de la que es su característica más representativa, puesto que en todos los casos, y sea cual fuere la versión de los mismos, los métodos combinados vendrán siempre configurados por la conjugación de los cuatro sistemas básicos anteriormente descritos.

Los distintos enfoques metodológicos hasta ahora considerados, ponían de manifiesto dos ideas fundamentales. Por un lado, la realidad de no poder alcanzar un rango de conclusión o certeza de la máxima fiabilidad. Por otro, la existencia de unos sistemas y herramientas de análisis capaces de otorgar los más altos valores de precisión a determinados cálculos y evaluaciones de identificación/eliminación. Además, llegado este momento, debe quedar ya totalmente asumido el hecho de que cualquier conclusión alcanzada mediante la utilización exclusiva de alguna de las cuatro aproximaciones generales citadas carecerá, en cualquier caso, del máximo rigor y eficacia que hoy en día debe exigirse a la técnica forense de identificación de hablantes.

Partiendo de estas tres sencillas premisas, y teniendo en cuenta los diferentes sistemas de análisis ya esbozados es hora de exponer la que es considerada solución metodológica más idónea.[I.R. n1, 7 y 8]

Cuando cualquier procedimiento científico aborda un objeto de estudio con carácter variable suele utilizar el mayor número de perspectivas de estudio, para así poder otorgar un alto grado de objetividad a las conclusiones alcanzadas. En aquellos casos, en los que todos los enfoques utilizados apunten en la misma dirección, podremos decir que nos encontramos en la mejor disposición para poder conocer y definir las características reales del problema analizado. En nuestro caso concreto, dicho planteamiento es insoslayable. Al margen de las dificultades relacionadas con el diseño de una metodología de trabajo, lógica consecuencia de diversos factores ya mencionados: naturaleza multidisciplinar, variabilidad intra-personal, etc, hemos de enfrentarnos con otros duros obstáculos tales como los diferentes pesos específicos otorgados a nuestra evidencia científica por los distintos sistemas legales, formación y cualificación de

nuestros expertos, etc. Por estas razones, la necesidad de establecer unas referencias metodológicas estandarizadas se plantea como un objetivo prioritario para intentar reducir en la medida de lo posible las frecuentes ambigüedades de criterio en nuestra comunidad científica.

Tomando en consideración estas reflexiones, no parece haber duda de que la filosofía actual del análisis de voz con fines identificativos - en este específico ámbito de investigación - deberá sustentarse, indefectiblemente, en la utilización de un "método combinado": *)por qué utilizar un único procedimiento de análisis, cuando puede ser complementado y corroborado por otros de distinta o similar naturaleza?*

En marzo de 1.999, fue celebrada en Wiesbaden una reunión de expertos en Acústica Forense europeos, como consecuencia del proyecto de estandarización desarrollado para tal materia por el P.C.W.G (Police Co-operation Working Group) de la Unión Europea. Meses después, en junio del mismo año, tuvo lugar en Madrid el segundo Ameeting≅ del grupo de trabajo para habla y audio forense de la red ENFSI (European Network for Forensic Sciences Institutes). En ambas reuniones, los métodos "combinados" de identificación forense de locutores fueron señalados como los más funcionales, eficaces y fiables. Igualmente, fue consensuada la definición del ámbito y sistemas de análisis que quedan integrados dentro de los mismos: *" se consideran métodos combinados de identificación forense de locutores, aquellos que entre sus aproximaciones de estudio incluyen, al menos, el enfoque perceptivo-auditivo, el análisis acústico (sonográfico, oscilográfico, espectrográfico) y el análisis fonético-lingüístico. "*

La aceptación institucional de los "métodos combinados" como los más idóneos resulta de extraordinaria importancia. Debe considerarse como el punto y final formal al tradicional cisma metodológico fonetistas/ingenieros. Por otra parte, esta nueva concepción no excluye en ningún caso el complemento de análisis representado por los enfoques semi-automáticos y automáticos. Muy al contrario, el espíritu de los asistentes a las citadas reuniones se manifestaba totalmente abierto a este tipo de opciones, si bien, en el momento actual se considera que solamente pueden ser apreciadas como un elemento complementario a los tres sistemas básicos ya referidos.

Sin embargo, el hecho de estimar como mejor solución la utilización de una metodología combinada, no tiene otro significado que el de señalar cuales son los ingredientes básicos que no deben faltar en el desarrollo de los distintos modelos forenses de identificación de hablantes. A partir de aquí cada laboratorio o experto, en función de sus posibilidades, necesidades, entorno legal, etc. diseñan el modelo que estiman más adecuado.

El margen de flexibilidad que propicia la utilización de una metodología combinada requiere de la lógica validación de sus sistemas y procedimientos de análisis. Para ello, dichos sistemas y protocolos deberán ser analizados y evaluados en profundidad por los distintos miembros de nuestra comunidad científica, a través de los proyectos y grupos de trabajos generados a tal efecto [I.R. n1 7 y 8]. Superada esta etapa de exploración, estaremos en condiciones de iniciar el camino hacia la obtención de unas referencias y protocolos de análisis de uso común.

CAPÍTULO III

UN MODELO DE IDENTIFICACIÓN FORENSE DE LOCUTORES PARA EL NUEVO MILENIO

III.1.- EJES DIMENSIONALES DEL MODELO

III.1.0.- Introducción

El modelo de identificación forense que vamos a presentar, ha de ser incluido en lo que hemos denominado entorno metodológico "combinado". En líneas muy generales, podemos decir

que se trata de un modelo combinado estándar que se complementa con un sistema de reconocimiento automático de modelado de clases fonéticas por mezclas de Gausianas (GMM).

Como ya hemos comentado en el capítulo anterior, un método combinado ha de contemplar, como mínimo, las perspectivas de análisis perceptivo- auditiva, acústica y fonético-lingüística. Decíamos también, que partiendo de diferentes conjugaciones sobre tal estructura básica podrían considerarse otras aproximaciones de análisis complementarias, de carácter automático o semi-automático. Pues bien, la utilización conjunta de enfoques automáticos o semi-automáticos y otros análisis de identificación clásicos no es algo novedoso. El "TOSI III" ya fue utilizado junto con el análisis auditivo-perceptivo y el espectrográfico por el Dr. Tosi. Con estos mismos sistemas - perceptivo y sonográfico - el Instituto Nacional de Investigación de Ciencia Policial de Japón viene utilizando un sistema de reconocimiento automático basado en el alineamiento temporal dinámico de la señal (DTW) [Suzuki et al., 1997]. Otro método de similar estructura al japonés ha sido desarrollado por Wojciech Majewski y Czeslaw Basztura en la Universidad Técnica de Wroclaw, Polonia [I.R. n112].

Podríamos referir algunos ejemplos más en los que sistemas de reconocimiento automático o semi-automático (DIALECT, SIVE, IDEM) están siendo considerados como un complemento a los enfoques clásicos de identificación forense de locutores. No obstante, hasta el momento presente no ha sido reportado un método forense de idénticas características al que vamos a describir.

La singularidad de nuestro modelo no vendrá únicamente avalada por el hecho de usar unos sistemas de análisis concretos. Es más, podemos asegurar con absoluta rotundidad que la utilización de unos u otros sistemas de análisis no tendrá sentido alguno si éstos no están sustentados en una clara definición y parametrización de unos objetos de estudio, acompañados de los criterios, herramientas y normas de procedimiento que sean más idóneos para la observación, selección y valoración de los mismos.

En síntesis, nuestro modelo combinado estará basado en la apreciación del mayor número de componentes, elementos, factores y contextos que estructuran las emisiones habladas. Todo ello será abordado desde distintas perspectivas y opciones de análisis en orden a obtener una estimación comparativa lo más amplia y fiable posible.

Por otro lado, no debemos olvidar que partiendo de estos presupuestos -ya sean enfocados a éste u otro método combinado- hemos de prestar especial atención a la formación y práctica del experto, quien siempre trabajando en equipo desempeñará un papel trascendental tanto a la hora de seleccionar y evaluar los distintos parámetros, como en el momento de elaborar e interpretar los resultados obtenidos.

III.1.1- Ámbito de aplicación del modelo

Las conclusiones de identificación/eliminación alcanzadas a través de la utilización del presente modelo serán absolutamente válidas para ser apreciadas como evidencia o indicio ante los tribunales de justicia, y por consiguiente, para servir como ayuda a las investigaciones desarrolladas por cualquier agencia o cuerpo de las fuerzas de seguridad, oficinas del ministerio fiscal, etc.

La estructura de nuestro modelo -con una aplicación enfocada a la *lengua Castellana*- podrá ser adaptada sin dificultad a cualquier otra lengua. No obstante, resulta oportuno comentar la existencia de un criterio unánime entre la gran mayoría de expertos y asociaciones forenses del campo, en lo que se refiere a trabajar en identificación de hablantes cuando los mensajes objeto de estudio no han sido producidos en la lengua materna del analista, o cuando éste no tiene un sólido conocimiento sobre dicha lengua.[I.R. n1 9], [I.A.F.P., s/f]

En términos generales, y desde un enfoque metodológico combinado, podemos constatar una posición de consenso en cuanto a que la idoneidad para la realización de informes periciales

judiciales en otras lenguas o dialectos distintos al nativo, corresponde a aquellos expertos que dominen las distintas normas de Fonética Acústica en las que se materialicen las diferentes bases de articulación de tales lenguas. Ese mismo es nuestro criterio, si bien, con carácter excepcional sería admisible emitir una opinión sobre registros en otra lengua, solamente en el caso de colaborar con las investigaciones de unidades policiales o fuerzas de seguridad. No es recomendable trabajar en este sentido de cara a la Autoridad Judicial.

III.1.2.- Objetos de estudio

Entendemos la voz como una realidad físico-acústica dimensionada por cuatro ejes: tiempo, presión o intensidad acústica, frecuencia y resonancia. Dichas referencias físicas tienen su correspondiente correlato a nivel perceptivo: duración, sonoridad, tonalidad y timbre. Estos ocho elementos fundamentales conformarán la estructura básica de análisis y serán complementados con distintas aproximaciones de estudio sobre las diversas características, elementos y rasgos que afectan e integran la supra estructura que denominamos habla.

De acuerdo a este planteamiento, cualquier experto o grupo de expertos que desarrolle un análisis de identificación forense de locutores, habrá de abordar los siguientes objetos de estudio:

III.1.2.1.- *COMPONENTES FÍSICO/PERCEPTIVOS BÁSICOS (VOZ)*

- FRECUENCIA SONORA
- PRESIÓN O INTENSIDAD ACÚSTICA
- TIEMPO
- ESTRUCTURA O ESPECTRO DE RESONANCIA

- TONÍA
- SONÍA
- DURACIÓN
- TIMBRE

III.1.2.2.- *COMPONENTES IDIOLECTALES DEL HABLA*

A/ Características Fonoarticulatorias (serán analizadas tanto referidas a alófonos como a otros grupos fónicos o estructuras de discuro de distinta dimensión).

A.1.- ***Fenómenos articulatorios*** :

- MODO DE ARTICULACIÓN
- LUGAR DE ARTICULACIÓN
- PRODUCCIONES O EVENTOS DE TRANSICIÓN
- RASGOS DISTINTIVOS PROSÓDICOS
- RASGOS DISTINTIVOS INTRÍNSECOS :

- Rasgos de Sonoridad : vocálico/no vocálico; consonántico/no consonántico; compacto/difuso; tenso/laxo; sonoro/sordo; nasal/oral; interrupto/continuo; estridente/mate.

- Rasgos de Tonalidad : grave/agudo; bemolizado/no bemolizado; sostenido/no

sostenido.

- ALTERACIONES DE DICCIÓN :

- Alternancias, Metátesis vocálicas y consonánticas, Sustituciones, Reducciones, Epéntesis (Prótesis, Elemento Esvarabático, Paragoge), Hiatos, Sinalefas, Asimilaciones y Disimilaciones, etc.

- ALTERACIONES DISFUNCIONALES : Dislalias (sigmatismos, rotacismos, para-rotacismo, lambacismo, ...) mudas de voz, Disfemia (tartamudeo), Bradilalias, Taquilalias, Disfonías Hipercinéticas, Hipocinéticas, alteraciones rino-faríngeas, etc.

A.2.- ***Modulación flujo respiratorio***:

- CADENCIA O RITMO RESPIRATORIO
- NIVEL DE INTENSIDAD RELATIVA DEL FLUJO FONATORIO

A.3.- ***Rasgos y secuencias de ataque/extinción articulatoria***.

B/ Características de orden lingüístico.-

B.1.- ***Nivel Morfo-sintáctico*** *:*

- CÓDIGOS DE CONSTRUCCIÓN DEL DISCURSO, etc.

B.2.- Nivel Léxico-Semántico

- VOCABULARIO
- RECURRENCIAS
- RECURSOS RETÓRICOS
- ELEMENTOS PARA LINGÜÍSTICOS
- ENTONACIÓN, ACENTO, etc.

C/ Características de la expresión comunicativa.-

C.1.- ***Alteraciones de la producción natural de emisiones:***

- *C.1.a.- de carácter involuntario*:

- PATOLOGÍAS
- FENÓMENO DUBITATIVO, ETC.

- *C.1.b.- de carácter voluntario*:

- VOZ DISIMULADA, ENMASCARADA, GRITADA, SUSURRADA,...
- VOZ PROCESADA, ETC.

- *C.1.c.- Elementos no vocales añadidos*:

- DE AMBIENTE ACÚSTICO (ruidos, efecto cocktail-party,etc)
- DE CANALES DE TRANSMISIÓN O TRANSDUCCIÓN,
- DE FACTORES DE REGISTRO O REPRODUCCIÓN (soportes, frecuencias de muestreo digital, etc.)

- **C.2.-** ***Rasgos suprasegmentales o prosódicos***

- ENTONACIÓN (MELODÍA/MONOTONÍA)
- ACENTO
- RITMO

- C.3.- ***Ratios Elocutivos***

- FLUIDEZ DE ELOCUCIÓN
- PAUSAS
- VELOCIDAD DE ELOCUCIÓN
- RATIO SILÁBICO (TEMPO)
- RATIO ARTICULACIÓN, ETC

- C.4.- ***Componente Psicolingüístico***

- FACTORES EMOCIONALES (furia, miedo, ansiedad, etc.)

III.1.2.3.- ***COMPONENTES DIALECTALES Y SOCIOLECTALES DEL HABLA***

- UBICACIÓN DIASTRÁTICA (Sociolingüística, Jergas, factor cultural)
- UBICACIÓN DIATÓPICA (Lenguas, Dialectos,acentos, códigos alternancia/influencia,etc.)
- PAUTAS DE RELACIÓN COMUNICATIVA (Proxémica, dialectos verticales, etc.)
- CARÁCTER MÁS O MENOS NORMATIVO, ETC.

III.1.3.-Clasificación y comparación de referencias de análisis

III.1.3.1.- Clasificación de referencias de estudio

Los mencionados objetos de estudio, desglosados, supondrán en su totalidad las que denominaremos *referencias base ó BR* para el análisis, cada una de las cuales podrá incluir a su vez diferentes *posibles referencias de decisión o PDR*. Por ejemplo, la tonía o referente perceptivo de la frecuencia fundamental de una voz, constituiría una de esas *BR* y sus correspondientes *PDR* podrían serían tres: Grave, Agudo y normal.

De la misma forma, las *BR* las clasificaremos en tres grupos distintos:

1.- *Referencias Globales o GR*.- Aquellas que debido a su propia naturaleza han de ser necesariamente consideradas como un todo: Timbre, componentes emocionales o sociolingüísticos, etc.

2.- *Referencias concretas o SR*.- Las que son apreciadas como aspectos singulares al margen de su posible ubicación en otra supra estructura más global. Sería el caso de las estructuras acústicas formanticas, que individualmente implican multitud de estimaciones, y a su vez son componentes de la cualidad vocal o timbre.

3.-*Parámetros mensurables o PA*.- Son las referencias básicas restantes. Su carácter específico responde a la sencilla, rápida y precisa determinación de los mismos en valores numéricos y estadísticos, utilizando una muestra de datos muy elevada.(Valor de F_0,

Jitter, Shimmer,etc.)

III.1.3.2.- Categorización de referencias

Las PDR en razón a su mayor o menor carácter individualizador deberán agruparse en tres niveles:

1.- *NIVEL 1 ó SIR*.- Valor identificativo estándar.
2.- *NIVEL 2 ó HIR*.- Alto valor identificativo.
3.- *NIVEL 3 ó MIR*.- Máximo valor identificativo.

III.1.3.3.- Clasificación de contextos comparativos

Las muestras objeto de comparación van acompañadas de unas condiciones más o menos adversas que propician una mayor o menor dificultad en el momento de emitir una decisión. De acuerdo a ello, dividiremos los contextos comparativos en tres clases:

*A/ FAVORABLES O NEUTROS***.-** Aquellos que no dificultan la toma de una decisión de identificación/eliminación.

B/ *DESFAVORABLES***.-** Los que debido a deficiencias de tipo cualitativo o cuantitativo de las muestras (*condiciones* α**)** o a la existencia de factores de discrepancia entre las mismas - plano expresivo, ratios elocutivos, distinto discurso,etc - (*condiciones* β), afectan de alguna forma a la formulación de una decisión.

C/ *MUY DESFAVORABLES***.-** Cuando concurren, en al menos una de las muestras, las anteriormente denominadas condiciones α *y* β.

III.1.3.4.- Tipos de tareas de comparación de las BR

Para obtener los diferentes resultados que nos permitirán elaborar una conclusión final de identificación/eliminación, se establecen tres tipos de comparaciones entre muestras:

- *TIPO I*.- Incluye comparaciones de dos clases: *Dubitada Vs Dubitada* e *Indubitada Vs Indubitada*. Este tipo de comparaciones se efectúan para obtener los valores

estándares de las BR -fundamentalmente los referidos a SR y PA- dentro de una misma emisión.

- *TIPO II*.- Comparación típica *Dubitada Vs Indubitada* para estimar los valores coincidentes o no coincidentes entre ambas.

- *TIPO III*.- Comparación de decisiones *coincidentes del TIPO II Vs Referencias Normativas* de la base de articulación analizada. Para otorgar una mayor o menor singularidad a la PDR en función de la desviación existente con su correspondiente valor normativo.

III.1.3.5 Márgenes de admisibilidad para BRs.

El mayor o menor número de BRs consideradas en una comparación estará en relación directa al grado de concreción alcanzado en los niveles finales de conclusión. No obstante, han de establecerse unos márgenes mínimos y óptimos de admisibilidad en función del tipo de BR evaluada. Los márgenes que a continuación desglosaremos solo serán aplicables a los sistemas de análisis básicos de los métodos combinados. Dada la ausencia de referencias documentales al respecto (excluyendo los estándares del VIAAS de la I.A.I.) la definición de dichos márgenes se fundamentará básicamente en la experiencia acumulada por el equipo de expertos del laboratorio de acústica forense del Cuerpo Nacional de Policía durante diez años de trabajo en casos reales. Los límites cualitativos y cuantitativos del sistema automático que utilizaremos en nuestro modelo, han sido establecidos por el equipo de diseñadores de la aplicación, tras los correspondientes estudios experimentales (posteriormente serán comentados)

A.- MÁRGENES CUALITATIVOS

Dejando al margen los ya abordados contextos de comparación en los que se valorarán en conjunto tanto las referencias de la calidad y cantidad de las muestras en sí mismas, como la influencia que sobre su correcta apreciación puedan ejercer otros factores que las acompañan, habrán de establecerse ciertos criterios en lo referente a la propia cualidad de la estructura acústica de la muestra. En este sentido, los márgenes de suficiencia cualitativa para el cálculo de PA y sus correspondientes estadísticos los marcarán, por una parte las referencias cuantitativas que se detallarán a continuación y , por otra, las propias necesidades técnicas de las aplicaciones utilizadas para dicho cálculo.

En lo relativo a SRs, y teniendo en cuenta que entre ellas se incluyen gran parte de los índices acústicos y fonoarticulatorios más importantes, una referencia a considerar pudiera ser el límite de los 2 Khz [VIAAS, I.A.I., 1991]. En los estándares marcados por el VIAAS de la I.A.I. se recomienda no abordar comparaciones espectrográficas cuando no exista información útil de señal de voz por encima de los 2000 Hz . Este límite orientativo puede soslayarse en el caso de ciertas referencias globales o GRs. Por ejemplo, determinadas características lingüísticas, patológicas o sociolectales podrán ser detectadas tomando como único condicionante la existencia de una suficiencia a nivel cuantitativo.

B.- MÁRGENES CUANTITATIVOS

- ***Márgenes de admisibilidad para GR y PAs*:**
(*Válidos para comparaciones de los tipos I,II y III*)

- *Margen mínimo (MM)* **:** Entre 10 y 60 seg. de muestra neta (mn)
- *Margen suficiente (SM)* **:** Entre 1 y 2 minutos (mn)
- *Margen óptimo (OM)* **:** Más de 2 minutos muest. neta.

- ***Márgenes de admisibilidad para SRs***:

A/ Para comparaciones del tipo I:

- *Margen mínimo* **:** Entre 5 y 10 referencias básicas (BR)
- *Margen suficiente* **:** Entre 10 y 15 BR.
- *Margen óptimo* **:** Más de 15 BR.

B/ Para comparaciones del tipo II :

- *Margen mínimo* **:** 10 referencias básicas (BR)
- *Margen suficiente* **:** Entre 10 y 15 BR.
- *Margen óptimo* **:** Más de 15 BR.

C/ Para comparaciones del tipo III :

- *Margen mínimo* **:** 10 referencias básicas (BR)
- *Margen suficiente* **:** Entre 10 y 15 BR.
- *Margen óptimo* **:** Más de 15 BR.

En relación con los márgenes de admisibilidad referenciados, resulta oportuno efectuar una apreciación. En el caso de aquellos relativos a las referencias globales (GR) y parámetros (PA), la diferenciación entre margen mínimo y margen suficiente podrá permitir desestimar la realización de un cotejo pericial tras proceder a una comparación del Tipo II, si se ha partido de un margen de calidad mínimo. Es decir, si analizada la calidad de una grabación dubitada no se le llega a otorgar el margen denominado "suficiente", podrá rechazarse el emitir una conclusión pericial una vez se disponga del correspondiente registro indubitado. Si en el proceso de evaluación de calidad se adjudicase un margen suficiente o superior, deberá procederse a emitir una conclusión pericial.

III.1.3.6.- Tipos de resultados de comparación

Los resultados derivados de una comparación cualquiera (TIPO I, II ó III) se expresarán de acuerdo a una de las tres posibilidades siguientes:

1.- *COINCIDENTES (MAT)***.-** Cuando la referencia comparada se encuentra dentro de los márgenes aceptables de variabilidad intrapersonal (curva PSS) de las emisiones habladas -para el caso de comparaciones del Tipo I y II- o coincide con los valores considerados normativos para las del Tipo III).

2.- *NO COINCIDENTES (NMA)*.- Cuando la referencia comparada no se encuentra dentro de los márgenes aceptables de variabilidad intrapersonal de las emisiones habladas -para el caso de comparaciones del Tipo I y II- o no coincide con los valores considerados normativos para las del Tipo III).

3.- *NEUTROS (NEU)*.- Cuando el resultado de la comparación no puede situarse con claridad en un nivel coincidente o no coincidente.

La determinación de los márgenes que delimiten la variabilidad intra/inter personal debieran ser establecidos para cada BR del castellano a través de los correspondientes estudios experimentales o trabajos de campo de alta muestra.

Los resultados de comparación no coincidentes podrán ser a su vez *"válidos" o "no válidos"* para su estimación. Se considerarán no válidos cuando las condiciones de contexto sean "desfavorables del tipo β" o "muy desfavorables".

Por otra parte, en el caso de las referencias concretas (SR), los *coeficientes de corrección* aplicados en comparaciones del Tipo III, serán de carácter *positivo para los resultados no coincidentes válidos*. Es decir, cuanto más alejada esté una BR de su referente normativo mayor valor identificativo poseerá.

III.1.3.7.- Niveles de Conclusión

Realizadas las necesarias comparaciones en cada una de las aproximaciones de estudio desarrolladas, y aplicados los coeficientes correspondientes a cada uno de los criterios arriba relacionados, los resultados obtenidos quedarán comprendidos en alguno de los siguientes niveles de conclusión:

- NIVEL DE IDENTIFICACIÓN
- NIVEL DE ALTA PROBABILIDAD
- NIVEL MEDIO-ALTO DE PROBABILIDAD
- NIVEL INCONCLUSIVO
- NIVEL MEDIO-BAJO DE PROBABILIDAD
- NIVEL DE BAJA PROBABILIDAD
- NIVEL DE ELIMINACIÓN.

Los niveles de identificación y eliminación nunca pueden ser entendidos como categóricos (100%), sólo representan el máximo nivel de certeza posible en una tarea forense de identificación de locutores.

Los anteriores niveles de probabilidad son absolutamente representativos de las reglas de decisión por escala de probabilidad verbal que en la actualidad son utilizadas por los expertos forenses del ámbito metodológico combinado.[I.R. n1 9 y 10]

III.1.4.- Sistemas de análisis

Conocidas las referencias y objetos de estudio, pasemos a enumerar y definir los sistemas de análisis en los que estos objetos son distribuidos para su óptima apreciación de acuerdo a la metodología propuesta en nuestro modelo:

III.1.4.1.- Perceptivo auditivo sobre referencias globales (GRs)

Se trata de un estudio comparativo sistemático de percepción auditiva, por alófonos, grupos fónicos y otras estructuras discursivas de rango superior. Como ya hemos explicado, el proceso perceptivo es algo más complejo que la simple audición. Por tal motivo, en este estudio juega un papel fundamental la experiencia y entrenamiento del experto para discriminar y contextualizar ciertas características de las emisiones habladas en un locutor determinado. Entre otras BR analizadas, pueden citarse la ubicación de la base de articulación y otras peculiaridades propias del sociolecto, las realizaciones a nivel suprasegmental, recursos retóricos, uso de las funciones expresivo-comunicativas del lenguaje, patologías, el timbre, ciertos ratios elocutivos, etc. Todo ello, se concreta documentalmente mediante el uso de la transcripción fonética.

De los cinco enfoques de estudio que vertebran nuestra propuesta, el que nos ocupa es percibido por los no expertos como el más accesible y vulnerable. Accesible, en cuanto a la aparente sencillez de comprensión del propio procedimiento de análisis; y vulnerable, en lo que respecta a su posible cuestionamiento por parte de aquellos que desconocen las dificultades reales de ésta aproximación de análisis. Por este motivo, merece la pena detenerse en los pormenores de este enfoque y desmenuzar todas las circunstancias que rodean al mismo para poderlo apreciar en su verdadera dimensión. De forma adicional, aportaremos una nueva opción de análisis (A.P.R.E.S.) que contribuirá a incrementar de forma notable el grado de objetividad

de este análisis perceptivo.

III.1.4.2.- La percepción auditiva en condiciones forenses.

Antes de adentrarnos en lo que será nuestro principal argumento -el sistema de análisis perceptivo-auditivo por expertos forensesB conviene detenerse un momento y examinar cómo dicha práctica es desarrollada en determinadas circunstancias por personas no expertas.

Existe un área de trabajo en el entorno de la Acústica Forense en la que se aborda el reconocimiento perceptivo-auditivo por víctimas y testigos de hechos delictivos. Nos estamos refiriendo a aquellas situaciones en las que un determinado acto de habla u otro tipo de sonido es escuchado por dichos sujetos sin que éstos hayan tenido la oportunidad de visualizar simultáneamente la fuente emisora de tal sonido (amenazas o extorsiones telefónicas, delitos sexuales o atracos por encapuchados, etc.). En estos casos, dado que el registro sonoro en cuestión se alojará en la zona de memoria a largo plazo de la víctima o testigo, será obligación del experto forense diseñar un protocolo de reconocimiento auditivo a través del cual se garantice el mayor índice de credibilidad en las posibles decisiones de identificación o eliminación alcanzadas por el sujeto. En estas situaciones, la comparación siempre se efectúa entre un modelo de locución a largo plazo y estímulos a corto plazo de la voz o voces de los posibles sospechosos. La parcela de la Acústica Forense que se ocupa de esta problemática es conocida a nivel internacional con distintos conceptos : AVoice line-up≅, AEarwitness Identification≅ o Ruedas de Reconocimiento de Voz (RRV). [Broeders y Rietveld, 1995 y 1996; Clifford, 1980; Hammersley and Read, 1983; Hollien H. et al.,1982,1995 y 1996; Huntley y Pass, 1993; Künzel,1994; Ladefoged 1978; Papcun et al. 1989]

Lógicamente, la participación del experto en este tipo de procesos no incide directamente en lo que es la pura práctica perceptiva que caracteriza una típica tarea forense de identificación por la voz. Cuando el experto

desarrolla su habilidad comparativa a nivel perceptivo lo realiza de forma metodológica y utilizando su memoria a corto plazo.

Desde los estudios pioneros de la doctora Frances McGegee en el emblemático caso del secuestro del hijo del coronel Charles Lindberg, [McGegee, 1937, 1944] se han desarrollado diversos estudios experimentales que han evaluado la eficacia del reconocimiento auditivo forense, por parte de sujetos expertos y no expertos y ante diferentes objetos, situaciones y factores. [Brown, 1979; Braun, 1996; Elaad, Segev y Tobin 1.998; Greenwald, 1.979; Hartmann, 1.979; Hollien H., 1990; Huntley, 1992; Koenig, 1986; Köster J.P. 1981; Kreiman et al., 1990; Ladefoged P. y Ladefodged J, 1980; Neiman et al., 1990; Nolan, 1990; Reich y Duke,1979; Rosenberg, 1973; Rothman, 1977; Saslove y Yarmey, 1.980; Shipp et

al., 1992; Smrkovsky, 1976; Stevens et al. , 1968; Tosi y Greenwald,1978; Tosi, 1979; Van Lancker y Kreiman 1987]. Sin entrar a analizar las distintas valoraciones expuestas en dichos estudios -todas ellas consecuencia de diferentes hipótesis y diseños de trabajo difícilmente comparables entre sí- resulta en cambio muy conveniente, llegado este momento, detenerse en lo que se consideran dos importantes axiomas. Por una parte, la indudable utilidad que este sistema de análisis representa para el experto forense, quien en muchas ocasiones, y aun contando con otras herramientas o aproximaciones de estudio de alta objetividad, encuentra en su propio oído el instrumento ideal para poder centrar y conducir su atención sobre determinadas circunstancias o eventos de la señal analizada.

Casualmente, ese factor Aatención≅ unido al de Aaprendizaje≅ o entrenamiento del experto, conformarán los basamentos que caracterizan y particularizan cada uno de los procesos perceptivos en cada individuo, se trate o no de un experto forense. Éste, es el fundamento de nuestro segundo axioma: resulta imposible controlar todas las circunstancias que intervienen o rigen cada uno de los procesos de codificación/comparación de estímulos, por lo que será extremadamente complejo establecer una referencia experimental que nos permita clasificar las diferentes habilidades perceptivas en cada individuo.

)Se encuentra entonces el experto en mejor predisposición que el no experto a la hora de precisar una decisión en la comparación de registros sonoros?

Con toda probabilidad, y en términos generales, podemos decir que sí. Si bien, como ya hemos razonado, estaríamos ante un claro error si nos pronunciásemos afirmativamente usando términos absolutos. Perfectamente, podría darse la circunstancia de que un profano, ante una situación determinada y unos estímulos concretos, alcanzase una decisión de superior rango de acierto que la de un sujeto entrenado.

Pero dejando al margen la capacidad auditiva óptima -la cual se presupone en el experto forense- tratemos de desglosar aquellas circunstancias que diferencian y aventajan su habilidad perceptiva respecto del no profesional. En primer lugar, y como referencia fundamental, el análisis perceptivo del experto es siempre desarrollado de forma metodológica y sobre un tipo específico de señal : la señal de audio forense. Esta señal, habitualmente degradada en su propia naturaleza por la transferencia telefónica, distorsiones, solapamientos, distintas clases de ruido y otros efectos del canal, en ocasiones se presenta acompañada de otras circunstancias no deseadas que pueden resultar críticas a la hora de llevar a cabo el análisis perceptivo : voz disimulada, imitada, susurrada, gritada, etc. Pues bien, estas especiales características

representarán para el experto forense su Apan de cada día≅ y para enfrentarse a ellas desde el enfoque más óptimo, ha de seguir un período de entrenamiento específicamente proyectado en las dificultades de tal entorno.

Durante su fase de adiestramiento, el experto adquiere una serie de hábitos perceptivos sustentados en una formación multi-disciplinar (Acústica, Psicoacústica, Fonética, Lingüística, Patologías y procesos de producción del habla, procesos de percepción auditiva, etc.) que le capacitan para obtener un conocimiento riguroso de las distintas estructuras (lingüístico-fonético-acústicas) a las que deberá enfrentarse en la casuística real. Una vez finalizado su entrenamiento, puede considerarse educada su percepción como la más aséptica y objetiva posible, estando así en la mejor disposición

para discriminar aquellas claves acústicas que considere más relevantes. De forma adicional, el experto Ba diferencia de lo que ocurre con el no profesional- se instruye en el manejo de distintas herramientas y opciones de análisis complementarias a lo que es el simple uso del oído. Dichos elementos, le permitirán una óptima selección y adecuación de la señal de cara a su consiguiente estudio de percepción auditiva.

III.1.4.3.- El análisis perceptivo-auditivo por expertos: objetos de estudio y procedimientos.

Como ya hemos comentado, en el actual ámbito de identificación de locutores forense carece de sentido la interpretación o utilización no complementaria de cualquiera de sus aproximaciones de estudio. Lógicamente, la perspectiva que nos ocupa no constituye una excepción. De hecho, las peculiaridades detectadas a nivel auditivo adquieren su verdadera entidad y peso identificativo cuando son dimensionadas respecto a sus correspondientes referencias acústicas, fonéticas, lingüísticas, etc. Es decir, imaginemos que durante nuestro análisis auditivo percibimos la especial realización de una A**s**≅. De forma inmediata, y en la medida de lo posible, procederemos a ubicar el fenómeno desde nuestro particular enfoque técnico. En un primer paso, asociamos dicha realización del fonema fricativo alveolar /s/ con una realización alofónica concreta, la cual a su vez, será comparada con otras realizaciones de fonemas iguales que aparezcan en las muestras analizadas, en similares y distintas distribuciones, para finalmente obtener una localización de la misma a nivel idiolectal, sociolectal e incluso patológico. Para ello, contaremos con herramientas y opciones de análisis que nos complementarán y facilitarán la tarea. Sonogramas de banda ancha pondrán de manifiesto los índices acústicos que caracterizan el evento analizado (en este caso podríamos

visualizar la iniciación de las estridencias y distribución de energía de las mismas) corroborando o no aquella sensación auditiva que nos informaba de un determinado grado en la realización articulatoria del fonema respecto a su adelantamiento/retraso, lugar, modo o tensión articulatoria , carácter mate o estridente, etc. En otros casos, análisis acústicos por espectrogramas de banda estrecha nos revelarán peculiaridades transitorias o permanentes

de los armónicos que integran los sonidos vocálicos, mostrándonos o confirmándonos aquellas circunstancias detectadas por nuestro oído que están directamente relacionadas con la propia naturaleza de la cualidad de voz .

El uso de estas opciones de representación gráfica de la señal, para constatar o descartar las discriminaciones auditivas efectuadas, unido a sus posibles interpretaciones complementarias desde otros enfoques de estudio (fonético, lingüístico, patologías del habla, etc.) es el único camino que conduce hacia una correcta y completa ejecución de lo que denominamos estudio perceptivo auditivo por expertos.

No puede decirse que exista un procedimiento normalizado o más deseable que otro para efectuar una práctica idónea del análisis auditivo. En algunos casos, dicho procedimiento estará determinado por el tipo de equipos de reproducción y análisis de los que disponga el experto. Por supuesto, el habitáculo donde se lleve a cabo la audición ha de reunir Ben la medida de lo posible- las mejores condiciones de aislamiento acústico e índices de reverberación. La utilización o no de auriculares para la escucha es una opción muy personal y dependiente en muchas ocasiones de las propias características del asunto objeto de estudio. Para extensas cantidades de registro, ésta opción puede incrementar el cansancio o estrés auditivo con la consiguiente pérdida de atención a nivel perceptivo; si bien, puede transformarse en una alternativa de gran utilidad para la escucha de eventos sonoros concretos, o en el caso de condiciones deficientes de aislamiento acústico del recinto de trabajo.

Dada la especificidad de éste ámbito forense, hoy en día es muy difícil encontrar un laboratorio o experto que no disponga de unas buenas herramientas para efectuar una audición en óptimas condiciones, desde el punto de vista de los procesos de grabación/reproducción. Otra cuestión, son las posibilidades y necesidades reales de cada laboratorio. En cualquier caso, aplicaciones de análisis y procesado digital de audio que hace dos décadas eran bastante inasequibles e incluso inaccesibles (caso del sonógrafo por ser considerado en su momento tecnología de uso militar) son hoy en día softwares muy económicos al alcance no sólo de cualquier experto sino de cualquier usuario doméstico. En este sentido, resulta totalmente necesario

disponer de estas aplicaciones de procesado y visualización gráfica (oscilograma,

sonograma, espectros FFT & LTAS, etc.) para adecuar y constatar las distintas características de la señal. Es un hecho frecuente Bespecialmente en los laboratorios de carácter público- la masiva afluencia de casos de estudio, algunos de ellos, incluyendo ingentes cantidades de grabaciones; ante esta situación, las opciones de análisis y procesado de señal en tiempo real se presentan como imprescindibles.

Teniendo en cuenta estas consideraciones, trataremos de describir el Aqué≅ y el Acómo≅, es decir los objetos de análisis y el procedimiento de trabajo. Empezando por este último, ya hemos avanzado que no puede hablarse de un procedimiento estandarizado. Por norma general, el experto se encontrará ante una grabación dubitada o anónima y otra u otras de carácter indubitado perteneciente(s) a uno o varios sujetos conocidos. Debe entenderse, que en la previa evaluación de calidad de las muestras objeto de cotejo (convenientemente etiquetadas y registradas) el experto ha determinado la existencia de una suficiencia en las mismas a nivel cuantitativo y cualitativo. A partir de aquí, o de forma simultánea a la realización de ésta tarea de evaluación preliminar (que normalmente se lleva a cabo sobre la muestra dubitada, pues de no haber unos mínimos suficientes en ella, no resulta necesaria la obtención de la indubitada) el experto toma buena nota de :

11.- todas las circunstancias elocutivas que estima relevantes desde un punto de vista identificativo, alteraciones incluidas (voz simulada, imitada, factor emocional, plano expresivo, posibles efectos de agentes tóxicos: drogas, alcohol, etc.). La materialización documental de esta labor se debe efectuar mediante el uso de la transcripción fonética, otra de las herramientas de indudable funcionalidad para la ejecución de distintas etapas del proceso.

21.- cualquier otro tipo de sucesos de registro que puedan afectar de alguna forma a la cadena de custodia o al posterior trabajo pericial (posibles daños, manipulaciones, distorsiones, solapamientos, alteraciones, etc.) .

En esta misma fase, cuando se dispone de las transcripciones de los registros dubitados facilitadas por los investigadores o la autoridad judicial, se procede a la revisión de las distintas correspondencias entre el material de grabación que se posee y las circunstancias reseñadas en las mismas: contenido, interlocutor al que se atribuye cada fragmento, número de soporte, cara, pasos o counter, etc.

Antes de iniciar la propia práctica perceptiva (nos referiremos a la auditivo-visual) , la señal de ambas muestras ha de ser pre-adecuada en similares condiciones de rango dinámico,

rango de frecuencia, velocidad de play-back, amplificación, transducción etc. Incluso, con un carácter excepcional, alguna o ambas señales pueden resultar sensiblemente procesadas (filtro de efecto telefónico, filtro para ruidos localizados, APRES*) aunque la norma general [VIAAS IAI, ; I.R. n110] recomienda no solo la no conveniencia, sino la inhibición del análisis auditivo en tales circunstancias.

Una vez igualadas las condiciones para una correcta dosificación de los estímulos, y en el caso de contar con una muestra indubitada del mismo contexto que la dubitada, se procederá a efectuar la comparación auditivo/visual a corto plazo (sonogramas de banda ancha) generalmente, por fragmentos no superiores a los dos segundos de duración por cada una de las muestras. Evidentemente, este marco temporal podrá variar en función del elemento o acto de habla objeto de análisis : alófono, grupo fónico, frase, etc. El establecimiento de umbrales mínimos de duración es igualmente complicado, aunque sí se observa una pérdida de eficacia a nivel auditivo cuando se utiliza el bombardeo continuado en bucle para fragmentos con duraciones inferiores a 300 ms. En estos casos, al producirse una descontextualización de eventos locutivos -respecto a los sucesos precedentes y consecutivos de la cadena hablada- puede generarse una percepción dificultosa o errónea de los mismos.

Cuando la comparación es efectuada con muestras de distinto contexto, se procede de idéntica forma, con la excepción necesaria de intentar localizar distribuciones articulatorias similares en ambos objetos de

estudio, para así poder complementar el análisis desde una perspectiva fonética. La existencia del mismo contexto y similar plano expresivo entre muestras facilitará enormemente el trabajo del experto forense. Por el contrario, distintos contextos y diferentes planos de expresividad incrementarán el grado de dificultad en la emisión de sus decisiones (especialmente en aquellas referidas al análisis fono-articulatorio y acústico del habla). No obstante, desde el enfoque de análisis lingüístico, la utilización de muestras de habla espontánea (distinto contexto) resulta insoslayable, pues ésta es la única vía para poner de manifiesto multitud de rasgos individualizadores del habla que aportarán una valiosa contribución al conjunto de las estimaciones.

En lo que concierne a cuáles son los objetos de estudio sobre los que tiene una especial incidencia la aproximación perceptivo auditiva, comentar antes de nada que la simple apreciación de los mismos no constituye por sí sola una referencia en el momento de emitir una decisión de comparación. De la misma forma, la constatación de similitudes o disimilitudes respecto a dichos rasgos, entre las muestras dubitada e indubitada carece de un valor definitivo. Para alcanzar una decisión de la máxima fiabilidad será necesario considerar

y combinar toda una serie de factores, además de los ya referidos: peso identificativo del rasgo, condiciones cualitativas y cuantitativas de la grabación analizada, existencia o no del mismo contexto comparativo en cuanto a contenido y plano expresivo, etc.

Dado que el proceso de evaluación de parámetros -tema crítico y complejo- ya ha sido comentado en un apartado anterior, pasaremos a detallar aquellos elementos de estudio sobre los que directa o indirectamente se desarrolla el análisis perceptivo forense.

Probablemente, las referencias de identificación biométrica por la voz con un carácter más elemental, son aquellas que nos aportan datos sobre la posible edad y sexo del hablante. Pero al margen de estas informaciones de rango general, nuestro análisis ha de sistematizarse clasificando las distintas características de la emisión de habla para, de esa manera, y en la medida de lo posible, lograr personalizarla. Para ello, incluiremos en un primer

apartado aquellas peculiaridades propias del sociolecto que de forma combinada son abordadas desde la perspectiva lingüística: ubicación diatópica (lenguas, dialectos, acentos, códigos de alternancia o influencia, etc.); factor sociolingüístico (jergas, dialectos verticales, etc.); códigos de construcción del discurso en sus distintos niveles (morfo-sintáctico, léxico-semántico, etc.); usos de la función expresivo-comunicativa (recursos retóricos, fluidez y velocidad elocutiva, elementos para lingüísticos, realizaciones a nivel suprasegmental : entonación, acento, ritmo, etc.). En un segundo grupo, citaremos aquellas características idiolectales que de forma conjunta son analizadas desde los enfoques fono-articulatorio o acústico. Es decir, rasgos de ubicación de la base de articulación (modo y lugar de las realizaciones, rasgos distintivos prosódicos e intrínsecos, etc.); habla defectuosa por causas involuntarias (patologías, tartamudeos, rotacismos, etc.) o voluntarias (voz disimulada, imitada, procesada, etc.)

En la actualidad, algunos de los índices mencionados, fundamentalmente ciertos ratios elocutivos (silábico, articulatorio) o indicadores suprasegmentales, pueden ser calculados de forma instantánea Bjunto a sus correspondientes valores estadísticosB con un altísimo grado de muestreo y, por lo tanto, de precisión. De la misma forma, pueden derivarse similares valores representativos de los componentes fundamentales del sonido del habla (F_0, amplitud de energía acústica, Jitter, Shimmer, etc.). Todos estos parámetros, aunque precisos, siempre tienen un valor relativo (con mayor o menor relevancia identificativa) en función del resto de circunstancias que acompañan a lo que consideramos pura realización del habla.

Indudablemente, la referencia reina en la identificación forense de locutores, por su poder individualizador, su alto nivel de invariabilidad y la circunstancia de constituir el correlato a nivel perceptivo de lo que se considera estructura básica del habla, no es otra que aquella denominada *cualidad de voz o timbre*.

III.1.4.4.- A.P.R.E.S., una opción de análisis

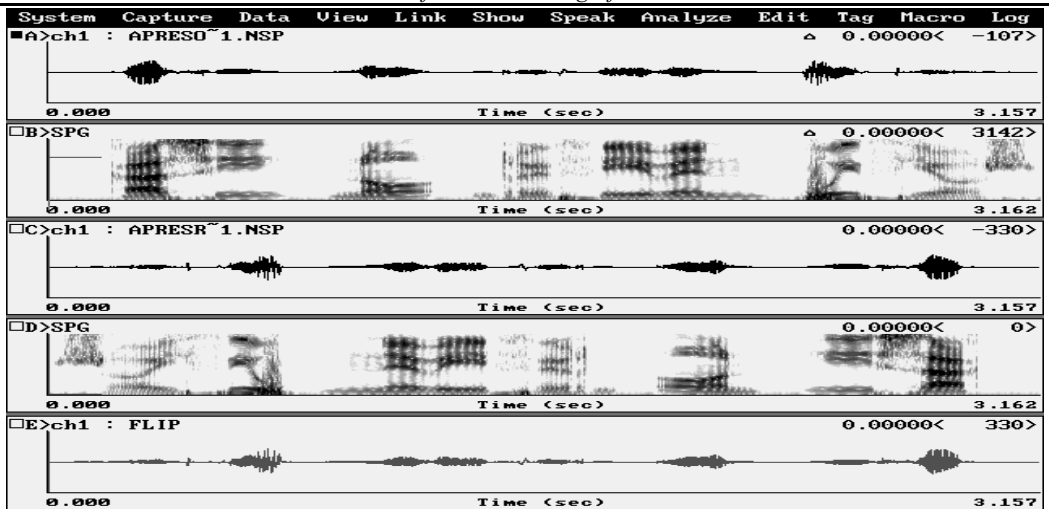
complementaria para la objetivación del proceso de percepción auditiva de la cualidad vocal.

El timbre que caracteriza una fuente emisora de sonido viene dimensionado por dos componentes básicos. Por una parte, los diferentes armónicos fundamentales que dicha fuente es capaz de generar y, por otra, la naturaleza y estructura de su propia cavidad de resonancia. De la misma forma, existen unas cualidades del timbre con un carácter altamente estable que están directamente relacionadas con el número, distribución y estructura acústica de sus componentes armónicos. Sin embargo, otras de sus características - aquellas relacionadas con los fenómenos articulatorios de ataque, permanencia y extinción - presentan una naturaleza más transitoria.

En el particular caso de la producción acústica del habla, el timbre vendrá siempre referenciado en base a los armónicos fundamentales generados en los pliegues glotales y al consiguiente efecto de resonancia al que dichos armónicos son sometidos en la cavidad resonante del tracto vocal. Dicha cavidad, aun poseyendo una estructura variable que puede ser modificada por la distinta disposición y funcionalidad de algunos de sus órganos (lengua, mandíbula, labios, etc.) está directamente relacionada con la responsabilidad de conferir a la cualidad vocal sus rasgos más individualizadores.

El análisis forense de la estructura espectral de la voz no sólo es proyectado sobre los componentes armónicos de la misma. La onda compleja que representa una emisión de voz incluye otros elementos pseudo periódicos y aperiódicos que también son objeto de análisis. Es decir, la apreciación perceptiva de la estructura básica de la cualidad de voz, va más allá de lo que consideramos simple percepción del timbre.

El oído entrenado del experto es capaz de discriminar la estructura básica de la voz de aquellos otros efectos y eventos sonoros simultáneos que acompañan y caracterizan las grabaciones forenses. No obstante, para lograr su objetivo, ha de alcanzar previamente el mayor grado de asepsia posible en relación a los distintos factores que subjetivizan el propio proceso perceptivo, ya sean éstos de una naturaleza exógena o endógena. De una u



otra forma, los efectos añadidos a la simple señal de habla (factor canal, ruido, respuesta telefónica, cocktail-party, etc.) junto con otras referencias de carácter contextual (planos semántico, emocional, expresivo) proporcionan siempre cierta dosis de subjetividad a la evaluación auditiva de la cualidad de voz. Por este motivo, cualquier medio que permita objetivar en alguna medida este tipo de estimaciones debe ser tomado en consideración.

La opción de análisis que hemos dado en denominar **A.P.R.E.S.(Aural Perception on Reverse Speech)** debe entenderse como una aportación de objetivación al clásico proceso perceptivo-auditivo de la cualidad vocal. En síntesis, no consiste en algo más que una audición sistemática de registros de habla en orden temporal invertido.

En la actualidad, diversas aplicaciones digitales de análisis o procesado de audio posibilitan la realización instantánea de este "reverso" temporal de la señal, que no debe ser confundido con la más habitual tarea de la inversión en fase. Dado que la ejecución técnica del trabajo no presenta mayor dificultad, el planteamiento inmediato es poner de manifiesto las aportaciones de funcionalidad que podremos obtener mediante la utilización de esta opción de análisis.

La pretensión prioritaria de este nuevo enfoque es incrementar el grado de objetivación de las apreciaciones perceptivo-auditivas sobre la cualidad vocal a través de la descontextualización de la misma, en relación a diversos factores y planos de referencia que dimensionan las emisiones habladas. Por tanto, no estamos ante una propuesta alternativa al análisis perceptivo clásico sino ante una opción complementaria específicamente diseñada para lograr una conceptualización más objetiva de la estructura espectral de la voz.

Los efectos de la descontextualización de mensajes hablados sobre los procesos de percepción de los mismos, han sido evidenciados ante distintas circunstancias [Miller e Isad, 1.963], [Pollack y Pickett, 1964], [Reddy, 1.976] pudiéndose constatar sensibles diferencias de codificación para similares eventos locutivos, cuando éstos se enfrentan a distintas situaciones de contexto. En nuestro caso, al proceder con el reverso de la señal de habla, la inmediata consecuencia perceptiva hemos de relacionarla con el cambio drástico de la propia naturaleza del estímulo. La cadena hablada en su desarrollo temporal natural se presenta como un estímulo "estructurado" con unos rasgos distintivos y unas constancias perceptivas familiares para el receptor. Sin embargo, el mensaje revertido se transforma en un estímulo de carácter "ambiguo" con un alto grado de abstracción, puesto que pierde diversos ejes de referencia a muy diferentes niveles (suprasegmental, semántico, emocional, sociolectal, etc).

De la misma forma que esta alteración de la cualidad del estímulo carece de sentido cuando el análisis del habla se realiza desde otras perspectivas de estudio, en el caso que nos ocupa, adquiere una especial relevancia. Al encontrarnos ante un estímulo de carácter ambiguo o no estructurado, los factores ligados a la experiencia individual del sujeto

receptor -conocimientos, intereses, motivaciones, actitudes- cobran un papel protagonista en el trabajo de interpretación del mensaje sonoro. Pero al producirse una ausencia de referentes, el proceso perceptivo comienza a desarrollarse sin la participación de ciertos elementos que caracterizan las tareas perceptivas clásicas. El *carácter cognitivo* de la percepción [Arnheim 1986], o lo que es lo mismo, las puertas de asociación a la experiencia [Booth, J., 1978] y los principios que rigen la separación de los distintos objetos auditivos en el ánalisis de la escena auditiva (ASA)[Bregman 1990] no llegan a establecerse. Lo mismo ocurre con los posibles fenómenos de *anticipación* (pudiera percibirse lo que se espera percibir y no lo que en realidad es) y *predisposición perceptual* basados en el hecho de que la percepción no es un input pasivo, sino un proceso activo de construcción o síntesis de conceptos [Neisser 1981].

En el caso concreto de la percepción de emisiones de habla se eliminan incluso factores tan delicados como aquellos que pueden derivarse de los efectos de *restauración fonémica* [Warren, 1970]. Warren, demostró experimentalmente que el oído -con el fin de otorgar un sentido semántico al conjunto del mensaje- trata de reconstruir aquellos intervalos en los que se produce la ausencia de información de un fonema sustituyéndolo por otro, que puede o no, ser coincidente con el originalmente emitido. Según Warren, este fenómeno se genera con más asiduidad cuando existe algún tipo de ruido -ingrediente característico de las condiciones forenses- en dicho intervalo de ausencia de información vocal.

En definitiva, podemos afirmar que la percepción de emisiones de habla mediante A.P.R.E.S. nos presenta la cualidad vocal desnuda, como si se tratase de una supraestructura que trasciende a la suma de sus distintos componentes. Un conjunto de rasgos sonoros que flotan en el aire carentes de otro significado que el de su propia identidad como estructura acústica.

Ya hemos mencionado que la realización técnica del reverso de señal no presenta mayor dificultad que la simple ejecución de una opción de edición; no obstante, sí resulta pertinente la recomendación de ciertas pautas de procedimiento para alcanzar buenos resultados en la consiguiente práctica perceptiva.

Antes de proceder con una tarea de A.P.R.E.S., hemos de decidir en que casos y ante que condiciones puede o no resultar oportuna. La aplicación de este procedimiento en casos forenses reales no necesariamente ha de efectuarse de forma sistemática. Cada experto debe determinar cuando puede resultar práctica su utilización. En este sentido, y siempre respetando la discrecionalidad del experto, nuestra experiencia nos dicta que precisamente en aquellas ocasiones en las que la confusión se apodera del criterio del analista, al efectuar una comparación auditiva a corto plazo en la forma tradicional, puede ser un excelente momento para desarrollar nuestra propuesta.

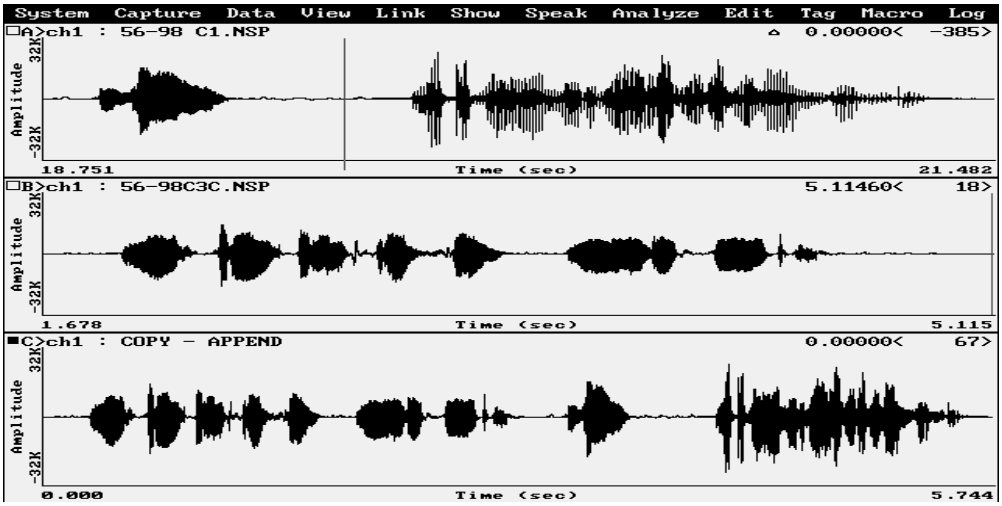
Antes de proceder a su reverso, las muestras objeto de cotejo requieren de una adecuación, fundamentalmente en sus referencias de amplitud y nivel de salida (se presuponen unas condiciones óptimas del canal de reproducción de la señal y de aislamiento acústico del recinto).

A diferencia de lo que ocurre en la comparación auditiva sobre señal no revertida, los fragmentos de señal seleccionados han de tener una duración superior. No olvidemos, que el objetivo de nuestro procedimiento es obtener una sensación de la cualidad vocal en emisiones en las que existe una ausencia de importantes claves de referencia perceptiva; por este motivo, necesitaremos unos mínimos cuantitativos de una duración superior a los típicos 2 segundos de la comparación clásica a corto plazo. A ser posible, debemos de pensar en fragmentos netos de habla de al menos 10 segundos, o superiores. Asimismo, resultará crítica la eliminación por edición digital de las pausas existentes entre los distintos grupos fónicos (nos estamos refiriendo a pausas de 200 ms o de mayor duración). La presencia de elementos acústicos no deseados (ruidos, efecto cocktail-party...) ha de evitarse a la hora de seleccionar las muestras para la comparación. En el caso de procederse ante una situación de estas características -por ejemplo, con un clásico ruido blanco de fondo- puede resultar más recomendable añadir mediante edición el elemento enmascarador a la muestra no contaminada, que tratar de eliminarlo de aquella que lo posee.

En una primera fase, los estímulos ya revertidos son administrados de forma independiente. Es decir, una vez pre-adecuadas las muestras nos concentraremos primeramente en una de ellas mediante estimulación por

repetición en bucle, sin pausas. Seguidamente, realizaremos idéntica tarea con la otra muestra. Una vez familiarizados con ambos estímulos, procederemos a su comparación a corto plazo de forma alternativa o *continuada*.

La *comparación continuada* sobre registros revertidos es en nuestra opinión la que debe ser utilizada en último lugar en el caso de combinarse



con otra de carácter alternativo. Para conseguir una comparación sobre muestras continuadas, los registros dubitado e indubitado una vez adecuados y revertidos se pegan mediante edición digital uno a continuación del otro (por fragmentos y sin pausa entre los mismos). Finalizada la tarea de procesado se procede a la audición de las señales editadas por repetición

en bucle y sin pausas.

Cuanto mejor sea la adecuación entre ambas señales, menos dificultades encontrará el experto a la hora de construir su opinión. El procesado de reverso sobre la señal produce un ruido residual añadido (matiz metálico) y cierta distorsión de la cualidad original, distorsión que no puede considerarse crítica desde un punto de vista perceptivo. Sin embargo, la eliminación de pausas intra e inter emisiones es una labor fundamental para evitar distraer la atención perceptiva sobre la auténtica naturaleza de nuestro objeto de estudio. Este postulado es fácilmente demostrable a través de la aplicación a estímulos de carácter sonoro de una de las cuatro leyes gestálticas definidas para la organización de los estímulos visuales. Nos estamos refiriendo a la A*ley de la buena continuación*≅**,** ley, que afirma que *A…los puntos que al conectarse, den lugar a líneas rectas o a una curvatura suave tienden a agruparse perceptivamente…*≅ . En este sentido, ya ha sido demostrada experimentalmente [Warren, Obusek y Acroff, 1.972] la existencia de una buena continuación auditiva por parte del receptor, cuando éste percibe los estímulos sonoros sin la presencia de intervalos de silencio o pausas.

Esta referencia de la Gestalt no es la única que refrenda la funcionalidad de nuestra opción de análisis. La *Aley de la similaridad*≅, por la que los *Aelementos similares tienden a agruparse*≅ fue igualmente analizada en relación a los estímulos auditivos. Los psicólogos Bregman y Campbell [Bregman y Campbell, 1.971] observaron que se producía una segregación en la percepción del flujo auditivo cuando los estímulos alternos (en nuestro caso muestra dubitada e indubitada) eran suministrados en secuencias temporales no suficientemente rápidas. Este fenómeno, podría ser la explicación de porqué la administración de estímulos para una comparación mediante A.P.R.E.S. debe ser efectuada con la mínima dilación posible entre ambos.

El análisis por A.P.R.E.S., además de suponer las ventajas ya expuestas - consecuencia de la descontextualización- incrementa el valor del análisis perceptivo forense en situaciones independientes de texto, adjudicando una mayor relevancia a las muestras de habla de carácter espontáneo. Por otra parte, el uso de esta opción complementaria de análisis

diferencia claramente la labor del experto forense objetivando el carácter de sus apreciaciones perceptivas en relación a los no expertos (aspecto importantísimo de cara a la defensa de informes periciales ante las autoridades judiciales).

III.1.4.4.5.- PERCEPTIVO MIXTO FONOARTICULATORIO (AUDITIVO/VISUAL) SOBRE SRs.

Las conclusiones alcanzadas a través del estudio fonoarticulatorio serán una consecuencia de asociar las peculiaridades propias del idiolecto a referencias estándar de las diferentes realizaciones articulatorias de la lengua española. Dichas peculiaridades, son detectadas mediante la combinación de análisis de percepción auditiva (similar al efectuado sobre las GRs) y de representación gráfica de la señal.

Para poder llevar a cabo una correcta estimación de las características analizadas, resultará imprescindible que el experto posea un profundo conocimiento de la base de articulación objeto de estudio.

En esta fase de análisis y en el estudio acústico es donde se suelen poner de manifiesto aquellas características de la fase productora del habla (PDR) con mayor valor identificativo o discriminativo a nivel individual.

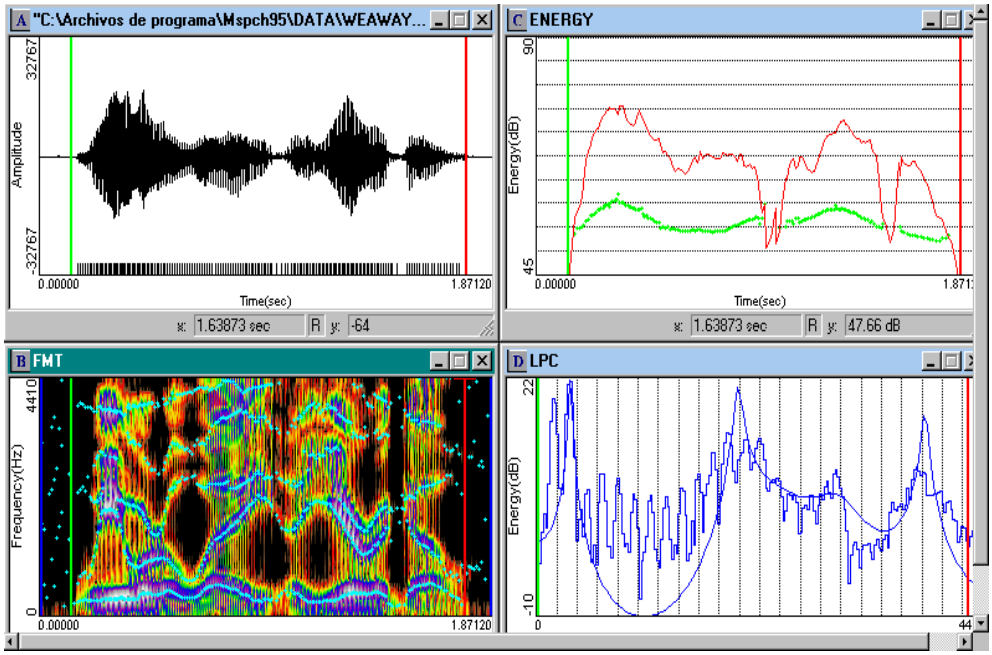
III.1.4.4.6.-ANÁLISIS ACÚSTICO SOBRE REP.GRÁFICAS DE LA SEÑAL

En esta etapa de estudio las grabaciones, dubitada e indubitada, son sometidas a un análisis comparativo a través de las distintas formas de representación gráfica de la señal de voz. Resonancias armónicas y otros índices acústicos producidos por las cavidades fonadoras en la realización de ambas muestras (transiciones, ataques, extinciones, intensidades relativas, áreas localizadas de energía, etc.) son apreciadas y mensuradas para obtener una estimación de similitud/disimilitud.

El estudio sobre representaciones gráficas de la señal suele deducirse de la observación de sonogramas de banda ancha, aunque también pueden utilizarse filtros de banda estrecha, oscilogramas, espectros a corto y largo plazo, etc. Las opciones de captura y análisis que

dimensionarán las distintas formas de representación gráfica del habla (intervalos temporales, ancho de banda del filtro, enventanado, rango de frecuencia, rango dinámico, escalas de color para gradación de energía, etc.) serán seleccionadas y combinadas por el experto en función de los distintos eventos acústicos que deban ser analizados.

III.1.4.4.7.- ANÁLISIS DE REFERENCIAS MENSURABLES (PA).



Existen parámetros mensurables de la voz que son estimados de forma complementaria a los resultados obtenidos con los otros sistemas de análisis. Debido al carácter variable del habla y

la diferencia que normalmente se produce entre los planos expresivos de la grabación dubitada e indubitada, solamente podrá otorgarse una alta relevancia identificativa a dichos parámetros, en el caso de hallarnos ante aquellos contextos comparativos que denominamos *Afavorables≅*.

El cálculo de estos parámetros de alta muestra (F_0 , Jitter, Shimmer, promedios espectrales o de energía, vAm, ratios de inarmonicidad , etc) y sus correspondientes valores estadísticos, es una tarea absolutamente automática. Sin embargo, la selección de los fragmentos de habla que depararán estimaciones válidas para su consideración con fines identificativos, sólo podrá ser efectuada por un experto.

III.1.4.4.8.- ANÁLISIS AUTOMÁTICO

Las cuatro anteriores aproximaciones de estudio que conforman el esqueleto de nuestro modelo se verán complementadas con una opción de análisis automática representada por un sistema de reconocimiento basado en el modelado de clases fonéticas por mezclas de Gaussianas (*Gaussian Mixture Model ó GMM*). [Reynolds, 1992].

) Por qué hemos seleccionado este enfoque automático y no otro de similar naturaleza, o de carácter semiautomático?) por qué hablamos de opción Acomplementaria≅?

Dadas las características del prototipo que utilizaremos en nuestro modelo práctico, en principio -y en tanto no sean testeadas exhaustivamente sus prestaciones para entornos forenses- no lo aplicaremos en la propia tarea de verificación (comparación entre dos muestras de habla para determinar si pertenecen o no al mismo locutor). Nuestro plantemiento interpretará los resultados aportados por este sistema como una opinión complementaria de identificación que relacionará nuestra muestra dubitada con una población de locutores entre los que se encontrará incluido nuestro candidato indubitado. Es decir, una vez obtenida una conclusión a través de los resultados alcanzados en los cuatro enfoques de estudio ya detallados, pediremos al sistema un refrendo a nuestra conclusión, solicitándole una valoración de similitud entre nuestra voz dubitada y una población indubitada concreta.

Para que dentro del marco de nuestro planteamiento metodológico, ésta u otra alternativa automática pudieran considerarse ante casos reales, debiera primeramente, estar referenciada a una base de datos de locutores suficientemente representativa en distintos niveles: sexo, intervalos de edad, condiciones de registro forense, sociolecto, dialecto, etc.; y en segundo lugar, tendría que ser validada experimentalmente para tener una clara idea de la eficacia del sistema ante distintas situaciones problema.

A la cuestión de porqué la preferencia de un sistema automático ante otro de los que denominamos semiautomáticos, responderemos con dos argumentos. Por un lado, ya hemos comentado suficientemente que el límite entre uno y otro entorno, es complicado de establecer. Todos los sistemas en cierta forma son semiautomáticos, aunque en anteriores capítulos, hemos convenido que cuando el sistema nos aporte una decisión de comparación, debe quedar enmarcado en el ámbito automático.

Por otra parte, también tuvimos oportunidad de citar algunos sistemas semiautomáticos que en la actualidad se están utilizando con fines forenses. En general, estas aplicaciones no hacen más que otorgar un mayor grado de funcionalidad y agilidad a métodos de identificación cuyas bases de análisis son muy similares a las de los enfoques de estudio tradicionales.

Atendiendo fundamentalmente a estas razones, consideramos que la utilización de una perspectiva de carácter más automático puede aportarnos:

1.- unos resultados fundamentados en unas opciones de parametrización distintas a las evaluadas en los enfoques de análisis clásicos.

2.- una decisión de comparación complementaria con un alto nivel de objetividad.

)Porqué un sistema de modelado de clases fonéticas basado en mezclas de gaussianas (GMM)?.

Para poder explicarlo mejor, haremos un alto en el camino y analizaremos brevemente las distintas generaciones de sistemas de reconocimiento automático con sus principales características y prestaciones. Una vez situados en este entorno, estaremos en mejor disposición para argumentar nuestra elección y explicar en detalle las distintas opciones que la integran. La mayor parte de los datos y valoraciones que aportaremos a continuación han sido proporcionados por los doctores Ortega García y González Rodriguez del Departamento de Ingeniería Audiovisual y Comunicaciones de la E.U.I.T. Telecomunicación de la Universidad Politécnica de Madrid, quienes contribuyen al enfoque automático de nuestro modelo aportando el prototipo de reconocimiento A*Identivox 2000"* (basado en modelado GMM).

III.5.5.1.- Clasificación básica de los SARL.

Desde sus comienzos hasta la actualidad, el reconocimiento automático de locutores ha deparado distintas generaciones de sistemas. En términos generales, dichos sistemas pueden ser agrupados en dos grandes vertientes: los denominados de comparación de patrones y aquellos

basados en la modelación de clases fonéticas. Los primeros, determinan la distancia de similitud entre un fragmento acústico concreto que se supone suficientemente representativo de la identidad de su emisor y un tramo patrón que es considerado referencia identificativa del hablante a reconocer. Los segundos, efectúan como tarea previa una discriminación de las distintas clases fonéticas de una locución para así modelar las características acústicas dependientes del locutor a través de las categorías fonéticas de los distintos alófonos. A continuación, se efectúa la comparación con independencia del texto pronunciado.

La servidumbre de la dependencia de texto que se produce en los sistemas de comparación de patrones junto a su ausencia de generalidad, representan una clara desventaja frente a las opciones de reconocimiento por modelado de clases fonéticas.

Al margen de estas dos grandes subdivisiones, hemos de citar los *sistemas basados en redes neuronales (Neural Networks ó NN*) [Lippman, 1987][Dayhoff, 1990] o arquitecturas conexionistas. Estos sistemas basan su poder clasificatorio más en la capacidad directa de discriminación de un locutor en relación a un conjunto dado, que en su capacidad de modelar la información a nivel individual. La topología de estos modelos de cálculo trata de emular los mecanismos de interconexión que se producen entre las neuronas cerebrales humanas. Se han propuesto diferentes arquitecturas de redes -perceptrón multicapa, NN de retardo temporal, funciones de base radial, etc.- que conjugadas con ciertas estrategias de agrupación (por ejemplo la binaria) posibilitan diversas potentes técnicas de clasificación como son la del Máximo global o la Binary Tree Search.

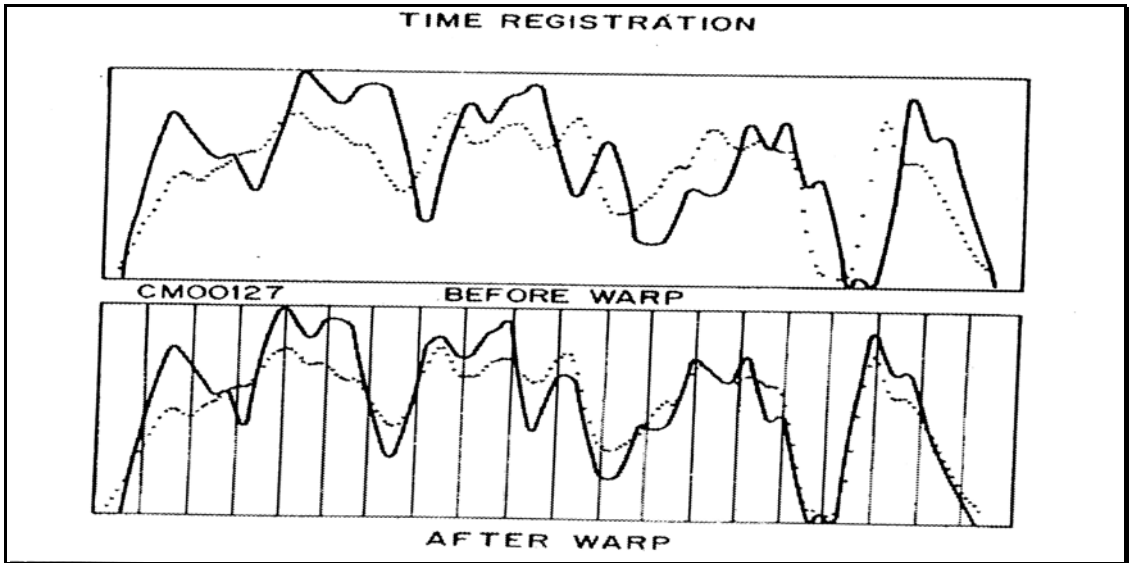
En general, los resultados de comparación de estos sistemas son similares a los que deparan los sistemas de cuantificación vectorial que posteriormente analizaremos, aunque en el caso de los NN basados en una sóla red el coste computacional de determinadas tareas (reentrenamientos de nuevos candidatos del sistema, por ejemplo) es mucho más elevado. No obstante, últimamente se están aportando soluciones de reconocimiento por la voz - tanto en tareas de identificación como de verificación - en las que parecen reducirse sensiblemente tanto los tiempos de entrenamiento del sistema como las capacidades de almacenamiento en memoria. Es el caso del sistema TESPAR/FANN (Time Encoded Signal Processing and Recognition / Fast Artificial Neural Networks) [King y Phipps, 1999].

Dentro de los sistemas basados en la *comparación de patrones*, podemos distinguir:

- *Sistemas basados en la comparación de tramos a corto plazo*: detectan y comparan tramos fonéticos de alta estabilidad (vocales abiertas, consonantes nasales) aplicando técnicas de correlación cruzada, coherencia, etc, para la medida de distancias. Los principales

inconvenientes de estos sistemas se relacionan con la enajenación de la información a nivel suprasegmental y la necesidad de supervisión en las tareas de segmentación.

suprasegmental. También son susceptibles de la utilización de medidas de distancia más robustas



en presencia de ruido.

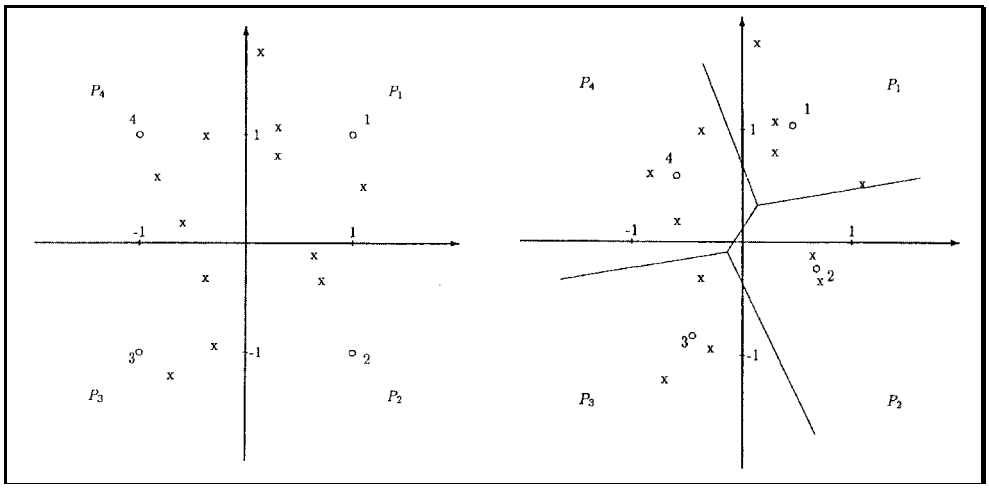
Los sistemas DTW han sido utilizados en algunas metodologías forenses como un complemento a otros análisis clásicos [Suzuki et al., 1997; Sistema SVL, 1998...].

- *Sistemas basados en la extracción de estadísticos a largo plazo de la señal:* extraen estadísticos de la señal en el dominio espectral (normalmente el espectro promedio) en tramos con suficiente duración. Su principal problema es el carácter demasiado genérico y por tanto poco discriminativo de las características que se derivan del promediado espectral a largo plazo.

Los esquemas de *modelado de clases fonéticas* pueden dividirse en tres grandes grupos de sistemas [Ortega y González, 1995]:

- *Sistemas basados en cuantificación vectorial o VQ:* alcanzan su auge de aplicación para el reconocimiento de locutores en la década de los ochenta. Hacen uso de las ventajas teóricas formuladas en la Teoría de la Información de Shannon, que establece que la cuantificación escalar es menos óptima que la vectorial. La cuantificación vectorial es considerada como una potente técnica de clasificación y discriminación [Soong, et al. 1987] [Gersho y Gray, 1991]. En una primera etapa del proceso de reconocimiento, las clases fonéticas que caracterizan las emisiones

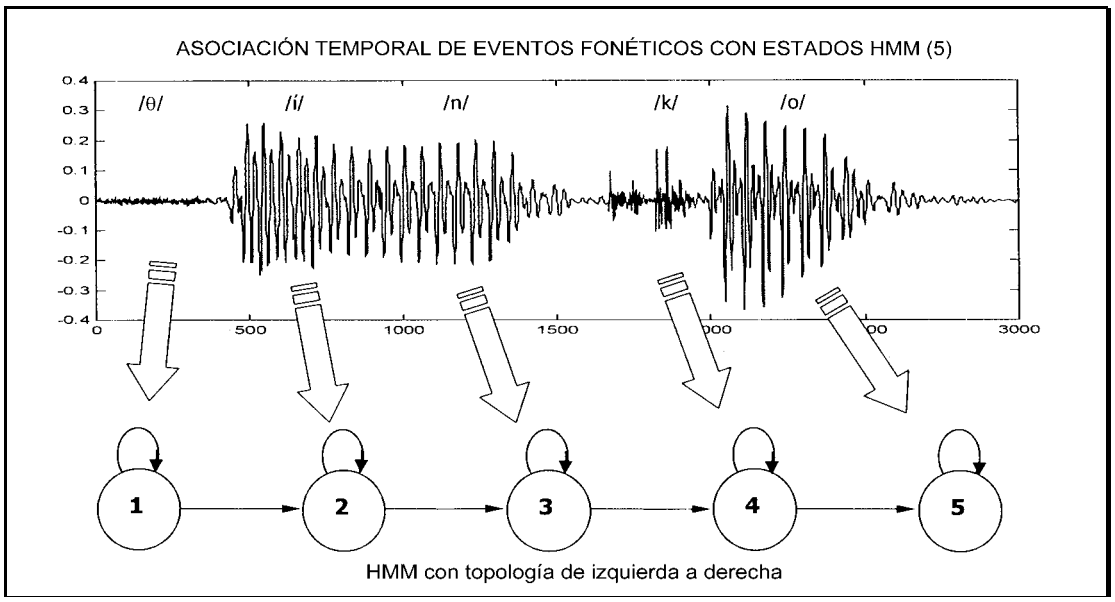
de un locutor(es) son representadas por un conjunto finito de vectores denominado libro de códigos o *codebook.* A cada uno de los vectores representativos de cada una de las clases o *clusters* a clasificar se les denomina centroides o *codeword*. Posteriormente, se efectúa el mismo proceso con la locución a comparar, asignando a cada uno de sus vectores representativos uno de



los centroides ya calculados (el más cercano en distancia).

Las ventajas más notables de los VQ vienen representadas por una reducción sensible de la capacidad de almacenamiento en el cálculo del análisis espectral y una reducción de la complejidad computacional en el cálculo de distancias (se puede usar cálculos tan simples como

la distancia euclídea o la de Mahalanobis). Sus inconvenientes más significativos están relacionados con la distorsión espectral por el error de cuantificación (al representar cada vector por un representante) y la relación inversa entre la necesidad de utilizar codebooks lo más

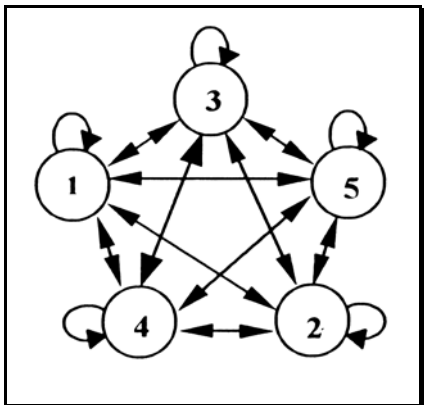


amplios posibles y los consiguientes problemas de almacenamiento .

- *Sistemas basados en Modelos Ocultos de Markov o HMM* : son redes de estados que intentan modelar el mecanismo de producción del habla. Están integrados por un conjunto de estados (asimilables a las distintas posiciones en las que puede configurarse el tracto vocal durante una locución) cada uno de los cuales desemboca en un conjunto de posibles salidas (asimilables a las posibles distintas realizaciones alofónicas). De alguna forma nos encontramos ante una discriminación fundamentada en la asociación fonema/alófono.

Mientras que las aplicaciones de reconocimiento del habla presentan una topología clásica de modelado conocida como Ade izquierda a derecha≅ (ver ilustración n162), los modelos utilizados por los sistemas HMM para caracterizar la identidad de un locutor independiente de

texto son los denominados *ergódicos,* en los que no existe una ordenación correlativa de las transiciones habidas entre los distintos estados del modelo y, por lo tanto, resulta factible cualquier combinación de transición entre estados.

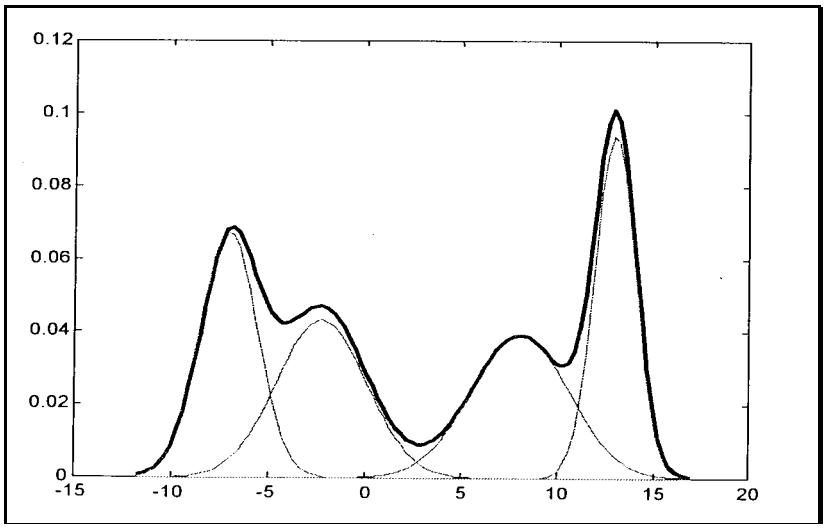


Cada HMM viene dimensionado por tres ejes de referencia: la matriz de probabilidades de transición entre estados, el vector de probabilidades de estado inicial y la distribución de probabilidades de salida u observación. Respecto a esta última, existe la posibilidad de trabajar con HMMs de observación discreta (DDHMM) o de observación continua (CDHMM). En general, los CDHMM modelan con mayor fidelidad y producen tasas más altas de identificación que los no continuos, si bien, en condiciones de trabajo en tiempo real los

discretos parecen resultar más ágiles.

La principal ventaja de los sistemas de reconocimiento basados en HMM respecto a otros tipos ya referidos, la constituye su gran versatilidad, tanto en lo que se refiere a los procesos de entrenamiento como a ciertas características variables de la muestra : duración, contenido fonético o lingüístico, contexto, etc . A todo ello, hemos de añadir su gran adaptabilidad a la variación de las condiciones de registro o del canal de transmisión y, lógicamente, su funcionalidad en condiciones independientes de texto.

- *Sistemas basados en modelos de mezclas de Gaussianas o GMM:* modelan los distintos vectores de parámetros de una locución dada, realizando una suma ponderada (mezcla) de funciones de densidad de probabilidad gaussianas. En la siguiente figura, apreciamos como una curva de función de densidad de probabilidad modela mediante una mezcla de cuatro funciones gaussianas una distribución experimental cualquiera.



Pueden entenderse como un sistema en el que se aglutinan las virtudes de aquellos otros basados en técnicas VQ y los denominados clasificadores gaussianos uni-modales. También pudieran apreciarse como CDHMMs de un sólo estado o CDHMMs ergódicos con probabilidades de transición entre estados equiprobables.

Sin embargo, utilizando GMMs podemos representar con un alto grado de fidelidad un amplio margen de distribuciones muestrales, tales como las de los diferentes coeficientes cepstrales que puede generar una locución concreta.

A diferencia de los HMMs ergódicos de varios estados, los GMMs no precisarán en la fase de entrenamiento segmentar en estados ni entrenar la matriz de probabilidades de transiciones. Además, en la etapa de reconocimiento, no será necesario buscar la secuencia de estados de máxima verosimilitud (algoritmo de Viterbi), sino que bastará con acumular las probabilidades que asocia el modelo con cada uno de los vectores instantáneos de entrada.

Además de las ventajas citadas, interesantes estudios comparativos sobre el rendimiento de diferentes técnicas de reconocimiento automático ante distintas circunstancias (procesos de entrenamiento, inputs de captura de muestras, SNRs, parametrización, factores de degradación, etc.) han contribuido a nuestra decisión de utilizar como sistema complementario de análisis un prototipo basado en modelado GMM. [Furui, 1994]; [Doddington et al., 2000]; [Ortega, 1996]; [Ortega y González, 1995]; [Vivaracho et al., 2000], etc.,

No obstante, el hecho de señalar cual será nuestra opción de modelado o la tarea de discriminación en que será utilizado el sistema (identificación/ verificación) no será información suficiente a la hora de referenciar nuestras conclusiones. También habremos de informar pormenorizadamente tanto sobre las condiciones de los registros analizados como de las distintas opciones de análisis seleccionadas: características de las poblaciones de locutores de referencia, técnicas de adecuación de la señal, duración de los tramos de entrenamiento, número de mezclas utilizadas, frecuencia de muestra, tamaño y solapamiento de las ventanas de análisis, tipo de enventanado, clase de parámetros de frecuencia, energía, etc., utilizados (LPCC. MFCC,...), uso

de coeficientes de regresión temporal o dinámicos (velocidadΔ, aceleraciónΔΔ), orden del análisis, etc.

Este tipo de datos relativos a las opciones concretas de análisis que serán utilizadas en la resolución de nuestro caso práctico, serán debidamente detallados en el capítulo siguiente.

CAPÍTULO IV

APLICACIÓN PRÁCTICA DEL MODELO

IV.1.- PRESUPUESTOS DE ANÁLISIS

IV.1.0.- Introducción

Ha llegado el momento de verificar la funcionalidad de nuestro modelo combinado. Para ello, partiremos de un supuesto delictivo en el que se incluirán ciertos elementos de estudio propios de nuestro ámbito de investigación. Lógicamente, dichos elementos contarán a priori con unos márgenes cualitativos y cuantitativos que permitan desarrollar su análisis a través de las diferentes perspectivas de estudio que conforman nuestro método. En este sentido, hemos de insistir en la especial dificultad que las grabaciones de audio forense representan para el buen rendimiento de los sistemas automáticos, motivo por el cual, nuestro planteamiento práctico tan solo interpretará los resultados obtenidos desde este enfoque como una orientación complementaria que corrobore en una u otra dirección otro tipo de conclusiones.

Las distintas fases del protocolo de trabajo que presentaremos a continuación, se corresponderán con la sistemática de acciones que habrán de desarrollarse para la plena ejecución de un estudio combinado de identificación de locutores. Es decir, no solo abordaremos las puras tareas de cálculo y análisis, sino también aquellas otras de carácter accesorio, aunque imprescindibles para la consecución de las más fiables resoluciones periciales: tomas de muestras indubitadas, adecuación y selección de parámetros, etc.,

IV.1.1.- Supuesto de investigación

En un establecimiento de hostelería situado a las afueras de un pueblo madrileño, la policía ha detectado ciertas actividades de tráfico de estupefacientes. Conocidos pequeños distribuidores o Acamellos≅ visitan con asiduidad el establecimiento, en el cual trabajan otros individuos también con antecedentes policiales por narcotráfico. El interés prioritario de los investigadores policiales se centra en descubrir las raíces del problema, o lo que es lo mismo, llegar a conocer los máximos responsables de esta trama delictiva que, como suele ser habitual, se encuentran a una respetable distancia de las evidencias y hechos más comprometedores.

Considerando los elementos expuestos, la autoridad judicial autoriza a las unidades policiales la intervención de la línea telefónica adjudicada al mencionado local. La accesibilidad a dicha línea está restringida a los trabajadores del establecimiento.

Una vez finalizado el plazo de intervención señalado por la autoridad judicial, los investigadores policiales cuentan con diversas grabaciones de las conversaciones sostenidas a través del teléfono en cuestión. Afortunadamente, muchas de estas conversaciones ponen de manifiesto la clara implicación al más alto nivel de uno de los empleados del local.

Posteriormente, y como consecuencia de otras investigaciones, la policía detiene a diez trabajadores del establecimiento, ordenando el juez su ingreso en prisión. Analizado el contenido de algunas de las conversaciones relacionadas con los hechos, la autoridad judicial procede al interrogatorio de los detenidos a fin de atribuir las correspondientes autorías.

Uno de los implicados, a quien se adjudican algunos de los contenidos que revelan su grado de participación activa y ejecutiva, niega reconocer su voz al ser reproducida en acto judicial. Ante esta disyuntiva, el juez ordena la identificación de los registros de habla atribuidos a dicho sujeto.

IV.1.2.- Evaluación previa de las muestras dubitadas.

El juez instructor ha seleccionado las conversaciones o fragmentos de discurso sobre los que desea sea realizada la pericia de identificación del locutor. Convenientemente acotadas, y guardando escrupulosamente los protocolos de cadena de custodia, las grabaciones objeto de estudio son remitidas al laboratorio forense para su evaluación a nivel cualitativo y cuantitativo; este requerimiento previo es imprescindible, pues no tendría sentido alguno continuar con otro tipo de procedimientos si ya en esta fase fueran observadas deficiencias críticas a uno u otro nivel.

Desde el mismo momento en que los soportes de grabación se encuentran en poder del experto, el proceso de estudio pericial puede considerarse iniciado. Preservada la cadena de custodia, las primeras observaciones se dirigirán hacia las características físicas del material remitido. Debidamente etiquetados y retirados los seguros de grabación en su caso, efectuaremos un examen visual de carcasas y emulsiones para detectar posibles daños o alteraciones físicas en los mismos.

Finalizada la inspección sobre los soportes de registro, continuaremos con la evaluación de las que ya son propias características de la señal para, de esta forma, poder establecer los correspondientes márgenes de admisibilidad en cada una de las muestras remitidas.

Al aplicar a nuestro supuesto de investigación los márgenes cuantitativos definidos en el modelo, y puesto que nos encontramos ante muchos minutos de grabación, tales márgenes -en principio- podrían situarse en el denominado *nivel óptimo,* tanto en lo relativo a las distintas clases de referencias (*GRs, PAs y SRs*) como en lo concerniente a las posibles alternativas de comparación : *Tipo I., II y III*. Sin embargo, no hemos de olvidar que los márgenes cuantitativos siempre estarán determinados por aquellos de tipo cualitativo, y viceversa. Es decir, no bastará con constatar una suficiencia cualitativa a nivel de estructura acústica, sino que también habrá de comprobarse que dicha suficiencia se materializa en un número de características o fragmentos de discurso que permitan deducir un nivel de admisibilidad. De la misma manera, la determinación de un margen cuantitativo positivo, implicará la previa necesaria suficiencia a nivel cualitativo.

Volviendo a nuestro caso de estudio, habíamos vislumbrado a priori una posible viabilidad en cuanto a la cantidad de señal disponible. Ahora, hemos de certificar si esa potencial suficiencia se confirma desde una perspectiva cualitativa.

IV.1.2.1.- Equipos de grabación, reproducción y análisis.

Antes de proseguir con la evaluación de admisibilidad de las muestras, merece la pena detenernos un instante y efectuar ciertos comentarios sobre las herramientas de trabajo que utilizaremos en dicho procedimiento.

Todo laboratorio de audio forense cuenta con diversas clases de equipos de grabación y reproducción para distintos formatos de soporte magnetofónico. Por regla general, dichos equipos -analógicos o digitales, portátiles o estacionarios- suelen ser de calidad profesional,

aunque ello no es óbice para que en un momento dado puedan presentar un desajuste de sus prestaciones técnicas. Por este motivo, resulta imprescindible la realización de periódicas operaciones de mantenimiento sobre dichos equipos para comprobar la óptima situación de sus referencias técnicas. Es muy importante disponer de unas herramientas de trabajo perfectamente calibradas o ajustadas para garantizar la más alta calidad en los distintos procedimientos de transferencia, análisis y almacenamiento de la señal. En este sentido, pueden tenerse en consideración los patrones de calidad descritos en los estándares del VIAAS de la Asociación Internacional de Identificación [VIAAS - IAI, 1991] alguno de los cuales, serán reseñados a continuación.

El alineamiento azimutal de las cabezas de grabación/reproducción debe ser observado, especialmente en aquellos formatos de bobina abierta o microcassette que trabajen a una velocidad de 2,4 centímetros por segundo (15/16 pulgadas por segundo) pues pueden producirse pérdidas de información en las más altas frecuencias registradas.

Frecuentemente, unas veces como consecuencia del propio uso y otras debido a la pobre calidad de los equipos utilizados, se producen errores de calibración tanto en la velocidad de playback como en la de grabación. Cuando estos errores -fácilmente detectables y corregibles- superan determinados márgenes, pueden llegar a provocar consecuencias fatales ya que inciden directamente en el desplazamiento del rango de frecuencia con las consiguientes falsas referencias de la estructura espectral. Afortunadamente, este tipo de desajustes puede ser detectado mediante la generación de tonos patrón (1Khz, 10 Khz) o la búsqueda de determinados componentes armónicos discretos de la señal que habitualmente acompañan a los registros forenses: frecuencia de red (50 Hz), tonos multifrecuencia de la línea telefónica, etc, . La búsqueda de estos componentes de referencia es bastante sencilla mediante análisis FFT. En general, una diferencia de velocidad superior al 3% entre las muestras objeto de comparación, imposibilitaría la obtención de parámetros y conclusiones fiables.

Además de las referencias de azimuth y velocidad, habrán de verificarse otras como el nivel de distorsión, flutter, respuesta en frecuencia para grabación y reproducción, nivel de grabación, etc. Concretamente, los estándares del Subcomité de Análisis Acústicos e Identificación de Voz de la I.A.I. recomiendan una revisión anual de los equipos y establece los siguientes márgenes de referencia:

- velocidad de playback dentro de +/- 0.5%.

- nivel de distorsión menor del 3% en 200 nWb/m.

- wow y flutter por debajo del 0.15%.

- respuesta en frecuencia para grabación/reproducción de 100 a 10.000 Hz (+/-3dB en 200 nWb/m).

- nivel 0 VU no superior a 250 nWb/m.

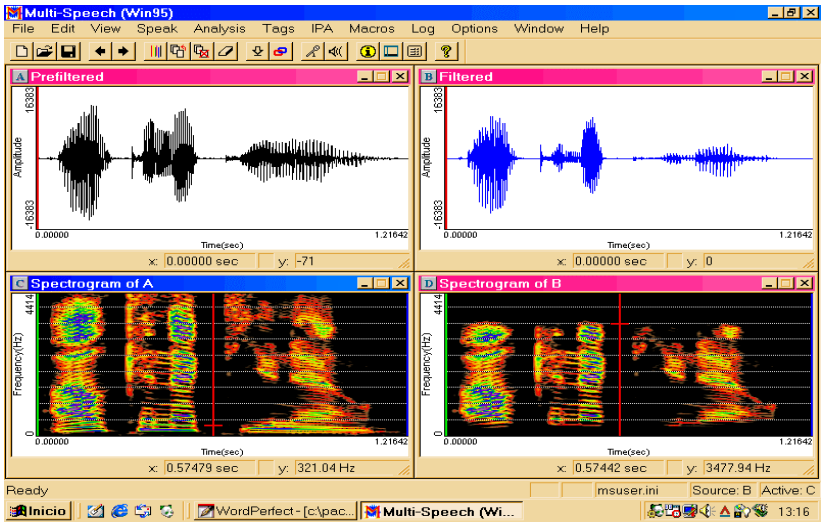
Los criterios arriba descritos encuentran su aplicación inmediata en los equipos de tecnología analógica. Ni que decir tiene, que dichos criterios son igualmente válidos para el entorno digital aunque, lógicamente, se presuponen implícitos en cualquiera de los actuales softwares y hardwares de análisis.

Realizadas estas necesarias puntualizaciones, regresaremos al examen preliminar de nuestras muestras dubitadas. Recordemos que en principio, parecía existir una suficiente cantidad de registros, aunque restaba confirmar si dichos registros presentaban un aceptable nivel de calidad.

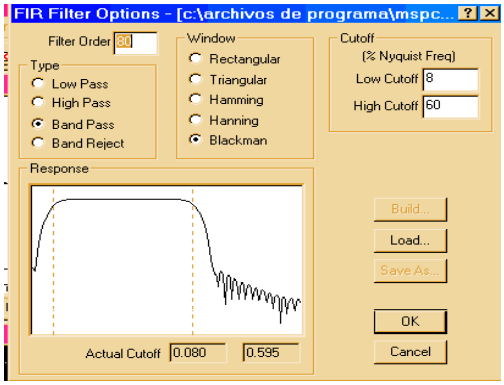
Para realizar una evaluación cualitativa con la máxima diligencia y agilidad hemos de contar con unas herramientas de análisis que nos permitan evaluar con alta precisión y en el menor tiempo posible, la mayor cantidad de señal. En este sentido, son de gran utilidad aquellas aplicaciones de análisis que posibilitan la audición y simultánea visualización de sonogramas en tiempo real: sonógrafos digitales KAY 5500, MEDAV 3000, CSL (rt), etc.; como ya fue referido, este Adisplay≅ dimensional es el que proporciona un estímulo más inmediato y estructurado de las diferentes referencias cualitativas de la voz.

La utilización de estas herramientas de representación y análisis para la obtención de parámetros de calidad, ha de enmarcarse dentro de los enfoques de análisis combinados básicos. En su momento ya comentamos, que las referencias de calidad de señal para el sistema automático que usaremos de forma complementaria, han sido establecidas por los diseñadores de la aplicación de reconocimiento. Posteriormente, cuando abordemos el análisis automático de nuestro caso práctico, serán especificadas.

Las primeras observaciones sobre nuestra señal ponen de manifiesto su naturaleza de transmisión. En nuestro caso, como ocurre en la casi totalidad de los casos forenses reales, el habla objeto de estudio proviene de una interceptación telefónica judicial. Esto supone una restricción inicial de información en rango de frecuencia que, para nuestro caso concreto (teléfono estacionario de línea) sería de un orden similar a la representada en la ilustración inferior. Si la interceptación telefónica se hubiese efectuado sobre un teléfono móvil la



restricción en frecuencia sería de un rango superior.



En el siguiente paso, examinaremos la presencia de ruidos, reverberaciones, distorsiones o posibles efectos de enmascarmiento o disimulo de los componentes de voz. Dado que también efectuaremos un análisis automático, este es un buen momento para el cálculo de referencias de la relación señal-ruido (SNR).

Si al desarrollar el proceso de evaluación de calidad a través de la audición y visualización en tiempo real de las muestras dubitadas, no detectamos de forma inmediata la presencia de alguno de los mencionados factores de degradación (ruido, distorsión, disimulo, solapamientos o enmascaramientos, alteraciones de velocidad o plano expresivo, etc.) habremos de situarnos en el estudio de la siguiente referencia de calidad: la ausencia de información en frecuencia por encima del margen de los 2KHz. Este límite cualitativo de la señal fue establecido por los estándares de la I.A.I. para la realización de evaluaciones de análisis desde el enfoque perceptivo-espectrográfico. Sin embargo, en la práctica de una metodología combinada puede darse la circunstancia de que ante la ausencia de información en frecuencia más allá de los 2000 Hz, sea perfectamente factible emitir un dictámen positivo de admisibilidad si se observan determinadas características en la muestra, que no necesariamente han de estar relacionadas con una suficiencia de índices a nivel sonográfico. Normalmente, este tipo de características se asocian a cualidades de voz muy particulares, habla patológica, presencia de elementos paralingüísticos, u otras peculiaridades conectadas a distintos fenómenos lingüísticos o fonoarticulatorios: recursos retóricos o expresivos, códigos de construcción sintáctica, léxico, rasgos prosódicos, alteraciones de dicción (metátesis, asimilaciones, etc.). Evidentemente, la conjunción de alguna de estas circunstancias con la referida falta de información acústica (por

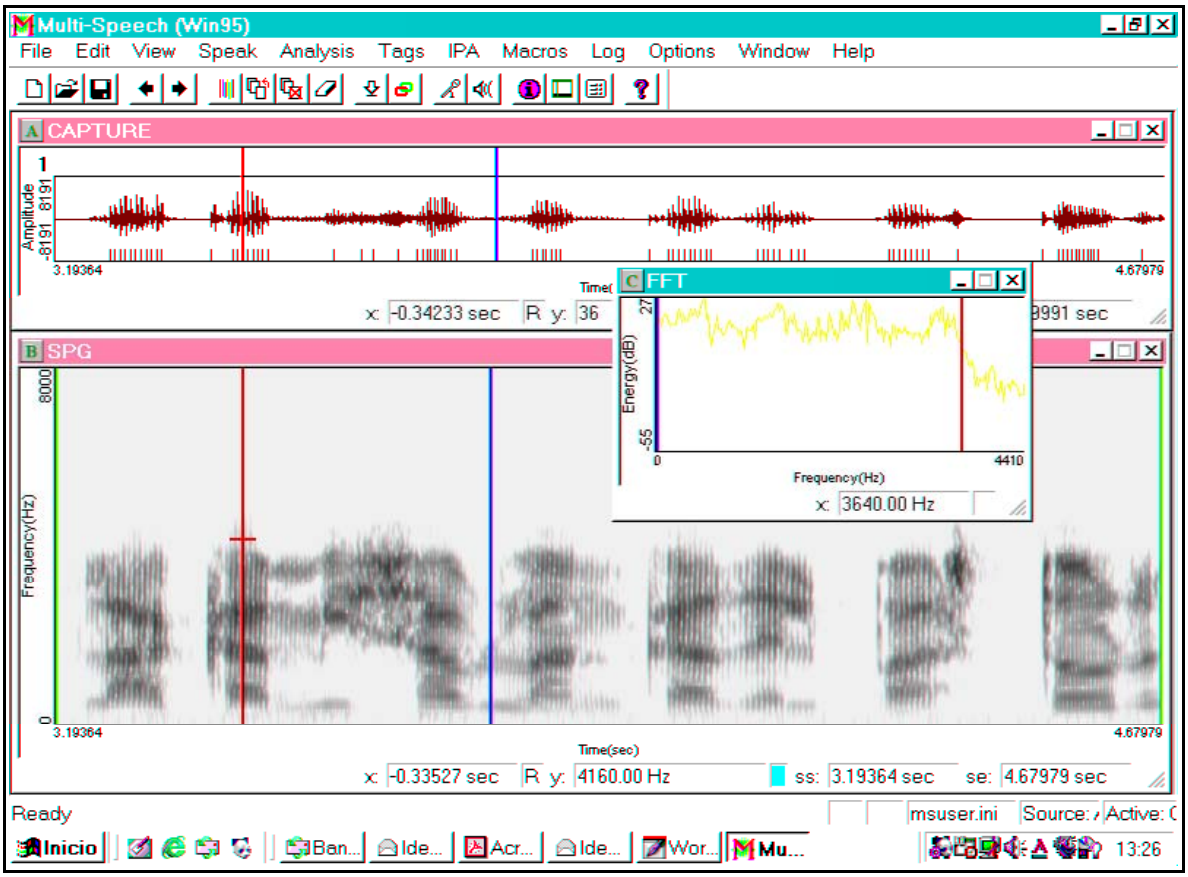
encima de los 2 Khz) no podrá conducir -en la gran mayoría de los casos- a un nivel de conclusión de la máxima certeza, aunque tampoco representará necesariamente un criterio de exclusión para la emisión de otro nivel de decisión. Es decir, será muy complicado alcanzar en nuestra escala de conclusión un nivel de Aidentificación≅ o Aeliminación≅ no disponiendo de señal acústica significativa por encima de los 2 Khz, si bien pudiera acontecer, que aun partiendo de esta situación, apareciesen diversas peculiaridades de máxima relevancia identificativa (spikes, patologías, etc.) en cuyo caso, sí podrían contemplarse tales niveles de conclusión.

Analizados los referidos parámetros de calidad en las grabaciones dubitadas de nuestro caso en cuestión -las cuales fueron registradas por la policía en cintas magnetofónicas de cassette a una velocidad de grabación de 4,7 cm/s- estamos en condiciones de corroborar el ***nivel óptimo*** de admisibilidad provisional que ya fue apuntado durante la evaluación cuantitativa de las mismas.

Las grabaciones dubitadas acreditadas por los expertos como válidas para su estudio comparativo, fueron seleccionadas de entre aquellos fragmentos de discurso señalados por la autoridad judicial a tal efecto; en nuestro caso, dichos fragmentos son atribuidos a uno de los empleados del local detenidos por la Policía.

He aquí la representación sonográfica de una muestra representativa de los registros dubitados seleccionados para su posterior cotejo:

Como podemos deducir de la ilustración arriba representada, nos hallamos ante unos registros dubitados con buenas referencias de calidad en cuanto a su estructura acústica:



ausencia de distorsiones y enmascaramientos, un aceptable nivel dinámico (entre 48 y 60 dB) y de SNR
(-6dB), índices de resonancia acústica útiles en un rango de frecuencia en torno a 300Hz-3.5KHz, etc; a todo ello, hemos de añadir la no existencia de voz simulada o alteraciones relevantes en ratios articulatorios, prosódicos o emocionales, así como la no presencia de otro tipo de perturbaciones consecuencia de los distintos procesos analógico-digitales de captura, transducción, grabación, reproducción, transmisión, codificación, etc.

IV.1.3.- Obtención de las muestras indubitadas

La etapa de evaluación de las grabaciones dubitadas ha deparado un resultado positivo. Ya conocemos las circunstancias de contexto en las que se circunscriben dichas grabaciones y,

por tanto, ha llegado el momento de procurarnos unas muestras indubitadas que nos dispongan en la mejor situación para realizar nuestros análisis comparativos. No olvidemos, que los mismos inconvenientes a los que el experto forense debe enfrentarse en el caso de los registros dubitados, pudieran acontecer para con las muestras indubitadas.

Uno de los pilares que sustentarán el éxito de nuestra metodología, lo constituirá el hecho de trabajar -siempre que las circunstancias lo permitan- con unas grabaciones indubitadas obtenidas mediante un protocolo que garantice la accesibilidad al mayor nivel de certeza posible en cada caso. Es decir, una vez conocidas la naturaleza y circunstancias que contextualizan las muestras dubitadas, seleccionaremos aquellos pasajes de las mismas que nos permitan, a través de un protocolo concreto de grabación, obtener unas muestras indubitadas con la estructura más idónea para su óptima apreciación comparativa.

Con todo esto, lo que deseamos subrayar es la importancia de trabajar con unos registros indubitados en las mejores condiciones ya que, en muchas ocasiones, no se le concede la importancia que merece a este aspecto, a pesar de que puede llegar a ser un elemento clave en el proceso de identificación.

A la hora de diseñar nuestro protocolo de toma de muestras indubitadas, hemos de contar de antemano con el carácter no cooperativo del sujeto objeto de la grabación. Por esta razón, el experto que selecciona el material dubitado, habrá de tener siempre en consideración este aspecto para elegir el material más inmune a esta habitual circunstancia.

En algunas ocasiones, las muestras indubitadas son aportadas directamente por la autoridad judicial o los investigadores sin que para la obtención de las mismas haya intervenido en forma alguna el equipo de expertos de identificación. En tales casos, la labor del experto -en esta fase de estudio- se limitará al hecho de valorar el factor cantidad y calidad de las muestras, de cara a su posible admisibilidad para el análisis comparativo. De acuerdo a nuestro planteamiento metodológico, ante este tipo de situaciones el experto forense parte con cierta desventaja si lo comparamos con aquellas otras en las que puede formar parte activa del proceso de obtención de muestras.

Un elemento que facilitará sensiblemente el estudio comparativo desde distintos enfoques (perceptivo, sonográfico, fonoarticulatorio, etc.) lo constituirá el hecho de contar con muestras de similar contexto a diferentes niveles: léxico-semántico, sintáctico, melódico-expresivo, ambiental, de transmisión, etc. ; lógicamente, lograr el máximo grado de adecuación posible entre muestras cotejadas, implicará una mejor disposición en el momento de emitir un *resultado*

de comparación.

De acuerdo a este razonamiento, el protocolo de toma de muestras indubitadas se inicia realmente durante la propia evaluación y selección del discurso dubitado. En dicha etapa, el experto habrá de localizar y diagnosticar los diferentes elementos de interés, siempre con la mente puesta en el momento de obtener los registros indubitados.

Pero regresemos a nuestro caso práctico. Una vez reseñadas y seleccionadas tanto las peculiaridades que caracterizan la propia naturaleza del habla dubitada, como sus circunstancias accesorias, procederemos a la obtención de las muestras indubitadas, de acuerdo al siguiente protocolo de actuación:

En presencia de la autoridad judicial, el sujeto a quien se atribuyen las grabaciones objeto de estudio, se somete voluntariamente al registro de distintas muestras de su habla de acuerdo al siguiente procedimiento:

11.- Se le facilita un texto de quince frases que ha de leer en tres ocasiones. Dichas frases coinciden con el contenido semántico de otras existentes en las grabaciones dubitadas (la duración aproximada de cada frase, en su realización dubitada, se situaba en torno a los dos segundos).

21.- A continuación, las mismas frases son pronunciadas por uno de los expertos actuantes, tratando de adecuarlas al contexto expresivo que fue utilizado en las emisiones dubitadas. El sujeto objeto de la prueba las repite en tres ocasiones, de un modo similar a como le son proferidas por el experto.

31.- Los pasos 11 y 21 se repiten mediante grabación por interceptación telefónica.

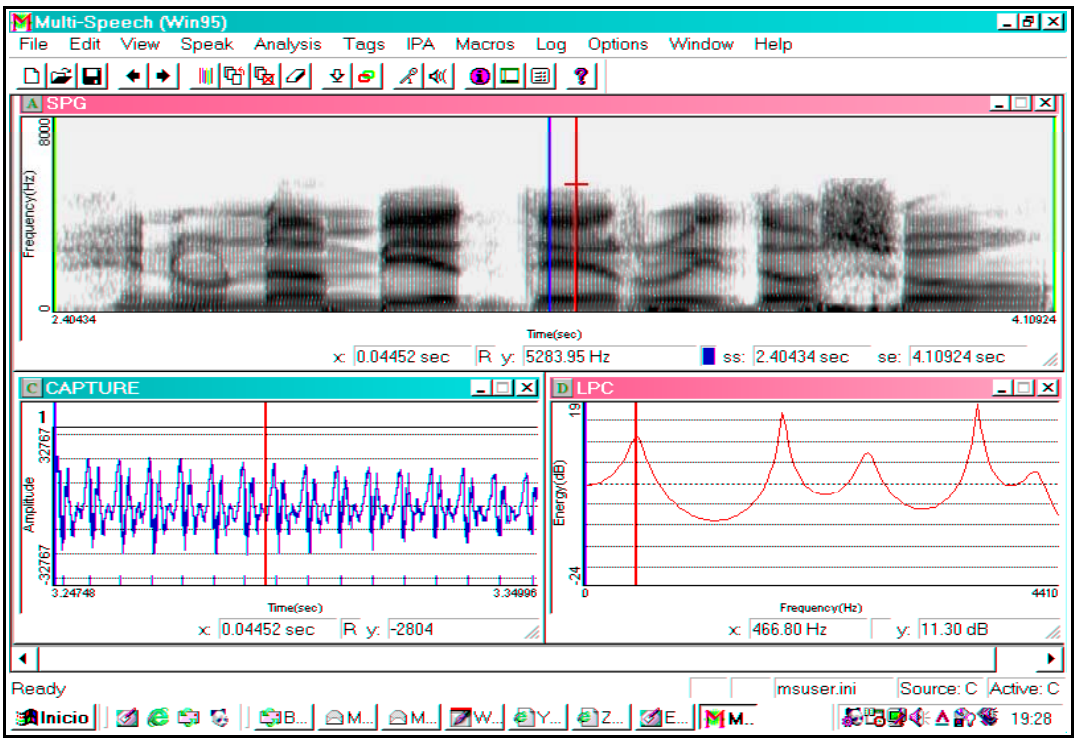
41.- Seguidamente, son registrados unos cinco minutos de conversación espontánea entre el experto(s) y el sujeto, sobre un tema intrascendente (deportes, lugar de nacimiento, etc.).

Como paso previo al propio acto de grabación, hemos constatado la viabilidad de las condiciones de aislamiento y reverberación del despacho o habitación donde dicho acto se realiza.

De la misma forma, hemos verificado que todos los medios y equipos de grabación funcionan correctamente.

Antes de finalizar el acto judicial, comprobamos que las grabaciones han sido correctamente registradas. Salvaguardando las correspondientes garantías de la cadena de custodia, las grabaciones indubitadas se trasladan al laboratorio forense para su análisis.

Recordemos que en nuestro caso, el habla dubitada no presentaba especiales dificultades en cuanto a su contexto expresivo-ambiental, por lo que el factor de no cooperatividad por parte del locutor no ha llegado a traducirse en un elemento de dificultad. Como es lógico, tampoco la calidad de la señal indubitada representará un obstáculo para la definición de su estructura acústica y los correspondientes rasgos fonoarticulatorios y lingüísticos. Observemos la buena calidad de dicha señal en uno de sus fragmentos:



Como cabía esperar, las referencias de calidad de la señal indubitada superan a las anteriormente definidas para la secuencia representativa de la dubitada: información hasta 5.500Hz , relación armónico a ruido (SNR) de -0.7 dB, rango dinámico entre 75 y 77 dB, etc.

De acuerdo a nuestro sistema de evaluación de parámetros, una vez alcanzado este punto, estamos en condiciones de definir la categoría del que será nuestro *contexto de comparación*. Puesto que en ambas muestras (dubitada e indubitada) no han sido halladas significativas deficiencias de tipo cualitativo o cuantitativo (*condiciones α*), o factores críticos de discrepancia entre las mismas (*condiciones β*) habremos de concluir en que nos encontramos ante un contexto de comparación de tipo ***favorable***.

IV.2.- DESARROLLO DE ANÁLISIS

IV.2.1.- Análisis perceptivo auditivo sobre referencias globales (GRs)

El objetivo principal de este enfoque de estudio es obtener una percepción general de la emisión hablada mediante la apreciación de los que consideramos sus índices de carácter global. Tales referencias, difícilmente pueden ser discriminadas y evaluadas mediante otros procedimientos de análisis; por esta razón, en esta fase de estudio jugarán un papel fundamental la experiencia y entrenamiento del experto.

La realización de este análisis perceptivo puede canalizarse a través de inputs auditivos exclusivamente (auriculares, equipos reproductores, etc.) o combinando dicha opción con la visualización sonográfica de la señal. Es más recomendable la utilización de esta segunda alternativa, pues de esta forma podremos, tanto corroborar ciertas características detectadas a nivel auditivo, como contextualizar debidamente otro tipo de informaciones accesorias a la señal

objeto de estudio (ruidos, interrupciones, distorsiones, etc.).

La ejecución de esta tarea requiere del mayor grado de concentración por parte del experto, además de las mejores condiciones de aislamiento y un uso adecuado de los equipos y opciones de análisis más idóneos: equipos de calidad profesional, correcta ubicación de los mismos, control de niveles de intensidad, corrección o eliminación de efectos no deseados, etc.

Las etapas a seguir en el desarrollo de esta labor perceptiva vendrán marcadas por la discrecionalidad de cada equipo de analistas, aunque es muy importante que dicha labor tenga un carácter sistemático.

En nuestro caso práctico, comenzaremos por la audición/visualización de distintos fragmentos de habla dubitada. Recordemos que en esta fase, los objetos de estudio son aquellas referencias que denominábamos *referencias globales o GRs*. Es decir, aquellas que debido a su propia naturaleza han de ser necesariamente apreciadas como un todo y no como la suma de sus distintos componentes. Nos estamos refiriendo a la ubicación de la base de articulación y otras peculiaridades propias del sociolecto, las realizaciones a nivel suprasegmental, recursos retóricos, el uso de las funciones expresivo-comunicativas del lenguaje, el timbre, ciertos ratios elocutivos, patologías, etc.

Una vez recogidas nuestras anotaciones sobre las características que consideramos más relevantes en el habla dubitada (para ello será de gran ayuda el uso de la transcripción fonética) repetimos idéntico proceso con las muestras indubitadas.

A continuación, detallaremos las peculiaridades coincidentes y no coincidentes, destacando como características de similitud entre ambas muestras, las siguientes:

- <u>*Ubicación de la base de articulación*</u>:

- En los dos casos nos encontramos ante un Acastellano estandar≅ en el que se incluyen algunos rasgos fonoarticulatorios propios de ciertos hablantes del área de Madrid. Entre otros cabe citar la realización del fricativo alveolar /s/ y del fricativo interdental /□/ -en posición postnuclear y seguidos del oclusivo velar sordo /k/- como fricativo postdorsovelar /ϒ/. Así ocurre en las realizaciones: /eϒkaβa_óra/, /i□kl,^érdo/, /kl,^óϒko/, /aϒkeá_o/ y /éϒke/. También

relacionado con dicha base de articulación, podría mencionarse la no realización sistemática del archifonema /D/ en posición final al articular AMadrid≅, o la utilización de diversos términos y expresiones, característicos de jergas juveniles que pueden enmarcarse en la citada referencia diatópica: *Achachi≅, Ada buti≅, Aes una jena≅, Aes una ful≅,Avente a la prospe≅,Apaletos≅*, etc.

- Igualmente se aprecia un sistemático laísmo y leísmo de persona (masculino), habitual en muchos hablantes madrileños: A*...tu le viste pero no te gustó*≅

- La articulación a nivel conversacional se realiza de forma relajada en general, aunque en ocasiones aparecen episodios enfáticos de relevante tensión. Se observa un para-rotacismo por mala vibración apical en la realización de grupos sinfones o directos dobles: *Aoctu**bre**≅, A**pre**ciosa≅, A**pro**mesa*≅, etc.

- *Otras características propias del sociolecto*:

- Desde otra perspectiva sociolectal, el hablante utiliza con cierta frecuencia términos característicos del ámbito profesional audiovisual: *Aframe≅, Aeditar≅, Avisionar≅, Arepicar≅, A...me encanta la steady...≅,* etc.

- *Recursos retóricos, función comunicativo-expresiva*:

- Gusto por el uso de refranes : *A...ya sabe, hombre prevenido...≅, A ...macho, al que madruga...≅, A...en boca cerrada...≅,* A *... más vale pájaro en mano...*≅etc.

- Utilización de los términos A*correcto*≅ y A*O.K.*≅ para subrayar afirmativamente situaciones conversacionales de acuerdo.

- Uso de la expresión A*...bueno, bueno*, ...≅ como fórmula de petición de turno conversacional ante una posición de desacuerdo con su interlocutor.

- Reiterada utilización del fonema vocálico /e/ en realización sostenida /e:/ , como recurso dubitativo en situaciones de respuesta o pronunciamiento comprometidos. La duración media de las pausas discursivas se sitúa en torno a los 0,6 sg.

- Habitual empleo de los términos A*titis*≅ y A*bólidos*≅ para referirse a las mujeres

o a los Acoches≅.

- Se observa una buena fluidez elocutiva y una entonación más bien melódica. También se aprecia riqueza de léxico y una correcta construcción a nivel sintáctico, lo que unido a la utilización de determinados términos y expresiones : *A...estaba salivando como el perro de paulov...≅, A...parece del gótico flamíjero...≅, Ano necesito un diseño de la bauhaus≅* pudiera asociarse a un nivel medio de formación académica en el locutor. En general, las pautas comunicativas y de expresión utilizadas en ambas emisiones pueden considerarse educadas:*A...perdona, pero eso no tiene mucho sentido...≅, Asería tan amable de informarme...≅* ; además, aparecen impregnadas de un subyacente optimismo y de frecuentes matices humorísticos, nunca chabacanos.

- <u>*Percepción del timbre*</u>

La cualidad de voz es percibida como limpia pues se presenta rica en componentes armónicos y carente de elementos fonatorios aperiódicos simultáneos. Además, nos encontramos ante un tono fundamental conversacional (SFF) [Hollien, 1990] más bien agudo, lo que unido a la anterior circunstancia, proporciona una sensación fonoestésica de voz juvenil. Por otra parte, la sonoridad del timbre se aprecia asténica; especialmente, en aquellos fragmentos de discurso conectados a episodios emocionales de desánimo.

Como características no coincidentes entre ambas muestras, pueden reseñarse las diferencias de la velocidad de elocución y valores medios del pitch en algunos tramos. Tales discrepancias se manifiestan en unos parámetros de mayor lentitud y más baja tonía en favor de la indubitada, aunque suelen desaparecer en aquellos fragmentos locutivos en los que los planos contextuales y expresivos de ambas muestras se adecúan. Por esta razón, las disimilitudes descritas carecen de relevancia y deben considerarse normales dentro de los que estimamos márgenes habituales de variabilidad intrapersonal. A título meramente ilustrativo, podemos señalar que el *ratio silábico* (n1 de sílabas por segundo) se halla en torno a las 4,5 sy/s en la dubitada y 2,6 sy/s en la indubitada, y los *ratios articulatorios* (n1 de sílabas por segundo, sin pausas) se sitúan en torno a 4,7 sy/s en la dubitada y 3,2 sy/s en la indubitada. Como hemos comentado, esta diferencia de valores se encuentra en los límites lógicos de variabilidad locutiva y no puede considerarse significativa en el presente estudio.

De forma complementaria (pues dada la buena calidad en general de las grabaciones no

resulta necesario) hemos efectuado un análisis de percepción auditiva en reverso (A.P.R.E.S.), entre algunos de los registros dubitados correspondientes a cuatro de los posibles interlocutores telefónicos de nuestro establecimiento, y las grabaciones indubitadas de toma telefónica. Como es lógico pensar, entre estos cuatro sujetos seleccionados se encuentran nuestro candidato y aquellos otros que, a nivel perceptivo, presentan una cualidad vocal más similar. Los resultados de identificación presentan la señal revertida atribuida a nuestro candidato como la más cercana al reverso obtenido de la voz indubitada.

IV.2.2.- Análisis perceptivo mixto fonoarticulatorio sobre SRs.

En nuestro anterior estudio sobre referencias globales, el análisis a través de la representación gráfica de la señal lo considerábamos algo opcional. Sin embargo, en el caso del estudio perceptivo mixto sobre referencias concretas, cualquier tipo de apreciación comportará necesariamente la utilización de dicha alternativa. Como ya explicábamos en el capítulo III, esta aproximación de análisis estará basada en la asociación de peculiaridades fonoarticulatorias idiolectales con sus correspondientes referencias estándar en la población.

En el enfoque mixto fonoarticulatorio, los objetos de estudio ya no son contemplados en su proyección más global. Ahora, nuestra atención se centrará en el entorno más próximo de cada rasgo articulatorio. Por ejemplo, si antes analizábamos la cualidad vocal como una única estructura, sin detenernos en las características específicas de los elementos que la integraban, ahora ha llegado el momento de examinar y describir dichos elementos en su dimensión acústico-articulatoria : distribución e intensidad de componentes acústicos, grados de sonoridad, oclusión, nasalidad, nivel de tensión articulatoria, alteraciones de dicción, transiciones e interacciones entre grupos fónicos, modo y lugar de articulación, etc.

Si en el análisis sobre GRs las tareas comparativas efectuadas eran del Tipo II y III, en el caso del análisis mixto habrán de incorporarse también las de Tipo I. Es decir, en una primera fase determinaremos las características y dimensiones físicas de cada rasgo, primero en la grabación dubitada y a continuación en la indubitada (Tipo I). Una vez acotada la naturaleza real de cada objeto de análisis, procederemos a la clásica comparación de los mismos entre la muestra dubitada y la indubitada (Tipo II). Obtenidos los valores de similitud/disimilitud entre las SRs de ambas muestras, otorgaremos el correspondiente peso identificativo a los resultados de comparación, mediante la referenciación de las mismas a los valores estándar establecidos para la población (Tipo III).

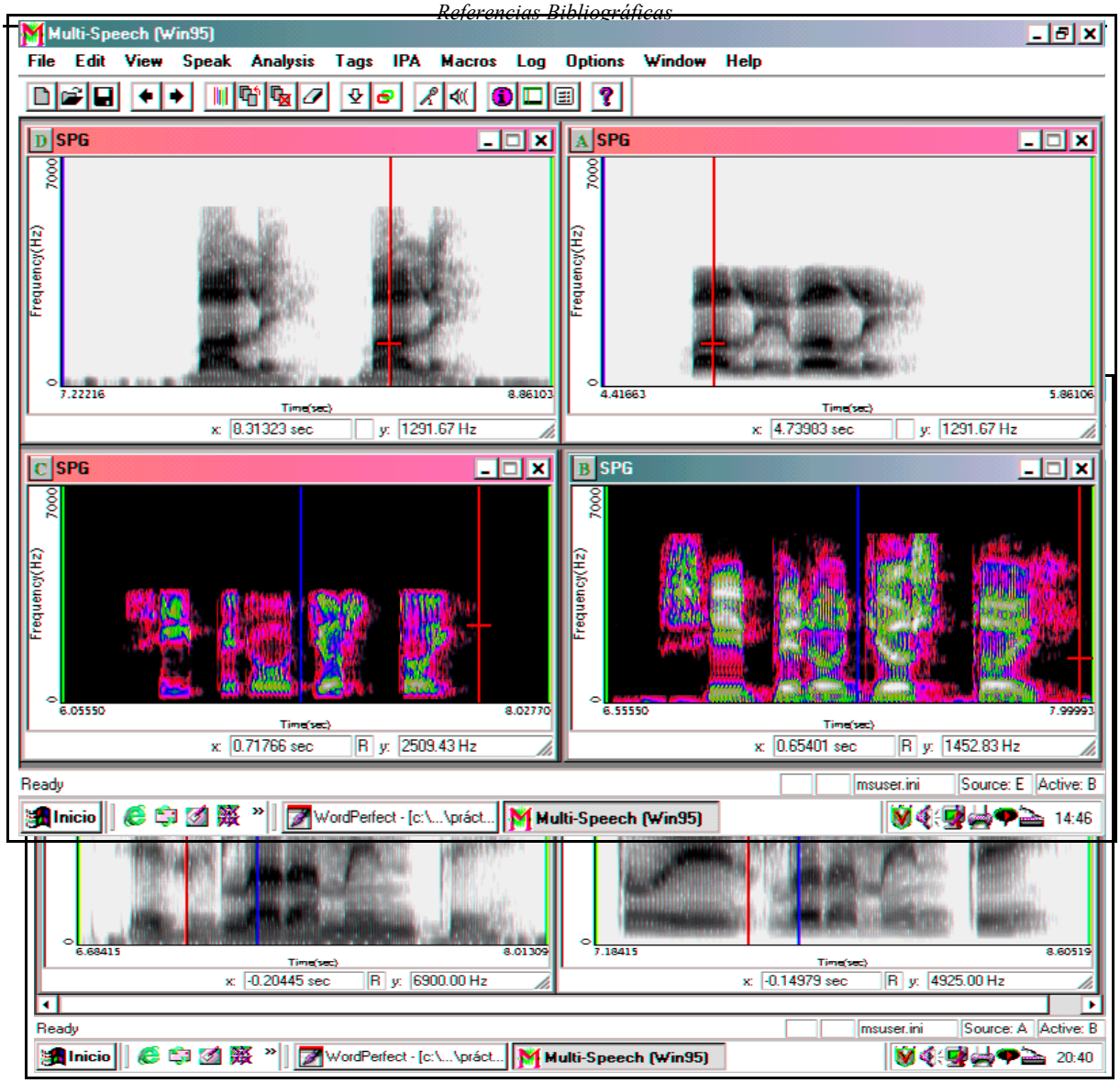
Es muy importante, reseñar la conveniencia de utilizar el mismo contexto semántico-expresivo en las comparaciones realizadas bajo esta perspectiva de análisis. Sin lugar a dudas, el mejor camino para lograr este objetivo es la utilización de las muestras indubitadas que obtuvimos mediante la repetición o lectura de aquellas frases previamente seleccionadas de las grabaciones dubitadas.

Trasladándonos de nuevo a nuestro caso práctico y, una vez dimensionadas y corrobaradas las referencias concretas (SRs) en cada una de las muestras, comprobamos la existencia de diversos aspectos fonoarticulatorios coincidentes entre ambas. Entre otros, podemos citar los siguientes:

- Como ya comentamos al ubicar la base de articulaciyn, se observa una realización del fricativo alveolar /s/ y del fricativo interdental /□/ **-en posiciyn postnuclear y seguidos del oclusivo**

velar sordo /k/- como fricativo postdorsovelar /ϒ/. Asн ocurre en las realizaciones /eϒ**ka**βa_ óra/, /i□kl,^érdo/, /kl,^óϒko/, /aϒkeá_o/ y /éϒke/.

- En las expresiones A*hay que* ***c****ontrolar*≅ y A*no es* ***c****laramente*≅, realiza con pronunciado grado de sonorización el oclusivo sordo velar /k,_/. En los sonogramas lo apreciamos en la no presencia de relevantes barras de explosión:



- Realización fonética cero del archifonema /D/ en posición final al articular A*Madri**d***≅

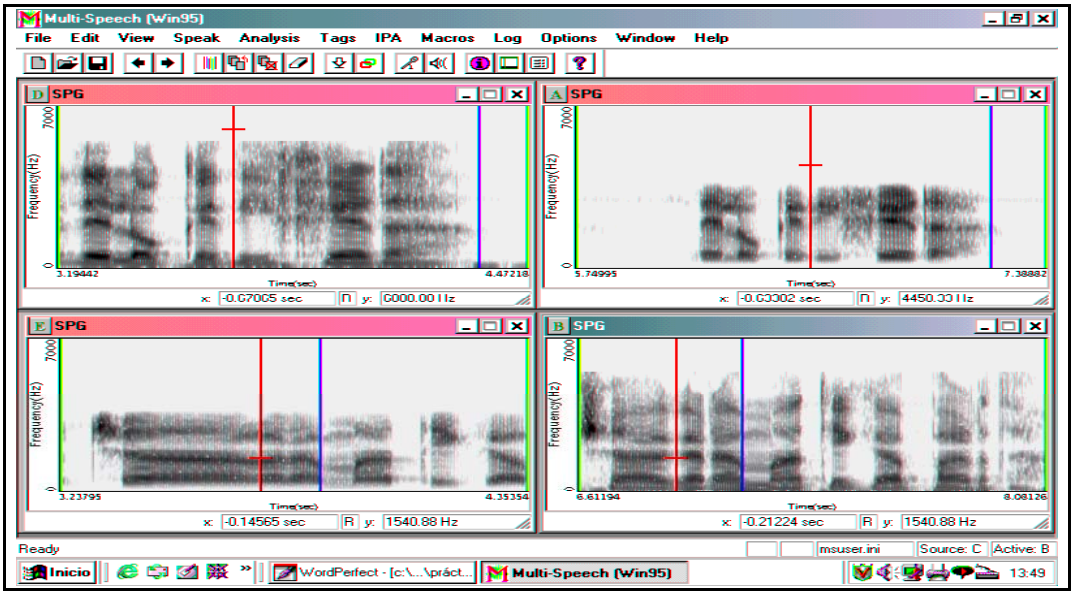
- Articulaciyn retrasada del fonema vocólico central /a,`/, en A***vale***≅**, aprecióndose el F2 a 1.290 Hz. (Ilustraciyn n**1 **70)**

- Realizaciyn monofonemótica de la secuencia vocólica /oe/, en la expresiyn A ***sи, pero yo estoy...*≅, aprecióndose el F2 a 1.408 Hz y el F3 a 2.5 Khz. (Ilustraciyn n**1 **70)**

- En la realizaciyn del fonema oclusivo velar sordo /k/, en la expresiyn A***que se quede...*≅, se aprecia un elevado grado de tensiyn articulatoria manifestado en relevante**

barra de explosiyn conformada en energнa difusa, abarcando todo el rango de frecuencia.

- En la realizaciyn del fonema oclusivo dental /t/, en la palabra A**total≅, a pesar de efectuarse con bastante tensiyn a nivel perceptivo, se aprecia dйbil barra de explosiyn sin apenas VOT.**



- Articulaciyn aspirada del fonema fricativo velar sordo /_h/ y fonйtica cero del fonema oclusivo dental /d/, por distribuciyn fricativo, en la expresiyn A**esta gente dice...≅.**

- Realización muy relajada y aspirada del fonema /s^{h}/ en posición implosiva, así ocurre en Ala**s** pasan≅, Apue**s** ya hace tiempo**≅,** Ate has≅ y Aeramo**s** nosotros≅.

- Realización tensa del fonema líquido /r/, en A*dije**r**a*≅, apreciándose un tiempo de oclusión de 24,7 ms con significativa energía (situada entre 1.700 y 3150 Hz) en el centro de la oclusión.

- Realiza relevante elemento esvarabбtico en la palabra A***gramбtica*≅, apreciбndose agrupaciones formбnticas a 650 Hz., 1.550 Hz. y 2.525 Hz.**

- Realización muy relajada del fonema /□/ en A*se**ñ**ora*≅.

- El fonema lнquido lateral palatal /_/, lo articula sistemбticamente como fricativo lingьopalatal central; asн ocurre en A***ll**orando*≅**,** A***ll**ueve*≅**,** A***all**н*≅**, etc.**

- Frecuentes asimilaciones de fonemas consonбnticos en situaciyn postnuclear con el fonema consecutivo. Se observa en *Ao/bt/uso*≅*, A corre/ct/o*≅*, Aa/dq/uirir*≅**, etc, .**

Descritas algunas de las peculiaridades coincidentes entre ambas muestras, hemos de acadir que en general no se han observado disimilitudes relevantes en aquellos tramos no afectados por las denominadas Acondiciones β≅**. Por otra parte, la presencia de ciertos elementos de discrepancia no significativos ha de considerarse algo normal, pues son la lygica consecuencia del factor de variabilidad intrapersonal.**

Por estas razones -al igual que se desprendна de las valoraciones obtenidas en nuestro anterior enfoque de estudio- estamos en condiciones de afirmar, que la conjunciyn de las referidas circunstancias en el desarrollo del presente anбlisis, nos reafirman en la direcciyn de que las muestras objeto de comparaciyn pueden corresponder a un mismo locutor.

IV.2.3.- Análisis sobre índices acústicos

Las diferentes opciones de representación gráfica de la señal sonora (oscilogramas, sonogramas, espectros, etc) permiten poner de manifiesto diversas características e índices acústicos que, en unos casos, corroboran las apreciaciones alcanzadas mediante otras alternativas de análisis y, en otros, describen en una referencia dimensional de mayor objetividad la verdadera naturaleza de los distintos componentes físicos de una emisión hablada. Estas son las aportaciones del análisis acústico. Una perspectiva de estudio que ha constituido por sí misma, y durante largo tiempo, una metodología exclusiva de identificación forense.

De la misma forma que la máxima eficacia del análisis fonoarticulatorio se supeditaba a la posibilidad de contar con unas muestras del mismo contexto, las prestaciones del estudio sobre índices acústicos se verán igualmente favorecidas por tales circunstancias. No obstante, en el caso de no poder contar con fragmentos de discurso de dichas características, habremos de localizar aquellas situaciones de la cadena hablada en las que encontremos el máximo nivel de adecuación en cuanto a una similar distribución en coarticulación, plano expresivo, etc.

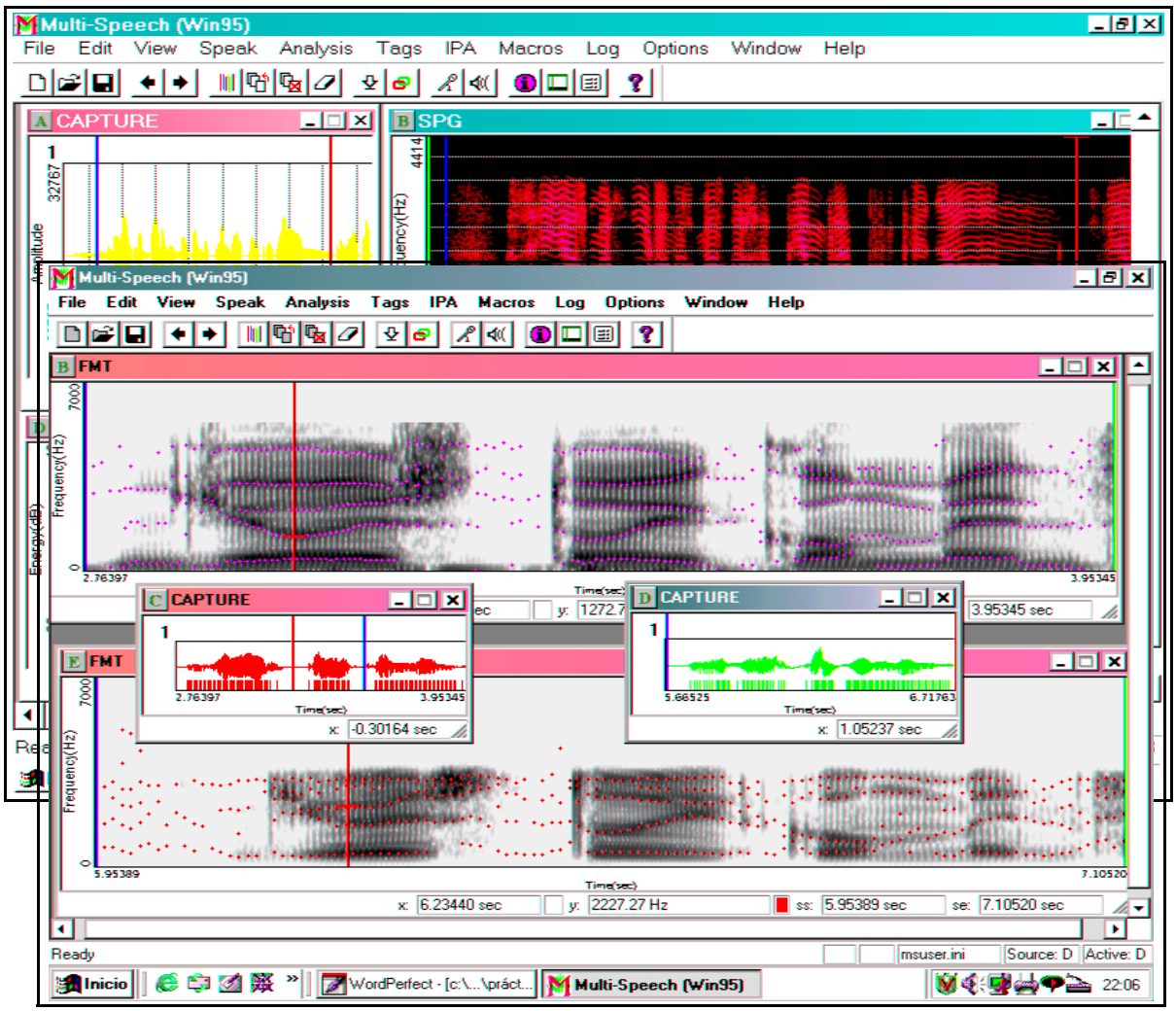
Puesto que en nuestro caso práctico disponemos de muestras indubitadas con el mismo contexto semántico-expresivo, haremos uso de las mismas para desarrollar el corazón del presente análisis. Ello no será inconveniente para la utilización de otra clase de registros si fuese estimado oportuno.

Al abordar el análisis perceptivo sobre GRs, dejamos a la discrecionalidad del experto la posibilidad de utilizar inputs de representación gráfica de forma simultánea al estímulo auditivo. Si por una u otra razón, durante dicha fase de estudio no se considera pertinente el examen de índices gráficos, este es un buen momento para complementar y objetivar aquellas apreciaciones perceptivas.

Recordemos que cuando definíamos la cualidad vocal de nuestras muestras, hablábamos de una voz limpia y rica en componentes armónicos.

En la siguiente ilustración, observamos en detalle la estructura acústica de dichos componentes, en dos fragmentos representativos de ambas grabaciones:

Como puede apreciarse en el gráfico anterior, la estructura de los componentes armónicos es uniforme y no presenta en general elementos de perturbación, por lo que la impresión obtenida



a nivel perceptivo queda plenamente ratificada tras el estudio acústico.

Constatados a través de este nuevo enfoque aquellos aspectos de las referencias globales que estimemos de interés, dirigiremos nuestro análisis sobre otros índices acústicos más específicos, normalmente, relacionados con las SRs.

El estudio acústico, evidencia con un carácter totalmente objetivo distintas características de la señal que no pueden ser apreciadas mediante otras aproximaciones de análisis. Situándonos ya en nuestro caso práctico, nos encontramos con nuestras grabaciones de voz proyectadas sobre distintos dominios y opciones de representación.

Efectuado el correspondiente estudio comparativo entre las muestras dubitada e

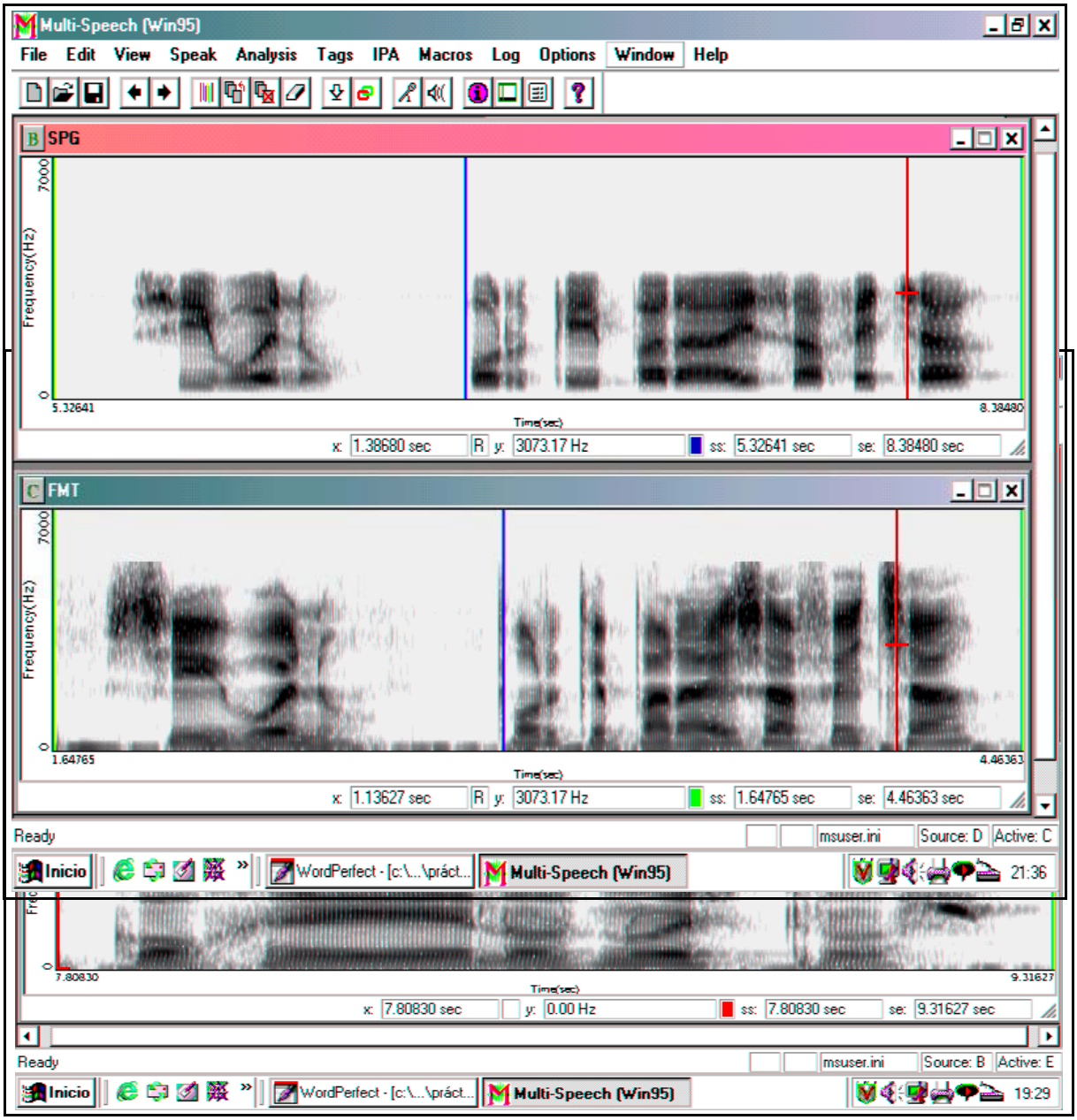
indubitada, nos encontramos con las siguientes circunstancias acústicas coincidentes:

- En la expresiyn A***yo estoy convencido***≅**, se aprecia similar transiciyn, en cuanto a trayectoria y altura frecuencial, del fonema fricativo central** /j,_/ **al vocõlico posterior /o/. Igualmente apreciamos similares VOT y similares conformaciyn y distribuciyn de energнa en las dobles barras de explosiyn del oclusivo velar sordo /k/:**

- En la ilustración n1 64 vemos representados los sonogramas de la emisión A*coge los nuestros*≅. Se aprecian los siguientes índices acústicos coincidentes:

1.- Similar trayectoria y altura frecuencial de las agrupaciones formánticas en la transición /we/ entre los respectivos F1y F2. En el vocálico /e/ se observan F1en torno a los 500 Hz, F2 a 1600 Hz. Igualmente, los patterns de la zona del F3 presentan idéntica estructura.

2.- En el grupo /tros/ el oclusivo dental se configura en barras de explosión dobles, con similar VOT (14,8 ms) y distribución de energía. El fricativo alveolar /s/ inicia sus estridencias en ambas muestras en torno a los 1200Hz, detectándose una agrupación muy densa de energía consecuencia de su carácter asibilado.



3.- El fricativo velar /ϒ/ se realiza con bastante tensión articulatoria, presentando similar distribución de energía en el eje de la frecuencia.

4.- En Anuestros≅ se detecta relevante elemento esvarabático con idénticas agrupaciones de energía armónica.

- En el gráfico anterior observamos los registros dubitado(B) e indubitado (C) de la

expresión A*sí bueno, porque para esa fecha*...≅ . Se aprecian los siguientes rasgos de similitud:

- Iniciación de estridencias en torno a 1.500 Hz del fricativo alveolar /s/. Así ocurre en A***S****í*≅ y A*e****s****a*≅. También en A*e****s****a*≅ , se observa muy similar distribución de energía.

- Articulación adelantada del fonema vocálico anterior alto /i/. Se manifiesta en un F2 a 2.200 Hz en A***s****í*≅. También se observa un F3 a 2700 Hz, resultando característica la interdistancia formántica al F2 (en torno a los 900 Hz).

- La transición (i-u-e) en A*sí bueno*≅ presenta muy similar estructura en cuanto a trayectoria frecuencial y duración.

- En A*bu****e****no*≅ la realización del vocálico anterior medio /e/ presenta estructuras formánticas a 640 Hz (F1); 1750 Hz (F2) y 2560 Hz (F3).

- En la realización del fonema nasal /n/, en A*bue****n****o*≅, se aprecian acumulaciones de energía a 1.600Hz. y 2.800 Hz.**-**

- En la transiciyn del fonema nasal /n/ al vocбlico /o/, se aprecian agrupaciones relativas de energнa conformadas en barra de explosiyn a 572 Hz, 1408 Hz y 2280 Hz.

- También en A*b****u****eno*≅, el fonema vocálico /u/ presenta un F1 a 554 Hz, un F2 a 1024 Hz (indicativo de articulación algo adelantada) y un F3 a 2603 Hz.

- Realización del oclusivo bilabial sordo /p/ con similar barra de explosión y similar duración de V.O.T. (en torno a 21ms en A***p****orque*≅).

- En A*p****o****rque*≅, el vocalico posterior /o/ presenta F1 a 597 Hz, F2 a 1100 Hz y F3 a 2800Hz.

- En la realizaciyn del fonema vibrante /R/ (archifonema), en A***po****r****qu****e*≅**, se aprecian agrupaciones de energнa con centros yptimos a 550 Hz. y 1.350 Hz.** El momento oclusivo presenta similar duración.

- Realización del oclusivo velar sordo /k/ en A***porque***≅ con una única barra de explosión con muy similar distribución de energía y duración de V.O.T.

- Lo descrito para el caso anterior, se produce igualmente con el oclusivo bilabial sordo

/p/ en A***p****ara*≅.

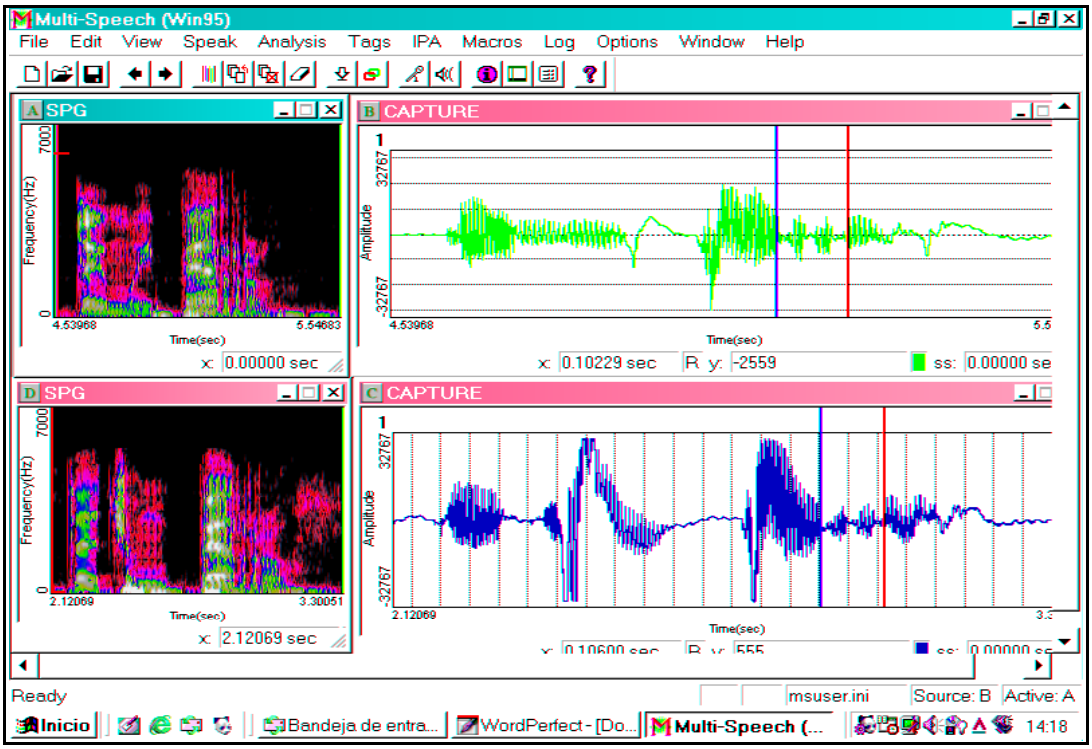
- También en A*p****a****ra*≅ el vocálico central /a/ presenta centros óptimos formánticos en 682Hz para el F1, 1536 Hz para el F2 y 2674 para el F3.

- En A***f****echa*≅, realización del fricativo labiodental /f/ con similar distribución de energía.

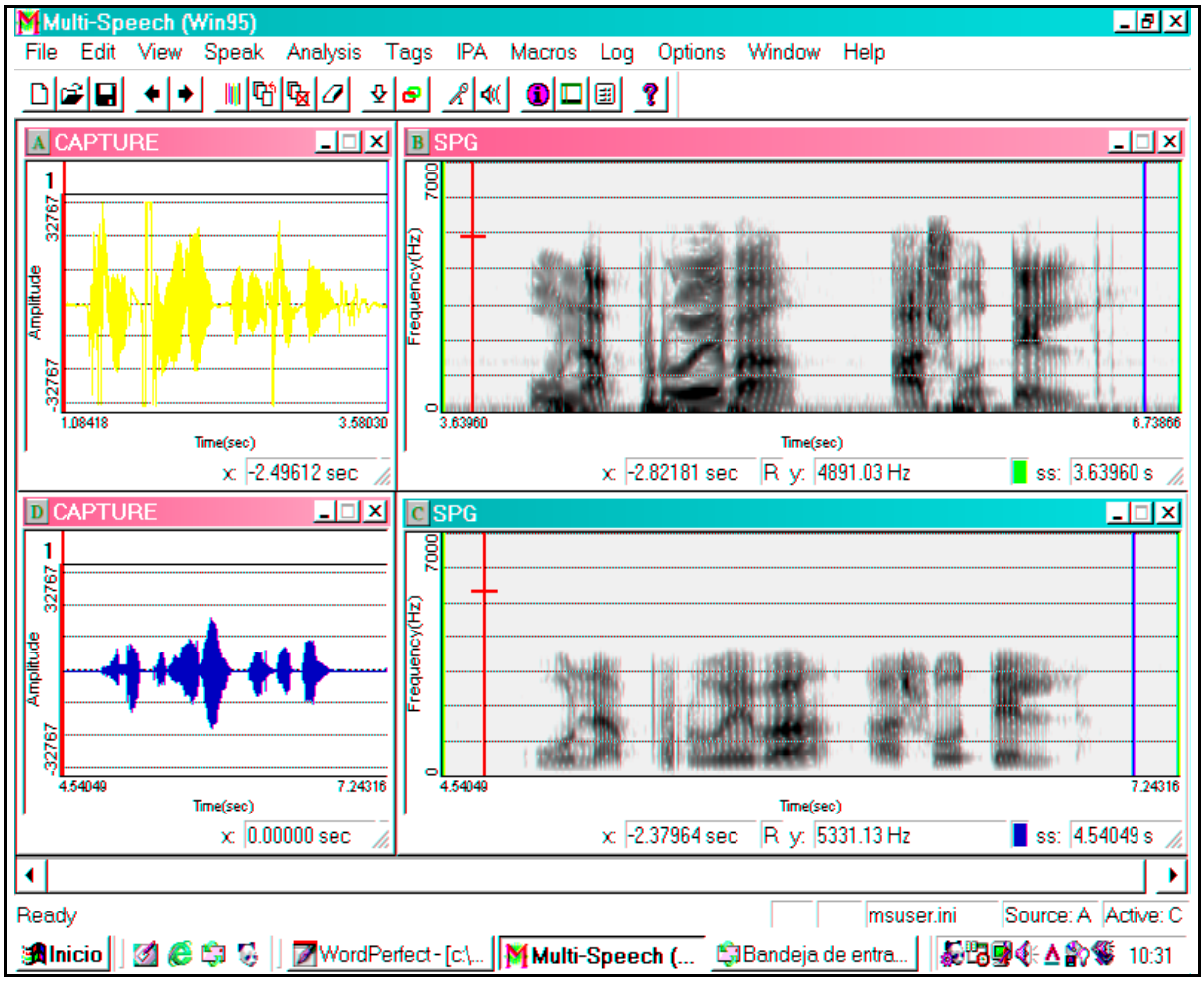
- También en A*f****e****ch****a***≅ los fonemas vocálicos /e/ y /a/ presentan respectivamente un F1 a 597 Hz y un F2 a 1621Hz.

- Realización del africado/ _,_ / con muy similar duración y estructura del momento fricativo y plosivo. Se observa una característica concentración de energía en el área de fricación a unos 3030 Hz (señalada por los cursores rojos en el gráfico).

- En el gráfico inferior observamos una realización enfática y muy tensa del vibrante múltiple /r,_/ en A*pe****rr****o*≅. Obsérvense las tres oclusiones en similar intervalo temporal (102-106 ms). La muestra indubitada obtenida para esta comparación se ha extraído de un fragmento de la conversación espóntanea.



- En la articulación implosiva del fonema líquido lateral /l/, en la expresión A*a su hotel*,



se aprecian acumulaciones de energía con centros óptimos a 1.600 Hz y 1700 Hz. Igualmente se observa significativa barra de oclusión final. (Ilustración n1 77).

- Realiza relevante elemento epentŭtico en la palabra A*p"rueba*≅, apreciŏndose agrupaciones formŏnticas a 480 Hz y 1150 Hz. La estructura y trayectoria de la transiciyn /we/ presentan el mismo pattern. (Ilustración n1 77)**.**

Como hemos podido comprobar, existen multitud de índices acústicos coincidentes que han de interpretarse como exponentes representativos de la totalidad existente en las grabaciones analizadas. Por tanto, partimos de la base de que las referencias adjudicadas a un rasgo concreto son extrapolables al resto de rasgos de su misma naturaleza, siempre que éstos se encuentren ubicados en su mismo entorno de distribución.

También han sido apreciadas ciertas características de no similitud en las que no nos

detendremos, pues siempre aparecen asociadas a diferencias de contexto en cuanto a su distribución, ámbito prosódico o plano expresivo.

Los resultados comparativos obtenidos en esta etapa de análisis arrojan un alto grado de similitud entre los registros dubitados e indubitados, refrendando las apreciaciones expresadas en los estudios perceptivo y fonoarticulatorio. A falta del cálculo de parámetros de alta muestra, y de la orientación complementaria que nos aportará el sistema de reconocimiento automático, todos los resultados de estudio señalan que las muestras de voz cotejadas pueden pertenecer a la misma persona.

IV.2.4.- Análisis de parámetros mensurables.

El hecho de que nuestra propuesta metodológica albergue un enfoque complementario de reconocimiento automático, determinará en cierta forma que el cálculo de determinados parámetros de alta muestra que pudieran incluirse bajo este epígrafe, carezca de especial significado. Recordemos, que cuando hablábamos de parámetros de alta muestra nos referíamos a aquellas referencias básicas (BR) cuyo carácter específico respondía a la sencilla, rápida y precisa determinación de los mismos mediante valores numéricos y estadísticos, partiendo de una muestra de datos muy elevada.

La utilización de una aplicación de reconocimiento automático de las características de nuestro prototipo, conlleva la selección de unas opciones de parametrización, fundamentalmente ligadas a parámetros de la información espectral (LPCC, MFCC). Ya hemos referido con anterioridad, que normalmente, la estructura acústica es el elemento básico de la voz más representativo de su carácter individual y, además, el menos influenciable ante la posible presencia de los típicos factores no deseados del audio forense: variabilidad intrapersonal, efectos del canal de transmisión, etc, . Sin embargo, el resto de componentes físicos fundamentales que dimensionan el sonido del habla, -amplitud, frecuencia y tiempo- se ven mucho más afectados por dichos factores de variabilidad aunque, en determinadas situaciones, algunos de sus parámetros representativos pueden adquirir una relevancia significativa.

No obstante, sí conviene insistir en la idea de que la pretensión de incluir en un modelo combinado ciertas estimaciones paramétricas, no puede ser otra que la de complementar los resultados obtenidos a través del resto de perspectivas de análisis.

Existe un amplísimo y casi ilimitado abanico de posibles referencias de alta muestra: de tipo espectral (LPC, FTRI, LSP, OSALPC, MFCC...) de frecuencia (F_0 , Jitter, RMS, NSH, RAP, PPQ, ...) de energía (Shimmer, LFBE, vAm, FATR, SAPQ,...) de tiempo (T_0, DVB, ...) de carácter acústico (LPCC, DALS, F_0 , ...) de patrones psicoacústicos (PLP, EIH, MFCC, ...) de

información estática, de fluctuación dinámica (Δcepstrum, ΔE, ΔF_0, $\Delta\Delta E$, $\Delta\Delta F_0$, ...) etc, . La consideración de unos u otros parámetros, ya sea de forma selectiva o combinada, proporcionará distintos tipos de información sobre las características y comportamientos de los diferentes componentes de la señal.

El problema surge cuando nos enfrentamos a una señal cuyas características y componentes manifiestan un altísimo índice de variabilidad. Un sistema de reconocimiento automático puede llegar a conjugar muy diversas opciones de parametrización; puede estimar las características de dichos valores en relación a sus factores de recurrencia o dinámica, pero al final, siempre surge algún elemento de degradación o alteración cuya naturaleza y comportamiento no pueden ser absolutamente previstos o delimitados. Esta es la principal razón por la que la intervención del experto, tanto en lo relativo a la selección de parámetros y tramos de muestra que prevé idóneos para su evaluación, como en la interpretación de los resultados de estimación, debe considerarse una aportación esencial.

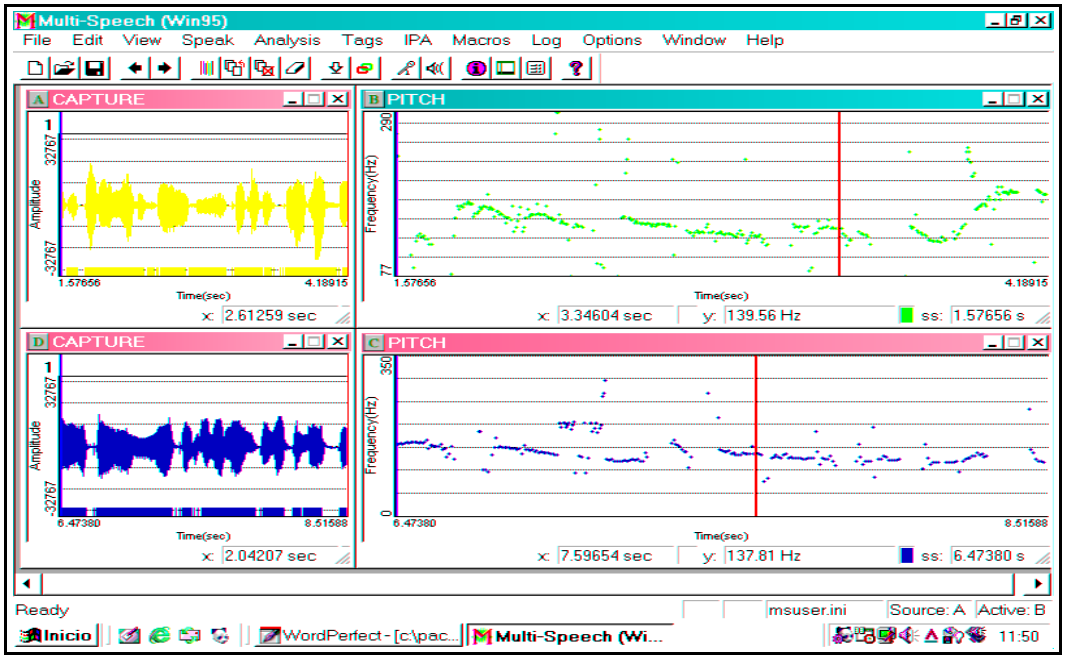
Dependerá de cada experto y, por supuesto, de cada objetivo de análisis (referencias de identidad vocal, diagnosticos de patologías, etc, ...) la viabilidad o no de proceder con el cálculo, así como la selección e interpretación de los parámetros en cuestión.

Hechas estas matizaciones, nos situamos de nuevo en nuestro caso práctico.

De forma representativa, y considerando concurren en las muestras analizadas las circunstancias adecuadas para proceder con su evaluación (concretamente nos referimos al marco expresivo-emocional), aportaremos algunas estimaciones estadísticas para intentar ubicar el valor del tono fundamental en su plano conversacional (SFF).

Para efectuar una estimación precisa de los valores del tono en las emisiones objeto de estudio, acotaremos aquellos fragmentos de discurso en los que a nivel perceptivo se observe una mayor adecuación inter muestras en sus contextos expresivo y emocional. Las muestras indubitadas han sido extraídas de la conversación espontánea.

En la ilustración n1 78, vemos un ejemplo del cálculo realizado con la aplicación de análisis Multispeech-95, correspondiente al estudio de dos fragmentos de grabación de unos 30



segundos cada uno.

Los datos estadísticos referidos al tono fundamental de la grabación dubitada han deparado los siguientes valores:

GRABACIÓN DUBITADA
Fuente: ventana A
Canal : 1
Frecuencia de muestreo: **11025**
Inicio de análisis (sec) **0.00000**
Fin de intervalo de análisis (sec) **29.99991**
Rango mínimo Pitch (Hz) **74.00**
Rango máximo Pitch (Hz) **297.00**
Longitud de frame (msec) De impulsos calculados
Análisis estadístico
Número de muestras: **2565**
Inicio de análisis (sec) : 0.0000
Fin de intervalo de análisis (sec) : 29.9999
Media de la frecuencia (Hz) : **146.22**

Media de Fo (Hz) : **141.86**
Media del período (msec) : **7.05**
Desviación estándar (Hz) : **28.88**
Mediana del Pitch (Hz) : **136.91**
Raíz cuadrada de la media RMS (Hz): **149.04**
Jitter (%): 4.677
GRABACIÓN INDUBITADA
Fuente: ventana A
Canal : 1
Frecuencia de muestreo: **11025**
Inicio de análisis (sec) **0.00000**
Fin de intervalo de análisis (sec) **29.99991**
Rango mínimo Pitch (Hz) **71.59**
Rango máximo Pitch (Hz) **297.97**
Longitud de frame (msec) De impulsos calculados
Análisis estadístico
Número de muestras: **3048**
Inicio de análisis (sec) : 0.0000

Fin de intervalo de análisis (sec) : 29.9999
Media de la frecuencia (Hz) : **138.56**
Media de Fo (Hz) : **136.32**
Media del período (msec) : **7.34**
Desviación estándar (Hz) : **18.81**
Mediana del Pitch (Hz) : **137.81**
Raíz cuadrada de la media RMS (Hz): **139.83**
Jitter (%): 5.786

Como puede observarse en la anterior tabla, existen diversos valores estadísticos vinculados al valor de la frecuencia de vibración glotal durante una locución. En nuestro caso de estudio, la referencia de interés es aquella que nos aporta el valor de tendencia de SFF en cada uno de los locutores. Dicha referencia, viene representada por el valor de la mediana, que como podemos apreciar se sitúa en 136,91 Hz para la grabación dubitada y en 137,81 Hz para la indubitada.

Esta simbólica, aunque significativa coincidencia, corrobora la posibilidad de que ambas emisiones hayan sido producidas por el mismo hablante.

Por otra parte, podemos observar que existe una diferencia en cuanto al valor de la desviación estándar inter muestras de unos 10 Hz . Esta variación del valor en dicha referencia - referencia indicativa de la fluctuación del tono y , por tanto, del carácter más o menos melódico de una emisión- debe considerarse algo normal pues la grabación dubitada suele producirse en un contexto más melódico que la indubita (aun utilizando la muestra de voz espóntanea).

IV.2.5.- Análisis de reconocimiento automático.

El hecho de incluir en nuestro modelo de estudio la perspectiva de un sistema de reconocimiento automático, no tiene otra pretensión que la de ofrecer una visión complementaria a la ya aportada en las anteriores aproximaciones de análisis. Por otro lado, la utilización de este nuevo enfoque también hemos de interpretarlo como la puerta que toda técnica forense debe mantener siempre abierta a las nuevas alternativas científicas.

El porqué de un sistema automático basado en modelos de mezclas de gausianas, ya tratamos de argumentarlo en el capítulo anterior. También en este capítulo explicábamos, que el test a desarrollar con el prototipo AIdentiVox 2000" será una tarea de identificación de tipo cerrado; o lo que es lo mismo, la comparación de una voz indubitada contra un número concreto de voces dubitadas.

Si nos retrotraemos a nuestro supuesto de investigación, recordaremos que la policía detuvo a diez empleados del local donde se hallaba el teléfono intervenido. Con posterioridad, el juez de instrucción llegó a determinar que, muy probablemente, tan sólo estos diez sujetos podrían haber tenido acceso a dicho teléfono.

Basándonos en este planteamiento, consideramos podría resultar interesante la realización de un test cerrado de identificación mediante el sistema automático. En este test, cotejaremos la voz indubitada de nuestro candidato contra los registros dubitados de los diez empleados. La utilización de muestras independientes de texto en el experimento, proporcionará otro interesante aliciente; especialmente, si tenemos en cuenta que el concurso de dicha circunstancia se revela como un componente de dificultad en la óptima ejecución de otros sistemas de análisis.

)Porqué hemos elegido un test cerrado y no abierto, y porqué identificación y no verificación?. Como veremos seguidamente, una tarea de verificación implica el establecimiento de unos umbrales que vienen determinados por la distancia existente entre dos curvas. Una, representa la posibilidad de que la voz de un candidato válido sea rechazada por el sistema (falso

rechazo). Y la otra, la probabilidad de que un impostor sea erróneamente reconocido (falsa aceptación). La definición de estos umbrales ha de efectuarse en base a una colección de modelos de voz que representen fielmente la teórica población de locutores en la que se enmarcaría nuestro candidato a la verificación. Nos estamos refiriendo a una base de datos que en su propia naturaleza resultaría amplísima, eso, sin tener en cuenta la circunstancia añadida de que cada hablante en sí mismo constituye otra población.

Pero partamos de la premisa teórica, de que ante suficientes cantidades de muestra

disponibles, lográsemos modelar con alta precisión toda una población de locutores (p.e. todos los hablantes gallegos). Admitiendo que ésta sería una labor imposible así planteada, no habríamos de olvidar tampoco, que para que nuestros modelos fuesen realmente robustos deberían contemplar otros importantes y habituales factores del audio forense: variabilidad del efecto canal, ruido, distorsiones, etc.,

Ante esta situación en la que el abanico de elementos a conjugar es tan extenso, no es difícil darse cuenta de la gran dificultad que supondrá llegar a validar como método exclusivo de ientificación, un sistema de reconocimiento automático de estas características. En nuestro caso concreto, sí podemos comentar que AIdentiVox 2000" ha sido testeado con muestras forenses de laboratorio ante distintos supuestos, en los que se han combinado algunas de sus alternativas de parametrización, entrenamiento, factor multisesión, normalización, etc, [Lucena y Díaz, 2000]. En este caso, los resultados se califican de esperanzadores aunque se recomienda una base empírica de mayor envergadura incluyendo tests con Agrabaciones reales≅. Al hilo de esta sugerencia, podemos decir que el prototipo del DIAC también ha sido sometido a la resolución de diversas tareas de identificación y verificación en casos forenses reales [I.R. n113]. A pesar de que el número de experimentos realizados en tal sentido sólo puede considerarse simbólico y, por lo tanto insuficiente, los resultados en principio aconsejan mucha cautela y una exhaustiva investigación para poder determinar ante que clase de registros y en que tipo de sesiones de trabajo la aplicación pudiera ofrecer unos resultados fiables.

En cualquier caso, y dado que en nuestro supuesto práctico contamos con unas grabaciones de buena calidad y un teórico conjunto cerrado de candidatos para llevar a cabo el correspondiente estudio, procederemos a la realización del test solicitando del sistema automático una opinión de identificación.

IV.2.5.1.- Características del sistema IdentiVox 2000

Para definir las características y prestaciones del prototipo que utilizaremos en nuestro análisis de reconocimiento automático nos remitiremos a las referencias descritas en [García Gomar et al., 2000].

Estamos ante un sistema de reconocimiento de locutores independiente de texto basado en modelos de mezclas de Gaussianas (GMMs, *Gaussian Mixture Model*s). Las principales

funciones del sistema son la capacidad de modelado de emisiones, realización de trabajos de identificación y verificación del locutor, y otros módulos auxiliares que incluyen técnicas de mejora de voz (sustracción espectral lineal y no-lineal), convertidor de formatos de ficheros y un segmentador de archivos de audio. Además, el sistema es capaz de realizar tanto normalización de canal como normalización de verosimilitud.

En síntesis, el sistema organiza sesiones de trabajo, donde los ficheros sonoros objeto de estudio pueden ser parametrizados utilizando LPCC *(Linear Predictive Cepstral Coefficient*s) o MFCC *(Mel-Frequency Cepstral Coefficient*s) . En cada sesión se diseña el número de mezclas de gaussianas para cada modelo y el tipo de normalización de canal o verosimilitud a emplear. Una vez establecidas las propiedades de cada sesión, el usuario elige los ficheros de audio con los que entrenar los modelos, y los segmentos de prueba de voz con los que se comparan frente a los modelos candidatos seleccionados (identificación) o contra un modelo específico (verificación). Los resultados se muestran en forma de relaciones de verosimilitud.

IdentiVox 2000 es una aplicación MDI *(Multi Document Interfac*e) y multihilo desarrollada bajo Windows con el entorno de programación *Visual C++* de *Microsof*t.Consta de dos partes bien diferenciadas, por un lado un núcleo de procesado, programado en su totalidad en ANSI C++, que implementa los algoritmos de procesado empleados; y por otro lado, una interfaz gráfica que posibilita la sencilla utilización en un entorno *Windows* por parte del usuario de toda la funcionalidad de este núcleo.Como aspectos a destacar en dicho núcleo de procesado podemos mencionar:

– *Normalización de canal* (CMN, *Cepstral Mean Normalizatio*n/RASTA), donde se busca la no influenciabilidad de la toma de audio empleada, es decir, eliminar el efecto que la respuesta del canal introduce en la señal. Esta normalización puede aplicarse cuando se parametrizan ficheros de audio en tareas como el entrenamiento, cálculo de umbrales para verificación, identificación, etc.

– *Normalización de verosimilitudes*, para que el resultado de verosimilitud obtenido en pruebas de verificación y de cálculo de umbrales sea relativo al modelo más próximo (normalización con mejor competidor) o a un modelo universal entrenado, obtenido en cada caso en relación a una base de datos de referencia (normalización con modelo universal).

Por otra parte, la interfaz gráfica diseñada para la plataforma *Win32* posibilita el trabajo con varias sesiones simultáneamente y en diferentes tareas de entrenamiento y reconocimiento. A su vez, cada sesión se define en razón a sus:

- Propiedades de parametrización y normalización de canal.

- Opciones de análisis: modos de trabajo para el cálculo de umbrales y normalización de verosimilitudes, condiciones de identificación, generación de informes y almacenamiento de ficheros de parámetros generados en la sesión.

- El conjunto de modelos de locutor asociados a la sesión.

Las propiedades de la sesión se seleccionan únicamente cuando ésta se crea, no pudiendo ser modificadas posteriormente. De esta forma, se asegura que todos los modelos del locutor generados dentro de una sesión, comparten las mismas propiedades de parametrización y normalización de canal/verosimilitud. De no ser así, los resultados de las pruebas de reconocimiento frente a distintos modelos no serían comparables.

Para el cálculo de la curva de FR *(False Rejectio*n) del modelo se asocia una serie de ficheros del locutor, distintos a los de entrenamiento.Para el cálculo de la curva de FA *(False Acceptanc*e) se contemplan varios métodos:

1.- *Modo Interno*: como locuciones de impostores se consideran los ficheros de audio para el cálculo de la curva de FR del resto de modelos de la sesión. Este modo será el utilizado en nuestro caso práctico.

2. - *Modo Externo Independiente*: las locuciones de los impostores son ficheros de audio asociados, de manera independiente, a cada modelo.

3.- *Modo Externo Común*: las locuciones de los impostores son ficheros de audio comunes a todos los modelos de la sesión.

También hemos señalado que el sistema incluye unas herramientas accesorias para facilitar la introducción, procesado y edición de los archivos sonoros utilizados. Como sabemos, la filosofía metodológica del análisis combinado básico (perceptivo-acústico-fonético) implica -en términos generales- la no utilización de opciones de procesado sobre la señal para no desvirtuar su auténtica cualidad. Sin embargo, en el caso del reconocimiento automático la ejecución de determinadas acciones de procesado que serían inviables en el análisis combinado clásico, no sólo no resultan perjudiciales, sino más bien pertinentes y recomendables.

Por tanto, ante ciertas situaciones de degradación de la señal (p.e. ruido) la aplicación automática, antes de utilizar los ficheros de voz, requiere de Auna limpieza≅ previa de los

mismos. Para realizar estas labores de filtrado o limpieza de ficheros de audio, el sistema dispone de una herramienta externa que además permite grabar, reproducir y visualizar dichos ficheros. La técnica empleada para la limpieza es la sustracción espectral (lineal o no lineal), y la estimación del ruido contaminante se puede llevar a cabo de forma manual o automática.

IdentiVox 2000 sólo trabaja con un único formato de audio: el formato WAV de un solo canal. También admite la utilización de ficheros de parámetros generados previamente, cuyo formato debe ser FPG. Si los ficheros disponibles son de un formato distinto, es necesario realizar una conversión previa. Por este motivo, el prototipo cuenta con un conversor de formatos que permite realizar conversiones múltiples de ficheros entre varios formatos de origen y destino (HTK, BDA, WAV estéreo, SUNAU8...).

En algunas ocasiones, cuando se dispone de poca cantidad de muestra para realizar tareas de entrenamiento, cálculo de umbrales, etc., el prototipo permite Atrocear≅ un único fichero para obtener múltiples tramos solapados o no, del mismo. Para realizar esta función, IdentiVox 2000 dispone de una herramienta externa (Atroceador≅) que materializa la segmentación simultánea de múltiples ficheros, fijando previamente el tamaño y solapamiento de los fragmentos a obtener.

IV.2.5.2.- Propiedades de la sesión de trabajo. Entrenamiento de modelos.

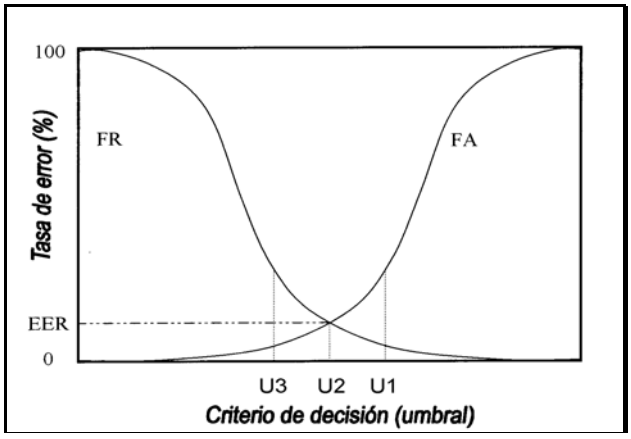
A diferencia de lo que ocurría con los otros sistemas de análisis de nuestro modelo, IdentiVox 2000 no tiene todavía predefinidos unos criterios de calidad o cantidad para regular la admisibilidad de las muestras. Por esta razón, las opciones de análisis en nuestra sesión de trabajo, vendrán marcadas por aquellas señaladas como más idóneas en los tests experimentales ya ensayados y comentados.

Lógicamente, la buena calidad y suficiente cantidad de la que partimos en nuestro supuesto práctico, constituirán un excelente caldo de cultivo para desarrollar el estudio comparativo automático en unas condiciones que podemos calificar de más que aceptables. No obstante, hemos de tener en cuenta que estos presupuestos cualitativos y cuantitativos de la señal no serán los que habitualmente encontremos en la casuística real.

El paso previo a cualquier trabajo de análisis será la digitalización de aquellos fragmentos de grabación que utilizaremos en el estudio. Una vez digitalizados, procederemos a la creación de los modelos de locutores, para lo cual utilizaremos unos 25 sg de grabación indubitada (voz espontánea) por cada locutor. La duración de cada fichero de test estará entre los 2,5 y 3,5 sg.

Una vez construidos los modelos, buscaremos un índice de su Arobustez≅ mediante el cálculo de las respectivas curvas de falsa aceptación y falso rechazo. Es decir, construiremos diez modelos (del M20 al M29) considerando como impostores para el cálculo de la curva FA en cada caso, al resto de modelos utilizados en la identificación (*Modo Interno*). La curva FR de cada modelo se establecerá con ficheros del mismo locutor modelado, aunque utilizando distintos actos de habla a las empleados en la construcción de dicho modelo (*ficheros de entrenamiento*).

Normalmente, el cálculo y representación gráfica de las curvas FA y FR nos otorga una medida de la eficacia del sistema en tareas de verificación. La expresión de tal eficacia se concreta en valores porcentuales de su tasa de error respecto de un rango de valores de verosimilitud. Como referencia base se utiliza la denominada tasa AEER≅ (Equal Error Rate), o lo que es igual, el error del sistema cuando el umbral de decisión es tal, que el porcentaje de falsas aceptaciones es igual al de falsos rechazos.



Como podremos apreciar, una vez calculadas las curvas FA y FR, IdentiVox 2000 establece tres umbrales por defecto: uno de ellos (U2) es coincidente con el valor de verosimilitud del EER; otro, (señalado como U3 ó umbral Aexigente≅) se traza en una referencia de verosimilitud de un 10% mayor que la correspondiente al EER; y un tercero, situado en un valor de verosimilitud del orden de un 10% menor que el definido para el EER (U1 ó umbral Arelajado≅).

Recordemos, que el sentido de calcular los umbrales y curvas FA y FR no se fundamenta en la intención de realizar un estudio de verificación, sino en comprobar cómo el sistema aprecia el comportamiento de cada uno de los modelos, en relación a nuestro modelo candidato (M 20). Esto es, conocer si los locutores que utilizaremos para el estudio de identificación se manifestarán ante el sistema como Aovejas≅, Acabras≅, Acorderos≅ o Alobos≅. Estos extraños términos, son los utilizados en el ámbito del reconocimiento automático [Doddington et al. 1998] para denominar las distintas clases de locutores, en función del grado de dificultad o tipo de error que pueden representar para la eficacia del sistema. Se llama Aovejas≅ a aquellos hablantes fáciles de identificar por el sistema. Si resulta problemática su identificación por el sistema se les denomina Acabras≅. Son Acorderos≅ los que tienen una voz muy fácil de imitar, por lo que otros locutores pueden ser equivocadamente reconocidos en su lugar. Y por último los Alobos≅, o aquellos hablantes que con facilidad son reconocidos equivocadamente por el sistema, en lugar de otros.

Efectuadas estas necesarias aclaraciones, pasamos a especificar las propiedades de nuestra sesión de trabajo y los detalles de entrenamiento y cálculo de umbrales para cada modelo. La selección de las siguientes opciones se ha realizado en virtud de aquellas que han sido señaladas como más funcionales en diferentes testeos a los que ha sido sometido el sistema: [González, 1999], [Lucena y Díaz, 2000].

- **NOMBRE DE LA SESIÓN DE TRABAJO**: ***SesionCD1***.

- **OPCIONES DE PARAMETRIZACIÓN**:

    - Normalización de canal: **CMN**

- Tipo de parametrización: **MFCC**
- Magnitud:
    - Número de filtros: **24**
    - HF:-1
    - LF:-1
- Ventana de análisis: **HAMMING**
    - Longitud de ventana: **32**
    - Solapamiento (%): **50**
- Coeficientes:
    - C0: SI
    - Energía: NO
    - Coeficientes Delta: **SI**
        - Amplitud: **2**
    - Coeficientes Delta-Delta: **SI**
        - Amplitud: **2**

- Preénfasis: **0.97**
- Frecuencia muestreo: **8000 Hz**

-**UMBRALES**:

- Modo Cálculo:
    - **Matriz Redundante**
    - **Modo Interno**
- Normalización de verosimilitud: NINGUNA

- **CARACTERÍSTICAS DE LOS MODELOS**:

MODELO 20

- Mezclas: 16

- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\020M1B01.wav
C:\archivos\audio\demo_locs\hombres\020M1B02.wav
C:\archivos\audio\demo_locs\hombres\020M1B03.wav
C:\archivos\audio\demo_locs\hombres\020M1B04.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav
C:\archivos\audio\demo_locs\hombres\020M1C02.wav
C:\archivos\audio\demo_locs\hombres\020M1C03.wav
C:\archivos\audio\demo_locs\hombres\020M1C04.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\020M1C08.wav
C:\archivos\audio\demo_locs\hombres\020M1B06.wav
C:\archivos\audio\demo_locs\hombres\020M1B07.wav
C:\archivos\audio\demo_locs\hombres\020M1B08.wav
C:\archivos\audio\demo_locs\hombres\020M1C05.wav
C:\archivos\audio\demo_locs\hombres\020M1C06.wav
C:\archivos\audio\demo_locs\hombres\020M1C07.wav
C:\archivos\audio\demo_locs\hombres\020M1B05.wav

Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\029M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1C03.wav
C:\archivos\audio\demo_locs\hombres\022M1B06.wav
C:\archivos\audio\demo_locs\hombres\022M1C01.wav
C:\archivos\audio\demo_locs\hombres\023M1B03.wav
C:\archivos\audio\demo_locs\hombres\023M1C01.wav
C:\archivos\audio\demo_locs\hombres\024M1B07.wav
C:\archivos\audio\demo_locs\hombres\024M1C03.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav
C:\archivos\audio\demo_locs\hombres\025M1C02.wav

C:\archivos\audio\demo_locs\hombres\026M1B09.wav
C:\archivos\audio\demo_locs\hombres\026M1C01.wav
C:\archivos\audio\demo_locs\hombres\027M1B07.wav
C:\archivos\audio\demo_locs\hombres\027M1C01.wav
C:\archivos\audio\demo_locs\hombres\028M1B04.wav
C:\archivos\audio\demo_locs\hombres\028M1C01.wav
C:\archivos\audio\demo_locs\hombres\029M1B08.wav

C:\archivos\audio\demo_locs\hombres\021M1B02.wav

Umbrales:

U1: -25.5757
U2: -24.3129
U3: -23.5348

MODELO 21

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\021M1C04.wav
C:\archivos\audio\demo_locs\hombres\021M1B02.wav
C:\archivos\audio\demo_locs\hombres\021M1B03.wav
C:\archivos\audio\demo_locs\hombres\021M1B04.wav
C:\archivos\audio\demo_locs\hombres\021M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1C02.wav
C:\archivos\audio\demo_locs\hombres\021M1C03.wav
C:\archivos\audio\demo_locs\hombres\021M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\021M1C08.wav
C:\archivos\audio\demo_locs\hombres\021M1B06.wav
C:\archivos\audio\demo_locs\hombres\021M1B07.wav
C:\archivos\audio\demo_locs\hombres\021M1B08.wav
C:\archivos\audio\demo_locs\hombres\021M1C05.wav
C:\archivos\audio\demo_locs\hombres\021M1C06.wav
C:\archivos\audio\demo_locs\hombres\021M1C07.wav
C:\archivos\audio\demo_locs\hombres\021M1B05.wav

Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\029M1C04.wav
C:\archivos\audio\demo_locs\hombres\020M1C02.wav
C:\archivos\audio\demo_locs\hombres\022M1B02.wav
C:\archivos\audio\demo_locs\hombres\022M1C01.wav
C:\archivos\audio\demo_locs\hombres\023M1B08.wav
C:\archivos\audio\demo_locs\hombres\023M1C01.wav

C:\archivos\audio\demo_locs\hombres\024M1B07.wav
C:\archivos\audio\demo_locs\hombres\024M1C02.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav
C:\archivos\audio\demo_locs\hombres\025M1C02.wav
C:\archivos\audio\demo_locs\hombres\026M1B08.wav
C:\archivos\audio\demo_locs\hombres\026M1C04.wav
C:\archivos\audio\demo_locs\hombres\027M1B07.wav
C:\archivos\audio\demo_locs\hombres\027M1C02.wav
C:\archivos\audio\demo_locs\hombres\028M1B05.wav
C:\archivos\audio\demo_locs\hombres\028M1C02.wav
C:\archivos\audio\demo_locs\hombres\029M1B08.wav
C:\archivos\audio\demo_locs\hombres\020M1B05.wav

Umbrales:

U1: -25.9804
U2: -24.2373
U3: -23.5429

MODELO 22

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\022M1C04.wav
C:\archivos\audio\demo_locs\hombres\022M1B02.wav
C:\archivos\audio\demo_locs\hombres\022M1B03.wav
C:\archivos\audio\demo_locs\hombres\022M1B04.wav
C:\archivos\audio\demo_locs\hombres\022M1C01.wav
C:\archivos\audio\demo_locs\hombres\022M1C02.wav
C:\archivos\audio\demo_locs\hombres\022M1C03.wav
C:\archivos\audio\demo_locs\hombres\022M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\022M1C08.wav
C:\archivos\audio\demo_locs\hombres\022M1B06.wav
C:\archivos\audio\demo_locs\hombres\022M1B07.wav
C:\archivos\audio\demo_locs\hombres\022M1B08.wav
C:\archivos\audio\demo_locs\hombres\022M1C05.wav
C:\archivos\audio\demo_locs\hombres\022M1C06.wav
C:\archivos\audio\demo_locs\hombres\022M1C07.wav

C:\archivos\audio\demo_locs\hombres\022M1B05.wav

Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\029M1B07.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1B05.wav
C:\archivos\audio\demo_locs\hombres\021M1C02.wav
C:\archivos\audio\demo_locs\hombres\023M1B08.wav
C:\archivos\audio\demo_locs\hombres\023M1C01.wav
C:\archivos\audio\demo_locs\hombres\024M1B06.wav
C:\archivos\audio\demo_locs\hombres\024M1C03.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav
C:\archivos\audio\demo_locs\hombres\025M1C01.wav
C:\archivos\audio\demo_locs\hombres\026M1B08.wav
C:\archivos\audio\demo_locs\hombres\026M1C05.wav
C:\archivos\audio\demo_locs\hombres\027M1B08.wav
C:\archivos\audio\demo_locs\hombres\027M1C03.wav
C:\archivos\audio\demo_locs\hombres\028M1B05.wav
C:\archivos\audio\demo_locs\hombres\028M1C01.wav
C:\archivos\audio\demo_locs\hombres\020M1B02.wav
C:\archivos\audio\demo_locs\hombres\029M1C06.wav

Umbrales:

U1: -28.0793
U2: -25.3335
U3: -24.1217

MODELO 23

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\023M1C04.wav

C:\archivos\audio\demo_locs\hombres\023M1B02.wav
C:\archivos\audio\demo_locs\hombres\023M1B03.wav
C:\archivos\audio\demo_locs\hombres\023M1B04.wav
C:\archivos\audio\demo_locs\hombres\023M1C01.wav
C:\archivos\audio\demo_locs\hombres\023M1C02.wav
C:\archivos\audio\demo_locs\hombres\023M1C03.wav
C:\archivos\audio\demo_locs\hombres\023M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\023M1C08.wav
C:\archivos\audio\demo_locs\hombres\023M1B06.wav
C:\archivos\audio\demo_locs\hombres\023M1B07.wav
C:\archivos\audio\demo_locs\hombres\023M1B08.wav
C:\archivos\audio\demo_locs\hombres\023M1C05.wav
C:\archivos\audio\demo_locs\hombres\023M1C06.wav
C:\archivos\audio\demo_locs\hombres\023M1C07.wav
C:\archivos\audio\demo_locs\hombres\023M1B05.wav

Ficheros Falsa Aceptación Independiente

C:\archivos\audio\demo_locs\hombres\029M1B09.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1B06.wav
C:\archivos\audio\demo_locs\hombres\021M1C03.wav
C:\archivos\audio\demo_locs\hombres\022M1B06.wav
C:\archivos\audio\demo_locs\hombres\022M1C02.wav
C:\archivos\audio\demo_locs\hombres\024M1B07.wav
C:\archivos\audio\demo_locs\hombres\024M1C02.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav
C:\archivos\audio\demo_locs\hombres\025M1C01.wav
C:\archivos\audio\demo_locs\hombres\026M1B08.wav
C:\archivos\audio\demo_locs\hombres\026M1C04.wav

C:\archivos\audio\demo_locs\hombres\027M1B07.wav
C:\archivos\audio\demo_locs\hombres\027M1C03.wav
C:\archivos\audio\demo_locs\hombres\028M1B05.wav
C:\archivos\audio\demo_locs\hombres\028M1C01.wav
C:\archivos\audio\demo_locs\hombres\020M1B04.wav
C:\archivos\audio\demo_locs\hombres\029M1C05.wav

Umbrales:

U1: -25.0265
U2: -24.077

U3: -23.7884

MODELO 24

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\024M1C04.wav
C:\archivos\audio\demo_locs\hombres\024M1B02.wav
C:\archivos\audio\demo_locs\hombres\024M1B03.wav
C:\archivos\audio\demo_locs\hombres\024M1B04.wav
C:\archivos\audio\demo_locs\hombres\024M1C01.wav
C:\archivos\audio\demo_locs\hombres\024M1C02.wav
C:\archivos\audio\demo_locs\hombres\024M1C03.wav
C:\archivos\audio\demo_locs\hombres\024M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\024M1C08.wav
C:\archivos\audio\demo_locs\hombres\024M1B06.wav
C:\archivos\audio\demo_locs\hombres\024M1B07.wav
C:\archivos\audio\demo_locs\hombres\024M1B08.wav
C:\archivos\audio\demo_locs\hombres\024M1C05.wav
C:\archivos\audio\demo_locs\hombres\024M1C06.wav
C:\archivos\audio\demo_locs\hombres\024M1C07.wav
C:\archivos\audio\demo_locs\hombres\024M1B05.wav

Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\029M1B04.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1B07.wav
C:\archivos\audio\demo_locs\hombres\021M1C03.wav
C:\archivos\audio\demo_locs\hombres\022M1B05.wav
C:\archivos\audio\demo_locs\hombres\022M1C01.wav
C:\archivos\audio\demo_locs\hombres\023M1B08.wav
C:\archivos\audio\demo_locs\hombres\023M1C04.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav
C:\archivos\audio\demo_locs\hombres\025M1C01.wav

C:\archivos\audio\demo_locs\hombres\026M1B08.wav
C:\archivos\audio\demo_locs\hombres\026M1C04.wav
C:\archivos\audio\demo_locs\hombres\027M1B07.wav
C:\archivos\audio\demo_locs\hombres\027M1C02.wav
C:\archivos\audio\demo_locs\hombres\028M1B06.wav
C:\archivos\audio\demo_locs\hombres\028M1C01.wav
C:\archivos\audio\demo_locs\hombres\020M1B04.wav
C:\archivos\audio\demo_locs\hombres\029M1B07.wav

Umbrales:

U1: -25.2555
U2: -24.2521
U3: -23.7283

MODELO 25

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\025M1C04.wav
C:\archivos\audio\demo_locs\hombres\025M1B02.wav
C:\archivos\audio\demo_locs\hombres\025M1B03.wav
C:\archivos\audio\demo_locs\hombres\025M1B04.wav
C:\archivos\audio\demo_locs\hombres\025M1C01.wav
C:\archivos\audio\demo_locs\hombres\025M1C02.wav
C:\archivos\audio\demo_locs\hombres\025M1C03.wav
C:\archivos\audio\demo_locs\hombres\025M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\025M1C08.wav
C:\archivos\audio\demo_locs\hombres\025M1B06.wav
C:\archivos\audio\demo_locs\hombres\025M1B07.wav
C:\archivos\audio\demo_locs\hombres\025M1B08.wav
C:\archivos\audio\demo_locs\hombres\025M1C05.wav
C:\archivos\audio\demo_locs\hombres\025M1C06.wav
C:\archivos\audio\demo_locs\hombres\025M1C07.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav

Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\029M1C05.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1B06.wav
C:\archivos\audio\demo_locs\hombres\021M1C02.wav
C:\archivos\audio\demo_locs\hombres\022M1B05.wav
C:\archivos\audio\demo_locs\hombres\022M1C02.wav
C:\archivos\audio\demo_locs\hombres\023M1B04.wav
C:\archivos\audio\demo_locs\hombres\023M1C01.wav
C:\archivos\audio\demo_locs\hombres\024M1B08.wav
C:\archivos\audio\demo_locs\hombres\024M1C03.wav
C:\archivos\audio\demo_locs\hombres\026M1B03.wav
C:\archivos\audio\demo_locs\hombres\026M1C01.wav
C:\archivos\audio\demo_locs\hombres\027M1B07.wav
C:\archivos\audio\demo_locs\hombres\027M1C03.wav
C:\archivos\audio\demo_locs\hombres\028M1B04.wav
C:\archivos\audio\demo_locs\hombres\028M1C01.wav
C:\archivos\audio\demo_locs\hombres\029M1B08.wav
C:\archivos\audio\demo_locs\hombres\020M1B04.wav

Umbrales:

U1: -25.6532
U2: -24.8126
U3: -24.771

MODELO 26

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\026M1C04.wav
C:\archivos\audio\demo_locs\hombres\026M1B02.wav
C:\archivos\audio\demo_locs\hombres\026M1B03.wav
C:\archivos\audio\demo_locs\hombres\026M1B04.wav
C:\archivos\audio\demo_locs\hombres\026M1C01.wav

C:\archivos\audio\demo_locs\hombres\026M1C02.wav
C:\archivos\audio\demo_locs\hombres\026M1C03.wav
C:\archivos\audio\demo_locs\hombres\026M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\026M1C08.wav
C:\archivos\audio\demo_locs\hombres\026M1B06.wav
C:\archivos\audio\demo_locs\hombres\026M1B07.wav
C:\archivos\audio\demo_locs\hombres\026M1B08.wav
C:\archivos\audio\demo_locs\hombres\026M1C05.wav
C:\archivos\audio\demo_locs\hombres\026M1C06.wav
C:\archivos\audio\demo_locs\hombres\026M1C07.wav
C:\archivos\audio\demo_locs\hombres\026M1B05.wav

Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\029M1C01.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1B06.wav
C:\archivos\audio\demo_locs\hombres\021M1C02.wav
C:\archivos\audio\demo_locs\hombres\022M1B04.wav
C:\archivos\audio\demo_locs\hombres\022M1C01.wav
C:\archivos\audio\demo_locs\hombres\023M1B02.wav
C:\archivos\audio\demo_locs\hombres\023M1C01.wav
C:\archivos\audio\demo_locs\hombres\024M1B05.wav
C:\archivos\audio\demo_locs\hombres\024M1C03.wav
C:\archivos\audio\demo_locs\hombres\025M1B04.wav
C:\archivos\audio\demo_locs\hombres\025M1C02.wav

C:\archivos\audio\demo_locs\hombres\027M1B07.wav
C:\archivos\audio\demo_locs\hombres\027M1C02.wav
C:\archivos\audio\demo_locs\hombres\028M1B04.wav
C:\archivos\audio\demo_locs\hombres\028M1C01.wav
C:\archivos\audio\demo_locs\hombres\029M1B03.wav
C:\archivos\audio\demo_locs\hombres\020M1B02.wav

Umbrales:

U1: -25.6921
U2: -24.2619
U3: -23.6317

MODELO 27

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\027M1C04.wav
C:\archivos\audio\demo_locs\hombres\027M1B02.wav
C:\archivos\audio\demo_locs\hombres\027M1B03.wav
C:\archivos\audio\demo_locs\hombres\027M1B04.wav
C:\archivos\audio\demo_locs\hombres\027M1C01.wav
C:\archivos\audio\demo_locs\hombres\027M1C02.wav
C:\archivos\audio\demo_locs\hombres\027M1C03.wav
C:\archivos\audio\demo_locs\hombres\027M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\027M1C08.wav
C:\archivos\audio\demo_locs\hombres\027M1B06.wav
C:\archivos\audio\demo_locs\hombres\027M1B07.wav
C:\archivos\audio\demo_locs\hombres\027M1B08.wav
C:\archivos\audio\demo_locs\hombres\027M1C05.wav
C:\archivos\audio\demo_locs\hombres\027M1C06.wav
C:\archivos\audio\demo_locs\hombres\027M1C07.wav
C:\archivos\audio\demo_locs\hombres\027M1B05.wav

Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\029M1C06.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1B05.wav
C:\archivos\audio\demo_locs\hombres\021M1C02.wav
C:\archivos\audio\demo_locs\hombres\022M1B04.wav
C:\archivos\audio\demo_locs\hombres\022M1C01.wav
C:\archivos\audio\demo_locs\hombres\023M1B07.wav
C:\archivos\audio\demo_locs\hombres\023M1C04.wav
C:\archivos\audio\demo_locs\hombres\024M1B02.wav
C:\archivos\audio\demo_locs\hombres\024M1C02.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav
C:\archivos\audio\demo_locs\hombres\025M1C02.wav
C:\archivos\audio\demo_locs\hombres\026M1B04.wav

C:\archivos\audio\demo_locs\hombres\026M1C01.wav
C:\archivos\audio\demo_locs\hombres\028M1B06.wav
C:\archivos\audio\demo_locs\hombres\028M1C02.wav
C:\archivos\audio\demo_locs\hombres\029M1B08.wav
C:\archivos\audio\demo_locs\hombres\020M1B01.wav

Umbrales:

U1: -25.19
U2: -23.9693
U3: -23.9235

MODELO 28

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\028M1C04.wav
C:\archivos\audio\demo_locs\hombres\028M1B02.wav
C:\archivos\audio\demo_locs\hombres\028M1B03.wav
C:\archivos\audio\demo_locs\hombres\028M1B04.wav
C:\archivos\audio\demo_locs\hombres\028M1C01.wav
C:\archivos\audio\demo_locs\hombres\028M1C02.wav
C:\archivos\audio\demo_locs\hombres\028M1C03.wav
C:\archivos\audio\demo_locs\hombres\028M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\028M1C08.wav
C:\archivos\audio\demo_locs\hombres\028M1B06.wav
C:\archivos\audio\demo_locs\hombres\028M1B07.wav
C:\archivos\audio\demo_locs\hombres\028M1B08.wav
C:\archivos\audio\demo_locs\hombres\028M1C05.wav
C:\archivos\audio\demo_locs\hombres\028M1C06.wav
C:\archivos\audio\demo_locs\hombres\028M1C07.wav
C:\archivos\audio\demo_locs\hombres\028M1B05.wav

Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\029M1C01.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav

C:\archivos\audio\demo_locs\hombres\021M1B06.wav
C:\archivos\audio\demo_locs\hombres\021M1C02.wav
C:\archivos\audio\demo_locs\hombres\022M1B05.wav
C:\archivos\audio\demo_locs\hombres\022M1C01.wav
C:\archivos\audio\demo_locs\hombres\023M1B03.wav
C:\archivos\audio\demo_locs\hombres\023M1C01.wav
C:\archivos\audio\demo_locs\hombres\024M1B03.wav
C:\archivos\audio\demo_locs\hombres\024M1C02.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav
C:\archivos\audio\demo_locs\hombres\025M1C01.wav
C:\archivos\audio\demo_locs\hombres\026M1B03.wav
C:\archivos\audio\demo_locs\hombres\026M1C01.wav
C:\archivos\audio\demo_locs\hombres\027M1B02.wav
C:\archivos\audio\demo_locs\hombres\027M1C01.wav
C:\archivos\audio\demo_locs\hombres\029M1B08.wav
C:\archivos\audio\demo_locs\hombres\020M1B04.wav

Umbrales:

U1: -25.997
U2: -24.6221
U3: -24.0836

MODELO 29

- Mezclas: 16
- Entrenado: SI

Ficheros Entrenamiento:

C:\archivos\audio\demo_locs\hombres\029M1C04.wav
C:\archivos\audio\demo_locs\hombres\029M1B02.wav
C:\archivos\audio\demo_locs\hombres\029M1B03.wav
C:\archivos\audio\demo_locs\hombres\029M1B04.wav
C:\archivos\audio\demo_locs\hombres\029M1C01.wav
C:\archivos\audio\demo_locs\hombres\029M1C02.wav
C:\archivos\audio\demo_locs\hombres\029M1C03.wav
C:\archivos\audio\demo_locs\hombres\029M1B01.wav

Ficheros Falso Rechazo:

C:\archivos\audio\demo_locs\hombres\029M1C08.wav
C:\archivos\audio\demo_locs\hombres\029M1B06.wav
C:\archivos\audio\demo_locs\hombres\029M1B07.wav
C:\archivos\audio\demo_locs\hombres\029M1B08.wav
C:\archivos\audio\demo_locs\hombres\029M1C05.wav
C:\archivos\audio\demo_locs\hombres\029M1C06.wav
C:\archivos\audio\demo_locs\hombres\029M1C07.wav
C:\archivos\audio\demo_locs\hombres\029M1B05.wav

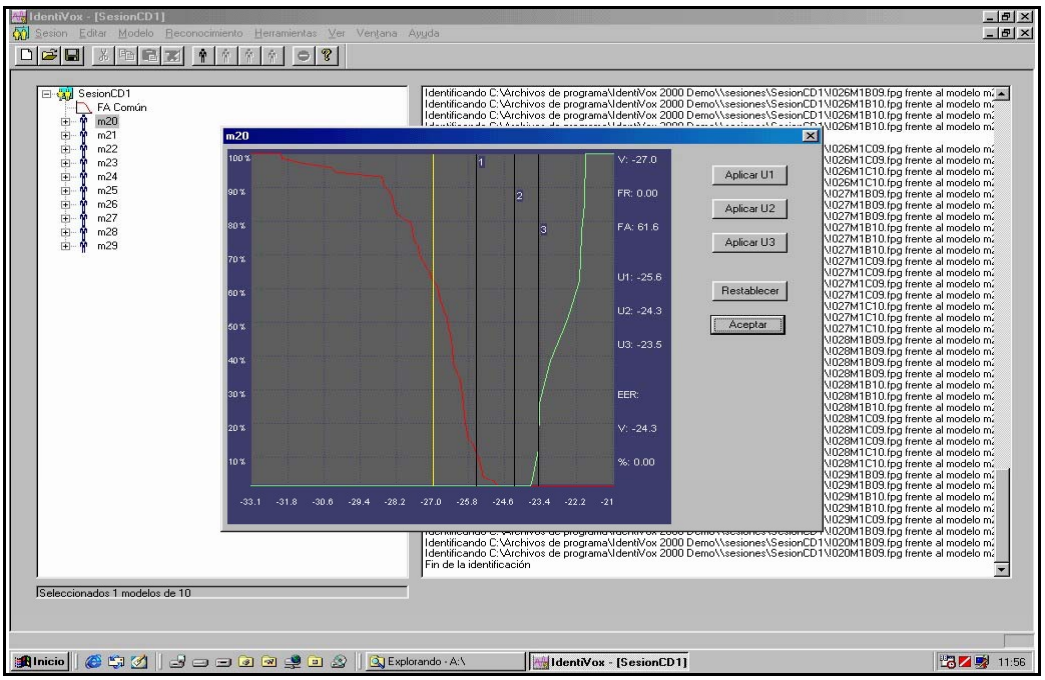
Ficheros Falsa Aceptación Independiente:

C:\archivos\audio\demo_locs\hombres\028M1C01.wav
C:\archivos\audio\demo_locs\hombres\020M1C01.wav
C:\archivos\audio\demo_locs\hombres\021M1B06.wav
C:\archivos\audio\demo_locs\hombres\021M1C03.wav
C:\archivos\audio\demo_locs\hombres\022M1B04.wav
C:\archivos\audio\demo_locs\hombres\022M1C02.wav
C:\archivos\audio\demo_locs\hombres\023M1B03.wav
C:\archivos\audio\demo_locs\hombres\023M1C01.wav
C:\archivos\audio\demo_locs\hombres\024M1B01.wav
C:\archivos\audio\demo_locs\hombres\024M1C02.wav
C:\archivos\audio\demo_locs\hombres\025M1B05.wav
C:\archivos\audio\demo_locs\hombres\025M1C01.wav
C:\archivos\audio\demo_locs\hombres\026M1B03.wav

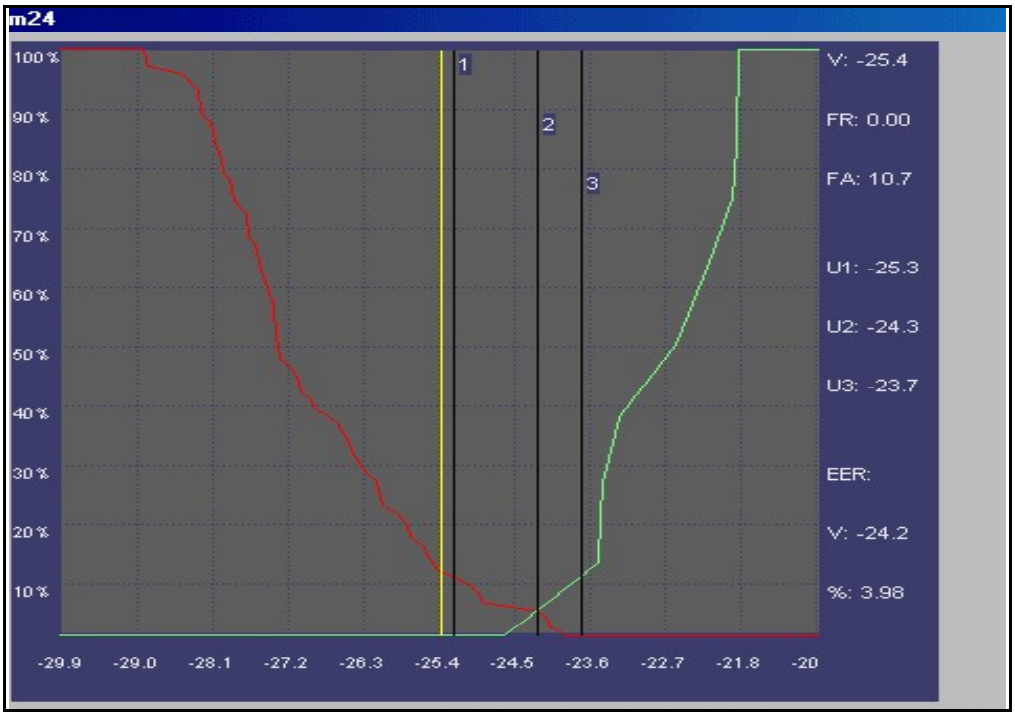
C:\archivos\audio\demo_locs\hombres\026M1B07.wav
C:\archivos\audio\demo_locs\hombres\027M1B06.wav
C:\archivos\audio\demo_locs\hombres\027M1C02.wav
C:\archivos\audio\demo_locs\hombres\028M1B05.wav
C:\archivos\audio\demo_locs\hombres\020M1B01.wav

Umbrales:

U1: -25.5977
U2: -24.2146
U3: -23.872



Detallados los datos relativos a archivos de entrenamiento, cálculo de curvas FA y FR y definición de umbrales, hemos podido comprobar por AModo Interno≅ que la estructura de robusted de cada uno de los modelos es muy consistente (Ilustración n1 80). En el peor de los



casos, (Ilustración n1 81) las curvas FR y FA presentan un valor del 3,98% de tasa de error en el EER, que no puede considerarse significativo.

IV.2.5.3.- Tests de identificación.

El cálculo de la curva FR en Amodo interno≅, ha permitido evidenciar una personalidad claramente diferenciada de cada uno de los diez modelos representativos de las diez voces indubitadas pertenecientes a los sujetos implicados en nuestro caso práctico. Por este motivo, la tarea de identificación de tipo cerrado que en un principio tan sólo íbamos a efectuar con fragmentos dubitados atribuidos a nuestro candidato (**m20** para el sistema), vamos a realizarla también con otros registros dubitados atribuidos al resto de locutores de los modelos competidores (los factores calidad y cantidad de estas muestras, son similares a los ya descritos en *IV.1* para las muestras problema).

Por tanto, llevaremos a cabo 10 tests de identificación cerrada, utilizando en cada uno de ellos 4 archivos de voz dubitada contra 10 modelos competidores. Pediremos al sistema que nos presente los resultados de comparación categorizándonos dichos modelos en razón de su puntuación en rango de verosimilitud (N-best). Al final, habremos efectuado 400 cálculos de verosimilitud.

TEST N1 1 .- DUBITADA: MUESTRA PROBLEMA , atribuida a m 20.

1.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**020M1B09.fpg:**
1.**m20**: -22.4224;
2.m29: -26.4191;
3.m26: -26.6185;
4.m23: -26.9433;
5.m27: -27.0438;
6.m28: -27.049;
7.m24: -28.269;
8.m25: -28.5923;
9.m21: -29.147;
10.m22: -31.8493;

2.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**020M1B10.fpg**:
1.**m20**: -22.3471;
2.m29: -25.8597;
3.m26: -26.4117;
4.m23: -26.75;
5.m28: -27.1526;

6.m27: -27.1764;
7.m24: -27.8352;
8.m25: -28.262;
9.m21: -28.4177;
10.m22: -32.0247;

3.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**020M1C09.fpg:**
1.**m20**: -22.6562;
2.m29: -26.3596;
3.m21: -26.3699;
4.m27: -26.4253;
5.m25: -27.1174;
6.m28: -27.1325;
7.m26: -27.336;
8.m24: -27.5749;
9.m23: -27.9914;
10.m22: -32.6749;

4.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**020M1C10.fpg**:
1**.m20**: -24.1015;
2.m29: -27.2119;
3.m27: -27.562;
4.m26: -27.6136;

5.m28: -28.2352;
6.m24: -28.2591;
7.m25: -28.4208;
8.m21: -28.4383;
9.m23: -28.4428;
10.m22: -33.6728;

TEST N1 2 .- Dubitada atribuida a m 21.

5.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1**!021M1B09.fpg**:
1**.m21**: -21.6324;
2.m25: -23.3467;
3.m24: -23.6387;
4.m28: -24.8158;
5.m27: -25.0846;
6.m20: -25.274;
7.m29: -25.6196;
8.m26: -26.6038;
9.m23: -26.9026;
10.m22: -27.2909;

6.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**021M1B10.fpg**:
1.**m21**: -20.8976;
2.m25: -23.3368;
3.m24: -23.4269;
4.m28: -24.7462;
5.m20: -24.8519;
6.m27: -25.1001;
7.m29: -25.3772;
8.m26: -26.3293;
9.m23: -26.5452;
10.m22: -27.5931;

7.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**021M1C09.fpg**:
1.**m21**: -24.0096;
2.m25: -26.3818;
3.m24: -27.6237;
4.m28: -27.7046;
5.m27: -28.2674;
6.m20: -28.4575;
7.m29: -29.6357;
8.m26: -30.6908;
9.m23: -30.9345;
10.m22: -32.5346;

8.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**021M1C10.fpg**:
1.**m21**: -23.1825;
2.m25: -25.6112;
3.m24: -25.7988;
4.m27: -26.5816;
5.m28: -26.6307;
6.m20: -27.4155;
7.m29: -28.3071;
8.m26: -29.3626;
9.m23: -29.7334;
10.m22: -30.4504;

TEST N1 3 .- Dubitada atribuida a m 22.

9.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**022M1B09.fpg**:
1.**m22**: -22.9381;
2.m24: -26.4189;
3.m27: -27.683;
4.m25: -27.867;
5.m26: -27.9593;
6.m23: -28.3458;
7.m29: -28.3983;

8.m28: -28.5319;
9.m21: -28.9749;
10.m20: -29.2306;

10.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1**!022M1B10.fpg**:
1.**m22**: -22.3898;
2.m24: -26.197;
3.m25: -27.0199;
4.m26: -27.4728;
5.m27: -27.6713;
6.m23: -27.773;
7.m28: -28.0169;
8.m29: -28.2953;
9.m21: -28.3916;
10.m20: -28.868;

11.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1**!022M1C09.fpg**:
1.**m22**: -24.5691;
2.m24: -28.9384;
3.m25: -29.3794;
4.m27: -29.5137;
5.m26: -29.6744;

6.m29: -30.3682;
7.m23: -30.4341;
8.m28: -30.6932;
9.m21: -31.3452;
10.m20: -31.9467;

12.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\\!**022M1C10.fpg**:
1.**m22**: -24.6376;
2.m24: -28.2263;
3.m25: -28.5029;
4.m26: -28.7052;
5.m23: -29.2965;
6.m28: -29.388;
7.m27: -29.5828;
8.m29: -29.8384;
9.m21: -30.1804;
10.m20: -30.8919;

TEST N1 4 .- Dubitada atribuida a m 23.

13.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\\!**023M1B09.fpg**:
1.**m23**: -21.6569;

2.m29: -24.4288;
3.m26: -25.6202;
4.m20: -26.5559;
5.m28: -27.3359;
6.m25: -27.6669;
7.m27: -28.1909;
8.m24: -28.2833;
9.m21: -29.7195;
10.m22: -30.1544;

14.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**023M1B10.fpg:**
1.**m23**: -22.0267;
2.m29: -25.2618;
3.m26: -26.1127;
4.m20: -26.6952;
5.m28: -27.7406;
6.m25: -28.1395;
7.m24: -28.4793;
8.m27: -28.5839;
9.m21: -30.2111;
10.m22: -30.5651;

15.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**023M1C09.fpg:**
1.**m23**: -23.4215;
2.m29: -25.3345;
3.m26: -26.9776;
4.m20: -27.1392;
5.m28: -28.2392;
6.m25: -28.5424;
7.m24: -28.7703;
8.m27: -28.9207;
9.m21: -29.6853;
10.m22: -31.5866;

16.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**023M1C10.fpg**:
1.**m23**: -24.3632;
2.m29: -26.6992;
3.m26: -27.1383;
4.m20: -27.4118;
5.m25: -28.1902;
6.m28: -28.4142;
7.m24: -28.6696;
8.m27: -28.7958;
9.m21: -29.5468;
10.m22: -31.1795;

TEST N1 5 .- Dubitada atribuida a m 24.

17.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**024M1B09.fpg:**
1.**m24**: -22.6905;
2.m27: -24.792;
3.m21: -26.0211;
4.m29: -26.4523;
5.m20: -26.6631;
6.m28: -26.751;
7.m25: -26.9207;
8.m26: -27.181;
9.m22: -27.3664;
10.m23: -27.4503;

18.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**024M1B10.fpg:**
1.**m24**: -21.5942;
2.m27: -23.9;
3.m21: -24.9336;
4.m28: -25.4713;
5.m20: -25.8151;

6.m29: -25.8243;
7.m25: -26.3783;
8.m22: -26.3837;
9.m26: -26.56;
10.m23: -27.0681;

19.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**024M1C09.fpg:**
1.**m24**: -24.0161;
2.m27: -26.1561;
3.m21: -27.2453;
4.m29: -28.193;
5.m20: -28.2912;
6.m28: -28.8568;
7.m25: -29.0832;
8.m26: -29.3582;
9.m23: -29.645;
10.m22: -30.2113;

20.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**024M1C10.fpg**:
1.**m24**: -24.687;
2.m27: -26.7756;
3.m21: -27.3888;
4.m20: -28.0407;
5.m29: -28.0884;
6.m28: -28.3353;

7.m25: -29.316;
8.m23: -29.6519;
9.m26: -29.7358;
10.m22: -31.541;

TEST N1 6 .- Dubitada atribuida a m 25.

21.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**025M1B09.fpg**:
1.**m25**: -23.8282;
2.m21: -26.2873;
3.m28: -26.7293;
4.m24: -27.4556;
5.m20: -27.7052;
6.m26: -28.6209;
7.m27: -28.6595;
8.m23: -28.9234;
9.m29: -29.1936;
10.m22: -30.5176;

22.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**025M1B10.fpg**:
1.**m25**: -23.8713;
2.m21: -26.5837;
3.m28: -26.8544;
4.m24: -27.429;
5.m20: -28.0965;
6.m27: -28.4517;
7.m26: -28.6159;
8.m29: -28.9135;
9.m23: -28.9191;
10.m22: -29.8701;

23.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**025M1C09.fpg:**
1**.m25**: -24.7181;
2.m21: -25.8386;
3.m24: -27.1274;
4.m29: -27.5486;
5.m26: -27.5848;
6.m27: -27.6378;
7.m20: -27.7101;
8.m28: -27.9682;
9.m23: -28.1406;
10.m22: -29.4502;

24.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**025M1C10.fpg:**
1.**m25**: -24.7895;

2.m21: -27.0489;
3.m24: -27.8776;
4.m28: -28.2858;
5.m20: -28.6079;
6.m27: -28.8222;
7.m29: -28.8531;
8.m26: -28.9156;
9.m23: -29.2204;
10.m22: -29.4148;

TEST N1 7 .- Dubitada atribuida a m 26.

25.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**026M1B09.fpg**:
1.**m26**: -21.9881;
2.m23: -25.3857;
3.m29: -25.4929;
4.m20: -25.6003;
5.m27: -26.5512;

6.m24: -27.3519;
7.m25: -27.3554;
8.m28: -27.8941;
9.m21: -28.1854;
10.m22: -28.3438;

26.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**026M1B10.fpg:**
1.**m26**: -22.6977;
2.m23: -25.621;
3.m29: -25.9052;
4.m20: -25.906;
5.m27: -26.9697;
6.m25: -27.6802;
7.m24: -27.8821;
8.m28: -28.4644;
9.m21: -28.7056;
10.m22: -28.934;

27.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**026M1C09.fpg**:
1.**m26**: -24.7049;
2.m23: -26.5249;
3.m29: -26.8183;
4.m20: -27.68;
5.m27: -27.934;
6.m25: -28.0773;
7.m24: -28.1774;
8.m22: -28.522;

9.m21: -28.5413;
10.m28: -28.7523;

28.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**026M1C10.fpg**:
1.**m26**: -24.6805;
2.m23: -26.2436;
3.m29: -27.0011;
4.m20: -27.0451;
5.m27: -27.4576;
6.m25: -27.5376;
7.m24: -27.6828;
8.m22: -28.0874;
9.m28: -28.1416;
10.m21: -28.3145;

TEST N1 8 .- Dubitada atribuida a m 27.

29.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**027M1B09.fpg**:
1.**m27**: -21.8082;
2.m24: -24.9097;
3.m20: -25.2332;
4.m29: -25.4244;
5.m26: -25.9226;
6.m21: -26.9403;
7.m28: -26.9537;
8.m25: -27.0058;
9.m23: -27.4971;
10.m22: -28.0844;

30.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**027M1B10.fpg**:
1**.m27**: -22.2194;
2.m24: -24.9241;
3.m20: -25.8782;
4.m26: -26.061;
5.m29: -26.1196;
6.m25: -26.417;
7.m21: -26.76;
8.m28: -26.8781;
9.m23: -27.5543;
10.m22: -27.6871;

31.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**027M1C09.fpg:**

1.**m27**: -24.5349;
2.m24: -27.2792;
3.m20: -28.0838;
4.m29: -28.6887;
5.m26: -29.0238;
6.m21: -29.4028;
7.m28: -29.7274;
8.m25: -30.0752;
9.m22: -30.492;
10.m23: -30.5633;

32.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**027M1C10.fpg**:
1.**m27**: -23.8774;
2.m24: -25.8387;
3.m20: -27.3741;
4.m21: -27.7891;
5.m29: -27.8471;

6.m26: -28.4726;
7.m28: -28.5216;
8.m25: -28.5472;
9.m22: -29.6431;
10.m23: -29.7893;

TEST N1 9 .- Dubitada atribuida a m 28.

33.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**028M1B09.fpg:**
1.**m28**: -22.235;
2.m20: -25.9938;
3.m24: -26.6631;
4.m21: -26.9282;
5.m25: -26.9853;
6.m29: -27.0208;
7.m27: -27.1789;
8.m23: -27.2352;
9.m26: -27.4558;
10.m22: -29.6502;

34.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**028M1B10.fpg:**
1.**m28**: -22.6942;
2.m20: -26.927;
3.m27: -27.3596;
4.m24: -27.4346;
5.m21: -27.6428;
6.m25: -27.7829;
7.m29: -28.7401;

8.m23: -28.8492;
9.m26: -28.8688;
10.m22: -30.722;

35.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**028M1C09.fpg**:
1.**m28**: -24.4064;
2.m21: -27.4055;
3.m24: -27.9316;
4.m25: -28.1017;
5.m27: -28.2435;
6.m20: -28.7729;
7.m29: -29.1656;
8.m26: -30.1265;
9.m23: -30.458;
10.m22: -31.1283;

36.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**028M1C10.fpg:**
1.**m28**: -24.5851;
2.m21: -28.4339;
3.m25: -28.7095;
4.m24: -29.0494;
5.m27: -29.1442;
6.m20: -29.7168;
7.m29: -30.941;
8.m23: -31.0121;
9.m26: -31.081;
10.m22: -32.2887;

TEST N1 10 .- Dubitada atribuida a m 29.

37.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**029M1B09.fpg:**
1.**m29**: -22.6521;
2.m23: -24.1954;
3.m26: -26.1633;
4.m20: -26.5994;
5.m24: -27.4308;
6.m27: -27.5324;
7.m28: -28.155;
8.m25: -29.0541;
9.m22: -29.9074;
10.m21: -30.2165;

38.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**029M1B10.fpg**:
1.**m29**: -22.035;
2.m23: -23.9386;

3.m26: -25.5726;
4.m20: -26.0227;
5.m24: -26.1489;
6.m27: -26.2962;
7.m28: -27.0842;
8.m25: -28.053;
9.m22: -28.7176;
10.m21: -28.8724;

39.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**029M1C09.fpg**:
1.**m29**: -23.3069;
2.m23: -24.1615;
3.m26: -25.6279;
4.m20: -26.1834;
5.m28: -27.1189;

6.m27: -27.1417;
7.m24: -27.452;
8.m25: -27.6514;
9.m21: -28.496;
10.m22: -30.2059;

40.C:\Archivos de programa\IdentiVox 2000 Demo\\sesiones\SesionCD1\!**029M1C10.fpg:**
1**.m29**: -23.1381;
2.m23: -24.5267;
3.m26: -25.8685;
4.m20: -26.353;
5.m25: -26.7903;
6.m28: -26.8652;
7.m27: -27.2113;
8.m24: -27.274;
9.m21: -27.5446;
10.m22: -29.7094;

Los resultados de los diez tests de identificación establecen un claro vínculo en cada uno los casos, entre los diferentes ficheros de voz dubitada y el modelo indubitado al que se atribuyen los mismos. Por tanto, podemos afirmar que de los diez locutores objeto de estudio, nuestro candidato ha sido identificado por el sistema como el sujeto competidor con una voz más parecida a la de las grabaciones dubitadas señaladas como muestra problema.

IV.3.- CONCLUSIONES DE ANÁLISIS.

A la vista de los resultados alcanzados tras el desarrollo de los cinco enfoques de estudio, es fácil deducir que la conclusión de nuestro análisis comparativo se situará en el que denominábamos A*nivel de identificación*≅.

Pero)de qué manera se transfiere la apreciación de los resultados obtenidos, a un nivel de ubicación concreto dentro de la escala de probabilidad verbal?. Antes de nada, hemos de reconocer que la casi totalidad de circunstancias de análisis consideradas apuntaban claramente en una dirección de identificación. Las grabaciones dubitadas e indubitadas del locutor candidato presentaban unas referencias óptimas de calidad y cantidad, e incluso el perfil de los modelos utilizados en el estudio de reconocimiento automático, podría tranquilamente calificarse como propio de un locutor Aoveja≅.[Doddington, 1998]

)Qué hubiera acontecido de haber surgido algún tipo de eventualidad adversa a uno u otro nivel?. Probablemente, dos consecuencias inmediatas hubieran sido la imposibilidad de ejecutar el análisis automático y, con casi toda seguridad, la no consecución de un nivel de máxima certeza en la decisión.

Planteamos estas cuestiones con la intención de centrarnos en lo que suele ser la realidad de la casuística forense, una realidad en la que la presencia de diversos elementos de dificultad constituye una constante.

)De qué forma evaluaremos la existencia y entidad de tales elementos?.)A través de qué procedimiento será apreciada su incidencia en relación a los objetos y sistemas de análisis?

Sin lugar a dudas, el único procedimiento posible será aquel que posea un carácter científico que posibilite la inferencia estadística sobre esos objetos y sistemas de análisis. Para lograr tal finalidad, habremos de comenzar por una descripción de criterios y una cuantificación de los vectores de ponderación. O lo que es lo mismo, una definición de los objetos de estudio y una adjudicación de pesos identificativos en los mismos para así poder proyectarlos en una estructura jerarquizada. La consideración de la mayor o menor incidencia de los distintos parámetros, habrá de extrapolarse a una población determinada mediante técnicas de reconocimiento de patrones que, por un lado, nos permitan la agrupación de rasgos o tipologías y, por otro, la discriminación de las diferentes variables que separan unas poblaciones de otras.

Seguidamente, para proceder con un análisis fiable, se recurrirá a técnicas multivariantes de análisis de datos para elaborar modelos de incertidumbre. En este sentido, pueden utilizarse modelos de componentes principales basados en el análisis de correlación o de covarianza,

análisis de correspondencias, de correlación canónica, y también análisis de la regresión para así poder estudiar las asociaciones entre parámetros según la naturaleza de los mismos.

Las técnicas estadísticas aplicables al desarrollo de los referidos aspectos, variarán en función del criterio metodológico utilizado. Desde hace unos años, y especialmente en el ámbito del reconocimiento automático, la interpretación de la inferencia bayesiana (asociada al cálculo de probabilidades de sucesos sobre informaciones conocidas) pretende traducirse en una definición de la probabilidad como una medida de la verosimilitud de la ocurrencia [Champod y Meuwly, 2000]. De hecho, el prototipo automático utilizado en nuestro caso práctico presentará próximamente una nueva versión denominada A*IdentiVox-LR*≅, la cual permitirá expresar los resultados de comparación en forma de A*likelihood ratios*≅.

Sin embargo, uno de los principales problemas a los que se enfrentan los sistemas de reconocimiento automático al enmarcar sus criterios de conclusión mediante técnicas bayesianas, es la insoslayable necesidad de contar con una población de referencia. Construir un modelo universal representativo de la plena diversidad de una comunidad lingüística, es una labor de altísima dificultad (por ejemplo, una base de datos de locutores que represente a la totalidad de los hablantes de Madrid). Además, dicha labor se torna aun más compleja, si tenemos en cuenta dos circunstancias añadidas. Por una parte, el Asimple≅ hecho de que todo locutor, en sí mismo, constituye una población. Y por otra, la existencia de numerosos factores que también habrían de considerarse para conjugar otras múltiples vicisitudes propias del audio forense: factor canal, ruido, etc, .

Las observaciones realizadas sobre reglas de conclusión en entornos automáticos, son perfectamente extrapolables al resto de aproximaciones de análisis. De ahí, la conveniencia del uso de escalas de probabilidad verbal que faciliten una más sencilla ubicación de los diferentes grados de certeza alcanzados por los expertos. No obstante, el hecho cuestionar la aplicación sistemática de técnicas bayesianas en nuestro marco de estudio, no será nunca una circunstancia excluyente para la utilización de otro tipo de estimaciones estadísticas que, evidentemente, también habrán de fundamentarse en los presupuestos anteriormente comentados.

En capítulos precedentes (*ver III.1.2 y III.1.3*) realizamos una detallada exposición, tanto de los que serían nuestros objetos de estudio, como de aquellas variables de referencia que dimensionarían su evaluación: márgenes de admisibilidad de muestras, contextos, tareas y resultados de comparación, peso identificativo, etc. . A través de dichas instancias hemos canalizado los resultados de nuestro análisis hasta desembocar en el llamado *nivel de identificación*. Sin embargo, algunos pormenores del proceso de evaluación como las referencias

de categorización o los coeficientes de ponderación, no han sido especificados. La justificación de esta carencia tiene una sencilla explicación. Como ya sabemos, la adjudicación de pesos u otro tipo de coeficientes ha de estar argumentada en unas bases empíricas las cuales, en nuestro caso, poseen un carácter reservado por estar vinculadas a la casuística del Laboratorio de Acústica Forense de la Dirección General de la Policía.

No obstante, estimamos que los elementos y características aportados, tanto en relación con los sistemas de análisis como con los procedimientos de evaluación, pueden considerarse suficientemente orientativos y clarificadores.

CAPÍTULO V

CONCLUSIONES GENERALES
Y
LÍNEAS DE TRABAJO FUTURO

V.1.- CONCLUSIONES GENERALES.

Dificultad del entorno de investigación forense.

Para evaluar de forma objetiva las prestaciones reales de una técnica forense de identificación, no existe otra alternativa que considerar la totalidad de sus circunstancias de contexto: marco legal del país donde se desarrolla, experiencia y capacitación de los expertos que la practican, objetos y sistemas de análisis que la integran, clases de tareas y tests elaborados, procedimientos de estimación de datos utilizados, organismos a los que se dirigirá el trabajo generado, garantías de control de calidad y cadena de custodia contempladas, etc, . Sólo estaremos en disposición de emitir opiniones en torno a la fiabilidad, eficacia o viabilidad de dicha técnica, cuando conozcamos en detalle las características de los citados ejes de referencia.

Si atendemos a la trascendencia y posibles repercusiones de las conclusiones de estudio, convendremos en la necesidad de establecer una clara diferenciación entre las técnicas de

identificación de hablantes en un ámbito de aplicación no forense, y aquellas otras que han constituido el núcleo del presente trabajo de investigación. A diferencia de lo que pudiera acontecer en otro marco de investigación en el caso de producirse una estimación equivocada, las consecuencias de un error en el ámbito forense pueden llegar a traducirse en una lamentable lesión de los derechos fundamentales de la persona.

Además de tener que desarrollar su trabajo bajo el peso de esta gran responsabilidad, los expertos en identificación forense de locutores (I.F.L.) han de superar muy diversos y complejos obstáculos. Algunos de ellos, relacionados con su etapa de formación: largos períodos de entrenamiento, carácter multidisciplinar de la técnica, habilidades auditivas y perceptivas, etc.; y otros, relativos a las propias características de los elementos de estudio: naturaleza variable de

las emisiones de voz, factores de degradación de la señal, carácter no cooperativo de los sujetos analizados, etc. A estas dificultades, habrán de añadirse otra serie de inconvenientes vinculados a la versatilidad de las aproximaciones de análisis, ausencia de estándares de referencia tanto en el propio contexto técnico como en el plano judicial, intereses institucionales enfrentados, desarrollo de investigaciones específicas, etc.

Mejor filosofía metodológica: los Amétodos combinados≅.

Desde sus comienzos en los años sesenta, la I.F.L. ha sido desarrollada a través de distintas propuestas metodológicas, como consecuencia de su carácter multidisciplinar. En algunas ocasiones, esta peculiaridad se ha visto acompañada de una falta de ética o profesionalidad por parte de determinados expertos. En otras, ha resultado afectada por conveniencias de tipo político o diferencias de criterio a nivel judicial. La conjugación de estos delicados elementos, ha deparado numerosos trastornos en el camino de consolidación de la técnica.

Afortunadamente, en el momento actual, la filosofía de los A*métodos combinados*≅ ha sido señalada en distintos foros internacionales de expertos como la opción más idónea para la práctica forense de identificación de voz. Dicha filosofía, parte de la utilización del análisis clásico (perceptivo-acústico-fonético) contemplando la combinación de cualquier otro enfoque de análisis que pueda aportar luz sobre el problema.

Aunque en casos muy concretos pudiera alcanzarse un resultado exitoso a través de una única perspectiva de estudio, el experto forense no ha de soslayar ninguna de las herramientas o enfoques que puedan otorgar a sus apreciaciones el mayor grado de precisión y objetividad.

Incidiendo en este sentido, hemos presentado una nueva alternativa teórica para el análisis perceptivo-auditivo, que hemos dado en llamar AA.P.R.E.S.≅ (*Aural Perception o Reverse Speech*). Su principal aportación se refiere al hecho de objetivar el proceso de percepción de la cualidad vocal, mediante técnicas que eliminan la influencia de otras informaciones inherentes a la propia estructura acústica de la señal de habla.

Presentación de un modelo combinado: análisis clásico y SARL basado en GMM.

Por vez primera se ha diseñado un modelo combinado que incluye un prototipo de reconocimiento automático basado en modelado por mezclas de gausianas. Este nuevo modelo se ha aplicado en la resolución práctica de un supuesto de investigación forense en el que se

planteaban una tarea de verificación y otra de identificación. Cada una de las fases de estudio han sido desarrolladas en detalle, especificando las diferentes opciones de análisis y circunstancias de contexto consideradas. Para ello, se ha tomado como referencia la propuesta de definición y evaluación de los objetos de estudio que previamente había sido elaborada como otra de las aportaciones científicas del presente trabajo. Dicha propuesta, ha sido complementada con una orientación general sobre los procedimientos de interpretación de resultados y formulación de conclusiones. A este respecto, la inferencia de diversas técnicas estadísticas en tales procesos, se plantea como algo fundamental.

De forma adicional a la propia ejecución de las distintas etapas de análisis, se ha efectuado una completa reseña de otras tareas accesorias, aunque igualmente necesarias: protocolos de toma de muestras y evaluación de calidad, adecuación de equipos y opciones de análisis, etc.

Síntesis de conclusiones.

Como corolario a los razonamientos arriba expresados, deseamos subrayar las siguientes ideas fundamentales:

- La filosofía de los métodos combinados proporciona una gran versatilidad de alternativas metodológicas perfectamente compatibles entre sí. Cada una de ellas, deberá ser desarrollada por unos expertos debidamente acreditados, habrá de sustentarse en unas bases científicas sólidas, y en unos sistemas de análisis y procedimientos de evaluación del máximo rigor.

- La futura estructura que pudiera presentar un modelo combinado estándar, podría corresponderse con la de una aplicación semiautomática de carácter abierto, en la que pudiera optarse por uno u otro tipo de enfoque de análisis y, consiguientemente, por una u otra fórmula de interpretación estadística de los resultados de estudio.

- El experto en identificación forense de locutores, además de guardar el más riguroso código de ética, habrá de estar convenientemente cualificado y específicamente entrenado ya que, hoy en día, su labor en distintas fases del proceso sigue siendo crucial.

- La práctica de cualquier técnica de I.F.L. de cara a la emisión de informes periciales para las instituciones de justicia, implicará necesariamente una precisa descripción de las tareas, aplicaciones, opciones y sistemas de análisis utilizados, así como de las correspondientes limitaciones que dichos sistemas puedan presentar en cada caso.

V.2.- LÍNEAS DE TRABAJO FUTURO.-

Las principales directrices de trabajo futuro relacionadas con nuestra técnica de investigación, pueden sintetizarse en las siguientes referencias:

Investigación y desarrollo:

- Creación de bases de datos de locutores que contemplen factores característicos de entornos forenses reales, para la construcción de modelos globales de referencia.

- Diseño de experimentos de investigación a partir de bases de datos de locutores representativas, para:

  - obtener valores normativos del habla en poblaciones concretas,

  - testeo de nuevas alternativas de análisis automático o semiautomático,

  - aplicación de técnicas estadísticas de cara a la interpretación de resultados de análisis y creación de protocolos de decisión: definición de criterios, cuantificación de vectores de ponderación, reconocimiento de patrones, etc.,

Área internacional.

- Continuación con las tareas de estandarización en los grupos de trabajo de expertos. Como acciones inmediatas, pueden plantearse: la exploración de diferentes modelos metodológicos, estado de laboratorios (control de calidad, cadena de custodia, etc), nivel de cualificación de expertos, estudio de protocolos de análisis y de elaboración de informes, etc. En una segunda etapa, se procedería a un proceso de normalización mediante la adopción de estándares o la validación de diferentes técnicas de identificación.

Ámbito judicial y formación de expertos

- Aproximación de la técnica forense de identificación de locutores a las distintas instancias del ámbito judicial (especialmente a jueces y fiscales).

- Reclamar de dichas instancias una unicidad de criterios traducida en una definición formal de estándares para la apreciación de éste u otro tipo de evidencia científica.

- Creación de centros o programas de formación para el entrenamiento, cualificación y actualización específica de los expertos en ciencias forenses.

REFERENCIAS BIBLIOGRÁFICAS

- [Abe, I., 1955] Intonational patterns of English and Japanese. Word, 11, pp. 368-398.
- [Abeles, M. y Goldstein, M. H., 1970] Funtional architecture in cat primari auditory cortex: columnar organization and organization according to depth. *Journal of Neurophysiology 33,* pp. 172-187.
- [Alarcos, E., 1974] *Fonología española*, Madrid: Gredos.
- [Alvar, M. y Quilis A., 1966] Datos acústicos y geográficos sobre la □ch□ adherente de Canarias. *Anuario de Estudios Atlánticos*, 12, pp. 337-343.
- [Aerospace Corporation, 1977] *Speaker identification.Program 7007 Final report, nı ATR-77 (7617-07)-1*. Aerospace C., El Segundo,CA.
- [Arheim R., 1.974] *Art and visual perception*, Ẽ2 ed.Berkeley : University of California Press.
- [Artemov, 1965] Ob intonema (sobre el entonema). *Phonética, 12.* Pp. 130.
- [Atal, B., 1972] Automatic speaker recognition based on pitch contour. *Journal of the Acoustical Society of America,*(*JASA*) 52: 1687-1697.
- [Atal, B., 1974] Effectiveness of linear prediction characteristics of speech wave for automatic speaker identification and verification. *Journal of the Acoustical Society of America*, 55(6) : 1304-1312. - Booth Davis, J. (1.978) The Psychology of music. London. Hutchinson.
- [Audiencia Provincial de Valencia, 1991] *Sentencia Nı 54 de 25 de febrero*. *Sección Ĩ*, Rollo 91/90, P.A. 260/90, J.Instrucción nı 2 de Valencia.
- [Békésy, G. von, 1960] *Experiments in hearing*. Nueva York: Mc Graw-Hill.
- [Becker, R.W., Clark F.R., Poza, F. Y Young, R.J., 1972] A Semiautomatic Speaker Recognition System. *Standford Research Institute Report ñ 1363*, Standford, CA.

- [Black, J. W., 1937] The quality of a spoken vowel. *Archives of speech 2*, pp.7-27.
- [Black, J. W., Lashbrook, W., Nash, W., Oyer, H.J., Pedrey, C., Tosi, O. y Truby, H., 1973] Reply to Speaker Identification by Speech Spectrograms: Some Further Observations, *JASA, 54*, pp. 535-537.
- [Böhme, G. y Hecker, G., 1970] Gerontologische Untersuchungen über Stimm-umfang und Sprechstimmlage. *Folia Phonial 22,* pp.176-184.
- [Bolinger, D.L., 1955] The Melody of Language, *Modern Language Form*, 40, p. 20.
- [Bolt, R.H., Cooper, F.S., David, E.C., Denes, P.B., Pickett, J.M. y Stevens, K.N., 1970] Speaker Identification by Speech Spectrograms, *Journal of the Acoustical Society of America*, 47, pp. 597-613.

- [Bolt, R.H., Cooper, F.S., David, E.C., Denes, P.B., Pickett, J.M. y Stevens, K.N., 1973] Speaker Identification by Speech Spectrograms: some further observations, *Journal of the Acoustical Society of America*, 54, pp. 531-534.
- [Bolt, R.H., Cooper, F.S., Green, D.M., Hamlet, S.L., Hogan, D.L., McKnight, J.G., Pickett, J.M., Tosi, O. y Undergood, B.D., 1979] *On the theory and practice of Voice Identification*, Washington, DC, National Academy of Sciences.
- [Booth, J., 1978] *The psychology of Music*, Hutchinson, London.
- [Braun A., 1.996] Age estimation by different listeners groups, *Fo. Linguistics,* 3: pp.65-73.
- [Braun A., 1.995] Fundamental frecuency, how speaker-specific is it?. *BEIPHOL 64: Studies in Forensic Phonetics,* p. 15
- [Bregman A. y Campbell J., 1.971] Primary Auditory stream segregation and perception of order in rapid sequences of tones, *Journal of Experimental Psychology,* 89: 244-249.
- [Bregman A., 1.990] *Auditory Scene Analysis,* MIT Press, Cambridge, MA.
- [Broeders A. y Rietveld A., 1.995] Speaker identification by earwitnesses, en A. Braun and P.Köster (eds.) *Studies in Fo. Phonetics,* Trier: Wissenschafterlicher Verlag,64: 24-40.
- [Broeders A., 1.996] Earwitness identification: common ground, disputed territory and uncharted areas, *Forensic Lingüistics,*3: 3-13.
- [Brown R., 1.979] Memory and decision in speaker recognition, *International Journal for Man-Machine Studies,* 11, 729-742.
- [Bunge, E., 1977] Automatic Speaker Recognition System Auros for Security Systems and Forensic Voice Identification. *Proced. International Conference Crime Countermeas,* Oxford, UK, 1-8.
- [Canfield, D.L., 1962] *La pronunciación del español en América. Bogotá*: Instituto Caro y Cuervo.
- [CAVIS project, 1985] , *proposal, Los angeles County Sheriff's Department*, Los Angeles CA.
- [Clifford B R, 1.980] Voice identification by human listeners: on earwitness reliability, *LawHuman Behavior,* 4, 373-394.

- [Champod y Meuwly, 2000] The inference of identity in forensic speaker recognition. *Speech Communication, Vol. 31,* pp.193-203.
- [Chlumský., 1956] La -s andaluza y la suerte de la -s indoeuropea en eslavo. *Publicaciones del ALEA. Tomo III, nı2*, Granada.
- [Cornut, G. y Lafón, J.C., 1960] Vibrations neuro-musculaires des cordes vocales et théories de la phonation. *J.F., O.R.L., IX*, 3, pp. 317-324.
- [Crystal, D., 1968] *What is linguistics?* London, Edward Arnold Ltd.
- [Culler, E.A., Coakley, J.D., Lowy, K y Gross, N., 1943] A revised frecuency-map of the guineapig cochlea. *American Journal of Psychology 56,* pp. 475-500

- [Daubert v. Merrel Dow Pharmaceuticals, 1993] *113 S. Ct. 2786*, 1993
- [Dayhoff, J.E., 1990] *Neural Network Architectures*, Van Nostrand Reinhold, New York.
- [Dejonckère, P.H., 1987] Physiologie phonatoire du larynx: le concept oscilloimpèdantiel. *Revue du laryng.*, Burdeos, 108, pp. 365-368.
- [Delattre, P., 1958] Les indices acoustiques de la parole. *Phonética, 2,* pp. 108-118.
- [Delattre, P., 1962] Le jeu des transitions de formants et la percepction des consonnes. *Proceedings of the 4th International Congress of Phonetic Sciences, Helsinki, 1961*. The Hague: Mouton.
- [Denes, P., 1959] *A preliminary Investigation of certain aspects of intonation*, p.106
- [De Pinto, O. y Hollien, H., 1982] Speaking fundamental frecuency characteristics of Australian women: then and now. *Journal of Phonetics 10*, pp. 367-375.
- [Diamond, I. y Neff, W., 1957] Ablation of temporal cortex and discrimination of auditory patterns. *Journal of Neurophysiology, 20*, pp. 300-315.
- [Documento ENFOPOL 80 (8943/95), 1995] *Documento de la Delegación Española al Grupo de Policía Técnica y Científica*. Director II. Consejo de la Unión Europea.
- [Documento ENFOPOL 144 (11641/95), 1995] *Documento de la Delegación Española al Grupo de Policía Técnica y Científica*. Director II. Consejo de la Unión Europea.
- [Doddington, G.R. (1985)] Speaker Recognition- Identifying People by their voices, *Proccd. IEEE, Vol. 73, nı 11, pp. 1651-1664.*
- [Doddington et al., 2000] "The NIST speaker recognition evaluation: overview, methodology, systems, results, perspective". *Speech Communication, Vol. 31,* pp.225-254.
- [Duffy, R.J., 1970] Fundamental frecuency characteristics of adolescent females. *Language and Speech 13*, pp. 14-24.
- [Durrant, J. y Jovrinic, J., 1977] *Bases of hearing science*. Baltimore: Williams and Williams.
- [Eimas, P. y Corbit, J., 1973] Selective adaptation of linguistic features detectors. *Cognitive Psychology 4*, pp. 99-109.

- [Elaad E., Segev S. y Tobin Y., 1998] Long-term Working Memory in Voice Identification, *Psycholog, Crime & Law, Vol.0,* pp. 1-16
- [Endress, W., Bambach, W. y Flösser, G., 1971] Voice spectrograms as a function of age, voice disguise and voice imitation. *JASA 49*, 1842-1848.
- [Falcone M.De Sario N., 1994] A PC based speaker identification system for forensic use : IDEM, *Proceedings of the ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*, Martigny, pp. 169-172.
- [Frye v. U.S., 1923] *293 F 1013* (D.C. Ct. App. 1923)
- [Fechner, G.T. , 1860] *Elementen der Psychophysik*. Leipzig: Breitkopf & Härtel, 1860.
- [Furui S., 1979] New Techniques for automatic Speaker Identification Using Telephone Speech. *Journal of the Acoustical Society of America (abstract)* Suppl. 1, 66: S35.

- [Furui S., 1981] Cepstral Analysis Technique for Automatic Speaker Verification. IEEE Trans. Acous. *Speech Signal Processing, Vol 29, nı 2* pp. 254-272.
- [Furui S., 1991] Speaker independent and Speaker Adaptive Recognition Techniques in *Advances in Speech Signal Processing* eds. S.Furui y MM Sondhi), Marcel Dekker, New York, pp. 597-622.
- [Furui S., 1994] An overview of Speaker Recognition Technology. *ESCA Workshop on Automatic Speaker Recognition, pp. 1-10.*
- [García, M., Ledesma, O., Martínez, E., Ortega, J. y González, J., 2000] □IdentiVox 2000": Una herramienta bajo □Windows□ para reconocimiento automático de locutores independiente de texto en aplicaciones forenses. *Actas del I Congreso de la SEAF*, pp.149-154.
- [Garrett,K.L.y Healey,E., 1987] An acoustic analysis of fluctuations in the voices of normal adults speakers across three times of day. *JASA 82*, pp. 58-62.
- [Gersho, A. y Gray R.M., 1991] Vector Quantization and Signal Compression, *Kluwer Academic Publishers*.
- [Gili Gaya, S., 1921] La □r□ simple en la pronunciación española, RFE, VIII, pp. 271-280.
- [Goldstein, E., 1984] *Sensation and Perception*. Wadsworth Publishing Company, Belmont, California
- [González, J., 1999] *Influencia y compensación del entorno acústico en sistemas de reconocimiento automático de locutores*, Tesis doctoral, ETSI.Telecomunicación, Universidad Politécnica de Madrid.
- [Gorban, II, 1997] Crime Automatic Speaker Verification and Identification (CASVI) system. *Proceedings of the 134th Meeting of the Acoustical Society of America.* 1p.
- [Gray, C.H. y Kopp, G.A., 1944] Voiceprint Identification. *Bell Telephone Laboratories Report,* pp. 1, 3, 13, 14, Bell Laboratories.
- [Greenwald,M.,1.979] The effects of decreased frecuency bandwidth on speker identification

by aural and spectrographic examination of speech samples. *Master Thesis*, Michigan State University, 1.979.

- [Hair, G. y Rekieta, T., 1972] Speaker identification final report.*Standf.Research Institute Report ñ 1363, Standford,CA*.
- [Hala, B., 1973] *La sílaba*. Madrid, 2̃ ed., C.S.I.C.
- [Hall, M.C., 1975] *Spectrographic Analysis of Interspeker and Intraspeaker variables of Profesional Mimicry*, Master Thesis, Michigan State University.
- [Hartmann D., 1.979] The perceptual identity and characteristics of aging in normal male adult speakers, *Journal of Communication Disorders*,12 : 53-61.

- [Hammersley R. y Read J., 1.983] Testing witnesses voice recognition: some practical recommendations, *Journal of the Forensic Sciencie Society,* 23: 203-208.
- [Hazen, B.M., 1973] Effects of different phonetics contexts on spectrographic Speaker Identification, *JASA, 54,* pp.650-660.
- [Helfrich, H., 1979] *Age markers in speech.* Sherer/Giles (eds.), 63-107.
- [Helmholtz, H.von, 1863] *On the sensations of tone as a psychological basis for the theory of music.* Nueva York: Dover.
- [Hennessy, J.J., 1970] *An analog of Voicprint identification*, Unpublished M.A. Thesis, Michigan State University.
- [Hirano, M., 1977] Structure and vibratory behavior of the vocal folds. *Dynamic aspects of speech production.* Tokio, University of Tokio Press.
- [Hjelmslev, L., 1966] *Le Langage*, Paris, Les Editions de Minuit.
- [Hollien H., 1977] Status report on □Voiceprint□ identification in the United States *Proceedings of the Carnahan Crime Countermeasures Conference*, Oxford U.K. pp.9-20.
- [Hollien H., 1.990] *The Acoustics of Crime,* New York: Plenum Press.
- [Hollien H., 1.996] Considerations of guidelines for earwitness lineups, *Forensic Linguistics,* 3: 14-23.
- [Hollien H. y Shipp, Th., 1972] Speaking fundamental frecuency and chronologic age in males. *JSHR 15*, pp. 155-159.
- [Hollien H. y McGlone, 1976] The effect of Disguise on Voiceprint Identification. *Proceedings of the Carnahan Crime Countermeasures Conference, University of Kentucky Press,* Lexington, KY, 1976.
- [Hollien, H., Childers, D.G. y Doherty, E.T., 1977] Semi-automatic Speaker Identification System (SAUSI), *Proccedings, IEEE, ICASSP* 26: 768-771.
- [Hollien H. y Majewsky W., 1977] Speaker Identification by long term spectra, under normal and distorted speech conditions. *JASA* 62 (4): 975- 980.
- [Hollien H., Majewsky W. y Doherty E.T., 1.982] Perceptual identification of voices under

normal, stress and disguised speaking conditions, *Journal of Phonetics,*10, 139-149.
- [Hollien H., Huntley R., Künzel H. y Hollien P., 1.995] Criteria for earwitnesses lineups, *Forensic Linguistics,* 2 : 143-153.
- [Houlihan, K., 1979] The effects of disguise on Speker Identification from Sound Spectrograms, *Current Issues in the Phonetic Sciences,* Amsterdam, J.Benjamins, B.V., pp. 811-820.
- [Hudson, A.I. y Holbrook, A. 1981] A study of the reading fundamental vocal frecuency of youngs black adults. *JSHR, 24,* pp. 155-159.
- [Huntley R.A., 1.992] Listener skill in voice identification, *presentación efectuada en la reunión de la American Academy of Forensic Sciences en Boston*, Massachussetts.

- [Huntley R.A. y Pass K., 1.993] Assesment of voice line-up procedures, *Program and abstracts of the 45th Annual Meeting of the American Academy of Forensic Sciences*,Colorado Springs, p. 102.
- [Husson, R., 1950] *Études des phonèmes physiologiques et acoustiques fondamentaux de la voix chantée*. *Thèse Sciences*, París.
- [I.A.F.P., sin fecha] Code of practice (International Association of Forensic Phonetics). Punto 6, apartados A y B.
- [Jones, D. , 1909] *Intonation curves*, Leipzig-Berlín.
- [Jackson-Menaldi, 1992] *La voz normal*. Ed. médica panamericana, Buenos Aires.
- [Jakobson, R., 1963] *Essais de linguistique générale*. París, Ed. De Minuit.
- [Jakobson, R., Fant, C., Gunnar, M. y Halle M., 1952] *Preliminaries to Speech Analysis. The Distinctive Features and Their Correlates*. Massachussets.
- [Kersta., L.G., 1962] Voiceprint Identification, *Nature* 196: 1253-1257.
- [King, R.A. y Phipps, T.C., 1999] TESPAR, a powerful new voice biometric for forensic applications. Paper for the First International Conference on Forensic Human Identification in The Millenium. London. (The Forensic Science Service).
- [Klevans R. y Rodman R., 1997] *Voice Recognition*, Artech House, Boston, London.
- [Koenig B.E., 1.986] Spectrographic voice identification: a forensic survey, Letter to the editor, *Journal of Acoustical Society of America, 79 :2088-2090.*
- [Kopelev, L., 1991] *Slake my sads*. Memoirs, Moskow: □Slovo□, en ruso. 1991.
- [Köster, P., 1.981] Auditive Sprechererkenung bei Experten und Naiven, en *Festschrift Wrangler,* Hambury: Helmut Buske, AG, 52: 171-80.
- [Koval, S., Kaganov, A. y Khitrov, M., 2000] *The chart of the standard expert actions and decision-making principles of forensic speaker identification*. Sin referencias de publicación.
- [Koval, S. y Krynov, S., 2000] *Practice of usage of spectral analysis for forensic speaker identification*.Sin referencias de publicación.

- [Kreiman J, Gerratt B. y Precoda K., 1.990] Listener experience and perception of voice quality, *Journal of Speech and Hearing Research,* 3: 103-115.
- [Krook, M., 1988] Speaking fundamental frecuency characteristics of normal Swedish subjects obtained by glottal frecuency analysis. *Folia Phoniat*, 40, pp. 82-90.
- [Künzel H., 1.994] On the problem of speaker identification by victims and witnesses, *Forensic Linguistics,* 1: 45-58.
- [Ladefodged P., 1.978] Expectation affects identification by listening, *UCLA Working Papers in Phonetics,* 41, pp. 41-12.

- [Ladefodged P. y Ladefodged J., 1.980] The ability of listeners to identify voices, *UCLA Working Papers in Phonetics,* 49, 43-51.
- [Lehiste, I., 1970] *Suprasegmentals.* Cambridge: The MIT Press, p. 95
- [Le Huche y Allali, 1993] *La Voz. Vol. 1: Anatomía y fisiología de la Laringe*. Ed. Massón.
- [Lenz, R., 1940] Estudios chilenos, *Bibliot. de Dialectol. Hispanoamericana, VI*, pp. 87-258.
- [León, P.R., 1970] Systématique des fonctions expressives de l intonation. *Studia Phonetica*, *Montreal,* Paris, Didier. Pp. 149-155
- [León, P.R., 1972] Patrons expressifs de l intonation. *Sympos. on Intonology, Praga*, pp. 57-74
- [Li, K. y Hughes, G., 1974] Talker differences as they appear in correlation matrices of continuous speech spectra *Journal of the Acoustical Society of America* 55: 883-887.
- [Liberman, A.M., Delattre, P., Cooper F.S. y Gerstman, L., 1954] The role of consonants-vowel transitions in the perception of the stop and nasal consonants .*Psychological Monographs, 379*, pp. 1-14.
- [Liberman,A., Cooper, F., Shankweiler, D. y Studdert-Kennedy, M., 1967] Perception of speech code. *Psychological Review 74,* pp. 431-461.
- [Lipeika A., Lipeikiene J, 1993] The use of pseudostationary segments for speaker identification. In: *Proceedings of the 3rd European Conference on Speech Communication and Technology.* Berlin, pp: 2303-2306.
- [Lipeika A., y Lipeikiene J, 1995] Speaker identification using vector quantization. *Informatica, 1995;* 6(2) : 167-180.
- [Lipeika A., Lipeikiene J. y Salna B., 1997] On usefulness of the LPC residue in speaker identification. *Proceedings of the international conference Biomedical Engineering,* Kaunas 1997: 61-64.
- [Lippmann, R.P., 1987] An introduction to Neural Nets, *IEEE Trans. on ASSP, Abril*. pp. 131-142.
- [Lucena, J.J. y Díaz, J.J., 2000] Evaluación del sistema de reconocimiento automático IdentiVox 2000, con la base de datos Ahumada. *Actas del I Congreso de la SEAF*,

pp.183-191.
- [Malmberg, B., 1955] *Studia Linguistica IX*, pp. 80-87
- [Malmberg, B., 1965] *Estudios de Fonética Hispánica*. Madrid: C.S.I.C.
- [Marescal, F., 1999] The Forensic Speaker Recognition Method used in the French Gendarmerie.*Paper presented in an experts meeting on Forensic Acoustics supported by the PCWG (Council of the European Union) celebrated in Wiesbaden,Germany.*
- [Matsui T. y Furui S., 1991] A text independent Speaker Recognition Method Robust Against Utterance Variations. *Proceed. IEEE International Conference, Acoust. Speech, Signal Processing*, S6.3, pp. 377-380.

- [Matsui T. y Furui S., 1992] Comparison of Text independent Speaker Recognition Methods Using V-Q- Distorsion and Discrete/Continuous HMMs. *Proceed. IEEE International Conference Acous. Speech Signal Processing*, II, pp. 157-160.
- [McCormick, 1984] Handbook of the Law of Evidence (USA), section 203 a 608. 3̃ ed.
- [McGehee F., 1.937] The reliability of the identification of the human voice, *Journal of General Psychology,* 17, 249-271.
- [McGehee F., 1.944] An experimental study in voice recognition, *Journal of General Psychology,* 31, 53-65.
- [McGlone, R.E. y Hollien, H. , 1963] Vocal Pitch characteristics of old women. *JSHR 6,* pp.164-170
- [McGurk, H. y Mac Donald, J., 1976] Hearing lips and seeing voices. *Nature 264,* pp.746-748.
- [Meuwly D., El-Maliki M.y Drygajlo (1998)] Forensic speaker recognition using gaussian mixture models and a bayesian framework.*RLA2C Workshop on speaker recognition by man and by machine: directions for forensic aplications,* Avignon., France.
- [Miller G. e Isard S., 1.963] Some perceptual consequences of linguistic rules, *Journal of Verbal Learning and verbal Behavior*, 2: 212-228.
- [Minnesota v. Constance Trimble] *Ramsey Co. Dist. Ct., St Paul, nı 24.049.*
- [Morris, R.J. y Brown, W.S. Jr., 1988] Aged related differences in F_0 and pitch sigma among females. *The IASCP bulletin 1/2,* pp. 36-39.
- [Navarro Tomás T., 1918] Diferencias de duración entre las consonantes españolas, *RFE,V*, pp.385-386.
- [Navarro Tomás T., 1948] *Manual de entonación española*. New York.
- [Navarro Tomás T., 1961] *Manual de pronunciación española*. Madrid: C.S.I.C.
- [Neiman G. y Applegate J., 1.990] Accuracy of listener judgements of perceived age relative to chronological age in adults, *Folia Phoniatrica,* 42: 327-330.
- [Neisser U., 1.981] *Procesos Cognitivos y Realidad*, Marova, Madrid.
- [Nolan F., 1.990] The limitations of auditory-phonetic speaker identification, en H. Kniffka

(ed), *Texte zu Theorie und Praxis Forens. Linguistik,* Niemeyer: Tübingen, 457-479.
- [Obrecht, D.H., 1975] Fingerprints and Voiceprint Identification*, Abstracts, Eight International Congress Phonetics Sciences*, Leeds, U.K., 215.
- [O´Connor, J., Gerstman, L., Liberman, P., Delattre, P. y Cooper F., 1957] Acoustic Cues for the perception of initial /w, j, l, r/ in English. *Word, 13*, pp. 24-43.
- [Ortega, J., 1996] *Técnicas de mejora de voz aplicadas a sistemas de reconocimiento de locutores.* Tesis doctoral, ETSI.Telecomunicación, Universidad Politécnica de Madrid.
- [Ortega, J. y González, J., 1995] Estudio comparativo de técnicas de identificación automática de locutores. X Symposium nacional de la Unión Científica Internacional de Radio URSI- 95, Valladolid.

- [O□Shaugnessy D., 1986] Speaker Recognition. *IEEE ASSP Magazine, 3,* nı 4, pp. 4-17.
- [Papcun G., Kreiman J, y Davis A., 1.989] Long-term memory for unfamiliar voices, *Journal of the Acoustical Society of America,* 85, 913-925.
- [People v. King] *case 13588, App. 2nd Div. 2*; Los Angeles, CA.
- [Perelló, J., 1962] *La théorie muco-ondulatoire de la Phonatión*. Ann. Otolárynx, 79, pp. 722-725.
- [Perelló, J. 1977] *Lexicón de Comunicología*. Ed. Augusta, Barcelona.
- [Perelló y Salvá, 1980] *Alteraciones de laVoz*. Ed. Científico-médica, Barcelona. Ẽ Ed.
- [Pollack I. y Pickett J., 1964] The intelligibility of excerpts from conversational speech, *Language and Speech,* 6: 165-171.
- [Pollack I. y Pickett J., 1964] Intelligibility of excerpts from fluent speech: Auditory vs structural context,*Journal of Verbal Learning and Verbal Behavior,* 3: 79-84.
- [Popov H.F., Linkov,A.N., Baicharov M.V., 1996] Personal identification by Russian speech phonograms on the automatized system DIALECT: *Manual for experts. In: Fesenko A.F. (Ed) Military unit nı 34435. Moscú.*
- [Potter, R., Kopp, G. y Green, H., 1947] *Visible Speech*. Van Nostrand, New York.
- [Prater y Swift, 1986] *Manual de terapéutica de la voz*. Masson -Salvat. Traducción J.Perelló.
- [Pruzansky, S., 1966] Pattern matching procedure for automatic talker recognition. *Journal of the Acoustical Society of America* 35: 354-358.
- [Quilis, A., 1963] *Fonética y fonología del español*. , Madrid, C.S.I.C.
- [Quilis, A., 1970] El elemento esvarabático en los grupos [pr, br, tr,..] Phonétique et Linguistique Romane (Melanges G. Straka) I, pp. 99-104.
- [Quilis, A. y Fernández, J.A., 1975] *Curso de Fonética y fonologías españolas*. Madrid, C.S.I.C. Ẽ ed.
- [Quilis, A., 1981] *Fonetica Acústica de la Lengua Española,* Madrid, Gredos.
- [Reddy R. 1.976] Speech recognition by machine: A rewiew. *Proceedings of the Institute of Electrical and Electronic Engineers,* 64: 501-531.

- [Reich, A. y Moll, K. y Curtis J.F., 1.976] Effects of selected vocal disguises upon Spectrographic Speaker Identification, *JASA, 60*: 919-925.
- [Reich, A. y Duke 1.979] Effects of selected vocal disguises upon speaker identification by listening, *Journal of Acoustical Society of America,* 66: 1023.
- [Reynolds D.A. 1992] A Gaussian Mixture Modeling approach to Text-Independent Speaker Identification. *Ph. D. Thesis, Georgia Institute of Technology*, 1992.
- [Reynolds D.A. 1994] Speaker Identification and Verification using Gaussian Mixture Speaker Models. *ESCA Workshop on Speaker Recognition*, pp. 27-30.
- [Robins, R.H., 1971] *Lingüística general. Estudio introductorio*. Madrid. Gredos.

- [Rodriguez, A., 1989] La construcción de una voz radiofónica. *Tesis doctoral*. CAVP, Uni. Autónoma Barcelona.
- [Rose, J., Brugge, J., Anderson, D. y Hind, J.E. , 1967] Phase locked response to low frecuency tones in single auditory nerve fibers of squirrel monkey. *Journal of Neurophysiology 30*, pp. 769-793.
- [Rosenberg A., 1.973] Listener performance in speaker verification tasks, *IEEE Transactions of Audio and Electroacoustics,* AU-21: 221-225
- [Rosenberg A. E. y Sambur, M.R., 1975] New techniques for automatic speaker verification, *IEEE Trans. Acoustics, Speech Signal Processing,* vol ASSP-23, pp. 169-176.
- [Rosenberg A. E. y Soong F.K., 1987] Evaluation of a Vector Quantization Talker Recognition System in Textindependent and text dependent modes. *Computer Speech and Language, 22, pp. 143-157*.
- [Rosenberg A. E. y Soong F.K., 1991] Recent research in automatic speaker recognition, in *Advances in Speech Signal Processing eds. S.Furui y MM Sondhi), Marcel Dekker, New York, pp. 701-737*
- [Rosetti, A., 1963] *Sur la théorie de la syllabe*. ̃2 ed., The Hague, Mouton.
- [Rothman H., 1.977] A Perceptual (Aural) and Spectrographic Identification of Talkers with Similar Sounding Voices, *Proceedings Internationa Conference, Crime Counter measures,* Oxford, UK, 37-42.
- [Rutherford, W. , 1886] A new theory of hearing. *Jou. of Anatomy and Physiology* 21, 166-168.
- [Saslove H. y Yarmey A., 1.980] Long-term auditory memory: speaker identification, *Journal of Applied Psychology,* 65: 111-116.
- [Sapir, E. , 1954] El lenguaje. Méjico, fondo de cultura económica. ̃1 ed. de 1921.
- [Saussure, F. 1955] *Curso de lingüística general*. Ed. Losada. Buenos Aires. Traducción de . Alonso.
- [Saxman, J.H. y Burk, K.W., 1967] Speaking fundamental frecuency characteristics of middle aged females. *Folia Phoniat, 19*, pp. 167-172.

- [Schafer, R.M., 1979] *Le paysage sonore*, París, J.C. Lattès.
- [Schubert, E., 1980] *Hearing: its function and dysfunction*. Viena: Springer-Verlag.
- [Shipp T., Qi Y., Huntley R. y Hollien H.(1.992)] Acoustic and temporal correlates of perceived age, *Journal of voice,* 6, pp. 211-216.
- [Schooneveld, C.H. van, 1961] *The Sentence Intonation of Contemporary Standard Russian as a Linguistic Structure*. S□- Gravenhage, p. 9
- [Silvestre, J.L. y Mac Leod, P., 1968] Le muscle vocal humain est-il asynchrone?. *Journal of Physiologie,* 5, pp. 373-389.
- [Sistema SVL, 1998] Sistema de verificación/identificación de locutores SVL. Documento presentado al laboratorio de Acústica Forense de la D.G.P.

- [Smrkovski L., 1975] Collaborative Study of Speaker Identification by the Voiceprint Method, *Journal of the AOAC, 48*, pp.453-456.
- [Smrkovski L., 1976] Study of speaker identification by aural and visual examination of non-contemporary speech samples, *Journal of Official Analytical Chemists* 59: 927-931.
- [Solzhenitsyn, A., 1968] *The first circle*. Traducción del ruso al inglés por T. Whitney. Harper & Row. New York.
- [Soong, F.K. et al., 1987] A vector Quantization Approach to Speaker Recognition. *AT&T, Tech. J., vol. 66*, pp. 14-26.
- [State vs. Cary] *Case 239 A. 2d 680, 685*; Elizabeth, NJ, 1968.
- [State vs. Rispoli and Straehle] *State of New York, Westchester County Court*; White Plains, NY, 1966.
- [Stevens K., Williams C., Carbonell J. y Woods B. 1.968] Speaker authentication and identification: a comparison of spectrographic and auditory presentation of speech material. *Journal of the Acoustical Society of America,* 44: 1596-1607.
- [Stevens S.S., 1936] □A scale for the measurement of a psychological magnitude: Loudness□. *Psychological Review*, 43, 405-416
- [Stoicheff, M. 1981] Speaking Fundamental Frecuency characteristics of non-smoking female adults. *JSHR 24*, pp.437-441.
- [Su, L. y Fu, K., 1973] Automatic Speaker Identification using nasal spectra and nasal coarticulation as acoustic clues. *Informe del Purdue University School of Electrical Engineering, TR-EE73-33. Air Force Office of Scientific Research Grant.* Purdue University, Lafayette, IN.
- [Sundberg, J., 1977] Acoustics of the Singing Voice, *Scientific American*, 236.
- [Suzuki, T., Tanimoto, M., Osanai, T., Kido, H. y Kamada, T., 1997] Speaker retrieval system for investigative operation. *Paper presented in the 82th Annual I.A.I. Conference*, Danvers, Massachusetts.
- [Thornwald, J., 1965] *Das Jahrhundert der Detektive*. Ed. Droemer, Zurich, 1954. Reeditado

en 1965, p. 53.
- [Tosi, O., Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J. y Nash, E., 1.972] Experiment on voice identification. *Journal of Acoustical Society of America 51* , pp. 2030-2043.
- [Tosi, O., 1.973] *Report to L.E.E.A.*, September, Michigan. (Ver: Tosi [1979], p.148)
- [Tosi, O., Pisani, R., Dubes, R. y Jaim, A., 1977] An objective method of Voice Identification. *Proceedings of the International Phonetic Sciences Congress,* Miami. Florida.
- [Tosi, O. y Greenwald M., 1978] Voice identification by subjective methods of minority group voices, *presentación del VII meeting de la IAVI en Nueva Orleans, LA.*
- [Tosi, O., 1.979] *Voice Identification. Theory and legal applications,* Baltimore, Md, University Park Press.

- [Tosi, O. y Nakasone H., 1989] A Computer Assisted Method of Voice Identification (Context Independent) *Journal of Forensic Identification*, 39(1), pp. 1-10.
- [Tosi, O. (sin fecha)] *A Voice identification Quantitative Ranking Scale*. MSU, East Lansing, MI.
- [Trager y Smith, 1935] *Outline of English Structure*, Norman, Oklahoma, p. 52
- [United States vs Wright] *Case 17 U.S. CMA 183, 37, C.M.R. 447*; CA, 1967.
- [U.S. Securities and Exchange Commission vs Klopp] Cleveland, OH, 1966.
- [Van Lancker D. y Kreiman J., 1987] Voice discrimination and recognition are separate abilities, *Neuropsychology,* 25: 829-34.
- [Vasilyev, V.A., 1965] El papel sintáctico de la entonación en inglés y ruso. *Phonetica, 12,* p. 137.
- [VIAAS, 1991] Voice comparison standards del Subcomité de Análisis Acústico e Identificación de Voz de la International Association for Identification. *Journal of Forensic Identification, Vol 41 (5),* pp. 373-386.
- [Vivaracho C.E., Ortega, J. y Romero L.A., 2000] Perceptrón Multicapa frente a Modelos de Mezcla de Gausianas en verificación automática de locutores. *Actas del I Congreso de la SEAF*, pp.85-90.
- [Warren R., 1.970] Perceptual restoration of missing speech sounds, *Science,* 167: 392-393.
- [Warren R., Obuseck C. y Acroff J., 1.972] Auditory induction of absent sounds, *Science,* 176: 1149.
- [Weber, E., 1834] *De pulse, resorpitione, auditu e tactu : Annotaciones anatomicae et physiologicae*. Leipzig: Khoelor, 1834.
- [Wever, E.G., 1949] *Theory of hearing*. Nueva York: Wiley.
- [Wolf, J., 1972] Efficient acoustic parameters for speaker recognition. *Journal of the Acoustical Society of America*, 51 pp. 2044-2056.
- [Young, M.A. y Campbell, R.A., 1967] Effects of Context on talker identification, *Journal of the Acoustical Society of America*, 42: p. 1250.

- [Zemlin, W., 1981] □*Speech and Hearing Sciencie*□, Prentice Hall, Inc., Englewood Cliffs, New Jersey.
- [Zwicker, E. y Feldtkeller, R., 1981] *Psychoacoustique*. Massón, París.

Las referencias señaladas como [I.R.] son informaciones reservadas propiedad de la Dirección General de la Policía.