Perception of natural and enhanced non-native contrasts in clear speech

Silvia Carmen Barreiro Bilbao
Universidad Nacional de Educación a Distancia (UNED)
España
If learning phonetic contrasts of a language is considered the result of selective attention, it then may be possible to redirect listeners’ attention towards non-native contrasts. In this paper we present an experiment carried out in order to investigate the perception of non-native contrasts by Spanish listeners, in particular, we want to analyse the effects of cue-enhancement on their perceptual performances of 25 different English words containing natural and enhanced sibilants, all presented in clear speech. The results of the experiments showed that the effects of the enhancement varied greatly among the consonants, and across the different vocalic contexts. Moreover, the results did not seem to correlate with subjects’ variables, such as subject’s age, time period of L2 learning or age of L2 learning, although error patterns could be related to the phonological system of Spanish listeners.

**Keywords**: Spanish perception; English sibilants; cue-enhancement.
1. Introduction

It is well known that an L2 learner will encounter perceptual difficulties when trying to differentiate speech contrasts which are not functionally distinctive in the mother tongue, or which are phonemic in the native language, but differ in the phonetic realisation in the L2.

Perceptual difficulties with non-native speech categories have been explained by different approaches, including the Perceptual Assimilation Model (Best & Strange, 1992), the Speech Learning Model (Flege, 1995), the Perceptual Learning Model (Pisoni et al., 1994) or the Native Language Magnet Model (Kuhl & Iverson, 1995). All these models varied greatly in relation to a deeper issue, that is, if the underlying neural mechanisms used in speech perception becoming finely tuned to the distinctive contrasts used in the native language is a reversible or non-reversible process. Nowadays, it seems that the second position is more broadly accepted: “perceptual difficulties are not due to a loss of sensory capabilities, but rather reflect perceptual attunement to phonetic information that is phonologically relevant in their native language” (Strange, 1995: 79). This is the approach followed in the present work—supported by Flege’s ideas (1995)—, that implies that if learning phonetic contrasts of a language is the result of selective attention (Pisoni et al., 1994), it is possible to redirect listeners’ attention towards non-native contrasts.

In all this process of redirecting attention, there are different variables that interact in the perception of non-native contrasts. First of all, the contrast factors, some contrasts are more difficult to differentiate perceptually than others (Lasky et. al., 1975; Werker & Logan, 1985; Flege, 1998), for instance, a lack of experience in L1 to voicing contrasts would imply a great perceptual difficulty in differentiating L2 phonemic contrasts based on voicing differences. Also, L2 learners seem to weigh differently the information available in the signal from the way L1 learners do in making a linguistic distinction (Strange & Jenkins, 1978; Underbakke et al., 1988; Yamada & Tohkura, 1992; Flege, 1984; Bohn, 1995; Hazan & Boulakia, 1993; or Flege & Eefting, 1987; and Williams, 1977, for Spanish studies). Leaving aside these contrast variables, there are also subject factors that do seem to contribute to the ability to distinguish non-native contrasts, as L1 background, L2 experience or age of the learner, to mention a few (as shown by MacKain et al., 1981; Ocke-Schwen Bohn, 1995; Mayo et al., 1997; Flege, 1998; or Bradlow & Pisoni, 1999).

Different auditory training techniques have been applied to improve listeners’ performance with non-native contrasts. One of the first works was carried out by Jamieson & Morosan (1986), in which francophone adults were trained in non-native VOT identification, using natural and synthetic fricative tokens. Since then, enhancement techniques have been applied to help in auditory training for different populations (L2 learners, hearing-impaired listeners, children with language disorders…) in clear speech or in noise (or Lombard) speech (Gordon-Salant, 1986; Hazan & Shi, 1992; Jamieson, 1995; Hazan & Simpson, 1996a, 1996b, 1997; Ortega-Llebaria & Hazan, 1999; Ortega-Llebaria & Huckvale, 2000; Skowronski & Harris, 2005).

Some of these techniques include cue-enhancement involving the amplification of selective regions that contain important cues in order to improve listeners’ intelligibility due to the fact that the manipulation of intensity ratios had shown the greatest effect on intelligibility in comparison with that obtained as a result of spectral or temporal enhancements (Revoile et al., 1986; Bunnell, 1990). It is worth mentioning the studies carried out by Hazan & Simpson (1998a, 1998b, 2000) which showed that all the listeners, including Spanish, improved the recognition of enhanced consonants in the presence of noise, showing an influence of their L1 background. Analysing the Spanish results in detail, it is worth mentioning that /p, b, g, f, v/ were the least
intelligible consonants. As far as the sibilant fricatives were concerned, /s/ had better results than /z/, whereas the palato-alveolar ones were not analysed. The results were an average over speakers and vowel contexts.

These encouraging results seem to open a new door for L2 training, and this made us think that it would be interesting to find out whether Spanish listeners benefit from the cue-enhancement in clear speech, and in a similar way to those reported in previous works, that is, the lower the initial intelligibility of the listener, the greater the improvement produced by the enhancement.

An experiment was set up to observe what perceptual benefits for Spanish listeners would have enhancing consonantal regions which contain a high density of acoustic cues to two phonemic contrasts in English, in clear speech: /s/ vs. /z/ and /s/ vs. /ʃ/. The difference between the alveolar fricatives is based on the “voicing” feature. This contrast is phonemic in English but phonetic in Spanish, that is, the voiced fricative is an allophone in Spanish that only occurs when followed by a voiced consonant, as in the word mismo. In the second contrast, the difference between the fricatives is founded on the “place of articulation” feature. The contrast is phonemic in English whereas the second phoneme does not exist in Spanish.

2. Method

The approach adopted to increase the intelligibility of clear speech was by enhancing salient acoustic cues known to encode phonemic contrasts in fricative consonants, using the same enhancement techniques described in the previous studies by Hazan mentioned above.

Stimuli were recorded in a sound-proof room and digitised at a 16 kHz sampling rate with 16-bit amplitude quantisation. The stimuli of this experiment were produced by a female talker with a RP accent. The digitised stimuli were, then, annotated using a waveform editing tool to mark the regions for amplification. These regions comprised the formant transitions at the release of the constriction, and the cues at release or of during the constriction, the weakest voicing cycles being given the most amplification. The same enhancement technique was applied to both contrasts, /s/-/z/ and /s/-/ʃ/.

Both the consonantal region and the formant transition regions at vowel onset were manipulated:

- a) For the transitions regions, the reduced amplitude, as the consonant constriction was released, was counteracted by amplifying the 3 cycles of the following vowel, by 4, 3, and 2 dB, respectively.
- b) The consonantal frication region was amplified by 6 dB.

Filtering was not used to change the spectral content of the regions which were perceptually relevant in order to make them more discriminable.

2.1. Test Material

14 CV(C)(C) English minimal pairs (a total of 25 different words) were used as the stimuli of the experiment (table 1). They comprised the consonants /s/, /z/ and /ʃ/ in prevocalic positions with /ɪ, e, æ, iː, uː, eɪ, au/.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Minimal pairs used for the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>/s/ vs. /z/</td>
<td>/s/ vs. /ʃ/</td>
</tr>
<tr>
<td>Sue/zoo</td>
<td>Sue/shoe</td>
</tr>
<tr>
<td>sip/zip</td>
<td>so/show</td>
</tr>
<tr>
<td>said/zed</td>
<td>seen/sheen</td>
</tr>
<tr>
<td>sap/zap</td>
<td>sip/ship</td>
</tr>
<tr>
<td>sink/zinc</td>
<td>said/shed</td>
</tr>
<tr>
<td>sing/zing</td>
<td>same/shame</td>
</tr>
<tr>
<td>seal/zeal</td>
<td>sin/shin</td>
</tr>
</tbody>
</table>

1 Formant-transition cues were amplified to counterpart the reduction in amplitude near the constriction.
Perception of natural and enhanced non-native contrasts in clear speech

Silvia Carmen Barreiro Bilbao

2.2. Listeners

Listener selection was mainly based on the fact that it was required word comprehension by the informants. 26 Spanish monolingual listeners with normal hearing were chosen as the subjects of the experiment. They all used Spanish (Castilian) as their L1 language. The average age was 26 years old, 11 years was the average of L2 learning, and the average length of studying English was 15 years. Therefore, none of them have problems of word comprehension.

2.3. Test Procedure

The group of Spanish listeners was tested in the Laboratorio de Fonética from León University (Spain). They listened to the DAT tape through headphones, presented in a comfortable level. The test called “Sue-zoo-shoe” comprised 200 stimuli (see appendix). Listeners had 4 blocks of 50 minimal pairs on their response sheet. In each block there were 2 repetitions of each word: 1 natural and 1 enhanced. The identification test was preceded by six examples of natural realizations of the three words contained in the test title to make the role of enhancement even more enlightening.

3. Results

3.1. /s/ results

In the natural condition (44 presentations per consonant per listener), the number of listeners who identified this consonant correctly above 90% of the times was 22 subjects, whereas in the enhanced condition (44 presentations) that number decreased slightly going down to 19.

As first impressions, it seemed that the enhancement was not beneficial for these Spanish listeners. Moreover, when analysed the data from the 7 people who had identification problems, it seemed that there was a significant difference between the natural and the enhanced according to the results of the t-test. In fact, the enhancement had decreased the intelligibility of this consonant, compared to the natural ones [t = 2.91, df = 6; p < 0.05].

On examining the data in relation to the adjacent vowel in all the speakers who had any problem of identification, it was revealed that the enhancement had different effects on the listeners depending on the vocalic context. Table 2 shows the number of people who improved, got worse or was not affected by the enhancement.

<table>
<thead>
<tr>
<th>Adjacent Vowel</th>
<th>Improvement</th>
<th>No effect</th>
<th>Worsening</th>
</tr>
</thead>
<tbody>
<tr>
<td>[æ]</td>
<td>2/8</td>
<td>0/8</td>
<td>6/8</td>
</tr>
<tr>
<td>[e]</td>
<td>2/5</td>
<td>1/5</td>
<td>2/5</td>
</tr>
<tr>
<td>[ɪ]</td>
<td>3/12</td>
<td>0/12</td>
<td>9/12</td>
</tr>
<tr>
<td>[ɨɪ]</td>
<td>2/13</td>
<td>1/13</td>
<td>10/13</td>
</tr>
<tr>
<td>[u]</td>
<td>0/6</td>
<td>2/6</td>
<td>4/6</td>
</tr>
<tr>
<td>[eɪ]</td>
<td>2/5</td>
<td>1/5</td>
<td>10/10</td>
</tr>
<tr>
<td>[oʊ]</td>
<td>0/10</td>
<td>0/10</td>
<td>0/10</td>
</tr>
</tbody>
</table>

It can be noticed that the enhancement decreased the intelligibility in all vocalic contexts. However, according to the t-test, there was not a significant alteration of the results produced by the enhancement, except when the voiceless alveolar sibilant was followed by the vowels [ɪ:], [t(12) = 3.12, at p > 0.05], [u:] [t(5) = 2.71, at p > 0.05], and [oʊ] [t(9) = 6.33 at p > 0.001].

3.2. /z/ results

The number of listeners who managed to identify the consonant above 90% or more of the times presented in the natural condition (22 presentations) was 22, almost the same number as in the enhanced one, 21 subjects. Therefore, it seemed that the enhancement may have a very mild effect, if any, on the listeners’ perception.

When the data from the 5 subjects with identification problems was analysed, it was clear that the enhancement had a benefit on the per-

---

2 The data was analysed manually.
3 It was considered that 90% correct would be the baseline to say that a listener had no perceptual problems in identifying the consonants.
ception of the consonant, although, as a whole, these improvement were not statistically signifi-
cant. Nevertheless, its effect differed across the vocalic contexts (table 3):

| TABLE 3 |
|-----------------|-----|-----|-----|-----|
|                | [æ] | [e] | [i] | [i:] |
| Improvement     | 4/6 | 3/5 | 3/9 | 1/1  |
| No effect       | 0/6 | 1/5 | 0/9 | 0/1  |
| Worsening       | 2/6 | 1/5 | 6/9 | 0/1  |

In all cases, except when followed by [i:], the enhancement increased the number of correct identification although in none of the contexts a significant difference was observed in the re-

3.3. /∫/ results

In the natural condition 20 listeners identified /∫/ correctly 90% or more of the 22 presenta-
tions of this consonant, whereas in the enhanced condition, there was a slight increase to 21 subjects. Consequently, only 6 subjects had not been able to identify the consonant within that percentage. Furthermore, the results of t-test with that group showed that the enhancement, as a whole, improved significantly the intelligibility of this consonant, compared to the natural ones, [t(5) = 3.29, significant at p > 0.05].

The analysis of the data by contexts showed that the effect of enhancement on listeners' per-
ception of the palato-alveolar consonant varied across the vocalic contexts (table 4):

| TABLE 4 |
|-----------------|-----|-----|-----|-----|
|                | [e] | [i] | [i:] | [u:] |
| Improvement     | 5/7 | 6/7 | 4/6 | 4/5  |
| No effect       | 2/7 | 1/7 | 1/6 | 1/5  |
| Worsening       | 0/7 | 0/7 | 1/6 | 0/5  |

It seemed that, overall, a consonant intelligibility benefit was observed across contexts, although there was little evidence of statistically-significant improvements except when the voiceless palato-alveolar sibilant was followed by the vowels, [e] and [i], [t(6) = 2.65] and [t(6) = 3.33 at p < 0.05], respectively.

4. Discussion

This experiment aimed to bring more light to the effects of cue-enhancement on improving consonant intelligibility for L2 learners in those cases where the target sound contrasts were likely to be difficult due to the listeners’ L1 back-
ground. In particular, the primary aim was to see the effects of cue-enhancement on Spanish lis-
teners’ sibilant identification. In view of the results of previous studies, it was initially predicted that there would be clear benefits from the cue-enhancement. However, the results of our experi-
ment differed slightly from the initial prediction, at least for the consonants under investigation in this study: the increase in intelligibility due to cue-enhancement was smaller than that obtained in studies such as Hazan & Simpson (1998a, 1998b, 2000), or even Ortega-Llebaria & Huckvale (2000).

As a whole, there was a general trend for higher intelligibility to be obtained for the enhanced conditions for /z/ and to a greater extend for /∫/, but the cue enhancement decreased the number of correct responses for /s/ (figure 1).

It seemed that the cue-enhancement was only beneficial for the consonants /z/ and /∫/ that were not phonemes in the listeners’ native language, bringing, consequently, more confusion than benefit in the case of /s/. Furthermore, as the initial intelligibility of the listeners was not low, the improvement produced by the enhancement was not as great as one would have been expected, although its effect on /∫/ was statistically-significant.

It could have been the case that the spectral characteristics of the adjacent vowel and,
perhaps the peculiarities of the consonant itself should have been taken into account when applying the cue-enhancement since its effects varied across consonants and contexts.

The analysis of the vocalic contexts (figure 2) revealed that the three consonants were more easily identified in some contexts than others, although little patterning was found across the consonants:

a) The enhancement was beneficial when the consonants /z/ and /∫/ were followed by /e/ whereas it had no effect on /s/ identification.

b) It was clearly detrimental when the adjacent vocalic element to /s/ was /ʊ/, but neutral for /∫/.

c) The identification of the alveolar fricatives worsened when followed by /ɪ/, in contrast to the palato-alveolar, /∫/.

d) When the following vowels were /æ, i;
u:/, /s/ scores worsened, in contrast to the improvement with the other two consonants.

e) When the voiceless sibilant was followed by /eı/ the scores became worse whereas they became better for the palato-alveolar /∫/.

The analyses of variance carried out on the intelligibility data to test for the effects of test condition (natural vs. enhanced) varied across the consonants and the contexts, as mentioned above.

A review of the articulatory and acoustic characteristics of the consonants in both languages to see other possible effects of L1 background on the outcome showed that in Spanish the voiceless sibilant /s/ tends to be described as having an apico-alveolar place of articulation, whereas in English, it is lamino-alveolar, that is, it is the blade, and not the tip, the part of the tongue that gets close to the alveolar ridge. This articulatory difference between both sounds could produce a clear perceptual difference. Furthermore, Spanish /s/ can have different realisations; moving further back or forward within the alveolar region, and it is sometimes taken as /ʃ/ for English listeners (Stockwell & Bowen, 1965; Quilis & Fernández, 1973; Nash, 1977; Finch & Ortiz Lira, 1982; Alarcos Llorach, 1983 [1950]; Mott, 1996 [1991]; and Navarro Tomás, 1999 [1918], among others).

Also, another potential cause for the reduction in the effect of cue-enhancement could have the level of amplifications. First of all, it has also been reported in the literature (Gurlekian, 1979, 1981) that there are intensity differences in the fricative of both languages: Spanish sibilants tend to be produced with more intensity than the English ones (around 6 dB). As far as English is concerned, previous works reveal that the alveolar sibilants are more intense than the palato-alveolar ones (Han-Yong, 1979; Jassem, 1987; Behrens & Blumstein, 1988). Secondly, as regard to voicing differences, voiced sounds are believed to have an overall lower intensity than their counterparts due to a lower air pressure in the vocal tract (Jassem, 1987), or most manuals like Cruttenden (2008 [1962]) or Roach (2009 [1983]). Therefore, certain key regions could have been not adequately amplified, being necessary to apply different levels of amplification across sibilants, or even consonants regions, in order to avoid an increase in errors, as happened in the past with other sounds (Hazan & Simpson, 1998a, 1998b; Ortega-Llebaria & Huckvale, 2000).

All in all, Spanish listeners were clearly affected by their L1 production and perception, and maybe, the amplification of the enhancement (precisely 6 dB in the consonantal portion) could have led to an increase in consonant confusion in the case of /s/. It would be interesting to analyse the production and perception of these listeners, when talking in each language. A possible additional task could be the analysis of the effects of native-language background on consonant intelligibility by examining the kinds of features that might be confusable on the basis of the L1 background in order to find out whether these were resolved as a result of the enhancements. Information Transfer analyses (Miller & Nicely, 1955; Wang & Bilger, 1973) could be carried out in order to determine how well these three consonants were recognised in terms of the features of voicing and place of articulation in the different conditions.

The statistical analysis of other intervening variables showed that the effect of enhancement on the words was not related to their frequency of use in the L2 language. There were very common words, such as so, which had more problems of being identified than less common words, like sap, for instance. Also, there was no correlation between the other subject variables analysed (length of L2 learning, age of L2 learning, or age of the listener) and the results obtained in this experiment. However, as also shown

---

4 This variability in its realisation may be due to the lack of phonemic contrast in the palato-alveolar region.
in previous studies, there was clear individual variability in the effect of enhancement, with some listeners showing little to no improvement in score in the enhanced condition whilst others showed significant increases.

Finally, it is fair to say that overall, there was not much perceptual confusion among the realisations of the chosen words for the tested Spanish ensemble\(^5\). Most of Spanish listeners obtained ceiling (or near-ceiling) intelligibility scores. The number of people who had difficulties in consonantal identification below the 90% mark varied between 5 and 7, varying slightly on the two conditions (natural vs. enhanced): For the alveolar sibilants, that number was a bit higher in the enhanced realisations than in the natural ones, whereas for the palato-alveolar was the opposite. Furthermore, as also shown in previous studies, there was clear individual variability in the effect of enhancement, with some listeners showing little to no improvement in score in the enhanced condition whilst others showed significant increases.

Its explanation could be two-fold: Firstly, the person who uttered the words to be recorded was an expert phonetician and a teacher, and may have pronounced the test-words too clearly\(^6\). And secondly, and most probable explanation, the fact that stimuli were not presented under degraded listening conditions (noise) but in clear speech could have also been beneficial to these listeners. To the light of previous works (Bradlow & Bent, 2002; Smiljanic & Bradlow, 2007; or Grynpas et al., 2011), clear speech is not only an effective enhancement strategy for native listeners but also for those with an extensive experience with the sound structure of the target language, as could have been the case of our informants whose level of English may have been higher than the one required to participating in such an experiment. It would be interesting then to investigate the perception of listeners with clear differences in their English level, and under different conditions (clear vs. noise).

In sum, this experiment produced only a limited set of results; therefore, more detailed experiments are needed for conclusive answers. It is therefore suggested that the future cue-enhancement techniques need to be tailored to the context, the consonant itself, and the target group (listeners’ language background) so that it is possible to redirect listeners’ attention towards non-native contrasts adequately.

Nevertheless, as the findings of this study demonstrated that the cue-enhancement technique could provide improvement in some cases, such technique should clearly be considered very useful for training L2 learners. Its correct application needs further study with more contrasts, listeners and contexts to be able to make a clear statement about the effects of cue-enhancement.

5. Bibliographic References


---

5 This is the reason why we decided not to test English speakers to establish a comparison. The results of the test would have not been enlightening due to a likely ceiling effect.

6 Markham & Hazan explained in detail the effects of talker-factors on speech intelligibility (2004) which could be applied to our experiment.


Hazan, Valerie L. & Adrian P. Simpson, 1997: “Enhancing in the intelligibility of natural VCV stimuli:


Pisoni, David B., Scott E. Lively, & John S. Logan, 1994: “Perceptual Learning of Nonnative Speech Contrasts: Implications for Theories of Speech Per-


6. Appendix

The Sue/zoo/shoe Test

You will hear one word at a time, please, underline the word you identify from each pair. Please, do NOT leave any blank answer. GUESS when you are not sure. Thank you!

BLOCK 1

1. same shame
2. so show
3. said zed
4. sin shin
5. Sue shoe
6. said zed
7. same shame
8. seal zeal
9. sap zap
10. seen sheen
11. sink zinc
12. sin shin
13. sin shin
14. said shed
15. same shame
16. seal zeal
17. same shame
18. sap zap
19. sap zap
20. Sue zoo
21. sap zap
22. so show
23. so show
24. sip zip
25. said shed
26. seal zeal
27. sing zing
28. sink zinc
29. seen sheen
30. seen sheen
31. sip ship
32. said zed
33. seen sheen
34. sip zip
35. sing zing
36. sink zinc
37. Sue shoe
38. sip zip
Perception of natural and enhanced non-native contrasts in clear speech

39. sing  zing
40. sip   ship
41. seal  zeal
42. sink  zinc
43. said  zed
44. sing  zing

BLOCK 2
1. same  shame
2. Sue   shoe
3. sin   shin
4. sink  zinc
5. sing  zing
6. seen  sheen
7. sing  zing
8. Sue   zoo
9. same  shame
10. so   show
11. sink  zinc
12. sin   shin
13. sap   zap
14. said  shed
15. Sue   zoo
16. sin   shin
17. said  zed
18. sap   zap
19. said  zed
20. seen  sheen
21. same  shame
22. said  zed
23. seal  zeal
24. sing  zing
25. sip   ship

BLOCK 3
1. sap   zap
2. same  shame
3. so    show
4. said  shed
5. seal  zeal
6. sip   zip
7. said  zed
8. seen  sheen
9. sin   shin
10. seal  zeal
11. seen  sheen
12. sink  zinc
13. sing  zing
14. sing  zing
15. same  shame
16. same  shame
17. sin   shin
18. sap   zap
19. sin   shin
20. sap   zap
21. so    show
22. sin   shin
23. Sue   shoe
24. sap   zap
25. said  zed
26. so    show
27. seen  sheen
28. Sue   shoe
29. sap   zap
30. same  shame
31. seal  zeal
32. sing  zing
33. sip   zip
34. seen  sheen
35. so    show
36. sap   zap
37. said  zed
38. seal  zeal
39. Sue   shoe
40. same  shame
41. said  zed
42. seen  sheen
43. sap   zap
44. same  shame
45. sing  zing
46. Sue   shoe
47. said  zed
48. sink  zinc
49. so    show
50. said  zed

BLOCK 4
1. same  shame
2. seen  sheen
3. seen  sheen
4. sink  zinc
5. sing  zing
6. said  zed
7. sin   shin
8. seal  zeal
9. sin   shin
10. said  shed
11. said  shed
12. seen  sheen
13. Sue   zoo
14. sink  zinc
15. seen  sheen
16. sing  zing
17. same  shame
18. Sue   zoo
19. sap   zap
20. seal  zeal
21. same  shame
22. said  zed
23. sap   zap
24. so    show
25. sap   zap
26. sip   zip
27. sip   zip
28. same  shame
29. sink  zinc
30. sin   shin
31. so    show
32. sip   zip
33. sink  zinc
34. said  zed
35. Sue   zoo
36. sip   zip
37. seal  zeal
38. seal  zeal
39. so    show
40. Sue   zoo
41. so    show
42. sing  zing
43. sin   shin
44. said  zed
45. sip   ship
46. Sue   shoe
47. Sue   shoe
48. sip   ship
49. sing  zing
50. sap   zap