Selective attention and English listeners' perceptual learning of the Polish post-alveolar sibilant contrast

Grant McGuire

University of California at Santa Cruz

Correspondence:

Grant McGuire
Department of Linguistics
University of California at Santa Cruz
Stevenson Faculty Services
Santa Cruz, CA, USA 95064
Email: gmcguir1@ucsc.edu
Telephone: 831-459-4988

Abstract

Speech categories are highly multidimensional and an understanding of how listeners weight these dimensions is vital. This study tests the hypothesis that weighting occurs through the allocation of attentional resources to a relevant dimension away from irrelevant dimensions (attention-to-dimension, Francis and Nusbaum 2002) using a training paradigm in which listener's learned the Polish post-alveolar sibilant contrast. Specifically, subjects were trained to categorize a highly naturalistic consonant place distinction in a two-dimensional sibilant fricative + vowel stimulus set varying by fricative and vocalic information. Subjects were trained to use either the fricative noise, the vocalic cues (e.g. formant transition), or both cues independently to classify the stimulus set. Results from discrimination and classification tasks indicated that subject's attention could be directed favor one cue over another and that such training resulted in increases in sensitivity across boundaries to trained dimensions as well as increases in sensitivity within category for difficult contrasts. However, no evidence for reduced sensitivity was found.

Introduction

A fundamental process in speech perception is the proper identification and weighting of the cues available in the speech signal. This is not a trivial matter as the speech signal is highly multidimensional; there are many aspects of the signal and the listener must determine which are the most valuable. Once properly weighted, listener's must evaluate those dimensions along which categories are disambiguated. In this vein, this paper addresses cue weighting by training listeners to attend to different dimensions of a sibilant contrast and examining how their sensitivity to those dimensions changes as a result.

Most phonetic contrasts are correlated with multiple aspects of the speech signal and many of these vary in tandem. For example, there are 16 correlates (or acoustic characteristics) of voicing for intervocalic stops in English (Lisker, 1986), each of which could potentially be useful to listeners when classifying them. It is up to the listener to determine which are the most useful and classify them accordingly. When presented with a non-native contrast, listeners often fail or have great difficulty in discrimination and identification tests, often due to a failure to attend to the proper cues.

A classic demonstration of a failure to adequately weight cues can be found in work on the English /r/~/l/ contrast. This is a contrast notoriously difficult for Japanese listeners, who struggle in discriminating and identifying the sounds even with training (Strange and Dittman 1984; Logan, Lively, and Pisoni 1991). The likely reason for this difficulty comes from their over-reliance on second formant frequency; which is in contrast to English listeners who heavily weight the much more reliable (for this contrast) third formant frequency (Miyawaki et al. 1975; Iverson et al. 2003). Similarly, in a perceptual learning study, Francis and Nusbaum (2002) demonstrate that English listeners primarily rely on VOT to classify the Korean initial stop contrast, but after categorization training they use F0 as well as other contrast dimensions – behaving much much more like native Korean speakers.

However, even for native categories cue-weighting is not fixed, but changes during development. A series of studies by Susan Nittrouer and colleagues (e.g. Nittrouer 1992, Nittrouer and Miller 1997, Nittrouer 2002) found that young children generally weigh transition cues heavily when discriminating place in sibilant fricatives (/s/ and /ʃ/), gradually reaching adult-like weighting of fricative noise as more informative than transition around 7-8 years old. Nittrouer (2002) further demonstrated that for the /f/ ~ /θ/ distinction children predictably used cues like adults, who weight formant transition much more heavily than fricative noise. Nittrouer's general conclusion is that children initially give more attention to dynamic cues rather than static ones; only through experience with the speech sounds and a growing awareness of phonetic structure do children re-weight the cues.

Knowing the proper cues and weighting them appropriately has certain consequences for the perceptual system. In visual perception studies it has been demonstrated that subjects show an increased sensitivity along dimensions which they are trained to use and a concurrent reduction in sensitivity along untrained dimensions they were previously sensitive to (Goldstone 1994, Livingston, Andrews, and Harnad 1998). Similar results have also been found for non-speech categories (Guenther, Husain, Cohen, and Shinn-Cunningham 1999), as well as speech categories, such as burst and formant transitions in English onset stops (Francis, Baldwin, and Nusbaum 2000). Results such as these have led researchers to propose attention-to-dimension (A2D) models of perception where the perceiver is aware of the relevant dimensions and learning requires changes in sensitivity to those dimensions (see e.g. Nosofsky 1986, Goldstone 1994). As noted by Francis and Nusbaum (2002), while such models account for many cases of perceptual weighting, perceivers may also begin attending to dimensions they were previously unaware of. With this caveat, the general concept that listeners direct attentional resources to relevant dimensions and that listeners' sensitivity to these dimensions increases is an appealing one that accounts for much of the perceptual learning literature.

If such models are in fact true, then specific training with a dimension should heighten sensitivity to that dimension while dimensions that are not used are reduced in sensitivity. To this author's knowledge, no study outside of the visual perception literature has demonstrated in a single set of experiments both the shift in weighting and concurrent changes in attention that these models predict. Therefore it is a goal of the current study to examine both concepts within individual subjects.

Place cues for fricatives were chosen as the object of this study as they have two main dimensions of contrast that can be manipulated and for the fact that English listeners have a demonstrated ability to use both of them. Specifically, fricatives have two basic dimensions available: frication noise, deriving from a narrow constriction in the vocal tract which produces turbulent airflow that results in aperiodic noise having a spectral shape dependent on the distance from the constriction to the lips (supplemented by higher frequency noise contributed by the airflow against an obstruction, in the case of sibilants), and the formant transitions, which are determined by the movement of the tongue into and out of the consonant gesture (Johnson 2003). English listeners differentially weight these cues depending on the contrast, preferring frication noise for the sibilant contrast /s/~/ʃ/ and formant transitions for /f/~/θ/ (Harris, 1955).

The specific contrast used in this study is the Polish post-alveolar voiceless sibilant contrast. This particular contrast is generally described as a retroflex /ʂ/ versus alveolopalatal /ɕ/ place contrast (Zygis & Hamann, 2003, Nowak 2006). These sounds are highly confusable for English listeners (Lisker 2001, McGuire 2007[chapter 3], Padgett and Zygis, 2010) and the formant transitions appear to be more salient than frication noise for these listeners (McGuire 2010).

The current study was designed as follows and was inspired by Goldstone 1994. In experiment 1 English listeners were trained to categorize two dimensional continuum of the /ʂ/~/ɕ/ contrast using only the variation in fricative noise spectrum and to ignore variation in the vocalic dimension. In experiment 2 a different set of listeners were trained to categorize using only the variation in the vocalic dimension. In experiment 3, listeners were trained to use both dimensions separately. In all three experiments the listeners first participated in a discrimination and classification pre-test to identify their initial sensitivity and weighting scheme for the stimuli. The design of the discrimination tests allowed for changes in sensitivity within categories and across category boundaries to be measured. After testing, subjects trained for one two three sessions to categorize the stimuli; this was then followed by post-testing which was identical to the pre-testing. A fourth experiment, acting as a control, uses only the testing conditions without training to compare the relative benefit of training to the previous experiments.

Experiment 1: Fricative dimension categorization

This experiment was designed to train English listeners to categorize a two-dimensional [ʂa] to [ɕa] continuum using only the fricative noise information. In this experiment tokens from a two-dimensional stimulus space were assigned to two categories separated such that only one dimension, based on frication noise, is necessary for categorization. Although subjects hear the full range of variation in both dimensions in training, only the fricative dimension was relevant to correct classification.

Materials

The stimulus set was the same as described in McGuire (2007) and consisted of a two-dimensional space varying in fricative noise and vocalic information. Specifically, several productions of [ʂa] and [ɕa] were recorded in a sound-proof booth by a male native speaker of Polish using a head-mounted microphone and a Marantz PMD670 solid state recorder at 44.1 kHz sampling rate. One example of each syllable was selected based on clarity and similarity to the acoustic analyses of Polish fricatives reported in Nowak (2006). The selected retroflex syllable had a peak located at 2890 Hz and an F2 onset at 1420 Hz with a midpoint F2 of 1280 Hz. The alveolopalatal had a peak located at 3890 Hz with an F2 onset of 1720 Hz and a midpoint F2 of 1320 Hz.

Each syllable was split in two at the boundary of the fricative and vocalic portions, determined by the onset of voicing. The two fricatives were brought to the same length by excising at zero crossings 32ms from [ɕ] in four 8ms chunks located at 20% intervals of the total length[1]. The vowels were modified using Praat (Boersma and Weenink, 2002) to have the same length, pitch, and RMS through PSOLA resynthesis. Both the fricative and vowel portions were then separately interpolated to form fricative and vowel continua consisting of 10 steps where each step was one of ten graded proportions in terms of intensity.

1 According to Nowak (2006) fricative duration differences are not consistent for this constrast.

That is, fricative step 0 consisted of 9/9 [ɕ] and 0/9 [ʂ], while step 1 consisted of 8/9 [ɕ] and 1/9 [ʂ], step 2 was 7/9 [ɕ] and 2/9 [ʂ], etc. The resulting fricative and vocalic portions were then concatenated into 100 CV syllables (Figure 1). There were no known unwanted artifacts produced by these manipulations.

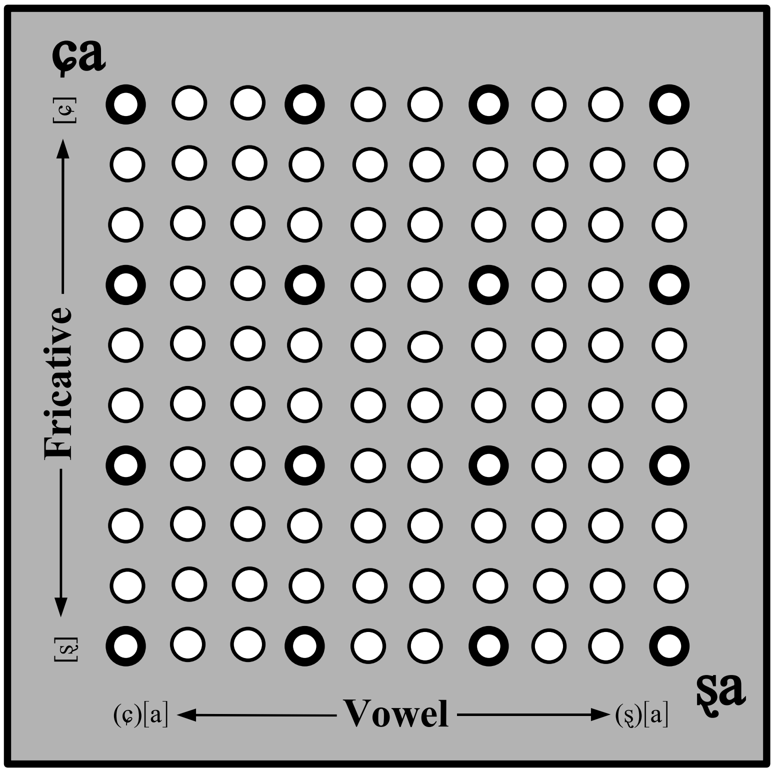

Figure 1: Stimuli design. Each circle represents a particular combination consonant and vowel from each continuum.

The interpolation of the signals allows for the full richness of the acoustic signal to be maintained while still manipulating the information present in the signal. The specific method used avoids the perception of multiple vowel signals by normalizing the F0 and durations. Further, the F2 locus for both productions, 1420 Hz, 10.73 Bark for the retroflex and 1720 Hz, 12 Bark for the alveolopalatal, is within the 3 - 3.5 Bark shown to be a trigger for the center-of-gravity effect (Delattre, Liberman, Cooper, and Gerstman, 1952; Chistovich and Lublinskaya, 1979; Xu, Jacewicz, Feth, and Kristamurthy, 2004) where two or more formants within a 3.5 Bark range can be perceptually merged into a single formant that is mediated by the relative amplitudes of those formants. The results of McGuire (2007) do indeed suggest that English listeners perceive such stimuli as natural speech and that Mandarin listeners also perceive these stimuli as natural and categorize them as similar to their nearly identical native sibilant contrast.

Subjects

Twenty-nine male and female subjects between the ages of 18 and 45 participated in the experiment. All were native English speakers with no experience with Polish or Mandarin Chinese. All reported normal hearing and were paid $10 per session.

Procedure

Subjects participated in three to five one-hour sessions held on consecutive days. The first session consisted of a discrimination test followed by classification familiarization and a classification test. The stimuli for the discriminating task were drawn from a 4 X 4 subset of the larger stimulus set (see Figure 2) where adjacent pairs were separated by three steps in either dimension. Pairs in the center of the fricative distribution were considered cross-boundary tokens and other pairs were considered within-category for the analysis below.

The discrimination task was a variant on a four-interval discrimination task. During each trial subjects were presented two pairs of sounds. One pair, the “different” pair, differed along a single dimension and the second pair, the “same” pair consisted of sounds identical to either the first or second sound of the “different” pair. That is, listeners were presented with AB AA, BA AA, AB BB, BA BB pairs, as “first pair different” trials. Only adjacent “different” pairs were tested such that the distance between each pair was three steps in a single dimension; larger and smaller distances were not compared. Subjects were asked to press a button on a five-button box corresponding to the “different” pair, either the first pair/leftmost button or the second pair/rightmost button. The two stimuli making up each pair were separated by 100ms while the two pairs were separated from each other by 500ms. There were 24 possible different pairs (12 differing in each dimension) x 2 presentation orders which resulted in 48 total different pairs. Four repetitions of each pair were used for a total of 192 trials; "same" - "different" order was balanced and internal pair order was not analyzed.

The classification testing consisted of a single random presentation of a syllable from the full set of 100 stimuli. Subjects were instructed to categorize each sound as category "A" or "B" by pressing the leftmost or rightmost buttons (respectively). All stimuli with alveolopalatal fricatives (fricative steps 0-4) were defined as category "A" while retroflex stimuli (fricative steps 5-9) were category "B". Thus only the fricative dimension was relevant for categorization. Subjects participated in six blocks of classification and were presented the entire stimulus set during each block. During the first block subjects received accuracy feedback; this was the classification familiarization period and deemed necessary for subject attention to the task. Positive feedback consisted of the cumulative accuracy score for that session while negative feedback consisted of the correct category label and cumulative accuracy for that session. Subsequent blocks comprised the classification test period and no feedback was given other than whether or not a response was detected.

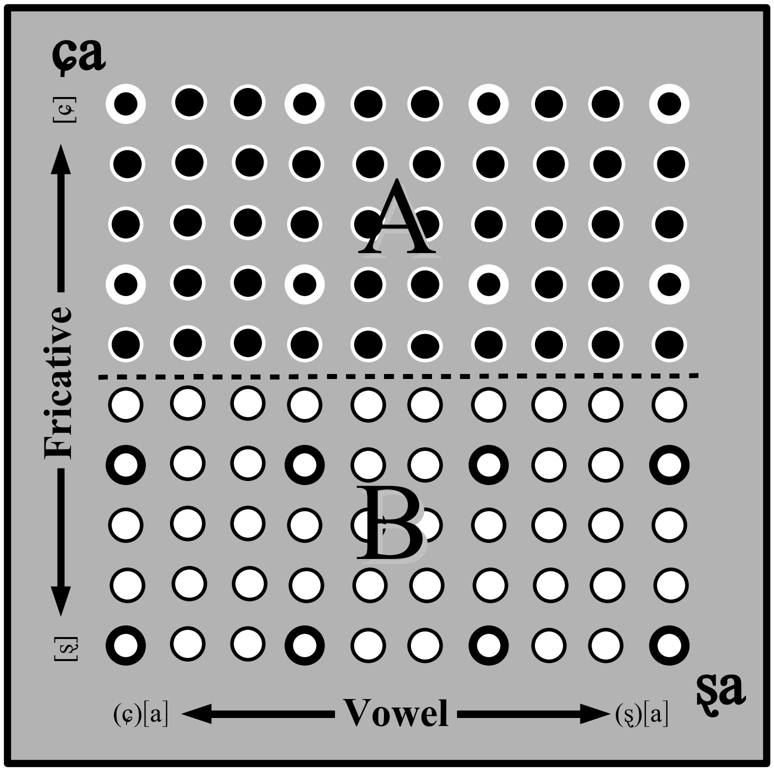

Figure 2: Category assignments for the stimuli in experiment 1. Heavy circles denote discrimination stimuli.

Sessions 2-4 comprised the training sessions. These consisted of six classification blocks with accuracy feedback. Subjects who achieved an overall session accuracy of greater than 85% or completed four sessions (three training) moved onto post-testing in the following session. Subjects who showed no major improvement by the middle of the third session were asked about their categorization strategy and encouraged to listen to the "first part" of the sound.

Post-testing consisted of one block of classification training (to act as a reminder), five blocks of the classification testing, and finished with the discrimination testing. Subjects were then debriefed and paid for their participation.

Results and discussion

Two subjects dropped out of the study prematurely and are not considered further. All accuracy scores from discrimination and classification were converted into a measure of sensitivity, *d′* units (Macmillan & Creelman, 2005). As a group, these subjects showed significantly higher sensitivity post-test than to pre-test in a paired *t*-test, discrimination: $t(647) = -4.91$, $p < 0.001$); classification: $t(25) = 5.93$, $p < 0.001$). However, since a major goal of the study is to assess changes in sensitivity due to learning, subjects who exhibited a change in categorization associated with learning were separated out. Specifically, the subjects were divided into three groups based upon their classification sensitivity to the extreme stimuli in the pre- and post-test, i.e. all stimuli having a fricative component within three steps of each endpoint. Six subjects had a pre-test *d′* greater than 2.0 and so were not considered to be potential learners as they were already at ceiling[2]. Of the remaining twenty-one subjects, seven did not demonstrate substantial improvement, having a classification $d' < 1$ on the final session and were classified as "non-learners"[3]. These subjects displayed two different patterns, subjects who showed no consistency in categorization and subjects who categorized using the vocalic dimension but were unable to switch to using the fricative dimension. The fourteen remaining subjects all achieved a final classification $d' > 2.0$, except four who all had a *d′* score > 1.0. These fourteen subjects were classified as "learners" and their results are analyzed in detail below.

Generally, the learners show a consistent initial categorization using the vocalic dimension but were able to switch to relying solely on the fricative dimension (Figure 3), showing a significant improvement in accuracy from pre-test to post-test in a paired *t*-test (t [13] = -7.6, $p<0.001$; mean pre-test d'=0.35, mean post-test d'=1.62).

---

2 All of these subjects were asked to return and categorize using the vocalic dimension (see experiment 2). Three agreed and showed equal or even greater accuracy using that dimension.

3 It is possible that these subjects could improve with more training or a different training paradigm. This was not attempted.

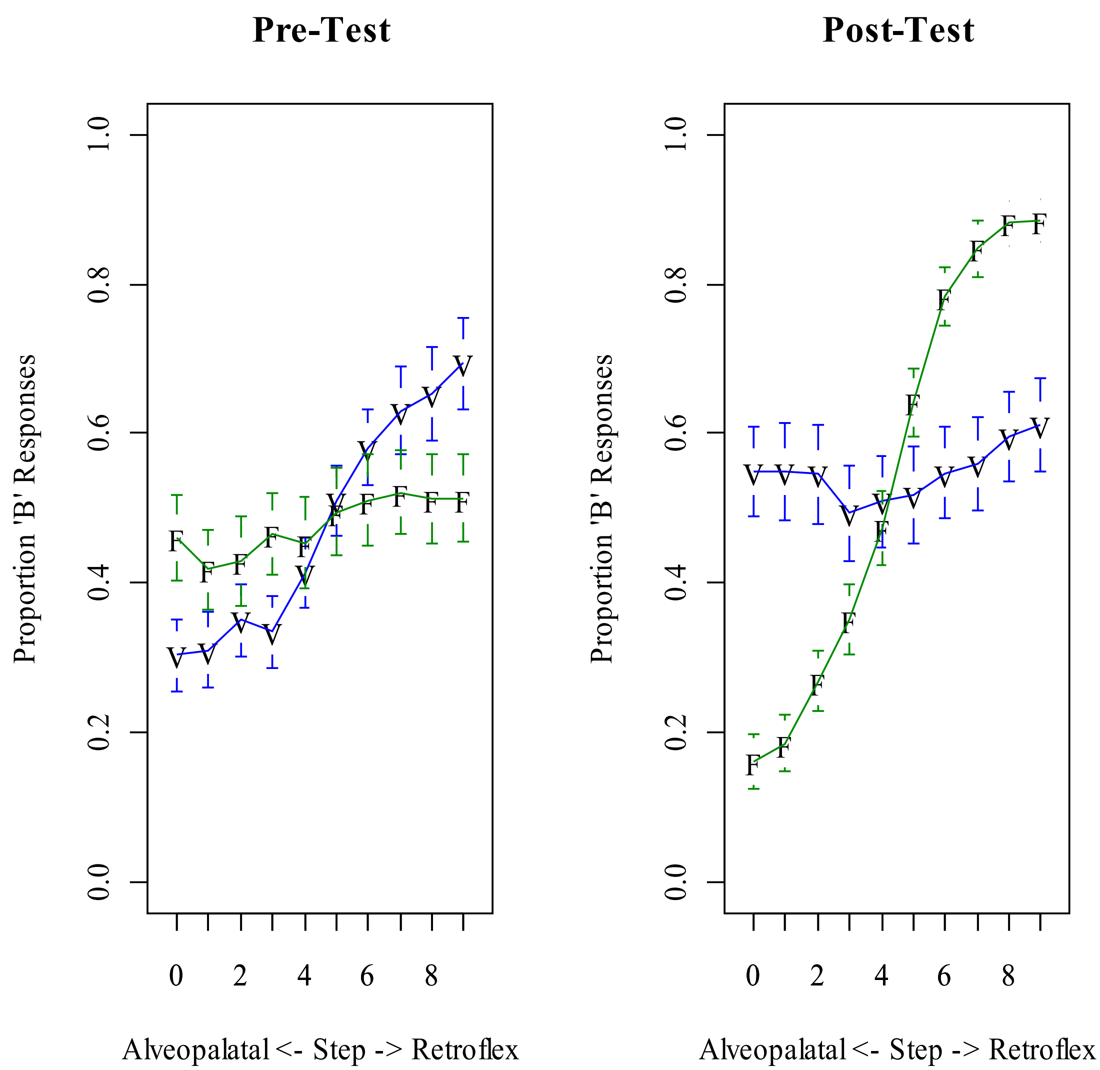

Figure 3: Cumulative classification in pre- (left panel) and post-test (right panel) for all learners. The vertical dimension represents the proportion of "B" label responses and horizontal dimension represents the step in the fricative, "F", or vocalic "V" dimension.

The discrimination results for learners were also converted to *d′* values. The overall mean d-prime was 0.59, the pre-test mean was 0.46, and the post-test mean was 0.72. These values were considerably lower than those values from the classification task. This resulted from the task and is due to the small differences in step size necessary to examine within category changes in perception While other studies examining the internal structure of a cateogry have been able to control step size for better within category discrimination (e.g. Guenther et al. 1999; non-speech), this was not possible for this experiment as natural phonetic categories were used. These were analyzed in a repeated measures ANOVA having the factors Test (pre-, post-), Dimension (fricative, vowel), and Boundary (cross-category, within category). Significant main effects were found for Test ($F[1,14]=15.89$, $p < 0.01$; pre-test d'=0.46, post-test d'=0.72) and Boundary ($F[1,14]=9.43$, $p < 0.01$; cross-category d'=0.70, within category d'=0.54). A significant interaction was found between Test and Dimension ($F[1,14]=6.62$, $p<0.05$, see Figure 4). The test and dimension interaction was explored further with Tukey post hoc tests. A significant difference between pre- and post-test *d*-prime scores was found for the fricative dimension only ($p < 0.001$). Additionally, only the post-test d-primes between the two dimensions were significantly different ($p < 0.05$).

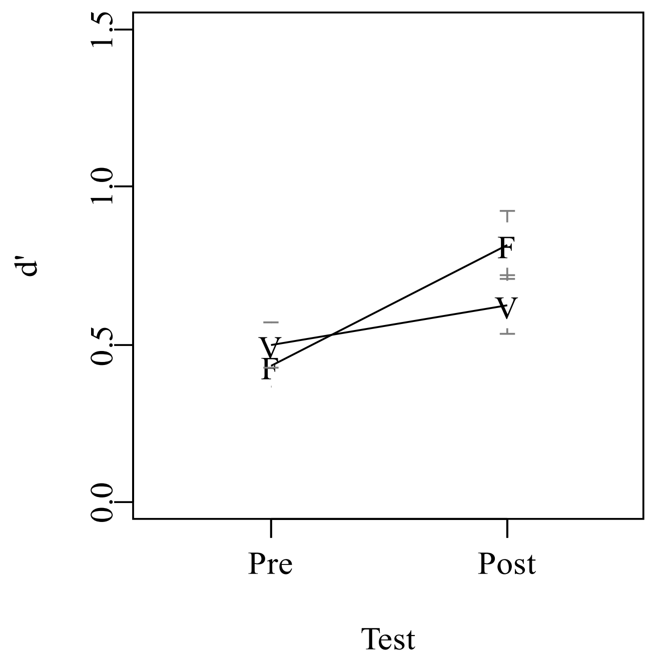

Figure 4: Fricative dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.

These results demonstrate that subjects could learn to rely solely on the fricative noise information for categorization, despite their initial reliance on the vocalic cues. Moreover, this resulted in a significant increase in sensitivity to the new dimension. Together, these results generally support Goldstone (1994) and Francis and Nusbaum (2002) and the A2D framework. There was also a high degree of variability in subject performance, with nearly a quarter of subjects failing to learn and another 20% able to perform at the task immediately with minimal exposure to the labels. The reason for these differences is not immediately clear. No evidence of within-category compression or significant change to irrelevant dimension was found. However, due to the already low sensitivities values, it is possible that a floor effect is present and no further reduction in sensitivity is possible.

Experiment 2: Vocalic dimension categorization

This experiment was conducted to direct subjects to categorize along the vocalic dimension. Because American English subject seem to be generally predisposed to categorize this dimension, within-category acquired equivalence might be expected as suggested by the results of previous experiments (Livingston, Andrews, and Harnad 1998, Goldstone 1998, Francis and Nusbaum 2002). Further, as subjects were biased against using the fricative dimension for categorization, subjects should show no change in sensitivity to that dimension or possibly even a decrease in sensitivity.

Methods

Stimuli

The stimuli for this experiment are the same as for the previous experiment.

Subjects

Twelve male and female subjects between the ages of 18 and 22 participated in the experiment. All reported normal hearing and were paid $10 per session. All were native speakers of English and had no exposure to Polish or Mandarin.

Procedure

The procedure was also identical to the previous experiment with two major exceptions. First, the category boundary was defined along the vocalic dimension so that only differences in the vocalic dimension were relevant to categorization (see Figure 5). Specifically this meant that stimuli with vocalic parts closer to the alveolopalatal end (v0-v4) were considered category "A" while those closer to the retroflex end (v5-v9) were category "B". Second, all subjects participated in only a single training session for at total of three sessions. This is due to all subjects achieving the criterion, as established in the first experiment, in one training session.

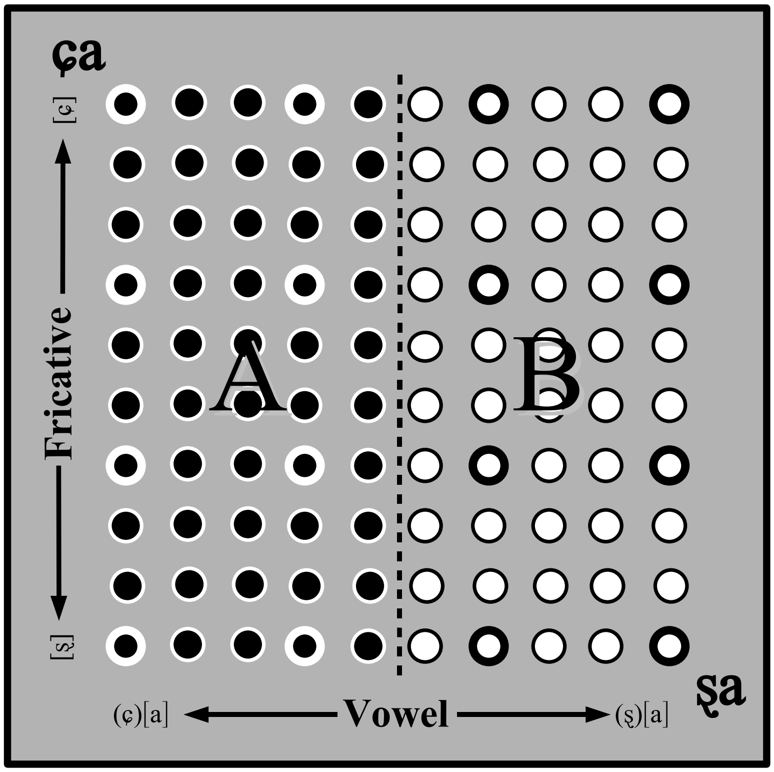

Figure 5: Category assignments of the stimuli for experiment 2, vocalic dimension categorizers. Stimuli marked with heavy circles were used in the discrimination task

Results and discussion

The classification results were analyzed as in the previous experiment. For this experiment, all subjects had high d' values in the post-test (>1) and are included in the analyses. Only classification sensitivity significantly improved from pre- to post-test ($t\,[11] = -2.53$, $p < 0.05$). As in the first experiment, subjects generally categorized consistently along the vocalic dimension. Here the primary effect of training was a sharpening of the category boundary (Figure 6).

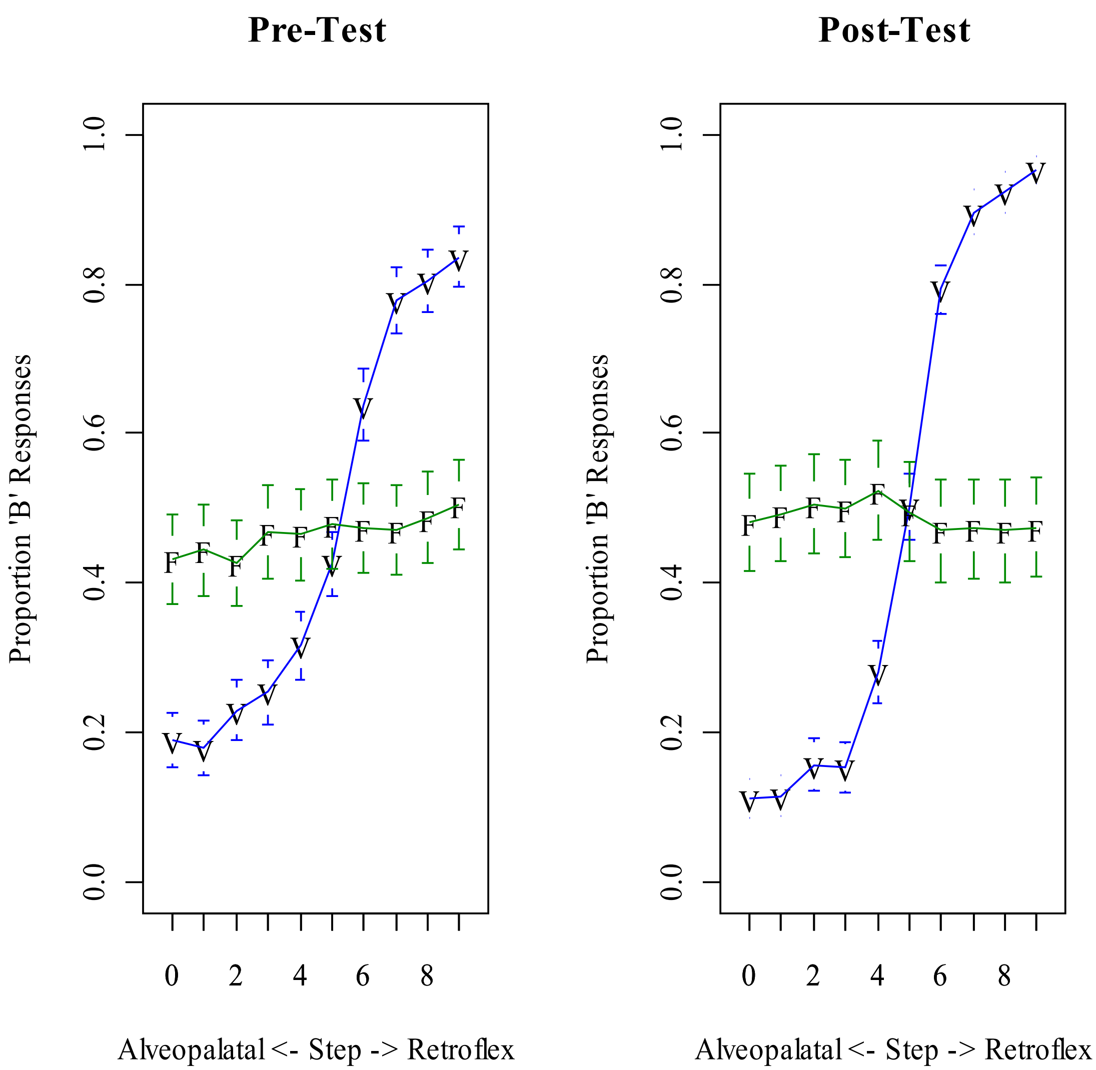

Figure 6: Cumulative classification in pre- (left panel) and post-test (right panel) for all learners. The vertical dimension represents the proportion of "B" label responses and horizontal dimension represents the step in the fricative, "F", or vocalic "V" dimension.

The discrimination results were also analyzed as in the previous experiment.. The Test factor (see Figure 7) did not reach significance (F[1,11]=4.1, p = 0.09; pre-test d'=0.59, post-test d'=0.76); the only significant main effect found was for Boundary (F[1,11]=32.68, p < 0.001; cross-category d'=0.97, within-category d'=0.53) indicating that subjects were overall more sensitive to cross-boundary tokens. Further, a two-way interaction between Boundary and Test (see Figure 8) was found to be significant ($F[1,11]=4.40$, $p < 0.05$). This interaction was explored further with with Tukey post hoc tests which demonstrated significant differences in means between pre-test and post-test for cross-boundary pairs ($p<0.01$) and between cross-boundary and within-boundary pairs for both pre-test ($p<0.05$) and post-test ($p<0.001$).

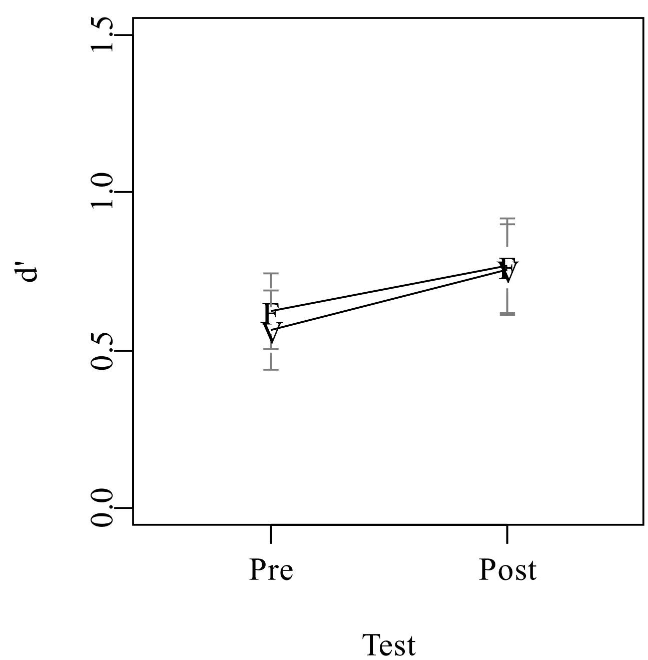

Figure 7: Vocalic dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.

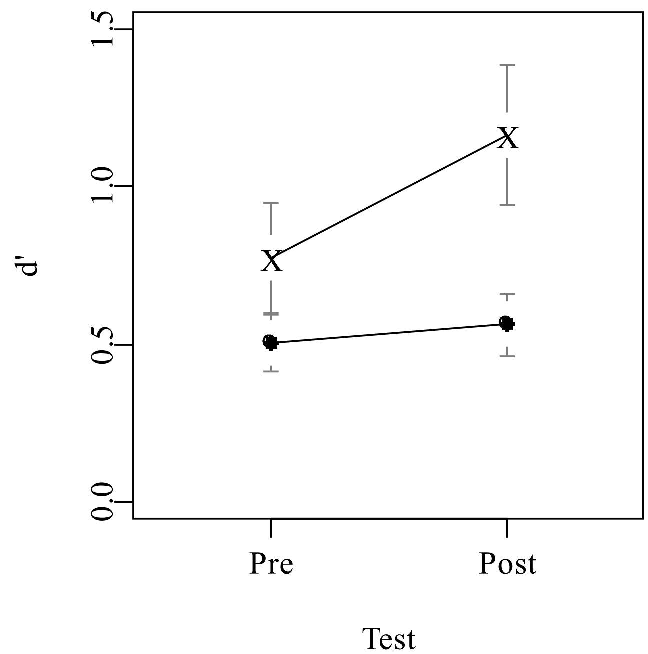

Figure 8: Vocalic dimension learner's change in sensitivity from pre-test to post-test cross-category "X" and within category "●".

These results do not support the initial hypothesis as there is no within-category compression. Indeed, there are no significant differences between the two dimensions. Instead, the primary effect of training seems to be a sharpening of the category boundary through heightened sensitivity to the boundary tokens for both dimensions.

Experiment 3: Two-dimensional categorization

This experiment is designed so that subjects were required to use both the fricative noise and formant transition dimensions independently. In this categorization task, subjects are asked to divide the stimulus space into four regions with the mid-point along each dimension acting as a category boundary (i.e. each quadrant.) Thus, subjects must use both fricative and vocalic information independently to make the categorization.

Methods

Stimuli

The stimuli were identical to those used in previous experiments.

Subjects

Sixteen male and female native speakers of English between the ages of 18 and 25 participated in the experiment. All reported normal hearing and were paid $10 per session. None had exposure to Polish or Mandarin.

Procedure

The procedure is identical to the previous experiment except that the stimulus space was divided into four categories such that participants had to use each dimension independently (see Figure 9). Category "A" consisted of alveolopalatal fricative and vocalic steps (f0-f4 + v0-v4). Category "B" consisted of alveolopalatal fricative steps with retroflex vocalic steps (f0-f4 + v5-v9). Category "C" consisted of retroflex fricative steps with alveolopalatal vocalic steps (f5-f9 + v0-v4). Category "D" consisted of retroflex fricative and vocalic steps (f5-f9 + v5-v9). Thus, "B" and "C" are "conflicting cue" categories representing unnatural combinations of cues, while “A” and “D” are "cooperating cue" categories representing the natural categories. The four leftmost buttons of the box were labeled “A”, “B”, “C”, and “D” respectively for labeling.

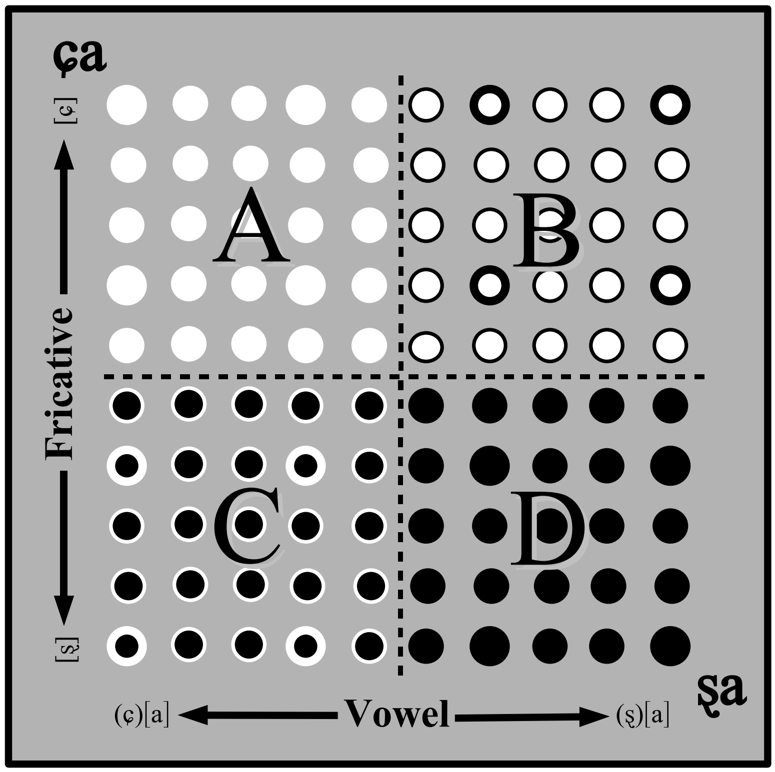

Figure 9: Category assignments of the stimuli for experiment 3, two-dimension categorizers.

Results

Overall, subjects as a group showed significant improvement from pre to post in both discrimination and labeling sensitivity, $t[359] = -5.82$, $p < 0.001$ and $t[15] = 8.62$, $p < 0.001$, respectively. Looking more closely at the individual performances, however, only eleven of the sixteen subjects consistently used the category labels in the post-test labeling, defined as an overall estimated $d' > 1$. As would be expected given the previous results, most subjects had difficulty initially with categories divided by the fricative dimension, i.e. A/C and B/D and the primary improvement was in learning to use the fricative dimension for categorization.. However, pooled across subjects, there was some initial consistency in use of the labels with just brief familiarization (see Figure 10). Training resulted in a general ability to use both dimensions independently for labeling as indicated by significant improvement in labeling from pre-test to post-test (t [10]=-7.42, $p<0.001$; mean pre-test d'=0.54, mean post-test d'=1.35). Pre-test accuracies for each category were: A=46%, B=46%, C=43%, D=36%. Post-test accuracies for each category were: A=63%, B=71%, C=75%, D=54%.

There were many different categorization patterns found among the subjects. The non-learners generally demonstrated a two-category division along vocalic dimension, conflating categories "A" and "C" against "B" and "D". However, the learners also demonstrated a variety of four-category representations (e.g. subjects 203 and 208; see Figure 11) and some subjects showed consistent use of only three categories (e.g. subjects 211, 212; see Figure 11). The large number of different categorization schemes prohibited further division of the learners into smaller groups.

Pre	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	AC	A	A	A	AB	AB	AB	B	B	BD
f1	A	A	AB	AB	A	AB	AB	B	B	BD
f2	AB	ABC	A	AC	AC	ABC	B	B	ABD	B
f3	AC	A	A	AC	A	AB	AB	BD	BD	BD
f4	A	AC	AC	AC	ABC	AD	AB	B	BD	BD
f5	C	C	AC	AC	AC	ACD	AC	BD	BD	ABD
f6	AC	AC	AC	C	AC	BC	BD	BD	BD	D
f7	AC	AC	C	C	AC	C	CD	D	D	D
f8	C	C	C	C	C	CD	CD	BD	BCD	D
f9	CD	CD	CD	CD	CD	D	D	D	D	D

Post	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	A	A	A	A	A	AB	B	B	B	B
f1	A	A	A	A	AB	AB	B	B	B	B
f2	A	A	A	A	A	AB	B	B	B	B
f3	AC	A	A	A	A	AB	B	B	B	B
f4	AC	AC	AC	A	A	AB	B	B	BD	B
f5	AC	AC	AC	AC	AC	A	AB	BD	BD	BD
f6	AC	AC	C	C	C	C	D	BD	BD	BD
f7	C	C	C	C	C	C	D	D	D	D
f8	C	C	C	C	C	C	CD	D	D	D
f9	C	C	C	C	C	C	CD	D	D	D

Figure 10: Cumulative classification in pre- (left panel) and post-test (right panel) for all learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Each square represents one stimulus and each letter indicates that for a particular stimulus > 25% of responses were for the corresponding category.

203	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	C	C	C	A	AC	AB	B	B	BC	B
f1	C	C	A	C	A	B	B	B	B	B
f2	C	C	AC	C	A	B	B	AB	B	B
f3	C	C	A	A	A	AB	B	B	B	AB
f4	C	A	AC	A	A	AB	B	B	B	B
f5	C	C	C	C	A	A	D	D	B	D
f6	C	C	C	C	C	AC	D	D	D	D
f7	C	C	C	C	C	A	D	D	D	D
f8	C	C	C	C	C	C	D	D	D	D
f9	C	C	C	C	C	CD	D	D	D	D

208	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	A	A	A	A	A	A	C	B	B	B
f1	A	A	A	A	C	B	B	B	B	B
f2	A	A	A	A	A	A	C	B	B	B
f3	A	A	A	AC	A	A	C	B	B	B
f4	A	A	A	A	D	C	C	B	B	B
f5	AC	C	C	C	C	C	C	BC	BD	B
f6	C	C	C	C	C	C	C	C	D	D
f7	C	C	C	C	C	C	C	D	D	D
f8	C	C	C	CD	C	C	CD	C	D	D
f9	C	C	C	C	C	C	D	CD	D	D

212	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	C	C	C	C	C	BC	BC	B	D	BD
f1	C	C	C	C	BC	BD	BC	B	B	B
f2	C	C	C	C	C	BC	BD	D	D	BD
f3	C	C	C	C	C	CD	BD	BD	B	B
f4	C	C	C	C	C	BDC	B	D	BD	BD
f5	C	C	C	CD	D	C	B	B	BD	BD
f6	C	C	C	C	C	C	BCD	D	D	CD
f7	C	C	CD	D	C	CD	D	D	D	CD
f8	C	C	C	C	C	D	CD	D	CD	CD
f9	C	C	C	C	CD	D	D	D	D	CD

211	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	A	A	A	B	A	B	B	B	B	B
f1	A	A	A	A	A	B	B	B	B	B
f2	A	A	A	A	A	B	A	B	B	B
f3	A	A	A	B	A	B	B	B	B	B
f4	A	A	A	A	AB	A	B	B	B	B
f5	A	AC	A	A	A	A	AB	B	B	B
f6	A	C	C	C	AC	A	AC	A	A	B
f7	C	C	C	AC	C	C	AC	A	AC	A
f8	C	C	C	C	C	C	C	C	C	C
f9	C	C	C	C	C	C	C	C	C	C

Figure 11: Examples of post-test classification spaces demonstrating a variety of categorizations of the stimulus space. Subject number is indicated in the upper left corner of each panel. See Figure 10 for an explanation of this graphic representation.

The discrimination results of learners (analyzed as before) show significant main effects for Test ($F[1,9] = 10.86$, $p<0.01$; pre-test $d'=0.58$, post-test $d'=1.03$, see Figure 12) and Boundary ($F[1,9] = 39.95$, $p<0.001$; cross-category $d'=1.16$, within-category $d'=0.63$), as well as a significant interaction between Test and Boundary ($F[1,9] = 5.54$, $p<0.05$; see Figure 13). Tukey post-hoc tests showed a significant increase in sensitivity from pre-test to post-test for both within-category pairs ($p<0.05$) and cross-boundary pairs ($p<0.001$). The two pair types were significantly different from each other in the post-test ($p <0.001$).

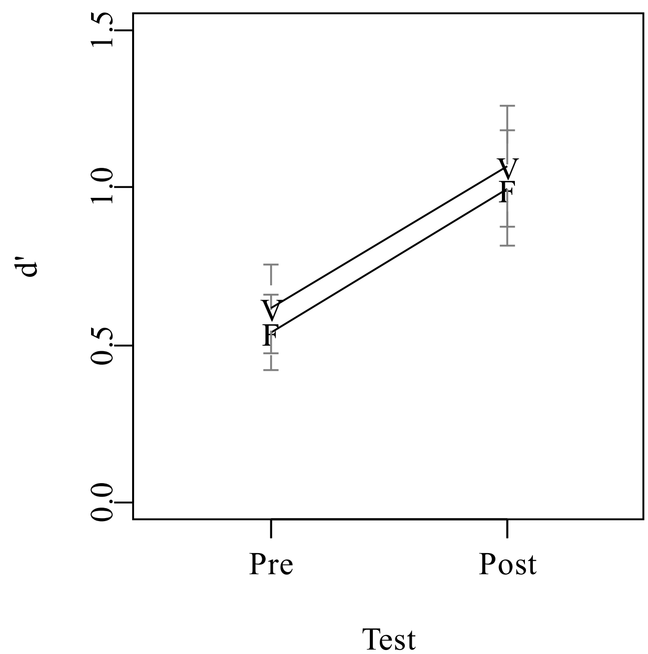

Figure 12: Two-dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.

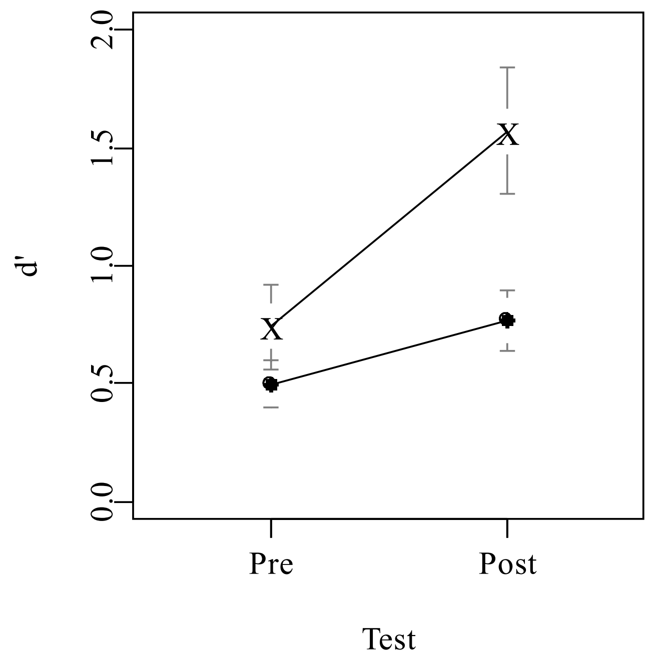

Figure 13: Two-dimension learner's change in sensitivity from pre-test to post-test cross-category "X" and within category " ● ".

Overall, a majority of subjects were able to attend to both dimensions separately and use both for categorization. These results further showed that training to use each dimension independently results in a general improvement in sensitivity to both dimensions even though the primary difficulty in categorization was learning to differentiate along the fricative dimension. Moreover, this improvement primarily affected cross-category sensitivity and to a smaller extent within-category sensitivity.

Experiment 4: Control

This control experiment examined how the exposure to the stimuli in the discrimination tasks alone affected sensitivity, if at all. In this experiment subjects were tested using the same procedure as the previous experiments, but no training or labeling experience of any kind was given. Thus any effects can be attributed to participating in the discrimination tests, which can be compared against performance in the previously described experiments.

Methods

Stimuli

The stimuli were the same as used in previous experiments except that only the testing stimuli were used.

Subjects

Fourteen male and female subjects between the ages of 18 and 23 participated. All reported normal hearing and were native speakers of English with no exposure to Polish or Mandarin. They were paid $10 per session for their participation.

Procedure

This experiment consisted solely of the two discrimination tests described previously, administered on separate consecutive days. No labeling was conducted.

Results and discussion

The discrimination results were analyzed as in previous experiments. There were no significant effects (Test factor: $F[1,12]=1.84$, $p = 0.2$; pre-test $d'=0.41$, post-test $d'=0.50$). The lack of a test effect suggests that the improvement seen in the previous experiments from pre- to post-test can be safely assumed to be due to the training, rather than simply exposure to the stimuli.

General discussion

The results presented here provide general support for A2D models, especially with the modification suggested by Francis and Nusbaum (2002). In this case, targeted training with sibilant fricative cues succeeded in directing attention to the specific dimension of contrast, which in turn resulted in perceptual warping such that listeners became more sensitive to that dimension. Within category sensitization is localized to the more difficult tasks, that is classification along the the frication dimension alone (experiment 1) as well as the frication dimension along with the vocalic dimension (experiment 3), suggesting that such difficult perceptual learning situations require sensitivity to within category variability. Conversely, when the contrast is easier to distinguish, as for the vocalic dimension, only further perceptual separation of the category members is necessary. These findings support much of the previous research into perceptual learning, but importantly, demonstrate how auditory dimensions of real speech categories are analogous to those of other perceptual domains (e.g. Goldstone, 1994).

Overall, there is only evidence for acquired distinctiveness in sensitivity, there was no evidence for desensitization of non-attended dimensions. However, the extreme difficulty of the discrimination task left little room for a reduction in sensitivity and it is possible the floor effects negate any acquired equivalence in the stimuli that may be present. In fact, it was hypothesized that the vocalic learners would show compression due to the ease of categorization using that dimension; instead there was no significant change. Similarly, it might be assumed that for the fricative categorizers the irrelevant vocalic dimension would show a loss in sensitivity. Though, again there was no effect. However, as mentioned above, the presence or absence of within category expansion can be interpreted as an effect of ease/difficulty of a dimension of contrast.

Additionally, it is of interest that warping of the perceptual space was found using such a highly sensitive psychoacoustic task. This contradicts the assumption that a speeded discrimination task bypasses lexical and language effects because it is a low-memory load task (Pisoni 1973, Pisoni and Tash 1973, Fox 1984, Johnson and Babel 2010). This suggests either that exposure to a contrast and training using a contrast has an effect at both very low cognitive levels as well as the higher, more categorical / symbolic level demanded by the labeling task or that the 4IAX task is not as sensitive as assumed.

The results of this study also give insight into the integrality of fricative noise and vocalic information for such a contrast. Specifically, English listeners do not seem to be affected by the degree of correlation between the fricative and vocalic cues, despite knowledge of such a correlation for their native sibilant contrast (Tomiak, Mullenix, and Sawusch 1987). Note that in the two-dimensional categorization experiment, subjects demonstrated no more difficulty in learning the conflicting cue categories composed of physically different chunks ("B" and "C") than the "natural" correlated categories ("A" and "D"). In fact subjects showed the least accuracy in labeling the correlated retroflex category "D" and assigned it the least area in the stimulus space. That subjects relatively easily learned to categorize speech stimuli that could be impossible to produce, and not physically generated by this particular talker's vocal tract, calls into question theories of speech perception that rely on individuals perceiving gestures (e.g. Liberman et al. 1967, Fowler 1986, Best 1995).

Overall, this study provides strong evidence in favor of categorization and cue weighting theories that view the process as directing attention to both new dimensions as well as previously used, but lowly weighted ones.

Acknowledgments

This work was supported by NIDCD grant R01 DC004421 and was first reported as part of a doctoral dissertation by the author. Thanks go to Keith Johnson, Mary Beckman, Susan Nittrouer, and members of the Berkeley and UC Santa Cruz phonetics and phonology reading groups for valuable input. Earlier versions of this paper were presented at the 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan 2006 and the Annual Meeting of the Linguistic Society of America 2007.

References

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 171–204). Baltimore: York Press.

Boersma, P., & Weenink, D. (2002). PRAAT. Amsterdam, NL: Institute of Phonetic Sciences. University of Amsterdam, NL.

Chistovich, L. A., & Lublinskaja, V. V. ~1979!. 'The "center of gravity" effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research 1*, 185–195.

Delattre, P., Liberman, A., Cooper, F., & Gerstman, L. (1952). An experimental study of the acoustic determinants of vowel color. *Word 8*, 195–210.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics, 14, 3-28.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human perception & performance*, *10*(4), 526-540.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62*(8), 1668-1680.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception & Performance, 28*(2), 349-366.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2),178-200.

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America, 106*(5), 2900-2912.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1, 1-7.

Iverson,P., Kuhl,P.K., Akahane-Yamada,R., Diesch,E., Tohkura,Y., Kettermann,A., Siebert,C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47-B57.

Johnson, K. (2003).

Johnson, K. & Babel, M. (2010). Perception of Fricatives by Dutch and English Speakers.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-361.,

Lisker, L. I. (1986). 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, 29, 3-11.

Lisker, L. I. (2001). Hearing the Polish sibilants [s š ś]: Phonetic and auditory judgements. In N. Grønnum,

& J. Rischel (Eds.), *Travaux du Cercle Linguistique de Copenhague XXXI. To honour Eli Fischer-Jørgensen* (pp. 226–238). Copenhagen: C.A. Reitzel.

Livingston, K.R., Andrews, J., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 24:3, 732-753.

Logan, J.S., S.E. Lively, & D.B. Pisoni (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *The Journal of the Acoustical Society of America*. 94, 1242-1255.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

McGuire, G. (2007). *Phonetic Category Learning*. Unpublished Dissertation ms. The Ohio State University.

McGuire, G. (to appear). Orthographic effects on the attention to phonetic cues. Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1975). An effect of language experience: the discrimination of /r/ and /l/ by native speakers of Japanese and English. Perception & Psychophysics, 18, 331–340.

Nittrouer, S. (1992). Age-related differences in perceptual effects for formant transitions within syllable and across syllables. *Journal of Phonetics, 20*, 1-32.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America. 112*(2), 711-719.

Nittrouer, S. & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America, 101(*4), 2253-2266.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39-57.

Nowak, P. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics, 34*, 139-152.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perceptual Psychophysics*, *13*, 253-260.

Pisoni, David B. & Jeffrey Tash. 1974. Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics*, *152,* pp. 285-290.

Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. Perception & Psychophysics, 36(2), 131-145.

Tomiak, G. R., Mullenix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. Journal of the Acoustical Society of America, 81, 755-764.

Zygis, M. & Hamann, S. (2003). Perceptual and acoustic cues of Polish coronal fricatives. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 395-398.

Zygis, M. & Padgett, J. (to appear). A perceptual study of Polish fricatives, and its relation to historical sound change. Journal of Phonetics, in press, corrected proof.

LIST OF FIGURES