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

Grant McGuire
grantmcguire@berkeley.edu

Department of Linguistics
University of California at Berkeley
Abstract

This study reports data from a series of training experiments exploring the warping of perceptual space due to the acquisition of phonetic categories by adults. In this study, 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 only fricative noise, only vocalic cues (e.g. formant transition), or both cues independently to label the stimulus set. Results from discrimination and labeling tasks indicated that subject’s attention could be directed favor one cue over another. Such training resulted in increases in sensitivity across boundaries to trained dimensions as well as increases in sensitivity within category for difficult contrasts. These results fit generally with general cognitive models of perceptual learning and demonstrate that learning in speech perception is a subset of general learning phenomena.
Introduction

Adults often have difficulty in discriminating and identifying non-native contrasts (Abramson and Lisker, 1970; Miyawaki, K., Strange, W., Liberman, Jenkins, and Fujimura, 1975; Best, McRoberts, and Sithole, 1988; Harnsberger, 2001; Mielke, 2003). One explanation for this difficulty is that listeners lacking the requisite perceptual experience with a given contrast may not attend to the necessary cues or weight them adequately (Francis and Francis and Nusbaum, 2002; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, and Siebert, 2003). Of crucial importance, therefore, is the extent to which training and laboratory experience can affect how listeners attend to cues and dimensions of contrast.

In general, attending to relevant information and dimensions is hypothesized to be a key aspect of general theories of perceptual learning (Gibson, 1964; Nosofsky, 1986; Goldstone, 1998). Children seem particularly sensitive to details that can be used for perception (see e.g. Werker and Tees, 1984; Mareschal, Powell, and Volein, 2003) and differing attention to cues affects their categorization (Maye, Werker, and Gerken 2002; French, Mareschal, Mermillod, and Quinn, 2004). In speech perception it has been shown that children acquiring their native language shift which cues they rely upon to make relevant distinctions. Specifically 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 weigh 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.

It has also been demonstrated that adult perceptual attention can be directed through training. For example, Goldstone (1994) demonstrated acquired distinctiveness for targeted dimensions of categorization using simple, abstract visual stimuli that differed in two dimensions. This study explicitly trained subjects to categorize a two dimensional stimulus set such that one group of subjects had to
categorize using a single dimension while ignoring variation in the other, a second group categorized using the previous group's irrelevant dimension, and a third group used both dimensions separately. There were two stimulus sets used for the experiment, one consisted of squares varying by the highly integral dimensions brightness and saturation and the second consisted of the highly separable dimensions size and brightness. Goldstone found that subjects could adequately categorize both stimulus sets using one or both dimensions independently. Importantly, this study also found a significant increase in sensitivity to the trained dimension for both stimulus sets and a decrease in sensitivity to the irrelevant dimension, but only for the perceptually separable set.

In speech perception, using natural categories, Francis and Nusbaum (2002) trained English listeners to identify Korean stop contrasts and examined subjects' attention to various cues before and after. The general finding was that subjects' changed their attention to specific cues where necessary. Moreover, sensitivity changed as a result such that both acquired equivalence and distinctiveness were present, but the effects varied by category. Specifically, they find evidence that contrasts having a high degree of distinctiveness before training show within-category compression while difficult contrasts show increases in sensitivities to the relevant dimensions.

These studies also note an effect whereby sensitivity to differences among stimuli changes as a result of perceptual learning. This is in line with the general finding that perceptual space around phonetic categories is warped such that fine phonetic variation is largely ignored and listeners make gross categorical judgments (e.g. Liberman, Hoffman, and Griffith, 1957; Eimas, Siqueland, Jusczyk, and Vigorito, 1971; Kuhl, 1991; Kuhl and Iverson, 1995). Such warping is a general phenomenon of perceptual learning (Gibson, 1969), and can be trained in both adults and children (Goldstone, 1994; Guenther, Husain, Cohen, and Shinn-Cunningham, 1999; Maye, Werker, and Gerken, 2002). In speech, much of this warping is due to native language phonologies where listeners' perceptual systems are tuned primarily to native language contrasts (Abramson and Lisker, 1970; Best et al., 1988; Mielke 2003, Iverson et al. 2003). Best et al. (2001) further refined this idea by showing that some non-native contrasts are easily discriminated by
listeners if that contrast can be mapped onto native categories; only when a single contrast is subsumed by a single native category do listeners have extreme difficulty discriminating that contrast. However, the adult perceptual system is malleable and new phonetic contrasts can be learned with differing degrees of success (Pisoni, Aslin, Perey, and Hennessy, 1982; Strange and Dittman, 1984; Jamieson and Morosan, 1986; Logan, Lively, and Pisoni, 1991; Case, Tuller, and Kelso, 2003; Iverson, Hazan, and Bannister, 2005; Guion and Pederson, 2007).

This warping can be understood as changing the perceptual distance between categories and specific changes in sensitivity to differences within and across categories. In speech perception, several studies have found heightened sensitivities across category boundaries rather than lessened sensitivities to stimuli within the category resulting from training (e.g. Strange and Dittman, 1984; Jamieson and Morosan, 1986). However, within-category reduction in sensitivity has been less commonly found. Notably Guenther, Husain, Cohen, and Shinn-Cunningham (1999) demonstrated such an effect for non-speech auditory stimuli. This study found decreases in sensitivity, specifically the Perceptual Magnet Effect (PME; Kuhl, 1991), using a kind of identification training where members of a category were identified from a list of potential competitors and discrimination testing using long inter-stimulus intervals. Moreover, acquired distinctiveness for the same stimuli was found when discrimination training was used, suggesting that training type affects the perceptual warping. However, although the stimuli were auditory in nature, they were not speech-like, instead consisting of filtered noise bands.

In a speech specific study, using natural and synthetic sounds from Hindi, Case, Tuller, and Kelso (2003) trained American English speakers using forced choice identification and passive listening on the Hindi dental vs. American alveolar stop contrast. MDS scores showed both acquired equivalence within category and acquired distinctiveness across categories over the course of training. Performance varied dramatically from subject to subject, with some showing improvement immediately, others only near the end of training.

The goal of the following experiments was to examine the perceptual learning of cues to a contrast
and apply the more general theories of perceptual learning and dimensional attention to speech. Of specific interest is how targeted training to attend to a specific dimension affects the perceptual space and how the two dimensions interact. It is also of great importance to ensure that the stimuli for this study are valid ecologically and therefore must be highly naturalistic rather than simplified synthetic tokens. To achieve these goals a design based on Goldstone (1994) was adapted to speech categories. Specifically, the Polish alveopalatal/retroflex voiceless sibilant distinction was chosen as the target of learning as it is difficult for English listeners to discriminate these sounds from each other as well as from native /ʃ/ (Lisker, 2001; Padgett and Zygis, 2007; McGuire, 2007) and offers relatively well understood cues to the contrast (primarily fricative noise and vocalic information). This distinction has several cues that Polish listeners rely upon with the primary ones being fricative pole frequency and F2 onset, along with slight post-consonantal vowel quality differences (Lisker, 2001; Zygis and Hamann, 2003; Nowak, 2006; Padgett and Zygis, 2007).

For English listeners, Lisker (2001) found difficulty in discriminating these sounds above chance in the context of full syllables, but much improved performance when presented with the fricatives in isolation. Using a speeded AX discrimination task, Padgett and Zygis (2007) found generally good discrimination for English listeners discriminating all Polish voiceless sibilant contrasts; Polish speakers generally performed at ceiling. McGuire (2007) found slightly different results where English listeners could reliably identify full syllables, but only using the formant transition information; fricative noise was not used consistently. The sum total of this perceptual research indicates that English listeners can identify and discriminate the Polish sibilant contrasts, albeit with some difficulty (depending on the task) and not using the full range information available in the signal.

Given these results the current study was designed as follows. In experiment 1 English listeners were trained to categorize using only the variation in fricative noise spectrum and to ignore variation in the vocalic dimension. In experiment 2 English listeners were trained to categorize using only the variation in the vocalic cues. In experiment 3, following Goldstone, subjects were trained to use both dimensions
separately. 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.

Because the acoustic signal is very rich in detail and the precise relationship between that information and what listeners rely upon for contrast judgment is not fully known, especially given the dearth of perceptual work with many fricative contrasts, this study relies on stimuli that do not greatly simplify the signal. Instead, the fricative noise and the vocalic (formant transitions plus steady state vowel) portions of the signal are treated as two temporally separate dimensions that may or may not have further crucial information within each dimension related to the spectral shape. That is, for this study, all aspects of the spectral shape are varied as a unit for each portion, it is the consonant versus vocalic spectral characteristics that are contrasted here and considered “dimensions”. It should also be noted that, though place of articulation information for the contrast is present in both portions, the exact perceptual relationship between the two dimensions is not fully known. Previous research suggests that, though temporally distinct, consonant and vowel place cues are integrated as a single unit (Repp 1978; Best, Morongiello, and Robson 1981; Tomiak, Mullenix, and Sawusch, 1987; Gow 2003).

Experiment 1: Fricative dimension categorizers

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 used 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 alveopalatal 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. 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 2)\(^2\). There were no known unwanted artifacts produced by these manipulations.

---

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

\(^2\) These can be listened to at this web address: http://www.linguistics.berkeley.edu/~mcguire/dissertation.html
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 alveopalatal, 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 Krishnamurthy, 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 contrasts.

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

Labeling consisted of a single random presentation of a syllable from the full set of stimuli. Subjects were instructed to categorize each sound as category “A” or “B” by pressing the leftmost or rightmost buttons (respectively). All stimuli with alveopalatal 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 labeling and were presented the entire stimu-
lus set during each block. During the first block subjects received accuracy feedback; this was the labeling familiarization period. 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 labeling 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 labeling blocks with accuracy feedback. Subjects who achieved an overall session accuracy of greater than 85% or completed four sessions 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 labeling training, five blocks of the labeling testing, and finished with the discrimination test. 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 labeling 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: $t(647) = -4.91$, $p < 0.001$; labeling: $t(25) = 5.93$, $p < 0.001$). 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 labeling 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$^5$. Of the remaining twenty-one subjects, seven did not demonstrate substantial improvement, having a labeling $d' < 1$ on the final session and were classified as “non-learners”$^4$. 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

---

$^5$ 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.

$^4$ It is possible that these subjects could improve with more training or a different training paradigm. This was not attempted.
switch to using the fricative dimension. The fourteen remaining subjects all achieved a final labeling $d' > 2.0$, 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 $d' = 1.62$).

![Figure 3: Cumulative labeling in pre- (left panel) and post-test (right panel) for all learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Black filled squares indicate > 67% labeling as "A". Gray squares indicate 33%-67% labeling as “A”. White squares represent < 33% labeling as "A".](image-url)
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 labeling task. This resulted from the task and is due to the small differences in step size necessary to examine within category changes in perception. 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 ($p < 0.05$).
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. Further, this resulted in a significant increase in sensitivity to the new dimension. 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 training. 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.

These results support those of Goldstone (1994) and Francis and Nusbaum (2002) in showing a significant increase to the trained dimension. If these dimensions are viewed as integral, then the behavior of the irrelevant dimension is as predicted by Goldstone (1994). However, if seen as separable, these results contradict those of Goldstone (1994) who found acquired equivalence for an irrelevant, highly separable dimension. Of course, such a comparison is complicated by the differences in the two experiments. First, this experiment uses speech categories rather than visual ones, so differences may be due to differences in these two modes of perception. Secondly, the two types of stimuli have a different status; in Goldstone's work the stimuli were squares having no particular meaning beyond the experiment, unlike this experiment where the stimuli may have special status. Third, and relatedly, Goldstone's stimuli were manipulated such that each pair of adjacent squares was equally discriminable in each dimension. As the current experiment used “natural” speech categories, this manipulation was not done. In fact, for this contrast, the vocalic dimension is generally privileged for English listeners.

Experiment 2: Vocalic dimension categorizers

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

The stimuli for this experiment are the same as for the previous experiments. The procedure was also identical to the previous experiment except that the category boundary was defined along the vocalic dimension and all subjects participated in only a single training session for a total of three sessions. Stimuli with vocalic parts closer to the alveopalatal end (v0-v4) were considered category “A” while those closer to the retroflex end (v5-v9) were category “B”. Only differences in the vocalic dimension were relevant to categorization (see Figure 5).
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.
Results and discussion

The labeling 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 labeling 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 labeling in pre- (left panel) and post-test (right panel) for all vocalic dimension learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Black filled squares indicate > 67% labeling as "A". Gray squares indicate 33%-67% labeling as “A”. White squares represent < 33% labeling as "A".
The discrimination results were further 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](image-url)

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

Figure 8: Vocalic dimension learner's change in sensitivity from pre-test to post-test cross-category "X" and within category "."
Experiment 3: Two-dimensional categorizers

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

The stimuli were identical to those used in previous experiments. 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 alveopalatal fricative and vocalic steps (f0-f4 + v0-v4). Category “B” consisted of alveopalatal fricative steps with retroflex vocalic steps (f0-f4 + v5-v9). Category “C” consisted of retroflex fricative steps with alveopalatal retroflex 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, while “A” and “D” are "cooperating cue" categories. The four leftmost buttons of the box were labeled “A”, “B”, “C”, and “D” respectively for labeling. 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.
Results and discussion

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 indepen-
dently 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. sub-


different 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. sub-
jects 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.
Figure 11: Examples of post-test labeling 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.](image-url)
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 affected cross-category sensitivity and to a smaller extent within-category sensitivity; there was no within-category compression.

Figure 13: Two-dimension learner's change in sensitivity from pre-test to post-test cross-category "X" and within category "."
These results support the findings of Goldstone (1994) for the two-dimensional learners in that increases in sensitivity were found for both dimensions. However, unlike Goldstone, who found the least improvement for two-dimensional learners, this experiment had the largest increases in $d'$ from pre-test to post-test. This discrepancy can be attributed to training length differences in the two studies. In Goldstone's study all subjects, regardless of categorization group, received the same amount of training. The goal of training in this study was equal performance across experiment. On average, listeners in Experiment 3 heard 1800 of training trials, listeners in Experiment 1 heard 600-1800, and listeners in Experiment 2 heard 200. This also may account for the significant improvement in the vocalic dimension in this experiment (and unlike Experiment 2) because subjects trained for a longer period. These effects of training can be further confirmed by examining the performance of subjects who receive no training.

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

The stimuli were the same as used in previous experiments. This experiment consisted solely of the two discrimination tests, administered on separate consecutive days. No labeling was conducted. 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 for their participation.

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.

Post-hoc analysis

As discussed earlier, there was no attempt to standardize the perceptual distances between the discrimination pairs; instead the natural distances were maintained. This is in contrast to Goldstone's study in which a separate experiment was conducted to establish uniform sensitivities with and across each dimension before training. Because of this, it is useful to know the extent to which subject's initial sensitivities to the Polish stimuli used in these experiments was biased or “warped”. In order to explore this, the following post-hoc analysis of the pre-test data was conducted to explore subjects' perception of the stimulus space.
Methods

The pre-test discrimination data from the previous four experiments was combined and analyzed as a whole. All subjects (n=79) were included regardless of performance on later training and post-test conditions. The stimulus pairs were classified as before by dimension, but also by region within the stimulus space and whether the two dimensions were primarily correlated or conflicting.

Results

The results were converted to $d'$ units as before; however the data was recoded. In this analysis each dimension was divided into three regions with the step 0-3, 3-6, and 6-9 pairs collapsed across each dimension. Specifically for each dimension it is assumed for the analysis that 0-3 represents alveopalatal within category discrimination, 3-6 represents cross-category discrimination, and 6-9 represents within category discrimination for retroflex cues. Further all pairs were divided into two groups, those that had correlated fricative and vocalic cues and those where the two cues were in conflict. A repeated measures ANOVA with the factors Dimension (fricative, vocalic), Region (alveopalatal, boundary, retroflex), and Correlation (correlated, conflicting) found a significant main effect for Region ($F[2]= 32.83, p < 0.001$) and a significant interaction between Dimension and Region ($F[2]=5.18, p < 0.01$). The Dimension by Region interaction is shown in Figure 14. Subsequent Tukey tests of means showed that the regions differ by dimension only for the retroflex cue ($p <0.01$). The alveopalatal and boundary regions differed significantly for both dimensions (fricative: $p<0.001$; vocalic: $p<0.001$) and the retroflex region differed from the boundary only for the fricative dimension ($p<0.01$). The difference between the alveopalatal and retroflex regions was significant for both dimensions (fricative: $p < 0.05$; vocalic: $p < 0.001$).
This analysis confirmed that English listeners do not perceive the stimulus space uniformly. They were least sensitive to differences within the alveopalatal region for both dimensions. For the fricative dimension, subjects were most sensitive to cross-boundary pairs while the retroflex pairs are the most easily discriminated in the vocalic dimension. No difference between discrimination of stimuli having correlated or conflicting cues was found.

Both dimensions exhibited a higher degree of sensitivity to retroflex tokens than alveopalatal. A likely explanation for this is that the retroflex tokens are most similar to the native /ʃ/ category (McGuire 2007b) and subjects were assimilating these tokens to the native category. Additionally, even though the previous
experiments demonstrate that English listeners show a bias towards using the vocalic dimension for categorization, this did not seem to be fully borne out by the results in this analysis as the listeners were more sensitive to the vocalic dimension only for the retroflex ends of the continuum.

General discussion

The results from this study demonstrate that targeted training directing attention to a specific dimension of contrast results in perceptual warping such that stimuli become more distinct across category boundaries and, under certain conditions, within category. The within category sensitization is localized to the more difficult tasks, fricative dimension learning alone and with vocalic dimension learning, 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. 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 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, although listeners could easily label using the vocalic dimension and had difficulty using the fricative dimension, this pattern was not fully present in the discrimination data; only for retroflex cues was this true. Quite possibly, however, this is a task related effect as the discrimination paradigm was necessarily highly sensitive, having short interstimulus intervals, in order to assess fine-grained within-category changes rather than broad category level changes in sensitivity (as might be shown by an ABX paradigm, for example.) This is supported by the results of Guenther, Husain, Cohen, and Shinn-Cunningham (1999) where a within category compression effect was found only when a long ISI, categorical discrimination task was used, as opposed to a short ISI discrimination task which revealed no effect of training.

However, it is also 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, Fox 1984, Johnson 2004, Johnson and Babel 2007). 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. Most directly, subjects were equally sensitive to correlated and conflicting cue stimuli as demonstrated by the post-hoc analysis. Additionally, 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).

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 phonetics and phonology reading group for valuable input.

References


York Press, 121-154


LIST OF FIGURES

Figure 1: Stimuli design. Each circle represents a particular combination consonant and vowel from each continuum.................................................................8

Figure 2: Category assignments for the stimuli in experiment 1. Heavy circles denote discrimination stimuli. .............................................................................................................10

Figure 3: Cumulative labeling in pre- (left panel) and post-test (right panel) for all learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Black filled squares indicate > 67% labeling as "A". Gray squares indicate 33%-67% labeling as “A”. White squares represent < 33% labeling as "A".................................................................12

Figure 4: Fricative dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions........................................................................................................14

Figure 5: Category assignments of the stimuli for experiment 2, vocalic dimension categorizers. Stimuli marked with heavy circles were used in the discrimination task.........................................................17

Figure 6: Cumulative labeling in pre- (left panel) and post-test (right panel) for all vocalic dimension learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Black filled squares indicate > 67% labeling as "A". Gray squares indicate 33%-67% labeling as “A”. White squares represent < 33% labeling as "A". .........................................................18

Figure 7: Vocalic dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions........................................................................................................19

Figure 8: Vocalic dimension learner's change in sensitivity from pre-test to post-test cross-category "X" and within category "."........................................................................................................20
Figure 9: Category assignments of the stimuli for experiment 3, two-dimension categorizers.

Figure 10: Cumulative labeling 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.

Figure 11: Examples of post-test labeling 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.

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 ".".

Figure 14: Sensitivity to each region of the stimulus space for the fricative "F" dimension and vocalic "V" dimension for all subjects' discrimination pre-tests.