Title
Native Language Experience Influences the Perceived Similarity of Second Language Vowel Categories

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 31(31)

ISSN
1069-7977

Authors
Farmer, Thomas
Liu, Ran
Metha, Neha
et al.

Publication Date
2009

Peer reviewed

Native Language Experience Influences the Perceived Similarity of Second Language Vowel Categories

Thomas A. Farmer[1,2] (taf22@cornell.edu), Ran Liu[1] (ral2012@med.cornell.edu), Neha S. Mehta[1] (nam2029@med.cornell.edu), and Jason D. Zevin[1] (jdz2001@med.cornell.edu)

[1]Sackler Institute for Developmental Psychobiology—Weill Medical College of Cornell University
Box 140, 1300 York Ave. New York, NY 10021
[2]Department of Psychology, Cornell University, Uris Hall
Ithaca, NY 14853 USA

Abstract

The dynamics of spoken word recognition are acutely sensitive to competition among similar-sounding words. Here, we take advantage of this sensitivity to examine the manner in which native Italian speakers who are late learners of English perceive English vowels. Native Italian speakers and native English speakers listened to recordings of naturally produced words (”pin,” ”pen,” and ”pan”) and used a computer mouse to select the matching stimulus from an array of two pictures. The same participants also performed a similarity judgment task. The perceptual similarity space for these vowel categories differed between groups, and these differences were also reflected in the dynamics of performance in the online measure. The results are largely interpretable in terms of models of second-language speech perception that predict performance from patterns of assimilation to native language categories. The results suggest, however, that there are also effects of graded differences in the perceptual similarity of categories as measured in native speakers of the target language.

Keywords: Speech recognition, intercategory variability, perceptual similarity

Understanding speech depends on the ability to converge on an interpretation of infinitely variable physical stimuli as consistent with a finite range of known categories. The perceptual abilities that support speech perception continue to develop into childhood, and possibly early adolescence (e.g., Hazan & Barrett, 2000; Nittrouer, Crowther & Miller, 1998), but show evidence of being shaped by native-language experience relatively early in infancy (Werker & Tees, 1984).

When people acquire a second language (L2) later in life, the traces of early native language (L1) learning are clearly discernible in both their perception and production of L2 speech sounds (Flege, 1991). This phenomenon can be understood in terms of the operation of domain-general perceptual and learning mechanisms. For example, the Speech Learning Model (Flege, 1992; 1995) argues that as L1 phonetic categories become highly developed through childhood and adolescence, they form stronger “attractors” for acoustically and perceptually similar L2 sounds. Thus, for late L2 learners, L2 phonemes will tend to be perceptually assimilated into highly similar L1 categories. Furthermore, the SLM proposes that the discrimination of L2 contrasts is predicted by the perceived similarity between each sound in the contrast and the closest corresponding L1 category.

Viewed this way, between- and within-language similarity highly related, as both are determined jointly by distributional information about the physical properties of stimuli and learned categories. Most studies of L2 speech perception, however, have focused on aspects of L1 performance that foreground discrete, categorical outcomes, obscuring graded and continuous processes revealed by more sensitive, dynamic measures (e.g., McMurray, Tanenhaus, Aslin, & Spivey, 2003). Here, we study native Italian speakers' performance on a three-way vowel contrast (/I/ - /ɛ/ - /æ/) using measures designed to capture graded and continuous effects of category similarity on the dynamics of performance.

The English vowels /ɛ/ and /æ/ are highly similar to the Italian vowel /ɛ/ whereas the English vowel /I/ is most similar to the Italian vowel /i/, both in terms of their acoustic properties and how they are perceived by Italian speakers (Flege & MacKay, 2004). These findings, interpreted in the context of the SLM, suggest that native Italian speakers’ performance on speech categorization tasks should reflect much poorer discrimination between /ɛ/ and /æ/ (they should assimilate to the same L1 category) than discrimination between the /ɛ/-/I/ or /æ/-/I/ pairings (each vowel assimilates to a different L1 category).

In studies of cross-linguistic speech sound categorization, L2 perception is typically assessed by comparing the performance of bilinguals in their L2 with the performance of monolingual native speakers of the same language. That is, monolingual native performance is viewed as the benchmark against which non-native performance is assessed, typically using discrete dependent variables such as error rate, for which native speakers are at ceiling. Thus, current approaches to L2 perception have focused only on the difficulties that non-native speakers experience in learning L2 rather than on the details of how non-native and native performance differ. While this approach can identify particular contrasts that are difficult for non-native listeners to perceive relative to native controls, it does not provide much insight into the nuances of native listeners’ perception and how these might relate to L2 perception.

We sought to characterize differences between Italian speakers and native English speakers in their perception of English vowel categories by combining metalinguistic perceptual similarity judgments with online, graded measurements of the speech categorization process. In perceptual similarity judgments, listeners are asked directly how similar they perceive stimuli to be, and these pair-wise judgments are analyzed with multidimensional scaling to map perceptual space (Shepard, 1980; Terbeek, 1977). This approach provides a rich description of the similarity of multiple stimulus classes that can be compared across groups of listeners. Metalinguistic judgments of stimulus similarity, however, may not reflect the processes that subserve speech perception under more ecological task goals.

Applying finer-grained, dynamic measures of online performance can provide a rich description of the relative similarity of speech sound categories without the drawbacks of metalinguistic judgments. For example, data from an eye-tracking study using a four-alternative choice paradigm have revealed graded patterns of similarity for sets of English speech sounds under conditions that produced essentially error-free performance in the choice data (Zevin, Farmer, & McCandliss, submitted). In that study, we found that although categorization responses were perfectly consistent, patterns of eye-movements to distractor stimuli before response execution were not randomly distributed, but depended on the phonetic similarity of the target and distractor.

A newer technique, “mouse-tracking” (Spivey, Grosjean, & Knoblich, 2005; see also Dale, Kehoe, & Spivey, 2007; Farmer, Anderson, & Spivey, 2007) permits us to observe the influence of confusability on the dynamics of response execution itself. Although individual saccadic eye movements can occasionally show some curvature (Doyle & Walker, 2001), individual movements of the arm and hand can show quite dramatic curvature (Tipper, Howard, & Jackson, 1997) that can be interpreted as the dynamic blending of two mutually exclusive motor commands (Cisek & Kalaska, 2005). Additionally, whereas eye-movement data allow for approximately 2-3 data points (saccades) per second, mouse-tracking yields somewhere between 30 and 60 data points per second. In light of the ability to record many data points per second, along with their ability to curve mid-flight in response to continuous dynamic competition between target and distractor, mouse movements have the unique ability to illuminate the spatio-temporal dynamics of the categorization process itself, not only the product of it. Thus, mouse-tracking complements the perceptual similarity data by providing analogously graded, continuous measures of performance in an online task. Together, these measures will permit us to make detailed observations of the influence of native language on the perceptual similarity space of vowel categories, and how this impacts performance in a task that approximates some aspects of language use "in the wild."

Method

Participants Sixteen right-handed native-English (EL1) speaking adults (with a mean age of 27.56 years, *SD*=4.91) and 19 native-Italian (IL1) speakers (mean age 31.22 years, *SD*=3.67) participated in this study. All spoke dialects in which “pin” and “pen” are treated as a minimal pair.

For native Italian speakers, the average Age of Arrival (AoA) was 25.61 years (*SD*=4.49), producing an average length of residence in the United States of 3.94 years (*SD*=3.26). Fifteen IL1 participants were from northern Italy, three were from Rome, and one was Sicilian. The IL1 participants reported using English an average of 64.39% (*SD*=23.48%) of the time while in the US.

Materials The stimuli were presented using Macromedia Director MX, and mouse movements were recorded at an average sampling rate of 40 Hz. The display resolution was set to 1024 x 768. Spoken target stimuli (pin, pen, and pan) were digitized using a SONY minidisc recorder, with each stimulus saved in a monaural 44,100 Hz WAV file with 10ms before and after the stimulus word using Praat, with 10ms rise/fall time to prevent background noise from mimicking a consonantal burst. For each word, five different recordings were selected so that for each stimulus type, variation in fundamental frequency, harmonicity, word duration, and overall amplitude were matched in order to minimize the influence of the idiosyncrasies of any particular recording on performance. The same stimuli were used in both tasks. See Figure 1 for details regarding the vowel portions of the stimuli.

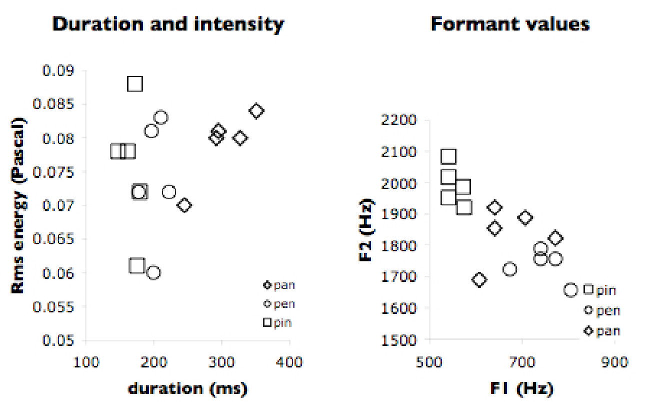

Figure 1: Acoustic measures of the steady-state vowel portions of the stimuli.

For the mouse-tracking portion of the study, images representing each critical item were prepared from public domain images found on the internet, resized to 2” (in the longest dimension) at a resolution of 72 dpi in Photoshop. On any given trial, one of the three target words was heard while viewing a scene that contained both a picture of the item denoted by the target word and a picture of one of the two remaining items (the distractor). One object appeared in the top left-hand corner of the display and the other appeared in the top right-hand corner (with the center of the image occurring at approximately 1.15” from its respective side of the display and 1.5” from the top). Additionally, at the bottom center of the display, a “Start” box approximately 1.5” x .33” appeared, along with a vertical transparent rectangle (.5”x 1”) centered above the start box.

Procedure Participants were asked to make themselves comfortable in front of the computer screen, adjusting the mouse to a location on the right-hand side that suited them. All participants first completed the mouse-tracking portion of the study, followed by the similarity judgment task.

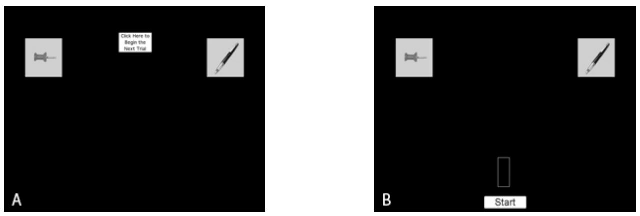

Figure 2: Example of the visual stimulus display for “pen” and “pin,” A. during the “preview” period and B. at the beginning of a trial.

Similarity Judgment. At the beginning of each trial, a start-box appeared at the bottom center of the screen, and four boxes appeared at the top of the screen. Each box contained a number between 1 and 4, signifying the degree to which the two stimuli to be heard sounded similar, and was accompanied by a verbal description (1=Exactly the Same; 2=Somewhat Similar; 3=Somewhat Different; 4=Completely Different). Upon clicking the start-box, it disappeared and the sound-file for the first target word began after a 500-ms delay, followed by the second target word. The stimulus onset asynchrony was 750ms, producing an approximately 500ms inter-stimulus interval. After the onset of the second target word, participants were able to click one of the four boxes to denote their similarity rating, thus ending the trial.

Mouse Tracking. At the onset of each trial, participants were presented with a black screen. After a 500ms delay, the target and the distractor appeared (at the top-left or top–right), each surrounded by a 2” x 2” light grey square. After another 500ms delay (used to provide a brief preview of the objects appearing in each location), a small box appeared in between the target and the distractor, containing the instructions “Click here to begin the next trial” (Figure 2A). Immediately after clicking the initiation box, a “Start” button appeared at the bottom center of the display (Figure 2B). Participants then clicked the “Start” button and moved the mouse up through the transparent rectangle. When the mouse entered the rectangle, the sound-file containing the target word was cued, signaling to participants the correct object on which to click. After clicking on either one of the two objects in the display, the display disappeared, thus ending the trial.

Stimulus presentation was arranged for each of the three possible target-distractor pairings (pin/pen; pin/pan; pen/pan), so that for each participant, each speech stimulus type occurred as both the correct and incorrect response, and with the corresponding images occurring in both possible locations. This set-up produced 60 experimental trials in a counter-balanced within-subject design: 3 (target-distractor pairing) X 2 (target 1 versus target 2) X 2 (target-left versus target-right) X 5 (spoken target-word token). Presentation order was randomized for each participant.

Results

Similarity Judgment Data

Rated similarity for each contrast was analyzed directly in one-way ANOVAs, with subjects (F_1) or items (F_2) as the random variables, and pair-type (pin/pen, pen/pan, and pin/pan) as the IV, revealing a significant effect of pair-type for the EL1 group, $F_1(2,30)=41.58$, $F_2(2,72)=104.18$, both p’s <.0005. Note that ratings were scaled so that low values reflect greater similarity, so that the lowest similarity was observed between pin and pan, and the greatest between pen and pan (Figure 3). Pair-wise comparisons on the EL1 data revealed reliable differences between the pen/pan pair (mean dissimilarity = 3.03) and the pen/pin pair (M=3.64), $t_1(15)=5.93$, $t_2(48)=9.74$, $p<.005$, and between pin/pen and pin/pan pairings (M=3.80), $t_1(15)=2.96$, $t_2(48)=3.42$, $p<.05$.

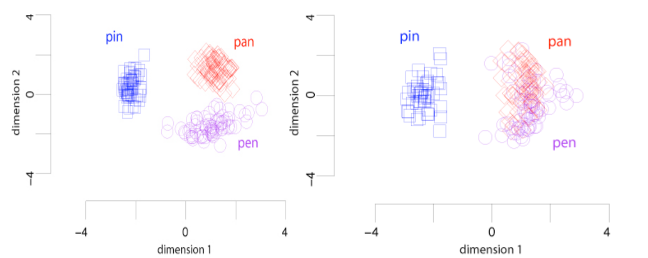

Figure 3: Perceptual distance between all stimuli for all EL1 (left) and IL1 (right) subjects Similarity judgments were entered into an isometric MDS analysis with two dimensions. Data are presented from all subjects in the same space

There was also a significant main effect of pair type for the IL1 group, $F_1(2, 26)=107.09$, $p < .0005$, $F_2(2, 72) = 355.12$, $p < .0005$. Like the EL1 speakers, there were significant differences in perceived dissimilarity between the pen/pan pairing (M=2.35) and both the pin/pen pairs (M=3.71), $t_1(13)=10.24$, $p < .0005$, $t_2(48)=19.17$, $p< .0005$, and the pin/pan pairs (M=3.77), $t_1(13)=10.85$, $p< .0005$, $t_2(48)=17.43$, $p < .0005$. Unlike the EL1 group, however, there was no significant difference in perceived dissimilarity between the pin/pan and the pin/pen pairs, both p’s > .10.

Mouse Movement Data

Data Screening and Coding Mouse movements were recorded for the entire duration of each trial, starting with the trial-initiating click on the Start button. Target-selection accuracy was high across each condition for the EL1 group, but was substantially lower for the IL1 group. The number of errors per each of the three pairings and each of the three target vowels is presented in Table 1.

As in previous studies (Farmer, Anderson, & Spivey, 2007), all remaining trajectories were visually inspected for obvious sporadic movements (loops, stops, etc.). Only one such trajectory was identified for the EL1 group, and 28 were identified in the IL1 group. Each error trial, along with each sporadic trajectory, was excluded from all further analyses.

Table 1: Error rates (incorrect object selected) across Conditions

	I/E Pair	I/A Pair	E/A Pair	Target "pIn"	Target "pEn"	Target "pAn"
EL1 Group						
Sum	6	0	7	6	3	4
(*M, SD*)	(.4, .7)	(0, 0)	(.4, .9)	(.4, .7)	(.2, .5)	(.3, .8)
IL1 Group						
Sum	11	5	99	13	55	47
(*M, SD*)	(.6, .8)	(.3, .6)	(5.2, 3.4)	(.7, 1.1)	(2.9, 1.8)	(2.5, 2.3)

As evident in Table 1, the task was considerably more difficult for the IL1 speakers. This difficulty was especially prevalent in the pen/pan condition, and in fact, five participants selected the incorrect object on over 40% of those trials. Due to the fact that they are basically at chance in that condition, their data were excluded from all subsequent analyses, leaving fourteen IL1 speakers.

Given the potential for trial-by-trial variability in trial duration, all analyzable trajectories were time-normalized to 101 time-steps by a procedure originally described in Spivey et al., 2005. All trajectories were aligned so that their first observation point corresponded to (0, 0). Then, across 101 normalized time-steps, the corresponding x,y coordinates were computed using linear interpolation.

Trajectory Divergence Analyses Mean trajectories for all three target/distractor pairs—with left-branching trajectories reflected in the y-axis—are shown in Figure 4. To determine whether any pair-wise divergences observed across the trajectories from each vowel contrast were statistically reliable, we first conducted a series of paired-sample t-tests. As per Spivey et al. (2005), we viewed the x-coordinates of the trajectories to be a strong index of spatial attraction toward either the competitor or the target object. As such, the t-tests were conducted across the x-coordinates of each possible vowel contrast, separately, at each of the 101 time-steps. In order to avoid the increased probability of a Type-1 error associated with multiple comparisons, and in keeping with Bootstrap simulations of such multiple t-tests on mouse-trajectories (Dale et al., 2007), an observed divergence was not considered significant unless the coordinates between the two vowel-contrast conditions diverged significantly (p<.05) for at least eight consecutive time-steps.

For the EL1 group, the x-coordinates elicited by the pen/pan contrast were significantly closer to the competitor (left-ward) object than for the pin/pan vowel contrast from time-steps 27-91 (all t's >2.13, p's <.05), or for the pin/pen contrast from time-steps 68-83, all t's > 2.12, p's <.05), consistent with the pen/pan contrast producing the highest degree of competition. Responses in the pin/pen contrast also diverged significantly from the pin/pan responses from time-steps 68-94 (all t's >2.13, p's <.05), indicating an intermediate degree of competition.

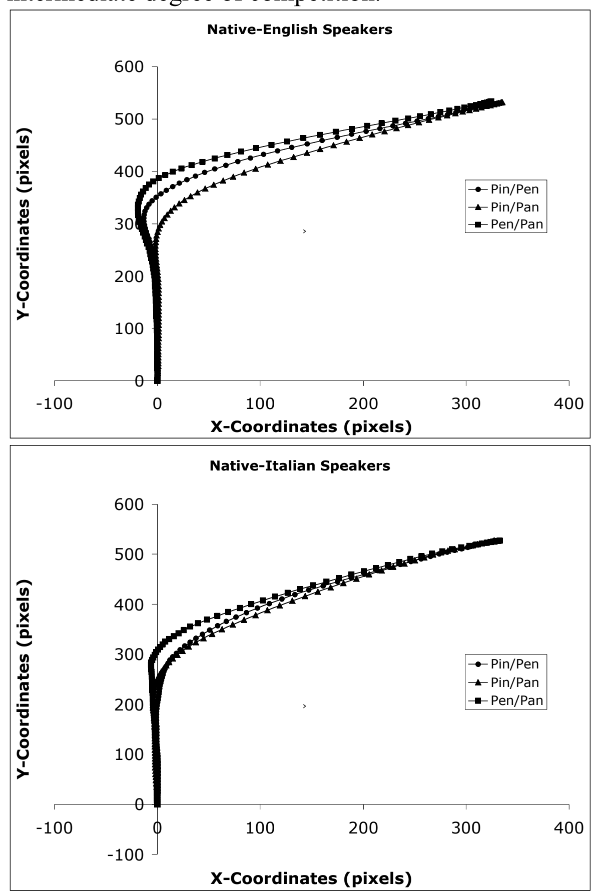

Figure 4: Mean trajectories for all three phonetic contrast conditions (with left-branching trajectories "flipped" for averaging).

Like the EL1 group, the pen/pan contrast was the most difficult for the IL1 participants. Average x-coordinates elicited by the pen/pan condition were significantly more left-ward (closer to the competitor / farther from the target object) than they were in the pin/pan condition, from time-steps 66-89 (all t's > 2.17, p's < .05) or in the pin/pen condition, from time-steps 67-94 (all t's > 2.19, all p's < .05). Unlike the EL1 group, however, there was no differential degree of difficulty in correctly categorizing the target between the pin/pen and pin/pan conditions, p's > .05 at each of the 101 time-steps.

General Discussion

The perceptual similarity space for front vowels differs strikingly between EL1 and IL1 speakers, and these differences were reflected in the dynamics of performance in an online word recognition task. Thus, in line with previous studies, the results presented here demonstrate a tight-knit relationship between the perceived similarity of vowel sounds and the ease with which they are categorized. Further, both the metalinguistic similarity judgments and online categorization task revealed graded, continuous patterns of perceived similarity that provide a rich description of how perception is shaped by both linguistic experience and acoustic properties of the stimuli.

For EL1 speakers, graded differences in vowel height influenced the perceptual similarity space, such that there were differences in similarity between adjacent pairs of vowels: "pen" and "pan" were judged to be more similar than "pin" and "pen" (though both were judged to be more similar than the non-adjacent pair, "pin" and "pan"). In the IL1 data, "pin" was equidistant from the two lower vowels. For both languages, however, the greatest similarity was found for the vowels in "pen" and "pan," with the similarity between these two categories being much greater for the Italian speakers.

Multidimensional scaling revealed how the similarity space for IL1 and EL1 speakers is organized differently (Figure 3). In both groups, "pin" is separable from the other two categories along the first dimension of the MDS solution, and "pen" and "pan" have different distributions along the second. For the IL1 speakers, however, the contrast along the second dimension is probabilistic rather than categorical. This is true even though participants included in this analysis were selected to have relatively high accuracy in identifying these vowels in the online task, and suggests that the IL1 speakers have failed to completely form a category boundary driven by this secondary dimension, although they are sensitive enough to it to distinguish these vowels from one another most of the time.

The similarity space derived from metalinguistic similarity judgments was also reflected in the dynamics of arm-movements during the online word recognition task. For EL1 speakers, a graded influence of similarity was observed, whereas for IL1 speakers, the contrasts between "pin" and the two other words did not differ in the degree of attraction toward the unselected competitor.

The graded measures of performance used here permit further insight into how perception differs between groups. First, the organization of the similarity space for EL1 participants shows that, although the stimuli were selected to vary along the single phonological dimension of "height," they are discriminated along two separable perceptual dimensions, and that the second of these dimensions is less strongly contrastive for both groups (and only probabilistically so for the IL1 group). The arm movement data reveal that stimuli contrasting on this second perceptual dimension are perceived in real time as more ambiguous by both groups, even when they are correctly identified. Thus, the measures go beyond showing that IL1 listeners are worse at some contrasts than others, to show that some patterns of difficulty are in fact latent in perceptual properties of the contrasts to which EL1 listeners are also susceptible. The overall pattern of results is highly consistent with the Speech Learning Model: Italian speakers' performance was strongly influenced by similarity between L2 and L1 categories, in that the two vowels most likely to assimilate to the same category showed the strongest evidence for competition, even when considering only trials on which accurate responses were made. Our data reveal another influence on L2 performance: The perceived similarity of categories for native speakers of the target language.

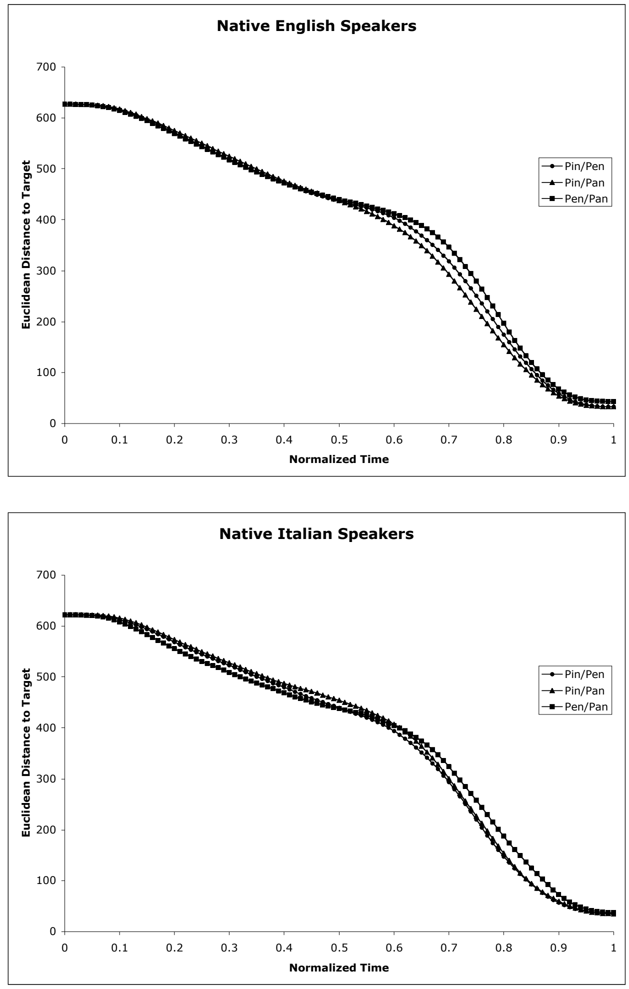

Figure 5: Mean Euclidean distance, across normalized time, to the correct target.

One aspect of the arm-movement data in Figure 4 that may seem incompatible with the SLM, however, is the fact that overall, the on-line categorization task appears to have been more difficult for the EL1 group. Indeed, the y-coordinates are much higher for the EL1 group, and the x-coordinates veer further to the left in the earlier part of the trial. In order to examine the x- and y-coordinate effects together, we calculated the Euclidean distance, across time, between the normalized trajectory and the center of the target object. In Figure 5, it is apparent that the task is not easier for the IL1 group, per se, but that differentiation between average trajectories occurred earlier (by approximately 18%) for the EL1 group than for the IL1 group. That is, competition between categories was more protracted for the non-native English speakers, reflecting different spatio-temporal properties of average responses, across groups.

For the most difficult pair studied here ("pen" vs. "pan") the Italian pattern could fairly be categorized as quantitatively weaker contrast along the same perceptual dimension that is least contrastive for English listeners. Thus, performance in this condition reflects an exaggeration of ambiguity that also has a clear influence on native performance. Interestingly, Italian speakers were less sensitive than English speakers to graded differences in inter-category similarity for these pairs. This can also be understood in terms of the pattern of assimilation to L1 categories. Because "pen" and "pan" are perceived—at least part of the time—as having the same vowel, they can be treated as equivalent, especially when they appear as graphically presented distracters along with an auditory stimulus that is assimilated to a different category. This apparent advantage for non-native speakers is striking, but may be illusory. The functional significance of maintaining inter-category similarity information during word recognition is not well understood, but it may be that divergence from native-like patterns of perceptual similarity puts non-native speakers at a disadvantage, for example when trying to understand speech under non-ideal conditions (Mayo, Florentine & Bus, 1997) or when attempting to accommodate for an unusual accent (Nygaard, 2005).

Acknowledgments

This research was supported by NIH grant R01-DC007694 and a Delores Z. Liebmann Fellowship to TF. We thank Bruce McCandliss for input on the experimental design, and Giuseppe Vezzoli for assistance in data collection.

References

Cisek, P., & Kalaska, J. (2005). Neural correlates of reaching decisions in dorsal premotor cortex. *Neuron, 45*, 801–814.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing exemplars. *Memory and Cognition, 35*, 15-28.

Doyle, M., & Walker, R. (2001). Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors. *Experimental Brain Research, 139*, 333-344.

Farmer, T. A., Anderson, S. E., & Spivey, M. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57*, 570-595.

Flege, J. E. (1988). The production and perception of speech sounds in a foreign languages. In H. Winitz (Ed.), *Human communication and its disorders: A review* (pp. 224–401). Norwood, NJ: Ablex.

Flege, J. E. (1991). Orthographic evidence for the perceptual identification of vowels in Spanish and English. *Quarterly Journal of Experimental Psychology, 43,* 701–731.

Flege, J. E. (1992). Speech learning in a second language. In T. Heubner, C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and application* (pp. 565-604). Timonium, MD: York Press.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.

Flege, J. E., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1-34.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6 to 12. *Journal of Phonetics, 28*, 377-396.

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.

McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research, 32*, 77-97.

Nittrouer, S., Crowther, C. S., & Miller, M. E. (1998). The relative weighting of acoustic properties in the perception of [s] + stop clusters by children and adults. *Perception and Psychophysics, 60*, 51-64.

Nygaard, L. C. (2005). Linguistic and paralinguistic factors in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception*. Oxford, England: Blackwell Publishers.

Shepard, R. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*, 390-398.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proc. National Academy of Sciences, 29*, 10393-10398.

Terbeek, D. (1977). WPP, No. 37: A cross-language multidimensional scaling study of vowel perception. In Working papers in phonetics. Los Angeles, CA: Department of Linguistics, UCLA.

Tipper, S. P., Howard, L. A., & Jackson, S. R. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition, 4*, 1-38.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7*, 49-63.

Zevin, J. D, Farmer, T. A., & McCandliss, B. D. (submitted). Confusion of English speech sounds by native and non-native English speakers in a visual-world experiment.