Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 100 (2007) 295–300
Brain and Language
www.elsevier.com/locate/b&l

The effects of inventory on vowel perception in French and Spanish: An MEG study

Valentine Hacquard *, Mary Ann Walter, Alec Marantz

Massachusetts Institute of Technology, USA

Accepted 20 April 2006
Available online 9 November 2006

Abstract

Production studies have shown that speakers of languages with larger phoneme inventories expand their acoustic space relative to languages with smaller inventories [Bradlow, A. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*(3), 1916–1924; Jongman, A., Fourakis, M., & Sereno, J. (1989). The acoustic vowel space of Modern Greek and German. *Language Speech, 32*, 221–248]. In this study, we investigated whether this acoustic expansion in production has a perceptual correlate, that is, whether the *perceived* distance between pairs of sounds separated by equal *acoustic* distances varies as a function of inventory size or organization. We used magnetoencephalography, specifically the mismatch field response (MMF), and compared two language groups, French and Spanish, whose vowel inventories differ in size and organization. Our results show that the MMF is sensitive to inventory size but not organization, suggesting that speakers of languages with larger inventories perceive the same sounds as less similar than speakers with smaller inventories.

Keywords: Magnetoencephalography (MEG); Mismatch field (MMF); Perception; Similarity; Vowel inventory

1. Introduction

It is well known that language-specific factors can affect and warp the perceptual distance between sounds. This explains for example why speakers of a particular language (e.g., English) easily differentiate two sounds if they cross a phonemic boundary (e.g., /l/ and /r/) while speakers of another language (e.g., Japanese) fail to differentiate those same sounds with the same ease (cf. Kuhl, 2000). While this warping of the perceptual space has been shown to operate at a local level (erasing boundaries between neighboring segments), the question arises as to whether language-specific factors can have an effect on perception at a more global level. This is the question we addressed in this study, by looking at the effects of the overall size and organization of vowel inventories on the perceptual space.

Production studies have shown how the crowdedness of an inventory can affect the acoustic distance between phonemes (Bradlow, 1995; Jongman, Fourakis, & Sereno, 1989), suggesting that speakers of languages with larger inventories (German or English) expand the acoustic space relative to speakers of languages with smaller inventories (Modern Greek or Madrid Spanish) by producing the same target phonemes acoustically further away from one another. However, little is known about the effects of inventory size on *perception*, due to the limitations of behavioral experiments (e.g., need to introduce noise to avoid plateau effects in confusion matrices (Peterson & Barney, 1952), task effects of offline similarity judgment tasks). Using magnetoencephalography (MEG), we tested the *expansion hypothesis* that there is a perceptual expansion of the acoustic space above and beyond the expansion in production.

MEG is a non-invasive brain-imaging technique which measures the strength of the magnetic fields associated with the electric currents in the brain and is able to provide a millisecond-by-millisecond picture of brain activity. The

* Corresponding author.
E-mail address: hacquard@mit.edu (V. Hacquard).

0093-934X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2006.04.009

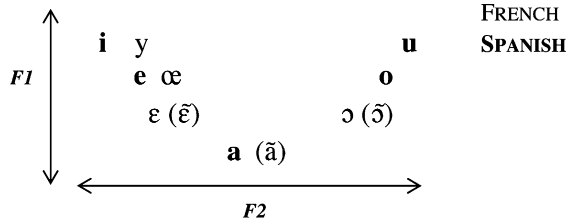

Fig. 1. Spanish (boldface) and French (boldface + normal) vowel inventories, schematized on a bi-dimensional acoustic space defined by the first two formants (nasals are represented in parentheses).

neural component we used as a measure of (dis-)similarity is the Mismatch Field (MMF), an automatic auditory brain response evoked by a deviant stimulus among a sequence of standards, peaking ~180–250 ms after the difference point between the deviant stimulus and the standard. This brain response is a good index of perceived linguistic similarity, with previous research having established its sensitivity to phonological category (Näätänen et al., 1997; Phillips et al., 2000; Sharma & Dorman, 1999).

The two languages compared here were French and Spanish, which differ in vowel inventory size and organization in crucial ways. The French vowel inventory is much larger than the Spanish one: it contains the five vowels of Spanish and an additional seven (Fig. 1). Furthermore, the extra French vowels are not uniformly distributed between the five Spanish ones. In order to isolate the effects of inventory, we controlled for acoustics and phonemic status: the same stimuli were used for both language groups (single acoustic values) and were designed to be native-sounding phonemes in both languages.

We looked at pairs of vowels using formant values that fell within phonemic boundaries for both languages and compared the mismatches elicited by these pairs. This paradigm allowed us to test whether the inventory size (number of vowels) had an effect on perceived similarity and/or whether the inventory organization did (distribution of phonemes within the vowel space).

2. Materials and methods

2.1. Participants

Twenty-three right-handed adults with normal or corrected-to-normal vision gave their informed consent to participate in the experiment (10 male and 13 female, evenly distributed between the 2 language groups). An additional 8 subjects were tested but did not show the normal MMF field distribution nor the normal M100 field distribution in a baseline tone test and were therefore not included in the analysis. Of the subjects analyzed, 10 were native speakers of French (Parisian), 13 were native speakers of Spanish (9 Argentineans, 4 Puerto Ricans). Following the experiment, participants read a list of sentences and nonce words to ensure that they spoke a dialect of French or Spanish that had the vowel inventories shown in Fig. 1. Participants were paid $20 for their participation.

2.2. Stimuli

Our stimuli were synthesized using the formant-based synthesizer HLsyn (Stevens & Bickley, 1991). We used a single token for each of the stimuli, with the formant values shown in Table 1. These values fall within the phonemic category for both languages, and were judged as native sounding by French and Spanish speakers prior to the experiment. We also verified with each subject post-experiment that the vowels sounded like vowels uttered by a speaker of his/her language.

2.3. Procedure

Subjects lay supine in a dimly lit magnetically shielded room in the MIT/KIT MEG laboratory. Stimuli were presented by Psyscope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993) binaurally over earphones. Evoked magnetic fields were recorded using a 93-sensor whole-head biomagnetometer array (Kanazawa Institute of Technology, Japan) for the first 7 participants. Sixty-three additional sensors were then added, offering 156 channels for the last 17 participants. Before beginning the experimental conditions, subjects listened to a 1 kHz tone presented 100 times, as a baseline test for the purpose of identifying and localizing the M100, an automatic auditorily evoked response, peaking ~100 ms post-stimulus onset, which produces a field pattern similar to that of the MMF.

To elicit mismatch responses we used a non-attentive paradigm. Participants were instructed to watch a silent movie, and not to pay attention to the sounds they heard through the earphones. The experiment consisted of two blocks: in the first block, the standard was /o/ and the two deviants were /a/ and /u/. In the second block, the standard was /e/ and the two deviants were /a/ and /ɛ/. Note that /o/, /a/, /u/, and /e/ are all phonemes in both languages, and while /ɛ/ is also a phoneme in French, /e/ and /ɛ/ are allophones in Spanish (they are both sounds produced by Spanish speakers, but do not form a contrast sufficient to establish a meaning difference). Stimulus presentation was randomized by Psyscope within each block, with a deviant to standard ratio of 15/105 (1050 standards and 100 of each of the two deviants for each block). The inter-stimulus interval randomly varied between 525 and 650 ms. Block 1 and Block 2 were presented in random order. Between blocks, the subject was given a break of self-determined duration. The whole experiment lasted approximately 50 min.

Table 1
Stimuli formant values

	*F*1	*F*2	*F*3
/o/	457	928	3077
/a/	737	1379	3080
/u/	284	847	3086
/e/	424	2114	3062
/ɛ/	553	1894	3082

Data were acquired at a sampling rate of 500 Hz. The raw data were subjected to a noise reduction routine to eliminate magnetic fields from external sources as measured with 3-axis magnetometer at a distance from the head. The data for each of the conditions were then averaged separately, in 600 ms windows keyed to the onset of the stimulus: 100 ms pre-, 500 ms post-onset. The averaged signal was subjected to a 30 Hz low-pass filter and adjusted to baseline using a 100 ms pre-stimulus interval.

2.4. Data analysis

The average waveforms for the standard and the deviants in each block were calculated by averaging the RMSs (Root Mean Square) of selected sensors covering the field pattern of the MMF component. The sensors were selected by visual inspection of the grand average across conditions and blocks for each subject. For each subject we chose 10 sensors in the left hemisphere, as the MMF has been shown to be strongly left lateralized when elicited by linguistic stimuli (Phillips et al., 2000). The mismatch waveforms were obtained by subtracting the standard from each deviant (/u/-/o/ and /a/-/o/ mismatches for Block 1, and /a/-/e/ and /ɛ/-/e/ mismatches for Block 2). The MMF peak amplitude was obtained by taking the highest peak amplitude value reached by this waveform subtraction during the MMF time window (130–230 ms) (as also done by Savela et al., 2003; Winkler et al., 1999).

3. Results and discussion

3.1. Predictions

3.1.1. Inventory organization

If inventory organization had an effect on perceived similarity, as measured by the strength of the mismatch response, both within-language and between-language variation would be predicted. Specifically, we would expect the Spanish /u/-/o/ mismatch to be roughly equivalent to the /a/-/o/ mismatch, as the relative space—relative in terms of intervening vowels—is the same between /o/ and /u/ as between /o/ and /a/ (see Fig. 1). On the other hand, we would expect the French /u/-/o/ and /a/-/o/ mismatches to differ. There is no intervening vowel between /u/ and /o/ in the French inventory, but // intervenes between /o/ and /a/; thus there should be a greater /a/-/o/ mismatch than /u/-/o/ for the French subjects. Between languages we would expect French and Spanish not to differ significantly for the /u/-/o/ mismatch but the French /a/-/o/ mismatch to be greater than the Spanish one. These predictions are in line with the /u/-/o/ and /a/-/o/ similarity values calculated for each language using Frisch's (1996) inventory-dependent metric, in which similarity is computed according to shared natural classes (these values differ for each inventory based on each language's inventory organization). The values calculated for this experiment were roughly equivalent for /o/-/u/ and /o/-/a/ in Spanish (.45 and .42), but /o/-/a/ were clearly less similar for French (.25), compared both to the French /o/-/u/ value (.48) and either Spanish value.

3.1.2. Inventory size

If inventory size had an effect on perceived similarity the two language groups should differ across the board, and they should differ in the same direction. The expansion hypothesis argues that speakers of languages with larger inventories expand their *perceptual* space in order to accommodate the larger number of segments, just as they expand their *production* acoustic space. This hypothesis thus predicts that the French speakers (with a larger inventory) should show larger mismatches (i.e., perceive segments as more dissimilar) than the Spanish ones. Note that the similarity values as given by Frisch's theory suggest that inventory size should not affect similarity.

3.2. Results

An ANOVA [2 (language) $\times$ 4 (vowel contrast)] was performed across both blocks and yielded a significant main effect of Language on MMF amplitude ($F(1,21) = 9.95$, $p = .005$). French subjects showed overall bigger MMF amplitude than Spanish subjects, suggesting that inventory size has an effect on the perception of similarity. Note that while the effect did not reach significance with the subject pool restricted to only those tested with additional sensors, it still showed a trend in the right direction ($F(1,14) = 3.38$, $p = 0.087$), with greater mean values for the French for each vowel pair. This suggests that the observed effect is not due to an asymmetrical distribution of subjects with respect to number of sensors. To ensure that French and Spanish speakers differed in their mismatch responses (i.e., how dissimilar two phonemes are to each other) and not in their response to sounds in general, and that the mismatch wave difference was not due to external factors, a one-way ANOVA was performed for each standard, which showed that French and Spanish speakers did not differ significantly in their response to the standards /o/ ($F(1,21) = 2.42$, $p = .134$) and /e/ ($F(1,21) = 1.111$, $p = .304$) in the MMF time window (130–230 ms).

Further post hoc comparisons showed a significant Language effect on MMF amplitude for /u/-/o/ ($p = .02$), /a/-/e/ ($p = .007$) only and no significant effect for /a/-/o/ ($p = .12$) and /ɛ/-/e/ ($p = .34$), though in each case the French mean response still exceeded the Spanish one (see Figs. 2 and 4).

A one-way ANOVA comparing the MMF amplitude of /a/-/o/ vs. /u/-/o/ for each language yielded no significant within-language effect of Condition (/u/-/o/ vs. /a/-/o/) for either French ($F(1,18) = .757$, $p = .396$) or Spanish ($F(1,24) = .01$, $p = .920$), contrary to the inventory organization predictions (Figs. 3 and 4).

3.3. Discussion

We found a main effect of language group on MMF amplitude, in line with the hypothesis that inventory *size* affects perceptual similarity. French subjects perceived the same vowels as more dissimilar than Spanish subjects,

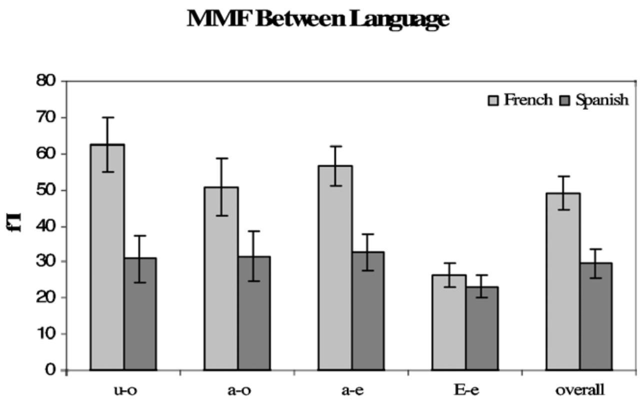

Fig. 2. Between-language comparison of the averaged mismatch wave peak amplitude for each condition, and overall.

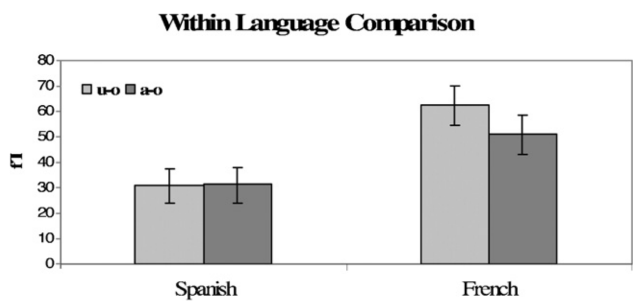

Fig. 3. Within-language comparison of the averaged mismatch wave peak amplitude for **/u/-/o/** and **/a/-/o/** for each language.

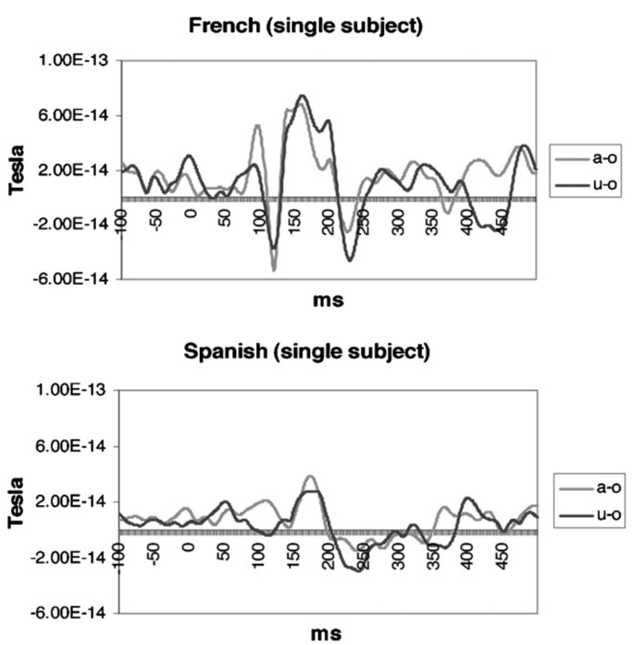

Fig. 4. **/u/-/o/** and **/a/-/o/** mismatch waveform for an illustrative subject of each language. No significant difference between condition for the **/u/-/o/** and **/a/-/o/** conditions for each subject, but the mismatch waves are overall larger for the French subject than the Spanish one.

suggesting that speakers of languages with larger inventories expand their perceptual space to accommodate their larger number of phonemes. If this is so, there would thus be a perceptual expansion above and beyond the expansion of the acoustic space attested in production. A possible explanation for this expansion comes from conflicting underlying tendencies in the production and perception of vowels, as evidenced in the cross-linguistic variation in vowel inventories. As we will see in the *Further Predictions* section, the lack of significance for the /ɛ/-/e/ pair (despite the difference in phonemic contrast) might support a particular view of this expansion.

3.3.1. Vowel inventories

The vowel inventory literature shows that, although there seems to be much cross-linguistic variation in inventory size and organization, some universal tendencies can be found. These tendencies have supported theories arguing that articulatory- and acoustic-based constraints shape vowel inventories. *Dispersion theory* (Flemming, 2002; Johnson, 2000; Johnson, Flemming, & Wright, 1993; Lindblom, 1986, 1990) articulates principles based on such constraints which predict that vowels will be maximally dispersed in the acoustic space, a prediction that has received some support in production studies (Bradlow, 1995; Jongman et al., 1989).

However, while there seems to be a trend for the production space to expand, Stevens (1972) shows that the extremities of the vowel space show more stability than the space within; that is, there is less acoustic variation in the production of point vowels (/i/, /u/, /a/) cross-linguistically than other vowels. It thus appears that a large inventory will have a tendency to disperse throughout the space, but that this dispersion is contained by the physical boundaries of that space at the extremes, in effect limiting the expansion at the production level.

This is where perception comes into play. Our results indicate that the perceptual space is stretched out for a language with a larger inventory, suggesting that languages with a large inventory compensate for the physical limitations of acoustic expansion in production by altering the perceptual space. The principles under Dispersion Theory drive speakers of languages with larger inventories to spread out their vowels in the acoustic space. However, this expansion is limited by physical constraints that prohibit the *production* of the point vowels beyond certain formant values. These limitations could then induce a distortion of the *perceptual* space in order to accommodate the numerous vowels of larger inventories, while still maximizing their distinctiveness (i.e., the perceived distance between vowels).

This warping of the perceptual space is schematized in Fig. 5: French and Spanish share the same absolute acoustic space: no vowel could be produced beyond what is physically possible. Although no formal production study of which we are aware has compared the French and Spanish production space, we hypothesize that the latter will be smaller, based on other comparative production studies. The two languages will then differ in their perceptual space, with the French one being more distorted in order to accommodate all of its vowels. Thus, the difference in perceptual distance between two vowels will be greater for French than Spanish.

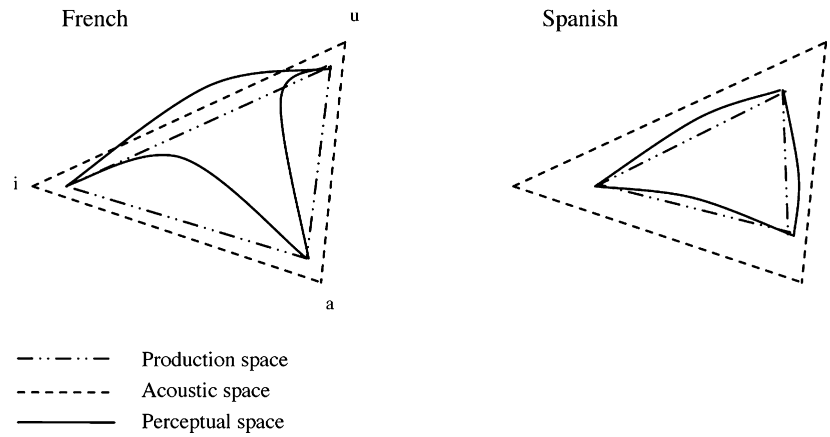

Fig. 5. Comparing the French and Spanish perceptual space: the more distorted the perceptual space, the greater the increase in distance between two phonemes as one moves towards the extremes of the space. The distance between two central vowels will be proportional smaller than the distance between a central vowel and a point vowel.

3.3.2. Further predictions—Mapping from the acoustic to the perceptual space

This model predicts that, overall, French speakers will perceive vowels as being more dissimilar from one another than Spanish ones. How does the perceptual space expand? The perceived distance between a given pair of vowels might increase proportionally, regardless of their position within the acoustic space. Alternatively, the distance might be enhanced to a greater degree toward the edges of the acoustic space. In that case, two central vowels produced with the same acoustic values in both languages will not seem to differ as dramatically as two point vowels. The results of the second block of this experiment seem to favor the latter option: we found that French speakers have bigger MMFs for /u/-/o/ and /a/-/o/ than Spanish speakers (i.e., perceive these sounds as more dissimilar), but that Spanish and French speakers do not differ significantly for a more central contrast, namely /ε/-/e/, *despite* the fact that /e/ and /ε/ are in phonemic contrast in French, but not in Spanish.

3.3.3. Other factors?

Given that phonemic status and acoustic values were kept constant, we have taken the effect to reflect the large difference in inventory size, an obvious distinction between the two language vowel inventories. As pointed out by a reviewer, another factor that might play a role is vowel token frequency. As the token frequency values in Table 2 show, /a/ (log probability = 3.83), is more frequent than /u/ (log probability = 5.44), itself more frequent than /o/ (log probability = 7.14) in French. In Spanish, /a/ (log probability = 3.09), is more frequent than /o/ (log probability = 3.53), itself more frequent than /u/ (log probability = 5.45). If token frequency played a role, we might expect within-language effects for either Spanish or French between the /u/-/o/ vs. /a/-/o/ contrasts, as also predicted if inventory organization were crucial. As reported above, we did not find such effects between these contrasts within either language, perhaps because they are too small to detect. It is not clear how a frequency-based account could explain the between-language contrasts. In French, the standard (/o/) is less frequent than either deviant, while in Spanish the standard is more frequent than /u/ but less frequent than /a/. Any frequency-based account of the larger mismatch response for the French as opposed to the Spanish speakers would also predict a difference in the mismatch between the two deviants in French not observed in this experiment.

Finally, we found no effect of inventory organization on the mismatch response. For both language groups the /u/-/o/ and /a/-/o/ mismatches did not differ significantly from one another. Furthermore, the /a/-/o/ mismatch did not differ significantly between languages. It would thus appear that either the effect of organization is too small to detect, or the MMF is not sensitive to it.

Table 2
Token frequencies in French and Spanish (French values from Hume and Bromberg, 2005, Spanish values obtained from the LEXESP Corpus (Sebastián et al., 2000), following the calculation procedure in Hume and Bromberg, 2005)

Phoneme	Log Prob	
	French	Spanish
o	7.14	3.53
a	3.82	3.09
u	5.44	5.45

4. Conclusion

While further pairs of languages and vowels will determine the ultimate validity of the expansion hypothesis, this study suggests that inventory size may indeed influence the amplitude of the mismatch field response, and therefore the perceived (dis-)similarity of phonemes, in line with the hypothesis that speakers of languages with larger inventories perceive the same sounds as less similar than languages with smaller inventories. We argued that these results reflected a warping of the perceptual space in order to accommodate the phonemes of a larger inventory, despite the fact that the actual acoustic distance remained the same across the two languages and that the vowels used were good prototypes of the phonemic categories. On the other hand, the relative distance between phonemes in terms of intervening sounds within each language had a definitely lesser effect, if any.

These results are in line with models of phonology that separate a (language-specific) perceptual space from a (universal) acoustic space. Just as limitations of the perceptual/articulatory systems constrain the make-up of grammars (phonological inventory), the grammar in turn constrains the perceptual/articulatory system.

Acknowledgments

We thank Corine Bickley for synthesizing our stimuli, and Adam Albright, Edward Flemming, Jim Harris, Colin Phillips, Liina Pylkkanen, Donca Steriade, and Linnaea Stockall for helpful discussion.

References

Bradlow, A. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*(3), 1916–1924.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: a new graphics interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257–271.

Flemming, E. (2002). *Auditory representations in phonology*. NY: Routledge.

Frisch, S. (1996). *Similarity and frequency in phonology*. Ph.D. Dissertation, Northwestern University.

Jongman, A., Fourakis, M., & Sereno, J. (1989). The acoustic vowel space of Modern Greek and German. *Language Speech, 32*, 221–248.

Hume, E., & Bromberg, I. (2005). Predicting epenthesis: an information-theoretic account. In *Seventh Annual Meeting of the French Network Phonology*.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language, 69-3*, 505–528.

Johnson, K. (2000). Adaptive dispersion in vowel perception. *Phonetica, 57*, 181–188.

Kuhl, P. K. (2000). Language, mind, and brain: experience alters perception. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 99–115). Cambridge, MA: MIT Press.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Orlando: Academic Press.

Lindblom, B. (1990). Models of phonetic variation and selection. *Phonetic Experimental Research*: Vol. 11 (pp. 101–118). Institute of Linguistics, University of Stockholm.

Näätänen, R., Lehtoskoskl, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175–184.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience, 12*.

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). The mismatch negativity and reaction time as indices of the percpetual distance between the corresponding vowels of two related languages. *Cognitive Brain Research, 16*, 250–256.

Sebastián, N., Martí, M. A., Carreiras, M., & Cuetos, F. (2000). LEXESP: Una base de datos informatizada del español. Barcelona: Servicio de Publicaciones de la Universitat de Barcelona.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America, 106*, 1078–1083.

Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In E. David & P. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). New York: McGraw Hill.

Stevens, K. N., & Bickley, C. (1991). Constraints among parameters simplify control of Klatt formant synthesizer. *Journal of Phonetics, 19*, 161–174.

Winkler, I., Lehtokoski, A., Alku, P., Vaino, M., Czigler, I., Csepe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7*, 357–369.