The role of orienting attention for learning novel phonetic categories

Susan G. Guion and Eric Pederson
University of Oregon

Authors' address:
Dept. of Linguistics
1290 University of Oregon
Eugene OR 97403-1290
USA

email: guion@oregon.uoregon.edu
epederso@darkwing.uoregon.edu

fax: 1-541-346-3917

Short title: Attention and phonetic category learning

Abstract. The current study examines the role of attention in phonetic category formation by experimentally manipulating endogenous orienting of attention. Two native English speaking participant groups were trained with an identical set of novel Hindi words containing unfamiliar consonants produced by multiple native Hindi speakers. Via instructions, the sound-attending group (N=37) was oriented toward phonetic information and the meaning-attending group (N=39) was oriented toward semantic information. A pretest and posttest assessed their learned ability to differentiate Hindi consonants and acquisition of word meaning. Participants oriented to phonetic information demonstrated greater learning of trained novel phonetic categories; whereas participants oriented to semantic information demonstrated greater learning of meaning. These results call for integration of attentional mechanisms into models of phonetic category formation (acquisition) as well as category modification (sound change). Different allocations of attentional resources by children and adult learners may account for much of their different success rates.

Keywords: *Adult second language learning, phonetic category, Hindi, consonant discrimination, attention, cross-language speech perception*

Introduction

The role of attention in learning is well established in many domains. For example, Curran & Keele (1993) differentiate learning of sequences resulting from attentional processes as opposed to learning from separate non-attentional mechanisms. In addition, many researchers have theorized about the importance of attention in second language learning (Schmidt, 1990; 2001; Tomlin & Villa, 1994; Robinson, 1995; Leow, 1997; 1998; Simard & Wong, 2001). On the other hand, the role of attention in phonetic category formation, whether learning ones' first categories natively in childhood or learning novel categories in a second language as an adult, is relatively undescribed. There is, however, some discussion in the literature about the role of attention in shaping perceptual space and, hence, phonological discrimination ability in first and second language acquisition.

Studies investigating infant development of perceptual discrimination have shown that a perceptual reorganization takes place within the first year of life. Most contrasts not found in the child's linguistic environment are better discriminated before about 10-12 months of age. After this age the ability to discriminate these contrasts decreases. These studies have argued that this is not a loss in auditory capabilities, but rather a reorganization of perceptual space (e.g., Best, McRoberts, & Sithole, 1988; Werker, 1989; Kuhl, 1991; Werker & Pegg, 1992; Kuhl, 2000)

Nosofsky (1986) proposed that perceptual dimensions are modified by experience. Namely, selective attention to particular perceptual dimensions changes similarity relationships such that differences along attended dimensions expand and differences along unattended dimensions contract. In other words, finer-grained distinctions are made along the attended dimensions. While Nosofosky's experiments were with visual categorization, a recent study by Guenther, Husain, Cohen, & Shinn-Cunningham (1999) in the auditory domain finds that it is possible to either increase or decrease discriminability with non-speech stimuli.

Based on Nosofsky's work, Pisoni, Lively, & Logan (1994) ascribe native adult phonetic processing as the consequence of attentive processes specifically aligned from prior linguistic experience. They suggest that, in principle, adult attentive processes can be modified through linguistic training (p. 156). This could have the effect of better discrimination of non-native phonetic contrasts.

In a series of experiments, Pisoni and colleagues (Logan, Lively, & Pisoni, 1991; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997) have demonstrated that laboratory training can effectively modify native Japanese listeners' ability to discriminate English contrasts not found in Japanese. In training they used stimuli produced by multiple talkers, claiming that the "modification of attention is … promoted by stimulus variability … Thus the role of stimulus variability is to provide a representative sample of possible exemplars so that changes in the relative weightings of different acoustic cues appropriate to the novel categories can take place." (Logan et al., 1991: 883).

The current study extends this work by directly examining the role of attention in phonetic category formation by experimentally manipulating attention during training. The benefits of variability across the exemplars of a trained category have been long established (Posner & Keele, 1968). So, like the above studies, the current study will employ a training paradigm using multiple talker stimuli. Categories formed with more varied stimuli are, in general, more robust and able to accomodate greater variety of novel stimuli.

Goldinger (1998), following the model of Hintzman (1986), finds that detailed, talker-specific episodic traces are stored in long-term lexical representations. Goldinger argues that these episodic traces constitute the basic substrate of lexical organization. Evidence of detailed information in phonetic categories is provided by Miller (1995), who found that within-category stimuli vary systematically in category goodness ratings and that even the best exemplars were subject to context-dependent effects. From this Miller concludes that the representation of phonetic categories includes fine-grained detail about phonetic form. These results suggest that whether or not there is development of an abstract or prototype category, there must be some degree of exemplar storage built into phonetic categorization.

These observations are incorporated into usage based models of phonetic categories (Johnson, 1997; Pierrehumbert, 2001; Bybee, 2001). While the details of subsequent processing and weighting of information vary in these models, each relies on a mechanism that derives information from all input equally, unmediated by attention. In other words, these models have the hearer/speaker in relatively passive roles with respect to the acquisition of exemplars.

Working to explain the mechanisms of sound change, Lindblom, Guion, Hura, Moon, & Willerman (1995), Guion (1996) and Bybee (2001) suggest that the acquisition of additional tokens for a phonetic category can modify the characteristics of that category. Lindblom et al. (1995) and Guion (1996) further argue that not all tokens are equal candidates for incorporation into the phonetic category. Only those tokens that were perceived during a "signal-oriented" mode (as opposed to a "content-oriented" mode) could be collected for incorporation and subsequent modification of a phonetic category.

In the current study, we seek to extend to language acquisition the insight that a signal-oriented mode is responsible for the collection of new exemplars. In other words, signal orientation leads to the creation of novel categories in addition to the adjustment of previously existing categories. The signal-oriented mode of Lindblom and colleagues was not defined in cognitive terms. Here, we situate the idea of signal orientation in terms of the cognitive mechanisms of attention.

Following Posner & Peterson (1990), the attentional system has several properties relevant to the concerns of this study. First, attention is anatomically separate from data processing systems. This means that the *collecting* of exemplars via the signal-oriented mode is independent from the *mapping* of these exemplars in terms of phonetic categories during speech perception.

Second, attention is divided into three systems: alertness, orienting, and detection. Alertness is conceived as the executive control modulating resources toward the orientation.

Orienting is the commitment of attentional resources to class-specific sensory stimuli. Orientation toward a class increases the chance of detection within that class. When a target is detected, the participant is aware of the signal and is processing it.

With in the study of second language acquisition, Tomlin & Villa (1994) argue that these three attentional mechanisms need to be distinguished. Specifically, they propose that the process of signal detection is necessary and sufficient for further processing and learning in second language acquisition. Similarly, Schmidt (1990; 2001) argues that "noticing" (equal to detection within selective attention) is necessary for learning a second language, but this learning can be implicit, i.e., without conscious abstraction or processing.

This study concerns the orienting of participants to particular features of complex verbal stimuli. Participant orientation can be controlled endogenously or exogenously (Posner, 1980). Exogenous orienting is the attraction of attention by external stimuli (e.g. flash of light). Endogenous orienting is the result of the individuals' directing their attention according to their goals or expectations. Endogenous orienting can be readily manipulated through directions to experimental participants. In other words, exogenous orienting is typically manipulated in an on-line manner, whereas endogenous orienting is typically manipulated off-line, i.e., separately from the presentation of stimuli. The current study manipulates endogenous orienting via varying instructions to encourage participants to orient more strongly to one or the other feature of identical stimuli, namely, to phonetic form or to the lexical semantics.[1] The control of orientation presumably increases 1) the likelihood of detection of the relevant acoustic cues, 2) the processing in short term memory, and 3) the transfer of the episode to long-term memory.

In this study, participants are divided into two groups, meaning-attending and sound-attending. In a pretest/posttest design, the effect of orienting attention will be investigated using training of novel Hindi words and Hindi consonant contrasts of varying difficulty. The hypothesis is that instructions to the participant prior to training will affect endogenous orienting of attention toward phonetic or semantic information during training. Specifically, participants oriented to phonetic information will demonstrate greater learning of trained novel phonetic categories; whereas participants oriented to semantic information will demonstrate greater learning of the meaning trained Hindi words.

Method

Participants

Seventy-six monolingual English speakers participated for course credit. None had lived in a non-English speaking region for more than six months, nor formally studied another language for more than three years, nor had any appreciable exposure to any Indo-Aryan language. Age ranged from 19-29 years. All reported normal hearing. Participants were randomly assigned to one of two groups: "sound-attending" and " meaning-attending". Each group had 13 males, and the sound-attending group had 24 females while the meaning-attending group had 26 females. Participants were unaware of the group assignment.

Materials

Minimal pairs were selected to represent the following Hindi stop consonant contrasts: [b]-[t̪] (voiced bilabial vs. voiceless unaspirated dental), [k]-[g] (voiceless unaspirated velar vs. voiced velar), [b]-[bʱ] (voiced bilabial vs. voiced aspirated/murmered bilabial), [k]-[kʰ] (voiceless unaspirated velar vs. voiceless aspirated velar), and [t̪ʰ]-[ʈʰ] (voiceless aspirated dental vs. voiceless aspirated retroflex). In addition ten words, "singletons", were included which were not part of a minimal pair contrast. The initial stop consonants in these words are different from those found in the minimal pairs. Table 1 lists the words used.

Table 1. Hindi words used in the study

Words in minimal pairs		**Singletons (not part of minimal pairs)**
[b] vs. [t̪]		
baːt ‘language’	t̪aːt ‘father’	dʱaːr ‘flow’
biːs ‘twenty’	t̪iːs ‘thirty’	dʱuːr ‘dust’
[k] vs. [g]		gʱol ‘solution’
kaːn ‘ear’	gaːn ‘sing’	gʱoʈ ‘polish’
kuːɻh ‘foolish’	guːɻh ‘secret’	paːs ‘nearby’
[b] vs. [bʱ]		pʌnd ‘advice’
beːl ‘vine’	bʱeːl ‘mixture’	pʰaːg ‘red powder’
bor ‘bangle’	bʱor ‘daybreak’	pʰaːɻ ‘split’
[k] vs. [kʰ]		pʰʌl ‘fruit’
kaːdʒ ‘task’	kʰaːdʒ ‘itch’	piːtʃ ‘rice water’
kaːl ‘time’	kʰaːl ‘skin’	
[t̪ʰ] vs. [ʈʰ]		
t̪ʰir ‘permanent’	ʈʰir ‘cold’	
t̪ʰok ‘quantity’	ʈʰok ‘blow’	

Four native Hindi speakers (two female) were recorded saying the words in Table 1. The words were presented to the speakers in Devanagri script and conventional Roman

transliteration. The speakers were recorded onto DAT tape repeating each word three. Words from three of speakers (two female) were edited for use in the training portion of the study. The three tokens of each word were digitized with a 22.05 kHz (16bit) sampling rate and edited into three separate files. The beginning of the files coincided with the onset of acoustic energy in the word and ended with the offset of acoustic energy. The digitized files were then normalized to 50% peak intensity.

Words from the fourth (male) speaker were used for the pre- and posttest. Two types of stimuli were created. The first type was used in the semantics test. The words were edited into files using the same procedure described for the first three speakers. However, only the second production was used for the semantics test.

The second stimulus type was used in the discrimination test. For these stimuli, consonant-vowel (CV) sequences were created from the minimal pair words only. The first, second and third productions of each word were used. First, vowel length was measured from the stop release (including aspiration and murmur) for each token. The shortest example of the vowel length in any of the six tokens for each minimal pair was the base line for truncating the other words in that minimal pair. The truncated vowels were ramped from 100% to 0% intensity over the last 20ms. The stimuli were normalized to 50% peak intensity. For each consonant contrast to be tested, this yielded six CV sequences with the same vowel duration (three productions of each member of the sound contrast).

The acoustic characteristics of the stimuli were investigated for minimal contrasts. Both the [k]-[g] and [k]-[kʰ] contrasts are distinguished by voice onset time (VOT) in a way unlike English. In word initial position (as with these stimuli), English typically has a VOT of 20ms vs. 80ms for the two-way voiced vs. voiceless velar plosive contrast (Lisker & Abramson, 1964). The [b]-[bʱ] contrast is differentiated by a period of breathy phonation, not found in English. The [t̪ʰ] vs. [ʈʰ] contrast uses a place distinction not found in English. Note that the [b]-[t̪] contrast is quite similar to the English place contrast between bilabials and alveolars.

Tables 2 through 5 report the results of the acoustic measurements. Speakers 1 and 3 are female and speaker 2 and the test speaker are male. Note that the VOT measures consistently differ for the [k]-[g] and [k]-[kʰ] contrasts. For the [b]-[bʱ] contrast, the segments are differentiated in two ways: 1) [bʱ] has less pre-voicing than [b] and 2) the pre-voicing for [bʱ] is followed by a period of breathy voicing.

The dental [t̪ʰ] vs. retroflex [ʈʰ] contrast is differentiated by the third formant (F3) transition of the following vowel. The F3 is lower after the retroflex consonant. The formant measures were made by centering an analysis window at the end of the second clear period in the vowel. Two spectra were calculated based on Fast Fourier Transform (FFT) and Linear Predictor Coefficients (LPC) and overlaid in the same window. The FFT used a 512 point window (43 Hz bandwidth) with an applied window length of 23 ms and the LPC used 26 coefficients. The measurement from the LPC spectra was checked against the FFT to be sure that a spurious formant was not recorded. No consistent differences in VOT or peak spectral frequency of the

burst were found between the dental and retroflex. However, the overall spectrum of the burst and aspiration was more compact for the retroflex segments.

Table 2. VOT (ms) averaged across stimuli for [k]-[g]

	[k]	[g]
speaker1	21	-165
speaker2	27	-99
speaker3	52	-109
test speaker	38	-149

Table 3.VOT (ms) averaged across stimuli for [k^h]-[k]

	[k^h]	[k]
speaker1	119	16
speaker2	96	28
speaker3	119	37
test speaker	121	33

Table 4. VOT (ms) and breathy period (ms) averaged across stimuli for [b]-[$b^ɦ$]

	[b] VOT	[$b^ɦ$]VOT	[$b^ɦ$] Breathy
speaker1	-156	-116	84
speaker2	-118	-88	73
speaker3	-118	-93	95
test speaker	-148	-133	72

Table 5. F3 (Hz) averaged across stimuli for [ʈʰ]-[tʰ]

	F3 of [i]		F3 of [o]	
	[ʈʰir]	[tʰir]	[ʈʰok]	[tʰok]
speaker1	2831	2791	2794	2716
speaker2	2474	2382	2638	2501
speaker3	3065	3006	3033	2987
test speaker	2373	2269	2415	2379

Procedure

Participants were run using a pretest/posttest design over two sessions using a computer with headphone audio presentation. Participants adjusted the presentation volume to a comfortable level before testing began. The first session contained a phonetics pretest, in which digitized audio was presented, and a semantics pretest, in which digitized audio and English written words were presented. The participants received no feedback on either test.

The second session was at least six days later (to ensure forgetting of the details of the pretest) and contained a self-paced training session (approximately 30-40 minutes), followed by a five minute break, followed by phonetics and semantics posttests (with trials identical to the pretest). For both the pretest and the posttest, all participants first took the phonetics test and then the semantics test.

Depending on their group assignment, participants were directed to pay attention to either phonetic or semantic information as follows:

At the end of the first (pretest) session, participants were scheduled to return and told either "In the next session, we will be learning to hear the differences between these Hindi sounds." (sound-attending group) or "In the next session, we will keep working with these Hindi words and their meanings." (meaning-attending group).

At the beginning of the second (training) session, the sound-attending participants were told: "Your task is to listen carefully to the *beginning* of each word and try to learn the difference between the Hindi sounds. You will notice that words with different meanings often have similar beginning sounds. However, the fact that they have different meanings, tell us that they are distinctive sounds in Hindi. During the course of this session, try to learn to distinguish between these sounds as best as you can. After this session, you will participate in another discrimination test like the one you did last week. And we will see if this training improves your discrimination of Hindi sounds."

At the beginning of the second (training) session, the meaning-attending group was told: "Your task is to try and learn the meaning of the Hindi words. Initially, it may seem intimidating, because there are a fair number of words. The orders of the words have been scrambled. But by the end of the session, you will have heard each word many times. After the session, you will participate in another vocabulary test, like the one you did last week. You will hear the Hindi word and then select the English translation. We want to see how well this training helps one to learn foreign language vocabulary."

Training. Both groups received same stimuli (with no feedback) in the form of triads of words grouped in three blocks. Each block had 40 triads produced by a different native Hindi speaker. Each triad contained a Hindi minimal pair exemplifying one of five initial-position phonetic contrasts in Table 1 in addition to one other word, a singleton, which was not part of a minimal pair, also listed in Table 1.

Each minimal pair was presented four times in each block, 12 times over the course of training. Each non-minimal pair word was also presented four times, but never with the same minimal pair. The order within each triad was pseudo-randomized. A separate fourth speaker was used for all the stimuli in the pre/posttests.

For each triad presentation, participants saw three written English translations without sound. Then each of these words was presented separately after the participants clicked a "next" button. For the individual word presentations, participants heard a single audio presentation of the first Hindi word and simultaneously saw its written English translation. They clicked a "next" button and this repeated for the second word, then another "next" and then the final of the three words. Again participants clicked "next" and the following triad began.

Phonetics test. The phonetics test was given both before and after the training. In both cases, it preceded the semantics test. An AXB categorial discrimination procedure was used to test the five Hindi consonant contrasts listed in Table 1. In this procedure, participants heard a series of three CV stimuli binaurally. Each stimulus was created from a different word production. Two of the stimuli were created from different productions of the same word; one stimulus was created from a production of a word with a different initial consonant, but with the same following vowel. The first (A) and third (B) tokens represented different phonemic consonants in Hindi. The middle (X) token, while physically different from both A and B, had either the same consonant phoneme as A or as B. Participants clicked either of two buttons "first" and "last" depending on whether they thought the middle token was more like the first or the last sound.

The test consisted of 80 trials in a two counter-balanced blocks with pseudorandom trial orders. Each sound contrast was tested 16 times: Eight times with CV tokens derived from one minimal pair and eight times with CV tokens from the other minimal pair. Each of the four possible orders (AAB, ABB, BBA, BAA) was presented equally for each contrast type with no two tokens co-occurring more than once. The inter-stimulus interval was 750ms and the inter-trial interval was 1500ms after response.

Prior to testing, participants had 16 practice trials covering two phonetic contrasts neither tested nor trained ([ɖ]-[g] and [p]-[pʰ]). As part of these practice trials, immediate feedback as to the correct answer was provided after each incorrect response.

In addition to the native English participants, two native Hindi speakers were run on the discrimination test (without training) to establish that the tested contrasts were indeed discriminable by native speakers. The four contrasts ([b]-[bʱ], [b]-[t̪], [k]-[g], [kʰ]-[k]) all were discriminated successfully above 94%. The [t̪ʰ]-[ʈʰ]contrast showed more difficulty at 81%, i.e. each Hindi-speaking participant made three errors on 16 trials; only one of these trials had an error by both participants. That trial did not receive substantially more errors than other trials by the native English participants.

Semantics test. The same semantics test was given both before and after the training. In both cases, it followed the phonetics test. In pilot work, participants had no trouble with the design of the semantics test, so no practice trials were given.

The test consisted of 30 trials testing the ability to select the correct translation of the 30 words in Table 1. In each trial, the participants were presented with one Hindi word token presented twice binaurally over headphones. The participants were instructed to choose one answer out of four simultaneously presented English translations. Sometimes the word presented was part of a trained minimal pair and sometimes the word was a trained singleton. The participants were not alerted as to trial type. In each minimal pair presentation trial type, there were four possible response types: (1) a correct answer, (2) an incorrect answer that was the translation of the other member of the minimal pair, (3) an incorrect answer that was the translation of a trained singleton, and (4) an incorrect answer that was a translation of a Hindi word, which was not trained. The serial order of the four response types was randomized across the trials. For example, if the participant heard [baːt] ‘language’, which was trained as part of a minimal pair (see Table 1), a possible set of responses would be: (1) language, (2) father, (3) fruit, and (4) mud.

In each singleton presentation trial type, there were four possible response types: (1) a correct answer, (2) an incorrect answer that was the translation of one member of a minimal pair, (3) an incorrect answer that was the translation of the other member of the same minimal pair, and (4) an incorrect answer that was a translation of a Hindi word, which was not trained. For example, if the participant heard [pənd] ‘advise’, which was trained as a singleton (see Table 1), a possible set of response would be: (1) advise, (2) permanent, (3) cold, and (4) curved.

Each written translation, including trained and non-trained, appeared in three trials and all the trained translations were the correct response one time. The untrained translations were always incorrect responses. Each translation was presented in conjunction with any other translation (except for the translations of minimal pairs) in only one trial. The correct response was equally distributed across the possible serial orders of the responses.

Two blocks with different pseudo-random trial orders were counterbalanced across the participants, with each participant receiving one order for the pretest and the other for the posttest.

Results and discussion

Phonetics test

Results. An ANOVA with the factors of Group (sound-attending vs. meaning-attending) X Session (Session 1 vs. Session 2) X Contrast ([b]-[bʱ], [b]-[t̪], [k]-[g], [kʰ]-[k], vs. [t̪ʰ]-[tʰ]) with repeated measures on the Session factor was performed on the number of correct responses (each participant heard a total of 16 trials for each contrast, for a possible 16 correct.)

The main effects of Session [$F(1,370)=15.03$, $p<.001$] and Contrast [$F(4,370)=367.36$, $p<.001$] were significant, indicating that overall, the participants responded more accurately in Session 2 and that some contrasts were responded to more accurately than others. In addition, the interaction of Contrast and Session was significant [$F(4,370)=4.05$, $p=.003$], indicating that some contrasts showed more difference between Session 1 and Session 2 than others. See Table 6 for mean proportions correct.

Table 6. Phonetics test: Mean proportions of correct responses for combined groups

Contrast	Session 1	Session 2	Mean
[b]-[bʱ]	.87	.92	.89
[b]-[t̪]	.98	.98	.98
[k]-[g]	.89	.91	.90
[kʰ]-[k]	.96	.96	.96
[t̪ʰ]-[tʰ]	.60	.64	.62
Mean	.86	.88	

The main effect of Group was non-significant [$F(1,370)=0.46$], as was the interaction of Group and Contrast [$F(1,370)=0.71$] and Group and Session [$F(1,370)=0.79$]. However, the three-way interaction of Group, Session and Contrast was significant [$F(4,370)=3.83$, $p=.005$], indicating that for some contrasts, the two experimental groups (sound-attending and meaning-attending) performed differentially on the Session 1 and Session 2 tests.

The three-way interaction was explored with a series of two-way ANOVAs investigating the Group X Session interaction on the number of correct responses for each contrast type individually. The only significant interaction was found for the [t̪ʰ]-[tʰ] contrast [$F(1,74)=6.31$, $p=.014$]. Figure 1 illustrates these results by plotting mean percent correct with standard error bars. Note that the sound-attending group showed a greater difference between Session 1 and Session 2, indicating a greater effect of training on discrimination accuracy than for the meaning-attending group.

A significant or near-significant effect of Session was also found for the [b]-[bʱ] [$F(1,74)=15.52$, $p<.001$], [k]-[g] [$F(1,74)=3.08$, $p=.063$], and [t̪ʰ]-[tʰ] [$F(1,74)=4.84$, $p=.031$] contrasts. As shown in Table 6, the Session 2 responses for these contrasts showed greater accuracy.

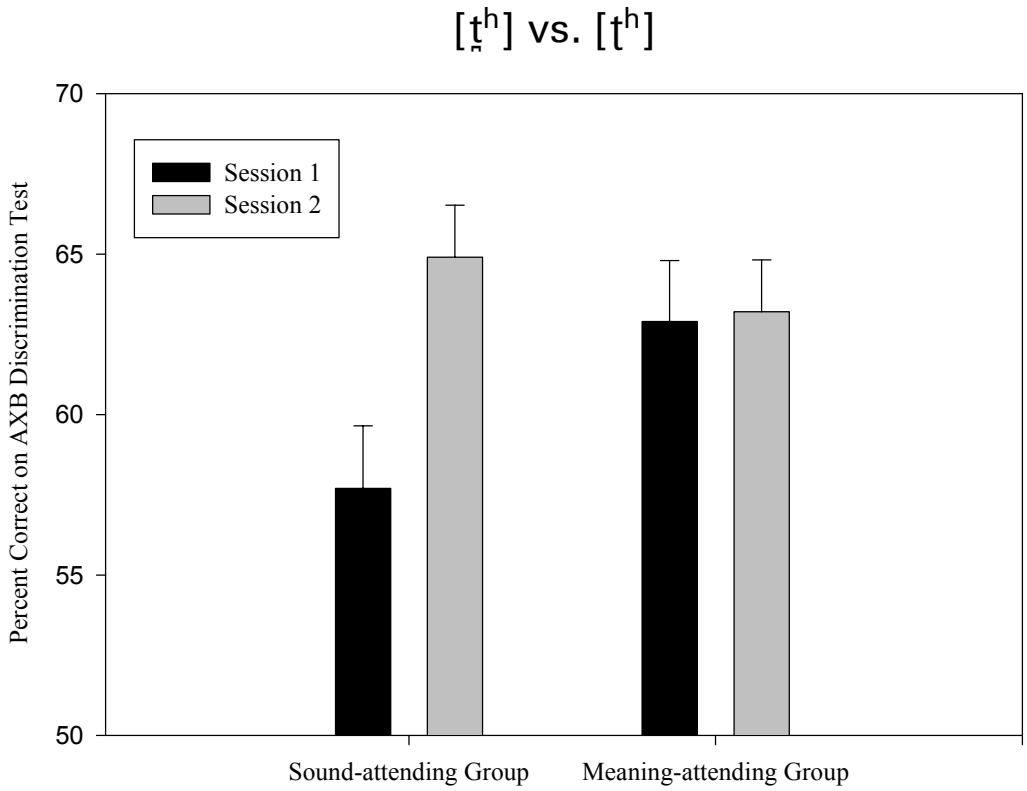

Figure 1. Phonetics Test: Group X Session interaction on the dental vs. retroflex aspirated voiceless stop contrast. (Mean percent correct on AXB discrimination test with standard errors)

Discussion. The overall improvement from Session 1 to Session 2 demonstrates an effect of learning during the training session. Some of the five contrasts tested, [b]-[bʱ], [k]-[g], and

[t̪ʰ]-[tʰ], were responded to more accurately overall in Session 2 than Session 1. These contrasts received lower overall proportion correct responses (.89, .90, .62 respectively) than the contrasts with no difference in correct responses from Session 1 to Session 2 (*viz*. .96 and .98 for [kʰ]-[k] and [b]-[t̪] respectively). Of the three contrasts that differed in correct responses from Session 1 to Session 2, only one exhibited differential improvement across the sound-attending and meaning-attending groups, namely the [t̪ʰ]-[tʰ] contrast. The sound-attending group responded more accurately at Session 2 than Session 1. The meaning-attending group showed no difference between Session 1 and Session 2.

Note that the [t̪ʰ]-[tʰ] contrast had the lowest correct responses overall when compared to the other contrast types. The finding that directions to orient attention to either semantics or phonetics only had an effect for the [t̪ʰ]-[tʰ] contrast could indicate that attention to phonetic form has a greater effect for more difficult contrasts (those with lower accuracy scores) than for easier contrasts. Alternatively, there could be a ceiling effect for all but the [t̪ʰ]-[tʰ] contrast, as scores for the other contrasts were quite high.

Nonetheless, the results for the [t̪ʰ]-[tʰ] contrast suggest that phonetic training in which the participant's attention was oriented to phonetic form facilitated learning of phonetic contrast to a greater extent than did orienting attention to learning the word meanings.

Semantics test

Results. The pretest was not expected to reveal results significantly different from random (i.e. close to 7.75 correct responses of 30). However, the sound-attending group performed marginally better ($F(1,74)=3.99$, $p=0.049$) than the meaning-attending group on the pretest. Accordingly, the two groups were compared in terms of their increase in performance from Session 1 to Session 2.

Overall, the combined groups went from 8.41 correct out of 30 trials in the pretest to 21.82 correct in the posttest, showing a robust learning effect. While both groups improved across the sessions, the meaning-attending group showed more learning than the sound-attending group, indicating that the instructions to attend to the meaning produced greater semantic learning. These results are shown in Table 7.

Table 7. Semantics test: Mean correct responses of 30 possible

	Session 1	Session 2
Sound-attending group (N=37)	8.86 (SD=2.11)	20.41 (SD=4.13)
Meaning-attending group (N=39)	7.97 (SD=1.77)	23.15 (SD=3.66)
Groups combined (N=76)	8.41 (SD=1.98)	21.82 (SD= 4.11)

The number correct for each participant was submitted to a Group (2) by Session (2) ANOVA, with repeated measures on Session. The main effect of Session was significant [$F(1,74)=793.03$, $p<0.001$]. In addition, the interaction of Session by Group was significant [$F(1,74)=14.71$, $p<0.001$].

In order to determine the extent to which phonetic learning played a role in the responses in the semantics test, we examined the distribution of the response types. Depending on whether the presented word was or was not part of a minimal pair, the available responses differed. For non-minimal pair words, all of the incorrect responses were translations of words with little phonetic similarity to the presented word. Conversely, words trained as part of minimal pairs had a close response alternative, namely the translation of a highly phonetically similar word.

In order to determine whether the sound-attending or meaning-attending Group performed better with words trained as part of minimal pairs, we examined the type of responses by Group. First, we examined the presentation of trained words that were *not* part of a minimal pair (singletons). The pattern of correct responses for just these words reflected the overall effect of greater learning by the meaning-attending Group. Table 8 presents the distribution of the responses by group.

Table 8. Semantics test: Mean responses for trained singletons by participant group.

	Response type		
Group	Correct singleton	Non-singletons[a]	Untrained
Sound-attending (N=37)	82.2%	12.2%	5.7%
Meaning-attending (N=39)	90.8%	4.6%	4.6%

[a]=Either translation of words trained as minimal pairs.

Second, we examined those trials testing a member of a trained minimal pair. Combining both groups, the participants overwhelmingly chose either the correct response or the incorrect minimal pair response. Unsurprisingly, trials testing words from the [b]-[t̪] contrast received largely correct responses and trials testing words from more difficult phonetic contrasts received relatively more incorrect minimal pair responses, see Table 9.

Table 9. Semantics test: Mean responses for trained minimal pair words (by phonetic contrast) for combined participant groups

	Response Type			
Phonetic Contrast	Correct	Incorrect of minimal pair	Singleton	Untrained
[b]-[bʰ]	69.7%	26.0%	2.3%	2.0%
[b]-[t̪]	93.1%	3.9%	2.0%	1.0%

[k]-[g]	46.4%	42.8%	8.9%	2.0%
[k^h]-[k]	67.4%	27.3%	4.3%	1.0%
[$t̪^h$]-[t^h]	52.0%	42.8%	3.3%	2.0%
Mean	65.7%	28.6%	4.1%	1.6%

As presented above for the phonetics test, there was a reliable interaction between Group, Session and Contrast type and a two-way interaction between Group and Session for the [$t̪^h$]-[t^h]contrast. Therefore, for the semantics test, we specifically examined the responses for minimal pairs containing the [$t̪^h$]-[t^h]contrast (which differentiated the two groups in the phonetics test).

The number of correct and incorrect responses for minimal pairs containing the [$t̪^h$]-[t^h] contrast by participant was submitted to a two-way ANOVA testing the effect of Group (2) by Response Type (Correct vs. Incorrect Minimal Pair). The main effect of Group was marginally significant [$F(1,74)=3.48$, $p=.066$], indicating that the meaning attending group gave marginally more correct and incorrect minimal pair responses and the sound attending group gave proportionally more incorrect non-minimal pair and untrained responses. No other effects were significant. In other words, even for the contrast that the sound-attending group learned better (as evidenced by the phonetics test), there was no apparent phonetic advantage for either group with this phonetic contrast while taking the semantics test.

Discussion. The results indicate that the meaning-attending group learned more word meanings overall, demonstrating that orientation of attention to semantics rather than phonetic form improved semantic learning. An analysis of the response types showed that most errors for both groups were with words trained as part of minimal pairs and the errors were largely the selection of the translation of the minimal pair counterpart.

Conclusions

The results of the phonetic discrimination task clearly show that instructions orienting attention to phonetic form enhance learning. This was demonstrated by greater improvement of the sound-attending group in the discrimination of the most difficult phonetic contrast. Conversely, orienting to semantics increases semantic learning. This was demonstrated by an overall higher correct response pattern for the meaning-attending group.

The phonetic pretest results showed that two of the five minimal pair discriminations ([b]-[$t̪$] and [k^h]-[k]) were at ceiling level of performance (98% and 96% correct respectively). For these contrasts, the posttest showed no improvement from training. Two other discriminations ([b]-[$b^ɦ$] and [k]-[g]) showed improvement from training, but still had high pretest scores (87% and 89% correct respectively). This suggests that all four of these discriminations are relatively straightforward for native English speakers. One interpretation is that one or both members of each of these four pairs readily map onto pre-existing English

phonetic categories. These English categories can then be used for the discrimination of the Hindi sounds (see Flege, 1995; Best, 1995). Of course, the actual cross-language mapping ultimately needs to be experimentally determined (as per Guion, Flege, Akahane-Yamada, & Pruitt, 2000). The fifth phonetic contrast ([t̪ʰ]-[ʈʰ]) is known to be difficult for English adults to discriminate (see Werker & Tees, 1984)—ostensibly because both of these Hindi consonants map onto the same English consonant: the voiceless alveolar stop [t].

Despite only about 30-40 minutes of self-paced training on 10 phonetic categories (5 minimal pair sets), there was clear improvement in the discrimination of non-English phonetic categories. Note, however, that the posttest was administered in the same session as the training. Most likely, the posttest scores would be lower if tested after a longer interval following the training. It remains to be determined how robust the differential learning by the two groups would be in this case.

Nonetheless, the improvement from relatively little training suggests that, with more substantial training, even difficult categories (such as [t̪ʰ]-[ʈʰ]) might be effectively learned with appropriate self-directed orientation to phonetic form.

In this study, there was no attempt to actively direct participant attention during learning. While further studies can seek online manipulation of attention, it is noteworthy that even participant-directed orienting of attention had a measurable effect. The experimental manipulation was simply varying the instructions to the participants. Presumably, these instructions created varying endogenously controlled orienting of attention across the two groups.

The instructions to the sound-attending group that they would be tested on phonetic form motivated the participants for greater orientation toward these forms. However the design of the training stimuli required some orientation to phonetic form even for the meaning-attending group because they had to learn phonetic form in order to distinguish the meanings of each element in a minimal pair. Similarly, while the meaning-attending group was instructed to orient their attention toward semantic information, both groups were equally and continually exposed to the translations of the spoken word stimuli and the translations of the Hindi words were available to aide differentiation of the phonetic form.

We cannot know the exact nature of the different attentional allocation precipitated by the different instructions. But its effects on both phonetic and semantic learning can be clearly seen. Presumably, the differing instructions gave different "weighting" to the importance of orienting to phonetic or semantic features of the stimuli during training presentations. These different "weightings" could either result in greater signal detection and/or greater manipulation in working memory. In other words, one simple possibility is that the orientation to phonetics increases successful detection of the relevant features of the sounds being trained. With greater detection of these features, there is greater chance of subsequent long-term storage, i.e. learning of the phonetic categories. Such an account can be applied more transparently to the case of phonetic learning than to semantic learning.

Another possibility is that both groups equally detect or notice both the phonetic features and the meanings of the words. However, because of the endogenously controlled orienting to either phonetic or semantic contrasts, each group will hold different aspects of the stimuli in working memory for different periods of time. Presumably, the longer a feature is manipulated or held in working memory, the greater the chance of "learning", i.e. creation of an abstract category or transfer of exemplars to long-term memory.

Implications for exemplar-based theories of phonetic categories

Work in exemplar-based learning commonly provides a simple model of exemplar traces. Essentially, all tokens of a category, which can be perceived, are equally available for learning. We feel there is need in such models for explicit mechanisms that provide for the *variable* detection, manipulation, and storage of episodic traces. Further work along the lines of this study should develop models about specific mechanisms by which exemplars are attended to and learned.

Bringing together the models of attentionally-driven learning with models of exemplar-based learning, we can begin to construct a psychological theory of phonetic category formation. Such a theory should ultimately explain the development of phonetic categories in first and second language acquisition as well as in historical sound change. With respect to language acquisition, this theory would more precisely explain the mechanisms underlying the poorer performance of adult (as opposed to child) phonetic category formation. A comprehensive theory of phonetic category formation must also explain phonetic category *modification*. That is, the modification of established categories in, for example acquiring a new accent and in sound change.

In addition to greater neurological plasticity than adults (or less entrenchment of the first language, see Flege 1995), we suggest that children may also, as a matter of course, have greater allocation of attention to phonetic features than adults normally do. This greater allocation of attention is the mechanism by which children better form categories in language learning as well as better modify their existing categories.

Adults largely rely on automatic processing of phonetic categories and, therefore, allocate relatively little attentional resources to phonetic form. However, in various contexts such as explicit language training, adults can increase attentional resources to phonetic form and thereby better modify their existing phonetic categories or develop novel ones.

Acknowledgements

We would like to thank the four anonymous Hindi speakers for recording our stimulus materials and Signe Rod for her invaluable assistance running the experiment. Partial funding for this work was kindly provided by the University of Oregon.

Endnote

[1] It is of note that the bulk of the attentional literature derives from experiments with visual stimuli. Even experiments in the auditory modality are principally about orientation and detection of certain classes of stimuli from among a variety of competing stimuli. Less studied is the presentation of complex stimuli where the participant must attend selectively to only some features of each stimulus. Attention to phonetic features within spoken full word presentation is such a case. Clearly in naturalistic acquisition of phonetic categories, people must selectively attend to certain components of complex input. Phones are overwhelmingly produced in a complex linguistic stream rather than in isolation.

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, MD: York Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology, 14*(3), 345-360.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Bybee, J. (2001). *Phonology and language use.* Cambridge: Cambridge University Press.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology, 19*(1), 189-202.

Flege, J. E. (1995). In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore, MD: York Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105* (2), 251-279.

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.

Guion, S. G. (1996). *Velar palatalization: coarticulation, perception and sound change*. Unpublished doctoral dissertation, University of Texas at Austin.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*(5), 2711-2724.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*(4), 411-428.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson, & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing*. NY: Academic Press.

Kuhl, P. A. (1991). Human adults and human infants show a "perceptual magnet effect" for prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50*(2), 93-107.

Kuhl, P. A. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science, 97*(22), 11850-11857.

Leow, R. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47*(3), 467-501.

Leow, R. (1998). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's (1994) fine-grained analysis of attention. *Applied Psycholinguistics, 19*, 133-159.

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista Di Linguistica, 7*(1), 5-37.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*(3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*(4), 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*(2), 874-886.

Miller, J. L. (1995). On the internal structure of phonetic categories: a progress report. In J. Mehler, & S. Frank (Eds.), *Cognition on Cognition* (pp. 333-347). Cambridge, MA: The MIT Press.

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

Pierrehumbert, J. B. (2001). Exemplar Dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency and emergence of linguistic structure* (pp. 137-157). Philadelphia: John Benjamins.

Pisoni, D., Lively, S., & Logan, J. (1994). Perceptual learning of nonnative speech contrasts: implication of theories of speech perception. In J. C. Goodman, & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 121-166). Cambridge, MA: MIT Press.

Posner, M., & Keele, S. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 2-25.

Posner, M. I., & Peterson, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25-42.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*(2), 283-331.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. (2001). Attention. P. Robinson (Ed), *Cognition and Second Language Instruction* (p. xxx-xxx). Cambridge: Cambridge University Press.

Simard, D., & Wong, W. (2001). Alertness, orientation, and detection: The conceptualization of attentional functions in SLA. *Studies in Second Language Acquisition, 23*(1), 103-124.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 183-203.

Werker, J. F. (1989). Becoming a native listener. *American Scientist, 77*, 54-59.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: models, research, implications* (pp. 285-311). Timonium, Maryland: York Press.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*(6), 1866-1878.