

69/2017

PERCEPTION OF VOICING IN ENGLISH FRICATIVES
BY SPANISH LISTENERS

Silvia Carmen Barreiro Bilbao

Universidad Nacional de Educación a Distancia

sbarreiro en flog.uned.es

Abstract

Speech contrasts which are not functionally distinctive in the mother tongue are difficult to differentiate perceptually for non-native listeners. A two-fold experiment was carried out in order to investigate the L2 perception by 26 Spanish listeners of English, with an English group used as a control. In the present paper, listeners were exposed to the English contrast /s/ vs. /z/, which is not phonemic in Spanish, in two synthesized words, *Sue* and *zoo*. The results showed that there were listeners who had already acquired the contrast with some language-specific differences in the use of some of the acoustic cues that signalled this contrast. Furthermore there was no correlation between the results obtained and the variables age of L2 learning, length of L2 learning, or listener's age. Final conclusions will be drawn from this research and the previous work devoted to place of articulation differences.

Key words: L2 voicing, Spanish listeners, acoustic cues, English sibilants

Barreiro Bilbao, Silvia Carmen. 2017.
Perception of voicing in English fricatives by Spanish listeners.
Círculo de Lingüística Aplicada a la Comunicación 69, 34-66.
<http://www.ucm.es/info/circulo/no69/barreiro.pdf>
<http://revistas.ucm.es/index.php/CLAC>
<http://dx.doi.org/10.5209/CLAC.55313>

© 2017 Silvia Carmen Barreiro Bilbao

Círculo de Lingüística Aplicada a la Comunicación (clac)

Universidad Complutense de Madrid. ISSN 1576-4737. <http://www.ucm.es/info/circulo>

Resumen

Los contrastes fonémicos que no son funcionalmente distintivos en la lengua materna son difíciles de diferenciar para oyentes no nativos. En el presente trabajo se exponen los resultados del segundo experimento de la investigación enfocada a analizar la percepción de contrastes no nativos por parte de 26 oyentes hispanohablantes, que se utilizaron como informantes. Un grupo de hablantes ingleses sirvió como el grupo de control. Los oyentes fueron expuestos al contraste inglés /s/ vs /z/, que no es fonémico en español, en dos palabras sintentizadas, *Sue* y *zoo*. Los resultados mostraron que parte de los oyentes habían adquirido el contraste; además, se apreciaron diferencias en el uso de ciertos índices acústicos que señalaban el contraste, posiblemente causadas por la influencia de la L1. Asimismo, los resultados no tuvieron correlación ni con la edad ni el tiempo de aprendizaje de la L2, ni la edad del participante. Las conclusiones finales recogen los resultados obtenidos en el presente trabajo así como las del trabajo sobre las diferencias basadas en el lugar de articulación.

Palabras clave: Sonoridad, oyentes españoles, índices acústicos, sibilantes inglesas.

Contents

1. Introduction 36
2. Method 42
 - 2.1. Test Material 43
 - 2.2. Listeners 44
 - 2.3. Test Procedure 45
 - 2.4. Test Analysis 46
3. Results 46
 - 3.1. Spanish Results 46
 - 3.1.1. *Both cues* Condition 47
 - 3.1.2. *Frication-duration cue* Condition 49
 - 3.1.3. Conclusions 50
 - 3.2. Comparison with English Results 50

4. Discussion 52

5. Final conclusions 54

References 57

Appendix 1: Data from the Spanish group in the (/s/ vs. /ʃ/) contrast 65

Appendix 2: Gradient values from MLE 66

1. Introduction

This paper presents the second part of a two-fold investigation on Spanish perception of non-native contrasts, in particular, how Spanish speakers perceive English contrasts that are not functionally distinct in the mother tongue, or which are phonemic in the native language but differ in their phonetic realisation in the L2.

In the previous part of this research (Barreiro, 2015), an experiment was carried out to analyse the perception of fricative *place* of articulation using /s/ vs. /ʃ/, a sibilant phonemic contrast in English but not in Spanish since the palato-alveolar is not present in the Spanish phonemic system. The results showed that perceptual cue weighting was language-dependent. In particular, the *spectral properties of the friction noise* were more powerful acoustic cues than the information provided by the *formant transitions*, especially for the Spanish group. Furthermore, there was great variation among individual Spanish listeners, which was not related to *length of L2 learning*, *age of L2 learning* or *age of the listener*, some of the personal factors that seemed to interact with the difficulty in perceiving non-native contrasts (Purcell & Suter, 1980; Bohn, 1995; MacKain et al., 1981; Bohn & Flege, 1990; Polka, 1991; Mayo et al., 1997; Flege, 1998; or Bradlow & Pisoni, 1999; Escudero & Boersma, 2004; Hall et al., 2004, Cebrian, 2009; Derwing & Munro, 2013; to mention a few).

The research that is presented here examines *voicing* distinctions in sibilant fricatives. It focuses on the acoustic cues contained in the signal which can be used by the listeners to identify a phoneme category as well as to apprehend the existence of a phonological contrast (Polka, 1991; Wright, 2001 & 2004; Kabak & Maniwa, 2007; McMurray & Jongman, 2011).

The presence of simultaneous cues (or *cue multiplicity*) in the signal implies redundant information, very useful in those cases when the signal is degraded as a result of the noise environment, or even more relevant, in L2 speech perception. Furthermore, cue multiplicity in L2 speech perception is closely related to the issue of the *perceptual weighting* given to the information provided by the acoustic cues when identifying a phoneme category. It seems clear from previous research studies that L2 learners weigh the information available in the signal differently from the way L1 learners do when making a linguistic distinction (e.g. Strange and Jenkins, 1978; Underbakke et al., 1988; Yamada et al., 1992; Flege, 1984; Hazan & Boulakia, 1993; Bohn, 1995; or Williams, 1977, McCasland, 1983, and Flege & Eefting, 1987, Broersma, 2005, Barreiro, 2014 for Spanish studies). This is an issue that we will also be tackling here.

In terms of voicing, English fricatives are classified into voiceless and voiced (/f, θ, s, ʃ, h/ vs. /v, ð, z, ʒ/) according to the *presence of phonation*, determined by the glottal pulsing (or excitation) that produces energy primarily at low frequencies (below 750 kHz). This acoustic parameter has traditionally been considered the most salient cue for voicing perception. Studies focusing on the role played by the glottal vibration to make voicing distinctions between fricatives have shown that the glottal pulsing is extended over a longer period for the voiced than for the voiceless fricatives (Hughes & Halle, 1956; Ingerman, 1960; Glicksman & Stevens, 1988; Stevens et al., 1992; Pirello et al., 1997; Maniwa et al., 2008, to mention a few).

Nevertheless, it is well known that in both initial and final positions voiced fricatives are not fully voiced, especially in the latter position (Stevens et al., 1992; Smith, 1997; Jesus & Shadle, 2002 & 2003; Dmitrieva 2014). Moreover, as the fricative place of articulation moves further back, the amount of devoicing increases proportionally (Jesus and Shadle, 2003): Alveolar fricatives appear to have a lesser preponderance of voicing throughout than labiodental fricatives (Pirello et al., 1997), which have been reported to remain fully voiced more often (Stevens et al., 1992). Therefore, if voicing in fricatives cannot only be characterized by the presence or absence of the glottal excitation, a complementary way of establishing differences between the so-called voiced fricatives and their voiceless counterparts needs to be identified.

Numerous studies prior to this have shown that this acoustic property tends to overlap with many other amplitudinal and, very especially, temporal cues (Maniwa et al., 2008; Raphael et al., 2011). Let us analyse this in detail.

The differences in *overall frication duration* (also called the *normalized duration*) between voiceless (or unvoiced) and voiced fricatives have been thoroughly studied in previous works. Although the study by Crystal & House (1988a) claimed no correlation between voicing and duration in fricatives, voiceless fricatives at a given place of articulation have been reported to be significantly longer than their voiced counterparts (Klatt, 1976; d Manrique & Massone, 1981; Álvarez González, 1982; Docherty, 1992; Stevens et al., 1992; Pirello et al., 1997; Smith, 1997; Jesus & Shadle, 2003) as “the maintenance of voicing during the articulation of a fricative constriction results in later onset and earlier offset” (Silbert & de Jong, 2008: 2770). In fact, when the overall frication duration is shortened, there is a clear perceptual shift from voiceless to voiced (Flege & Hillenbrand, 1986; Jongman et al., 2000), which appears to be more pronounced for the non-sibilants than for the sibilants (Harris, 1958; Jongman et al., 2000). This will also be analysed here.

With regard to the interaction of the overall frication duration with *context* variables, it has been pointed out that the fricative position within the word can also have an effect on fricative perception, both in isolated syllables and in running speech (Crystal and House, 1988a). According to some researchers, it is particularly significant for the voicing distinction in syllable-initial positions, with voiceless fricatives having longer noise durations than voiced fricatives (Álvarez González, 1982; Behrens & Blumstein, 1988; Baum & Blumstein, 1987; Giavazzi & Cho, 2008 & 2009). Other works, however, claim that syllable affiliation (onset versus coda position) does not correspond to reliable frication duration differences (Gordon, 1989; Pirello et al., 1997; Silbert & de Jong, 2008).

Therefore, the, not uncommon overlapping of the mean overall frication duration values between voiced and voiceless fricatives, in many cases has led to the *vowel duration* and *the relative duration of the vocalic and fricative segments* being analysed as other possible temporal acoustic cues in relation to the listener’s ability of making voicing contrasts in fricatives (Denes, 1955; Harris, 1958; House, 1961; Raphael, 1972; Cole &

Cooper, 1975; Soli, 1981; Hogan & Rozsypal, 1980; Baum & Blumstein, 1987; Crystal & House, 1988a & 1988b; Glicksman & Stevens, 1988; Jongman, 1985 & 1989; Gordon, 1989; Broersman, 2010; Raphael et al. 2011; and McMurray et al. 2011). According to these research works, when the adjacent *vowel* is relatively long the consonant is usually perceived as voiced whereas when the vowel is relatively short, it is perceived as voiceless. Nevertheless, although voicing perception seems to increase as fricative duration decreases in relation to the adjacent vowel, the relative weighting of this cue (vowel context) depends on the *context* (Wilde, 1993 & 1995; Giavazzi & Cho, 2008 & 2009). In particular, the duration of the vowel has a strong effect on fricative voicing perception in syllable-final position (Denes, 1955; Álvarez González, 1980; Raphael, 1972, 1981; Maxwell & Weismer, 1982; Smith, 1997; Broersma, 2010), and a weak effect, in syllable-initial position (Cole & Cooper, 1975; Baum & Blumstein, 1987).

It is worth mentioning that for a few studies (Hughes & Halle, 1956; Raphael et al., 2011), relative durations of vowel and frication are an even more relevant cue for identifying the voicing class of a fricative than the frication duration itself.

In short, the most salient temporal properties related to voicing in fricatives, and which have a greater effect on voicing than on place of articulation, seem to be the *overall frication duration*, the *previous vowel duration*, and the *relative duration between vowel and fricative segments*. Works by different researchers have had varying results and attributed different roles to these.

In relation to the amplitudinal cues, the *overall frication amplitude* is one of the differentiating properties for fricative voiced/voiceless pairs which has been studied in depth in previous studies. It has to do with the degree of muscular effort and breath involved in their articulation. The *fortis* /f, θ, s, ʃ/ are usually pronounced with a relatively more muscular effort and breath force than their *lenis* counterparts /v, ð, z, ʒ/. /h/ is usually included in the former. In general, voiced fricatives show lower overall frication amplitude than their voiceless equivalents (Ingerman, 1960; Balise & Diehl, 1995; Silbert & de Jong, 2008). This amplitudinal feature is particularly useful for differentiating between non-sibilants and sibilants, with a greater difference between non-sibilant voiceless and voiced fricatives than for the sibilant voiceless/voiced ones (Jongman et al., 2000).

Consistent with the observations of overall frication amplitude, voiceless fricatives seem to have significantly greater *relative amplitude* in relation to the vowel amplitude than their voiced counterparts (Ingerman, 1960). Furthermore, it is related to the fricative *place of articulation*, in other words, the difference in relative amplitude is significant between the voiceless and voiced labiodental and palato-alveolars, while it is of little significance for dentals and alveolars (Jongman et al., 2000). A number of other studies have concluded that this amplitudinal property is not a relevant acoustic cue for sibilants (Mann & Repp, 1980; Stevens, 1985; Behrens & Blumstein, 1988; Hedrick & Ohde, 1993). As far as the effects of *speaker* and other *context* variables are concerned, overall, it can be said that relative amplitude is moderately affected by all these variables (McMurray & Jongman, 2011); even *gender* produces voicing distinctions in relative amplitude, with greater values for males than for females (Jongman et al., 2000).

Other amplitude-relative measures that have lately been associated to voicing in fricatives include *harmonics-to-noise ratio* (Maniwa et al., 2008), *dynamic amplitude* (Jesus & Shadle, 2002), *low-frequency energy* (McMurray & Jongman, 2011), and *the comparison of the acoustic power in higher frequencies* (or ‘noise power’ above 750 kHz) *and lower frequencies* (or ‘voice power’ below 750 kHz) taken at different intervals for each fricative (Silbert & de Jong, 2008). In relation to this latter measure, the authors concluded that phonologically voiceless fricatives have higher noise power than phonologically voiced fricatives, especially in those intervals near the vowel, the difference being greater in onset than in coda position. There was also a significant interaction with *gender*, with a greater difference between voiceless and voiced fricatives in the low-frequency power measurements of the male speaker values than those of the female ones.

Briefly, then, the most salient amplitudinal properties related to voicing in fricatives seem to be both the *overall amplitude duration* and the *relative amplitude between vowel and fricative segments*, with studies by previous researchers varying in their findings on the degree of perceptual influence.

No definite results have been shown regarding the information provided by other cues such as *formant transitions*, which has been considered a secondary or additional spectral cue for fricative voicing perception (McMurray & Jongman, 2011). In

particular, it seems that the transitions of the first formant (F1) are wider when the vowel is adjacent to voiced fricatives than when adjacent to the voiceless cognates, at least in intervocalic position (Stevens et al., 1992). As far as the other two transitions are concerned, it has been pointed out that there is a significantly higher F2 onset for voiced palato-alveolars than for their voiceless counterparts (Jongman et al., 2000) and no voicing difference among the other places of articulation. Furthermore, the F2 onset frequencies between voiceless and voiced fricatives are more dependent on *vowel* context than the F3 onset frequencies (Wilde, 1993).

Other recent studies include analyses of the *spectral properties of frication noise* in relation to fricative voicing, showing a mild but significant effect of this acoustic cue, with a different distribution of energy across fricative spectra between voiceless and voiced fricatives. In this respect, voiceless fricatives are characterized by higher values for spectral mean, skewness and kurtosis than voiced fricatives, and, in addition, they have significantly less variance than voiceless ones (Jongman et al., 2000; Jesus & Shadle, 2002; Silbert & de Jong, 2008; Maniwa et al., 2009). Therefore, in comparison with voiced fricatives, the spectra of voiceless fricatives have a concentration of energy towards slightly lower frequencies and slightly better defined peaks at a significantly higher frequency (Jongman et al., 2000). There are also significant interactions between voicing and place, and voicing and prosodic context. Voiced labiodental fricatives particularly show higher spectral variance (standard deviations) than their voiceless counterparts, an effect not found in coronal alveolars (Silbert & de Jong, 2008). Also, voiceless onset fricatives have higher spectral means, with greater differences in variance (standard deviations) between these two places of articulation in onset position than in coda.

Finally, results from other techniques, such as the *spectral slopes of two regression lines* (Jesus & Shadle, 2002) reveal that the average regression always overlaps between voiced and voiceless fricatives, supporting the idea that the inherent spectral features of the frication noise are a secondary or additional cue to voicing perception.

In sum, the analysis of prior literature (which consisted of fewer research studies than those that focussed on the study of fricative place of articulation) has shown that, in general, both amplitudinal and temporal properties can be relevant acoustic cues for

fricative voicing perception, whereas the spectral properties tend to provide secondary or additional information. In other words, the *presence of phonation*, the *frication duration*, the *relative durations of vowel and fricative segments*, and the *frication amplitude* are considered more salient acoustic cues for fricative voicing perception than the spectral properties of *frication noise* and *formant transitions*.

Bearing in mind the results of the previous paper (Barreiro, 2015) and the results from prior literature, an experiment was carried out with the following aims:

- (a) To test (i) Spanish listeners' categorisation ability for an English phonemic contrast that does not occur in Spanish (/s/ vs. /z/); in Spanish the voiced consonant is an allophone that only occurs when followed by a voiced consonant.
- (b) To assess the stage of speech-perceptual development that listeners have reached, if any.
- (c) To find out (i) if voicing distinctions in fricatives can be perceived by using multiple cues, and (ii) which cue(s) had the greatest perceptual weighting for that L2 population.
- (d) To examine subject variables, in particular (i) inter-speaker variations in the use of perceptual acoustic cues on the identification of non-native contrasts, (ii) effects of intra-speaker variables (namely, *length of L2 learning*, *age of L2 learning* and *age of the listener*) on results, and (iii) language differences (Spanish vs. English) in relation to the acoustic information that is taken into account for English sibilant identification.

2. Method

A similar procedure was used to that described in the previous research (Barreiro, 2015).

2.1. Test Material

Synthetic speech was used to control what was in the signal, allowing the manipulation of the acoustic cues and the contrast of different conditions. Subsequently, it was possible to ascertain which cues had the greatest perceptual importance for a given listener, or group of listeners. The synthetic speech patterns used for the experiment were provided by *SPA* (**S**peech **P**attern **A**udiometer). A minimal pair, *Sue/zoo* had been produced by computer-generated synthesis with a very high quality. This enabled us know which speech pattern elements (or acoustic cues) were being used to establish a contrast. Furthermore, it allowed us to assess the stage of speech-perceptual development that the listeners had reached (Hazan et al., 1995).

Two acoustic cues were analysed, one amplitudinal and one temporal, namely, the *presence of phonation* and the *duration of the fricative noise*, respectively. The former cue, which will be called *voice bar* from now on, refers to the relative amplitude (or power) between the voice bar (energy at low frequencies) and the frication noise, whereas the latter cue, *frication duration*, comprises a temporal property of the frication noise. Both cues are present in the fricatives themselves, not in the surrounding sounds.

The acoustic cues had been carefully and individually manipulated. A continuum of six steps was created from the minimal pair, in which one or two cues had been changed in small equal steps. The acoustic cues that were varied in order to signal the contrast were the *frication duration* and the *voice bar*.

Two different test conditions were then constructed in which cues had been varied in the following terms:

- (a) Contrast signalled by changes in both cues: *The duration of the frication* was varied in six steps of 20 ms, going from /s/ to /z/, as can be seen in table 1.

Table 1. Parameters changed in the continuum of the *Sue/zoo* test

	Frication Duration (ms)	Voice bar Amplitude (dB)	Frication Amplitude (dB)
Step 1	225	30	60
Step 2	205	34	58
Step 3	185	38	56
Step 4	165	42	54
Step 5	145	46	52
Step 6	125	50	50

Different parameters were changed along the continuum with the values of the frication duration starting with 225 ms in the first step (heard as /s/) and finishing with 125 ms in the sixth and last step (heard as /z/).

As far as the *voice bar* was concerned, there was an amplitude increase of 4 dB along the voice bar continuum, starting at 30 dB, as the amplitude of the frication was gradually decreased in steps of 2 dB, from 60 dB (see table 1).

(b) Contrast signalled by changes in *the frication duration* mentioned above, with no the voice bar amplitude fixed at 30 dB and the frication amplitude, at 50 dB.

Random number tables were used in order to assign a random presentation of the material from *SPA* system. The material was then recorded in a *DAT* (Digital Audio Tape), with a level of recording of around -10 dB in both channels (right and left).

A forced-choice test was designed as an *identification test*. Therefore, there were a fixed number of presentations (in this case, 10) of the randomised stimuli, and the listeners were instructed to use a restricted set of responses (*Sue* or *zoo*).

2.2. Listeners

26 Spanish listeners were chosen as the participants of this experiment, the same group as in the previous study (Barreiro, 2015). They all used Spanish (Castilian) as their L1

language, and had their permanent residence in Spain. Information was gathered via a questionnaire, including their family language background, age of L2 learning and length of time that they had been studying English. We also included questions for self-assessment of their level of English (production and perception). They all reported having normal hearing.

The average age was 26 (s.d. 5) years old, the average age of starting L2 learning was 11 (s.d.4) years, and the average time they had been studying English was 15 (s.d. 4) years, including at least one year of instruction at university level. They all considered themselves to be *monolingual*, with an *intermediate* or *advanced* level of English comprehension and fluency. They were all willing to participate in the experiment, although they did not receive any money for doing so.

Control data was obtained from the same group of 20 native speakers of British English, working or studying at UCL, who had already listened to the same material with an adaptive procedure rather than a fixed number of presentations. Their average age was 24 (s.d. 7) years old. They also reported having normal hearing.

2.3. Test Procedure

Using the same procedure as the previous study, the group of Spanish listeners was tested in the *Laboratorio de Fonética* of León University (Spain). The laboratory was equipped with a Tandberg Educational Language System, comprising a master table from which we could control the recording material given to the listeners.

The *DAT* tape was played back on the master recorder and the subjects listened to the tape through headphones (with a total impedance 200 ohm) played at a comfortable listening level.

Initial instructions and clarifications were given in their mother tongue (Spanish). After the initial explanations, the *Sue/zoo* test with 120 stimuli was carried out. The test was preceded by six examples of the material (each extreme of the continuum was presented three times). Immediately after, listeners heard one stimulus every four seconds and were told to label it, and, if they were not sure which sound they had heard, to guess. Listeners were provided with a stimulus response sheet to record their judgements. They

had ten blocks of twelve *Sue-zoo* pairs on their response sheet and had to underline the identified word of each pair. Each block was presented with an interval of 8 seconds.

2.4. Test Analysis

The analysis of the data from the two acoustic cues under investigation (*frication duration* and *voice bar*) was carried out with the program *BASIC*, an identification test scoring program. The outcome was the number of *Sue* responses at each step.

The assessment of the mean identification functions (or *labelling curves*) was based on the overall shape of the curve, the values obtained at the extremes of the range, the phoneme boundaries (defined as the point on the continuum where each category is heard equally often), and the gradient or slope of the curve (Hazan & Fourcin, 1985; Hazan et al. 1995). The labelling curves plotted the average percentage of /s/ responses across the range of stimuli in each of the two test conditions.

Finally, the *MLE* (**M**aximum **L**ikelihood **E**stimate) was used to obtain a quantitative measure of phoneme boundaries and gradients.

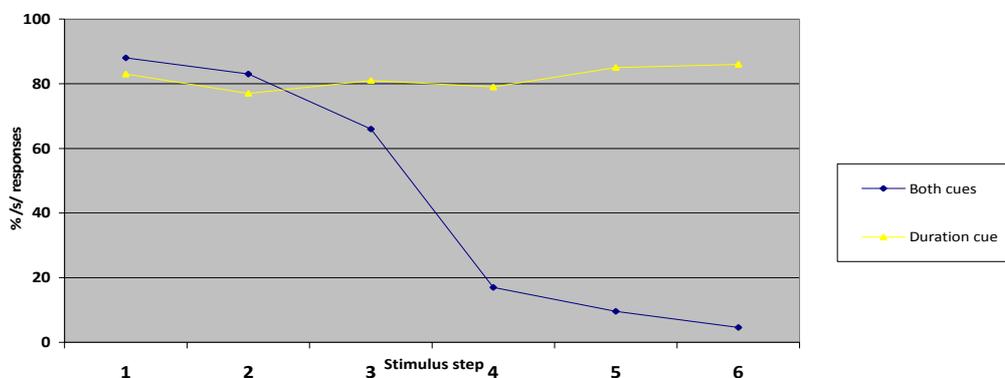
3. Results

Both the data of each member of the group of Spanish listeners (from *BASIC*) and the mean values of the gradients for the English and the Spanish groups (from *MLE*) for each of the test conditions can be seen in the appendix section (appendix 1 and 2, respectively).

3.1. Spanish Results

Figure 1 shows the mean identification functions obtained from the Spanish group's data in the *Sue-zoo* test for both test conditions.

Figure 1. Mean identification functions of averaged Spanish listeners in the test Sue-zoo test



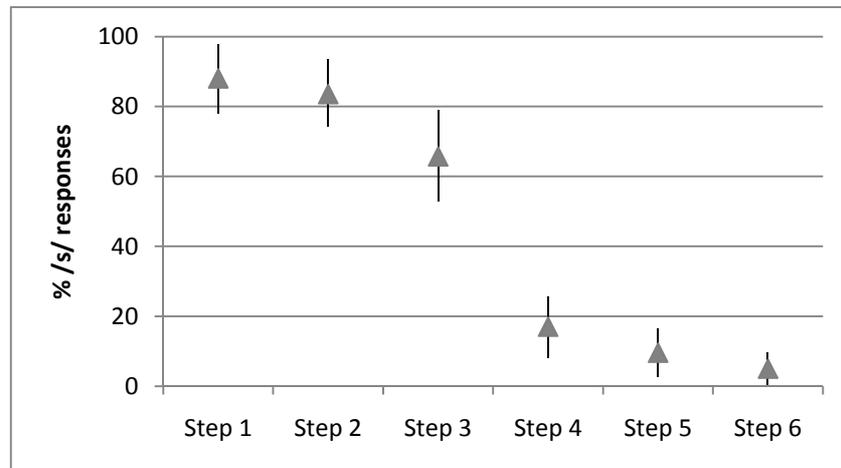
3.1.1. *Both cues* Condition

When both cues were present, the group of Spanish listeners was able to label one of the extremes of the ranges (10% or less), and almost reached top values on the other extreme. The responses to the intermediate stimuli, however, were not fully established. The mean identification function obtained could be described as *progressive*, following Hazan & Fourcin's terminology (1985: 328).

The Spanish listeners as a group identified the first two steps of the continuum as /s/ almost 90% of the time, and steps 5 and 6 were reported to be /z/ above 90%. There was a gradual perceptual shift from voiceless to voiced sibilant between steps 3 and 4. In particular, the phoneme boundary was situated at 3.011 (SD 0.870) according to the *MLE* results. This implied that this group of Spanish listeners perceived a change from the alveolar sibilant to the palato-alveolar one when the acoustic cues manipulated had the following values: (i) the *frication duration* was approximately 185 ms, (ii) the values of the *amplitude of the voice bar* was around 38 dB, and (iii) the *amplitude of the frication* was slightly less than 56 dB.

The analysis of the great variability in scores around the mean values showed great differences among the listeners (figure 2).

Figure 2. Percentage of /s/ responses (mean and standard deviation) of the Spanish listeners' results in the /s/-/z/ contrast when both cues are present



In the first place, the labelling functions of 14 subjects showed, in Hazan & Fourcin's words, "confident responses at the extremes of the range and consistent categorization of intermediate stimuli" (Hazan & Fourcin, 1985: 328), what is known as *categorical* identification functions: listeners have an ability to clearly divide the stimuli continuum into two categories. Furthermore, 5 subjects obtained mean identification functions that could be described as *progressive*, leaving the other 7 with *random* labelling curves. *Random* comprises many different types of labelling shapes, including those with a flat response around 100% /s/ identification, those completely random at the level of the phoneme boundary, and those with values scattered across the continuum.

In the second place, variability also affected the phoneme boundary, whose range of variation was between 2.226 and 4.302. For the majority of the Spanish listeners with progressive or categorical labelling functions (13 subjects), the perceptual shift from the voiceless to the voiced fricative occurred between steps 3 and 4. It was also found that the phoneme boundary had shifted to the left (between the step 2 and 3) for 5 subjects, and was situated further to the right (between the step 4 and 5) for 1 other subject.

Correlation analyses were carried out across the 26 speakers included in the experiment, relating the intra-speaker variables (i.e., *length of L2 learning*, *age of L2 learning* and *age of the listener*) and the type of labelling curve identified. According to the results of Pearson's tests ($r = -0.292$, $r = -0.029$, $r = -0.376$ at $p > .05$, respectively), no significant correlations were observed between any of these factors and the identification functions.

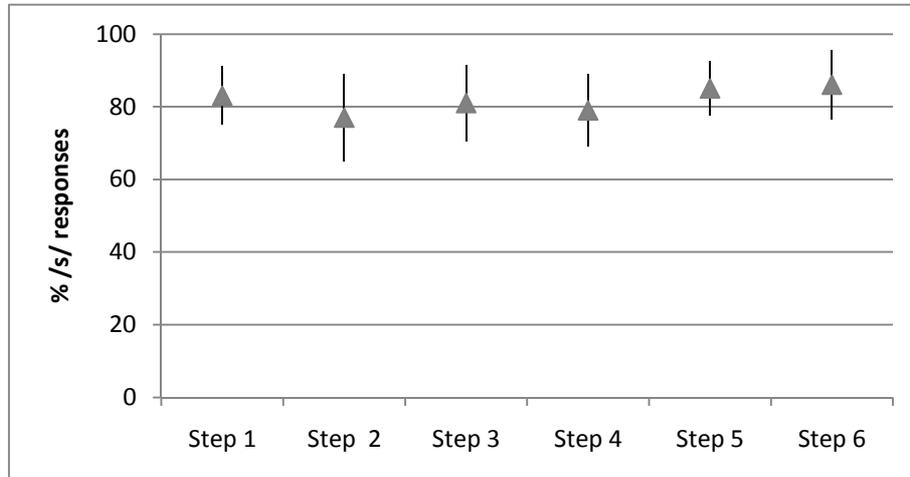
Finally, it was found that the number of categorical identification functions was greater within the group of listeners with an advanced level of English than within those with an intermediate level (57% vs. 43%), although it was not statistically significant ($r = 0.326$, $p > .05$).

3.1.2. *Frication-duration cue* Condition

When the voice bar was removed and the contrast was only cued by the duration of the frication, the Spanish listeners were unable to perceive *voicing*. The mean labelling functions obtained from the averaged Spanish listeners could be described as *random*. Furthermore, 7 listeners did not perceive any voicing at all, identifying the stimuli as /s/ over 90% of the time along the continuum.

Regarding inter-speaker differences, there was a much lower level of variability among the listeners' scores (figure 3) compared to that observed in the previous test condition. In fact, all 26 subjects showed *random* identification functions.

Figure 3. Percentage of /s/ responses (mean and standard deviation) of the Spanish listeners' results in the /s-/z/ contrast when only the frication-duration cue was present



As in the previous test condition, no significant correlations were seen between the intra-speaker variables and the type of labelling curve identified ($r = 0.146$, $r = 0.029$, $r = 0.376$ at $p > .05$, respectively). Furthermore, there were more listeners who showed categorical identification functions within the group with an advanced level of English than within the intermediate group (59% vs. 41%), although it was not statistically significant ($r = 0.280$, $p > .05$).

3.1.3. Conclusions

In view of the results, we concluded that the Spanish listeners were almost entirely reliant on the presence of phonation (voice bar) in establishing the voicing contrast and were not making use of the information given by the duration of the fricative noise. In other words, the amplitudinal property had a much greater effect than the temporal property on sibilant voicing identification by this L2 population.

The majority of the listeners (73.7% of the total number of subjects) showed a great degree of confidence in identifying the contrast when both cues were present. It can therefore be said that most of them either were able to confidently label the non-native phonemic contrast /s/-/z/ (53.8%), or had almost fully established the ability to process this contrast (19.9%).

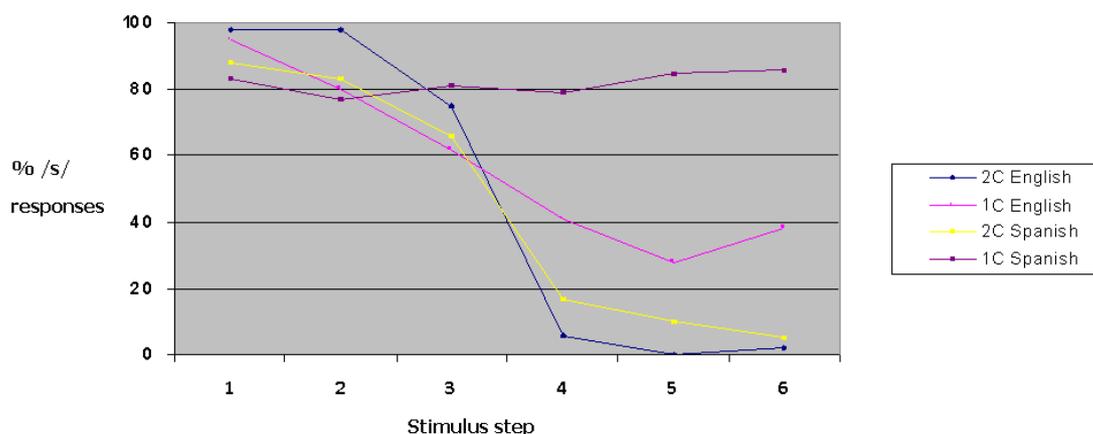
The fact that in the second condition (*frication-duration* cue only) none of the subjects managed to perceive a difference in the voicing of the sibilants was a clear indication that the cue did not provide the listeners with enough information about this contrast, even for those who had already acquired the ability to perceive it.

The result from the *General Linear Model* test carried out on the gradient values obtained for identification data to test for the effects of test condition (two-cue vs. one-cue) showed a significant difference between the mean slopes of the averaged listeners of the two conditions [$F(1,25) = 61.432; p < 0.0001$].

Finally, no significant correlation could be found between any of the intra-speaker variables and the results obtained in the present study.

3.2. Comparison with English Results

The mean labelling functions obtained from the data for the English group are shown in figure 4 together with those from the Spanish group.

Figure 4. Mean identification functions of averaged English and Spanish listeners in the *Sue-zoo* test

Leaving aside the variability among the listeners, in the first test condition, when both cues were present, the overall shape of the English labelling function could be described as *categorical*, as the curve was characterized by a plateau at both extremes of the range and a steep gradient at the boundaries, showing confident labelling and good processing ability. On the other hand, the Spanish group showed a *progressive* curve, with certain degree of uncertainty in the intermediate values. The fact that the extremes of the ranges were not reached in the Spanish labelling curve was due to the individual variability among the listeners' results. It was also found that both groups perceived a shift from a voiced alveolar sibilant to its voiceless counterpart in the same area, that is, between steps 3 and 4.

In the second test condition, when there was no voice bar and the contrast was only cued by frication duration, despite the great variability among listeners, the *random* mean labelling curves revealed that neither the English group nor the Spanish group could make a distinction based only on this cue. It can therefore be concluded that the information provided by this acoustic feature of the speech signal is not enough to perceive a voicing distinction between alveolar sibilants, regardless of the listener's mother tongue. However, the Spanish labelling curve showed a flatter response, with a progressively higher percentage of /s/ responses from step 2 onwards, showing that this cue provided even less information for the Spanish group than for the English group. It was also found that the labelling of endpoints was better in the English group: 60% /z/ response at shortest duration by 20% in the Spanish group.

4. Discussion

The general aim of this study was to achieve a better understanding of Spanish perception of the voicing of the English contrast /s/-/z/, which is not phonemic in the Spanish language, through the comparison with native English speakers.

Results of the experiments suggest that regarding the *Spanish listeners' categorization ability* in a non-native contrast, there was a clear indication that the ability to distinguish the sibilant /z/ from /s/ had been already acquired by the listeners. The mean labelling functions for the averaged Spanish listeners when both cues were present revealed that the group was able to confidently label the extremes of the range (above 90% or more) although the intermediate values were not fully established, showing some degree of uncertainty. Between step 3 and step 4 there was a gradual perceptual shift from one consonant to another.

The assessment of the *stage of speech-perceptual development* showed that the average stage of the listeners could be described as *progressive* - a misleading interpretation if we take into account the evident inter-speaker variation in the use of English fricative perceptual acoustic cues, which revealed a high number of listeners who achieved *categorical* identification functions. Therefore, the labels given to the speech-perceptual development of the averaged listeners have to be considered as an approximation of the average tendency.

The comparison with the English group indicated that, although most Spanish listeners seemed to perceive a clear shift from one category to the other, their results were worse, especially in the intermediate values, than those obtained from native speakers of English. The results clearly showed that native listeners exhibited a higher level of confidence when identifying the phoneme of this contrast, phonemic in English but phonetic in Spanish.

Analysis of the location of the *phoneme boundary* along test continua, revealed no clear language-specific changes in the contrast/s/-/z/. There was some variation, although the t-test results indicated that this was not statistically significant.

The use of synthetic speech, in which different acoustic cues were manipulated, helped us determine the *perceptual weighting* of those cues by the two groups of speakers *in native and non-native contrasts*:

- (1) Most listeners, irrespective of their L1, were able to establish the *voicing* contrast when both cues were present in the signal.
- (2) Both groups of listeners were almost completely reliant on the *voice bar* (phonation) cue in establishing the *voicing* contrast. Completely different results were obtained when the contrast was only cued by the *duration* of the *frication* noise. This temporal cue in isolation did not provide sufficient perceptual information to distinguish between the two sibilant consonants in either of the groups of listeners, native and non-native. This result supports the idea that when the overall frication duration is shortened there is a perceptual shift but it is not pronounced and significant for the sibilants (Harris, 1958; Jongman et al., 2000).
- (3) The information provided by the *overall frication duration* cue was not significant for either group, although it was a more salient cue for the English listeners than for the Spanish. This slightly different language-specific pattern of taking voice bar into account for sibilant identification means that the English listeners were able to rely on the information of other acoustic cues for a voicing contrast between phonemes with the same place of articulation, namely, the voiceless alveolar /s/ and the voiceless alveolar /z/.

These findings could also indicate that the Spanish group might have made a different use of an acoustic cue depending on the language. For instance, in Spanish, the voice bar has not been considered a robust cue when making voicing distinctions in stops (Martínez Celdrán, 1984). Nevertheless, these listeners used it in the L2 fricatives, at least in initial position. It is interesting to mention here that, in the same line of research, some previous studies have claimed that bilinguals seem to have a separate coding system, one for each language, and they shift from one to another depending on the language in use (Macnamara et al, 1968).

In relation to the effects of the *intra-speaker variables* (namely, *length of L2 learning*, *age of L2 learning* and *age of the listener*) that may have influenced the perception of non-native contrasts, our results should make us sceptical about the importance of such variables. No correlation was found between these factors and the Spanish listeners' stage of speech-perceptual development. As suggested in the previous study (Barreiro, 2015), some listeners who considered themselves as having an intermediate English

level and had never been abroad to an English-speaking country (or only for short periods of two or three months), seemed to have better perception than others who had studied longer, had been abroad for longer periods of time, or started to study English earlier.

5. Final conclusions

The arguments presented in this paper clarify and extend the analysis of a previous study (Barreiro, 2015) by shedding more light on the issue of Spanish perception of non-native contrasts, in particular, English sibilants. Therefore, our final conclusions will include references to both pieces of research.

Based on our preceding work and theoretical models, certain predictions were possible about Spanish listeners' perceptual behaviour when they are exposed to non-native contrasts with a different status in the L1 (/s/ and /ʃ/, on one hand, and /s/ and /z/, on the other). The voiced alveolar sibilant is an allophonic realisation of /s/ in Spanish, only occurring when followed by a voiced consonant (mainly plosive), whereas the palato-alveolar sound /ʃ/ does not exist in the Spanish language.

The fact that both languages (Spanish and English) have a different phonological system meant that our Spanish listeners might encounter certain perceptual problems. Language-specific perceptual differences were to be expected, due to the nature of the L1, the relationship between L1 and L2, as well other subject factors such as *year of experience with L2*, *age of subject* and *age of learning*.

The initial hypothesis in relation to the *Spanish listener's categorisation ability* was that there would be a better perception of the place of articulation contrast (/s/ and /ʃ/) than the voicing contrast (/s/ and /z/), as /ʃ/ would have been acquired as *new* category by native Spanish speakers of English (according to Flege's *Speech Learning Model*, 1995), or, even if a new category had not been created, listeners would generally do well discriminating sounds that had been assimilated into two different English categories (following Best & Strange's *Perceptual Assimilation Model (PAM)*, 1992, or the more recently developed Best & Tyler, *PAM-L2*, 2007). Our results seemed to confirm the research hypothesis: the Spanish listeners were able to label the palato-alveolar sibilant /ʃ/ more confidently than the voiced sibilant /s/ due to the different

status of those sounds in their L1 (non-existence vs. allophone). The lack of experience in L1 to voicing contrasts in fricatives, as between /s/ and /z/, implied a greater perceptual difficulty when differentiating L2 phonemic contrasts based on this difference (Lasky et. al., 1975; Werker et al., 1985; Mann, 1992; Polka 1995; Flege, 1998; Best et al., 2001; Johnson, 2004; Kabak & Maniwa, 2007; Boomersshine et al., 2008, to mention a few).

As a group, the Spanish listeners were in a progressive *stage of speech-perceptual* development for both contrasts. Nevertheless, the careful analysis of the evident inter-speaker variation in the use of perceptual acoustic cues showed that most Spanish listeners seemed to have reached a further stage in the /s/ vs. /ʃ/ than in the /s/ and /z/ contrast; the number of listeners who obtained *categorical* identification functions was higher for the *place of articulation* contrast than for the *voicing* contrast, 19 vs. 14. Also, there was a smaller number of random curves in the former contrast, 7 vs. 2, a finding that also supports the initial research hypothesis. However, discrimination tests would have enabled us to make a more valid statement about listeners' perception by establishing whether the group of listeners had developed the ability to perceive categorically.

The results of both experiments not only indicated which cues had the greatest perceptual importance in non-native contrasts for the Spanish group but also allowed us to highlight *language similarities between the Spanish and English groups*:

- (1) Most listeners, regardless of their first language, perceive a clear shift from one category to the other in both English phonemic contrasts, although the English listeners showed a slightly higher degree of confidence in the identification of phonemes.
- (2) There was a great inter-speaker variability among the listeners' results for both languages not only in the stage of speech-perceptual development but also in the phoneme boundaries which separated the categories of each phonemic contrast.
- (3) The location of the *phoneme boundaries* along test continua did not seem to be language-specific in either contrast. The slight variations in location across the experimental conditions were not statistically significant in either of the two languages.
- (4) Most listeners, regardless of their first language, were able to establish both English contrasts when both cues were present in the signal.

All the listeners had very similar results when the *place of articulation* contrast was only cued by the frication spectral characteristics. However, F2 transitions did not yield positive results. This cue in isolation did not provide enough information to distinguish between the two consonants for either of the groups, especially in the case of the Spanish group. It appeared to be an additional or secondary cue.

The primary acoustic cue for both groups in making *voicing* distinctions between alveolar sibilants was the presence of phonation (or voice bar) for both groups. The information coming from the temporal property of the frication noise was an additional or secondary cue at that syllable position. Since it is not significant in initial position, as had been observed in previous literature (Cole & Cooper, 1975; Baum & Blumstein, 1987), it would be worth carrying out a further analysis to discover whether frication duration in VC position has more perceptual weighting than in CV when identifying voicing of alveolar sibilants, especially with non-native sounds. In addition, the role of other additional cues, such as the duration of the adjacent vowel could also be analysed, following previous works (Denes, 1955; Harris, 1958; Raphael, 1972; Cole & Cooper, 1975; Soli, 1981; Baum & Blumstein, 1987; Crystal & House, 1988a & 1988b. Glicksman & Stevens, 1988; Jongman, 1989; Gordon, 1989; Raphael et al. 2011; McMurray et al. 2011).

These findings might also reveal certain *cross-linguistic differences*. It was found that the way L2 learners seemed to weight the information available in the signal was different from the way L1 speakers did this when making a linguistic decision. In particular, the Spanish group could have made a different use of an acoustic cue depending on the contrast and language. For instance, in Spanish, although the voice bar is not a strong cue when making voicing distinctions in stops, these listeners seemed to use it in the L2 fricatives. Furthermore, Spanish speakers did not use F2 transitions as cues for the perception of place of articulation contrasts of English, while these acoustic features seem to play a relevant role in identifying the place of articulation of Spanish consonants, as claimed by Moreno Llaneza (1990). It would also be interesting to investigate the effects of F3 transitions, a cue which was not manipulated in this study.

Finally, the analysis of the self-assessment showed no correlation between subject variables (namely, *length of L2 learning*, *age of L2 learning* or *age of the listener*) and

their stage of speech-perceptual development in any of two English contrasts. Some listeners seemed to have better perception than others who had studied longer, being abroad for longer periods of time and started studying English earlier. As mentioned in the previous paper (Barreiro, 2015), it may be that the Spanish listeners' immersion in the L2 was not sufficient to cause significant differences, as they all lived in Spain, and the contact with the L2 was limited to relatively short periods abroad. Moreover, they all began learning English after the age of 5, i.e., they were *monolingual*, and described themselves as such.

As suggested in the first part of this two-fold investigation (Barreiro, 2015), it would be interesting to analyse a more heterogeneous group with objective differences in their level of English (as opposed to the self-assessment criteria used here) in order to reveal significant effects on the results of the experiments. Other subject factors could be added, such as *gender* or *motivational* influences, to widen the scope of this two-fold investigation on Spanish perception of non-native fricative contrasts.

References

- Álvarez-González, J. A. (1980). *Vocalismo español y vocalismo inglés [Spanish and English vowels]*. Unpublished Doctoral dissertation, Universidad Complutense de Madrid.
- Álvarez-González, J. A. (1982). Duration of English consonants. *Anglo-American Studies*, II/1, 49-62.
- Balise, R. R., & Diehl, R. L. (1994). Some distributional facts about fricatives and a perceptual explanation. *Phonetica* 51(1-3), 99-110.
- Barreiro, S. C. (1994). *Análisis acústico descriptivo de fricativas español-inglés: estudio comparativo*. Unpublished PhD thesis. León: University of León.
- Barreiro, S. C. (2014). Fricatives revisited. In R. Monroy-Casas, R. & I. Arboleda-Guirao (Eds.), *Readings in English Phonetics and Phonology* (pp. 51-88). I. Valencia: IULMA, University of Valencia.

- Barreiro, S. C. (2015). Perception of place of articulation in English fricatives by Spanish listeners. *VIAL, Vigo International Journal of Applied Linguistics* 12, 37-62.
- Baum, S. R. & Blumstein, S. E. (1987). Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. *The Journal of the Acoustical Society of America* 82, 1073.
- Behrens, S. & Blumstein, S. E. (1988). Acoustic characteristics of English voiceless fricatives: a descriptive analysis. *Journal of Phonetics* 16, 295-298.
- Best, C. T., McRoberts, G.W. & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, 109, 775.
- Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.S. Bohn & M.J. Munro (Eds.), *Language experience in second language speech learning*. Amsterdam: John Benjamins.
- Best, C. T. & Strange, W. (1992). Effects of experience and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305-330.
- Bohn, O. S. (1995). What determines the Perceptual Difficulty encountered in the Acquisition of Nonnative Contrasts? *Proceedings of ICPhS 95*, Stockholm 4, 84-91.
- Bohn O. S. & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11, 303-328.
- Boomershine, A., Hall, K. C., Hume, E. & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. *Contrast in phonology theory, perception, acquisition*, 145-172.
- Bradlow, A. & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America* 106, 2074-2085.

- Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America* 117(6), 3890-3901.
- Broersma, M. (2010). Perception of final fricative voicing: Native and non-native listeners' use of vowel duration. *Journal of the Acoustical Society of America* 127(3), 1636-1644.
- Cebrian, J. (2009). Effects of native language and amount of experience on crosslinguistic perception. *The Journal of the Acoustical Society of America* 125, 2775.
- Cole R. A. & Cooper, W.E. (1975). Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America* 59, 1208.
- Crystal, T. H. & House, A. S. (1988a). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America* 83, 1553–1573.
- Crystal, T. H. & House, A. S. (1988b). Segmental durations in connected-speech signals: Syllabic stress. *The Journal of the Acoustical Society of America* 84, 1574–1585.
- de Manrique, A. M. B., & Massone, M. I. (1981). Acoustic analysis and perception of Spanish fricative consonants. *The Journal of the Acoustical Society of America*, 69(4), 1145-1153.
- Denes, P. B. (1955). Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America* 27, 761-764.
- Derwing, T. M. & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, 63(2), 163-185.
- Dmitrieva, O. (2014). Final voicing and devoicing in American English. *The Journal of the Acoustical Society of America* 136(4), 2174-2174.
- Docherty, G. J. (1992). *The timing of voicing in British English obstruents* (9). Berlin: Foris.

- Escudero, P. & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26(4), 551-585.
- Flege, J. E. (1984). The effects of linguistic experience of Arabs' perception of the English /s/ vs. /z/ contrast. *Folia Linguist* 18, 117-138.
- Flege, J. E. (1995). Second language learning: theories, findings and problems. *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Ed. W. Strange. Timonium MD: York Press.
- Flege, J. E. (1998). Second-Language Learning: The Role of Subject and Phonetic Variables. *Proceedings of the ESCA-StiLL 98*. Marholmen, Sweden, 1-8.
- Flege, J. E. & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15, 67-83.
- Flege, J. E., & Hillenbrand, J. (1986). Differential use of temporal cues to the/s-/ /z/contrast by native and non-native speakers of English. *The Journal of the Acoustical Society of America*, 79(2), 508-517.
- Glicksman, L. B. & Stevens, K.N. (1988). Some parameters affecting the perception of voicing for fricatives. *The Journal of the Acoustical Society of America* 84(S1), 58-59 (A).
- Hall, K. C., Hume, E., Johnson, K., & Reiter, A. (2004). The influence of language experience on perception: The case of Spanish and English. *OSU Hispanic Linguistics Colloquium*.
- Giavazzi, M. & Cho, H. (2008). *On the perception of voicing in strident fricatives*. http://web.mit.edu/magia/www/docs/ChoGiavazzi_2008.pdf.
- Giavazzi, M. & Cho, H. (2008). *Perception of voicing in fricatives*. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.370.725&rep=rep1&type=pdf>.
- Gordon, P. C. (1989). Contexts effects in recognizing syllable-final /z/ and /s/ in different phrasal positions. *The Journal of the Acoustical Society of America* 86, 1698-1707.

- Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1, 1–7.
- Hazan, V. L. & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech* 36, 17-38.
- Hazan, V. L. & Fourcin, A. F. (1985). Microprocessor-controlled Speech Pattern Audiometry. *Audiology* 24, 325-335.
- Hazan, V. L., Wilson, G., Howells, D., Miller, D., Abberton, E. & Fourcin, A.F. (1995). Speech pattern audiometry for clinical use. *European Journal of Disorders of Communication* 30, 116-123.
- Hedrick, M. S. & Ohde, R.N. (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustical Society of America* 94, 2005-26.
- Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America* 67, 1764.
- House, A. S. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America*, 33(9), 1174-1178.
- Hughes, G.W. & Halle, M. (1956). Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America* 28, 303–310.
- Ingerman, F. (1960). Formants as voicing cue in the perception of /s/. *The Journal of the Acoustical Society of America* 66, 1700.
- Jesus, L. M., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics* 30(3), 437-464.
- Jesus, L. M., & Shadle, C. H. (2003). Temporal and devoicing analysis of European Portuguese fricatives. *Proceedings of ICPHS 2003*. Barcelona: 779-782.
- Johnson, K. (2004). Cross-linguistic perceptual differences emerge from the lexicon. In A. W. Warren & S.H. Park (Eds.), *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in speech production and perception*. Sommerville/ MA: Cascadilla Press: 26-41.

- Jongman, A. (1985). Duration of frication noise as a perceptual cue. *Journal of the Acoustical Society of America* 77(S1), S26(A).
- Jongman, A. (1989). Duration of frication noise required for the identification of English fricatives. *Journal of the Acoustical Society of America* 85(4), 1718-1725
- Jongman, A., Sereno, J. A., Wayland, R. & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 118, 1252–1263.
- Kabak B. & Maniwa, K. (2007). L2 perception of English fricatives in clear and conversational speech: The role of phonemic, phonetic and acoustic factors. *Proceedings of ICPHS 2007*. Saarbrücken, 781-784.
- Lasky, R. E., Syrdal-Lasky, A. & Klein, R. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology* 20, 215-225.
- Mackain, K. N., & Best, C. T. & Strange, W. (1981). Categorical perception of English /ɾ/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2, 360-390.
- Maniwa, K., Jongman, A. & Wade, T. (2008). Perception of clear English fricatives by normal-hearing and simulated hearing-impaired listeners. *Journal of the Acoustical Society of America* 123, 1114–1125.
- Maniwa, K., Jongman, A. & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America* 125(6), 3962-73.
- Mann, V. A. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *The Journal of the Acoustical Society of America* 92, 711.
- Mann, V. A. & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction: III Acoustic observations. *Haskins Laboratories Status Report on Speech Research* 62, 177-85.
- Maxwell, E. M. & Weismer, G. (1982). The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops. *Applied Psycholinguistics* 3(01), 29–43.

- Mayo, L. H., Florentine, M. & Buus, S. (1997). Age of Second-Language Acquisition and Perception of Speech in Noise. *Journal of Speech, Language and Hearing Research* 40, 686-693.
- McCasland, G. P. (1983). The noise cues of English and Spanish fricatives. *Journal of the Acoustical Society of America* 73(S1), S30.
- McMurray, B. & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review* 118(2), 219-246.
- Moreno Llaneza, M. A. (1990). Transiciones vocálicas y punto de articulación consonántico. *Estudios de Fonética Experimental IV* (pp. 52-102). Barcelona: PPU, Laboratorio de Fonética, Facultad de Filología, Universidad de Barcelona.
- Pirello, K., Blumstein, S.E. & Kurowski, K. (1997). The characteristics of voicing in syllable-initial in American English. *The Journal of the Acoustical Society of America* 101, 3754.
- Polka, L. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America* 89, 2961-2976.
- Purcell, E. T. & Suter, R.W. (1980). Predictors of pronunciation accuracy: A re-examination. *Language Learning* 30(2), 271-287.
- Raphael, L. J. (1972). Preceding vowel duration as a Cue to Perception of the Voicing Characteristic of Word-Final Consonants in American English. *Journal of the Acoustical Society of America* 51(4), 1296-1303.
- Raphael, L. J., Borden, G. J. & Harris, K.S. (2011). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Baltimore: Lippincott Williams & Wilkins.
- Repp, B. H. & Mann, V.A. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction: II spectral factors. *Haskins Laboratories Status Report on Speech Research* 61, 65-84.

- Silbert, N. & de Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America* 123(5), 2769-2779.
- Smith, C. L. (1997). The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics* 25(4), 471-500.
- Soli, S. D. (1981). Second formant in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America* 70, 976-984.
- Stevens, K. N., Blumstein, S.E., Glicksman, L., Burton, M. & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *The Journal of the Acoustical Society of America* 91, 2979-3000.
- Strange, W. & Jenkins, J. J. (1978). The role of linguistic experience in the perception of speech. In R.D. Walk & H.L. Pick (Eds.), *Perception and Experience*. New York: Plenum Press, 125-169.
- Underbakke, M., Polka, L., Gottfried, T.L. & Strange, W. (1988). Trading relations in the perception of /r/-/l/ by Japanese learners of English. *Journal of the Acoustical Society of America* 84, 90-100.
- Yamada, R. A. & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/-/l/ by Japanese listeners. *Perception and Psychophysics* 52, 376-379.
- Werker, J. F. & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37, 35-44.
- Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics* 21(4), 289-297.
- Wilde, L. F. (1993). Inferring articulatory movements from acoustic properties at fricative vowel boundaries. *Journal of the Acoustical Society of America* 94, 1881.
- Wilde, L. F. (1995). *Analysis and synthesis of fricative consonants*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, U.S.A.

Wright, R. (2001). "Perceptual cues in contrast maintenance". In H. Elizabeth & J. Keith (Eds.), *The role of speech perception phenomena in phonology*. San Diego: Academic Press, 251-277.

Wright, R. (2004). A review of perceptual cues and cue robustness. *Phonetically based phonology*, 34-57.

Appendix 1: Data from the Spanish group in the (/s/ vs. /ʃ/) contrast

Two Cues											
Gradient	s.d.	Intercept	Ph. Boundary	Chi-Square	d.f.	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
-5.90695	2.0327	20.67434	3.5	0.0029	4	100	100	100	0	0	0
-2.48807	0.7027	6.968412	2.800726	9.0506	4	90	100	40	0	0	0
-0.28779	0.16337	0.503958	1.751109	12.4039	4	20	80	60	40	10	20
-0.7034	0.199	2.730696	3.882125	11.0068	4	90	50	90	60	40	0
-0.60777	0.1893	1.704806	2.805019	3.1619	4	60	70	60	40	10	10
-3.78463	1.1981	12.8241	3.388466	0.1	4	100	100	80	10	0	0
-0.4754	0.1753	2.053853	4.320244	8.3426	4	100	80	30	50	50	40
-0.71669	0.2046	1.962137	2.737767	3.2541	4	70	70	60	10	20	10
-2.42431	0.6787	7.030425	2.89997	2.972	4	100	80	60	0	0	0
-4.45726	1.5744	16.40337	3.680145	0.0399	4	100	100	100	20	0	0
-3.19983	1.0219	12.18408	3.807732	0.504	4	100	100	90	40	0	0
-2.61724	0.7406	9.690482	3.702559	2.3289	4	100	100	90	20	10	0
-0.20969	0.165	0.01951	0.093046	0.7245	4	50	30	40	30	30	20
-5.18424	1.7567	18.52268	3.572882	0.0096	4	100	100	100	10	0	0
-0.44655	0.1814	0.775388	1.736394	1.3304	4	60	50	30	20	30	10
-4.05516	1.4892	9.027823	2.226255	3.5192	4	100	80	0	0	0	0
-1.73968	0.4453	4.501036	2.587271	6.5808	4	100	70	30	0	10	0
-2.11142	0.5518	6.753478	3.198554	4.5264	4	100	80	80	10	0	0
-1.14807	0.2751	3.774134	3.287367	8.5066	4	80	100	70	10	10	10
-1.66606	0.4072	5.324988	3.196148	1.282	4	100	90	50	20	10	0
-1.80713	0.4538	5.413353	2.995553	7.9654	4	90	80	80	0	0	0
-2.42431	0.6778	9.939727	4.100029	2.9681	4	100	100	100	40	20	0
-4.06078	1.4839	13.10108	3.226248	3.5419	4	100	100	80	0	0	0
-3.72736	1.399	11.35057	3.045206	0.8951	4	100	100	60	0	0	0
-1.10697	0.2278	2.915818	2.634058	3.162	4	80	60	60	20	0	0
-3.83316	1.4275	12.01914	3.135572	1.7101	4	100	100	70	0	0	0
-2.3535			3.011		Mean	88.0769	83.4615	65.7692	17.3077	9.61538	4.61538

barreiro: fricatives 66

Duration Cue											
Gradient	s.d.	Intercept	Ph. Boundary	Chi-Square	d.f.	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
0.12812	0.2553	1.767886	-13.7989	6.3531	4	90	100	70	90	90	100
0.02241	0.2118	1.656662	-73.9093	4.3553	4	80	90	90	70	100	80
0.04767	0.1545	0.239299	-5.02038	0.7356	4	60	50	60	70	60	60
0.13887	0.1534	-0.35065	2.524986	8.6139	4	70	20	50	40	80	60
0.01407	0.1678	0.87886	-62.4545	4.3863	4	70	80	70	70	50	90
0	1840.523	21.2004	0	0	4	100	100	100	100	100	100
-0.31229	0.1758	0.189627	0.607208	4.5571	4	60	30	20	20	40	10
0.25807	0.1982	0.540619	-2.09483	4.1327	4	70	60	100	80	80	90
-0.04951	0.2228	2.047723	41.35618	4.5688	4	90	90	70	100	90	80
0	1840.523	21.2004	0	0	4	100	100	100	100	100	100
-0.18396	0.357	3.632324	19.74495	3.1249	4	100	90	100	90	100	90
0.06738	0.2124	1.503386	-22.3106	4.3595	4	90	90	70	80	80	100
0.10334	0.1615	0.336633	-3.25745	1.945	4	70	50	60	80	80	70
-1.14269	0.8679	8.742167	7.650548	0.9709	4	100	100	100	100	90	90
0.33668	0.1883	0.001396	-0.00415	2.6545	4	70	50	70	90	80	90
0.38203	0.2091	0.173622	-0.45447	3.6784	4	70	80	60	80	90	100
0.41014	0.1951	-0.21743	0.530127	3.3004	4	60	50	90	70	70	90
-0.56777	0.3457	4.749472	8.365166	1.2595	4	100	100	90	90	90	80
0.43765	0.2231	0.132965	-0.30382	1.047	4	60	70	90	90	90	90
0.05734	0.1958	1.29626	-22.6054	14.7221	4	100	50	100	60	90	90
-0.13479	0.198	1.982406	14.70757	9.355	4	90	80	100	50	90	80
0	1840.523	21.2004	0	0	4	100	100	100	100	100	100
-0.64212	0.803	6.848109	10.66489	2.9767	4	100	100	100	100	90	100
-0.27246	0.2899	3.440097	12.62616	8.2187	4	100	100	90	70	100	90
0.37421	0.1965	-0.01372	0.036656	3.8281	4	70	70	60	70	90	100
-0.64212	0.803	6.848109	10.66489	2.9767	4	100	100	100	100	90	100
-0.0045					Mean	83.4615	76.9231	81.1538	79.2308	85.3846	85.7692

Appendix 2: Gradient values from MLE

English listeners' results							
		Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Mean	Both cues	97.8022	97.8022	75.22436	6.153846	0	1.098901
Mean	Duration cue	94.91758	80.08242	62.22528	40.52198	27.88462	37.71978

Spanish listeners' results							
		Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Mean	Both cues	88.07692	83.46154	65.76923	17.30769	9.615385	4.615385
Mean	Duration cue	83.46154	76.92308	81.15385	79.23077	85.38462	85.76923

Received: December, 16, 2016

Accepted: February 17, 2017

Published: February 28, 2017

Updated: March 27, 2017