CATEGORICAL PERCEPTION OF NATURAL AND UNNATURAL CATEGORIES: EVIDENCE FOR INNATE CATEGORY BOUNDARIES[1]

JILL GILKERSON
gilkerson@psych.ucla.edu

The results reported here support the claim that naturally occurring phonemic contrasts are easier to acquire than unnatural contrasts. Results are reported from two experiments in which English speakers were exposed to non-native phonemic categories using a bi-modal statistical frequency distribution modeled after Maye and Gerken (2000). Half of the participants heard a distribution in which the category boundary was that of the Jordanian Arabic uvular/pharyngeal contrast, while the other half heard a distribution with an unnatural category boundary. Immediately after exposure, participants completed an A-X delayed comparison task, where they were presented with stimuli that crossed category boundaries. Results indicated that participants in the natural training group responded "different" to across-category pairs significantly more often than participants in the unnatural training group.

1. INTRODUCTION

The purpose of this research is to investigate whether adult acquisition of non-native phonemic contrasts is solely dependent on general learning principles or whether it is influenced by principles specific to natural language. Theories of innate grammar claim that humans are endowed with a genetically predetermined system specifically designed to facilitate language acquisition (cf. Chomsky 1965, among others). Nativists maintain that humans are "prewired" with sensitivity to linguistically relevant information in the input. Others have suggested that there are no learning mechanisms specific to language acquisition; it is facilitated by an interaction between statistical distribution of elements in the input and general learning mechanisms not specific to language (cf. Bates, 1979; Rummelhart and McClelland 1994, among others). Accord to the non-nativist approach, experience-dependent learning mechanisms play a primary role in language acquisition. If this were indeed the case, then language learners should focus primarily on the statistical distribution of speech elements in the ambient language and should be less sensitive to the quality (or naturalness) of the language input.

[1]I am grateful to Colin Wilson for his invaluable help during all phases of this experiment. I am also grateful to Pat Keating for advice on speech synthesis and helpful discussions, and to Henry Tehrani for assistance with the speech synthesizer. All errors are my own.

The current study tests these claims by investigating adult acquisition of non-native phonemic contrasts. Using synthesized speech technology, adults are trained to perceive a natural category boundary and an unnatural category boundary (i.e., a category boundary that does not occur in any of the world's languages). The question under investigation is the following: Given an equal amount of exposure in the input, will it be easier for adults to acquire a natural categorical distinction vs. an unnatural distinction?

2. BACKGROUND

2.1 *Studies Involving No Prior Exposure*

It is known from early categorical perception studies that, during a very early stage in language development, infants can perceive non-native phonemic contrasts with no prior exposure (Werker and Tees 1981, 1983; Werker, Gilbert, Humphrey and Tees 1981). These studies suggest that the ability to perceive such contrasts declines markedly during the second part of the first year of life.

Using a visually reinforced infant speech discrimination paradigm, Werker, Gilbert, Humphrey and Tees (1981) investigated whether English-speaking adults and 6-7 month-old infants were sensitive to the Hindi retroflex/dental distinction, a contrast that is not phonemic in English. They found that the infants could distinguish retroflex and dental voiceless stops, while adult native speakers of English had difficulty discriminating the phonemes.

Other studies have shown that infants acquiring English can perceive the French nasal/oral vowel distinction and the Czech strident distinction (Trehub 1976), as well as the Salish glottalized velar/uvular contrast (Werker and Tees 1983), while adult native speakers of English have difficulty perceiving these contrasts. Such findings are consistent either with the claim that infants are born with an innate sensitivity to naturally occurring phonemic category boundaries, or the claim that infants are sensitive to phonetic distinctions. Whether the skill these infants exhibit is phonemic or phonetic, the ability to distinguish non-native contrasts seems to be lost after the first year of life.

The categorical perception studies mentioned above do not involve any learning or training on the part of the adults or the infants. While these studies provide information infant and adult ability to distinguish non-native contrasts with no prior exposure, in order to gain information about the ability to acquire non-native contrasts and ultimately about the acquisition process, it is important to examine adult acquisition of non-native contrasts with some exposure.

2.2 *Adult Training Studies*

Several studies have suggested that adults can be trained to perceive category boundaries not evidenced in their native language (Bradlow Akahane-Yamada, Pisoni and Tohkura 1999; Lively, Pisoni, Yamada, Tohkura &Yamada 1994; MacKain et al. 1981). Bradlow, Akahane-Yamada, Pisoni and Tohkura (1999) trained native Japanese speakers to distinguish English /l/ and /r/ by repeatedly presenting them with minimal pairs and asking them to identify the phonemes; participants were rewarded monetarily for each correct answer. Participants in the experimental group took part in 15 1-1 ½ -hour long training sessions over a period of 3-4 weeks, while participants in the control group received no training. A posttest comparison showed that participants in the experimental group performed significantly better than participants in the control group on phoneme identification and production tasks 3 months after completing the training sessions. Results from this and similar studies indicate that, at least with intense training, adults have the ability to perceive phonemic category boundaries unattested in their native language.

While the studies involving no prior exposure show that adults cannot perceive non-native distinctions without training, the adult intensive training studies importantly show that adults do not loose the ability to perceive non-native distinctions provided that they are given sufficient exposure to the phonemes. However, participants in these studies made a conscious effort to "learn" the particular contrasts, and were reward monetarily for their success. This being the case, it is unclear to what extent such intensive training situations can generalize to natural language acquisition.

2.3 *Passive Learning of Non-Native Phonemic Contrasts*

Recent research has avoided this criticism through the use of short passive listening tasks. In a study designed to test whether adult sensitivity to non-native contrasts could be influenced by manipulating the frequency distribution in the input, Maye and Gerken (2000) report that adults can demonstrate knowledge of non-native category boundaries after only nine minutes of passive exposure. Their study investigated native English-speakers' perception of voiceless unaspirated /t/ (e.g., *st̲ay*) and voiced /d/ (e.g., *d̲ay*), which are not contrastive in English but are phonemic in Spanish, French and Japanese (Pegg and Werker 1997). The stimuli consisted of an eightpoint /ta/ – /da/ continuum, which varied the formant transition frequencies in the vowel. During a short training session, each participant listened to 192 tokens from the continuum whereby the distributional frequency of stimuli varied between two groups. Half of the participants were trained using a mono-modal distribution, in which tokens from the center of the continuum were presented four times as often as tokens from the edges of the continuum. The other group was trained using a bi-modal distribution, in which tokens near the endpoints were presented four times as often as tokens in the middle of the continuum.

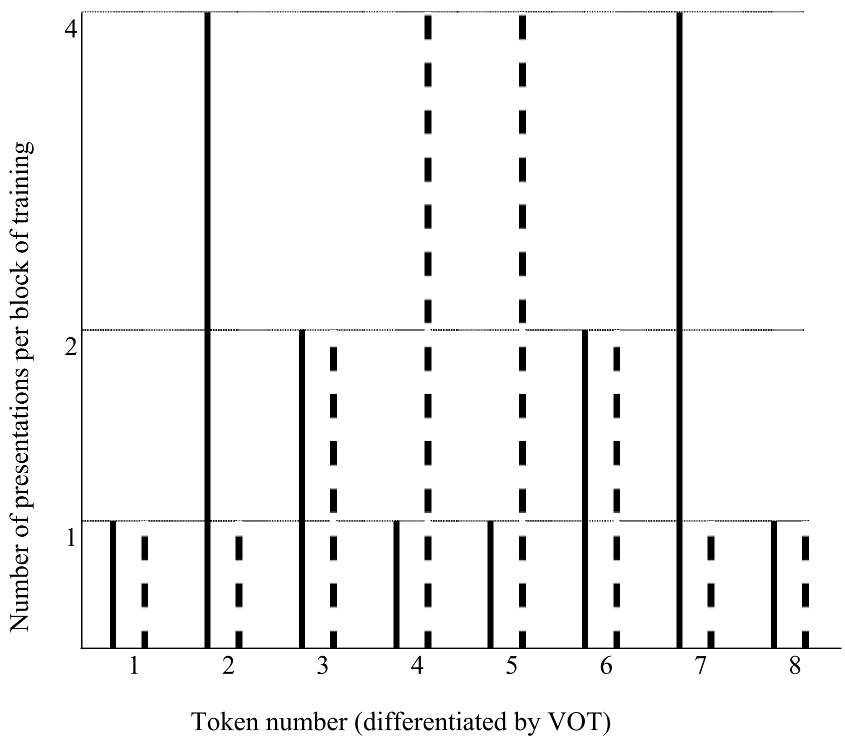

Figure 1. Stimuli presentation frequency during acquisition phase (Maye and Gerken 2000).

This frequency distribution was designed to investigate whether participants in the bi-modal training group and participants in the mono-modal training group would differ in their ability to categorically distinguish the non-native voiced /d/ – voiceless unaspirated /t/ contrast.

Results showed that participants in the bi-modal group were more likely to distinguish tokens at opposite edges of the continuum than participants in the mono-modal group. Thus, the results suggest that given only nine minutes of exposure to stimuli presented using a bi-modal statistical distribution, English-speaking adults can demonstrate sensitivity to non-native phonemic contrasts.

While the voiceless unaspirated /t/ – voiced /d/ distinction is not phonemic in English, the stimuli used in the above study were nonetheless quite similar to voiceless aspirated /t^h/ and voiced /d/, which are contrastive in English. This being the case, it is possible that the categorical perception of voiceless unaspirated /t/ and voiced /d/ reported in Maye and Gerken (2000) may have been facilitated by participants' knowledge of the English /t^h/ – /d/ distinction. The current study attempts to extend Maye and Gerken's results by investigating adult English-speakers' acquisition of the uvular – pharyngeal contrast (which is not phonemic in English) and an unnatural contrast unattested in any of the world's languages.

2.4 *Purpose of the Current Study*

Thus far, we have seen that infants demonstrate the ability to perceive non-native category boundaries while adults have difficulty distinguishing non-native phonemes without prior exposure. However, it is apparent that adults do not loose the ability to learn/perceive such contrasts, as they demonstrate knowledge of non-native category boundaries with even a small amount of exposure. While the adult training studies indicate that adults can learn to perceive non-native contrasts, it is not yet clear whether this ability is facilitated by genetically endowed language-specific acquisition capabilities, or general learning mechanisms.

The present study attempts to gain information about the mechanism behind adult acquisition of non-native contrasts by manipulating the quality of the phonemes in the input. Using speech synthesis techniques, it is possible to create unnatural category boundaries that are not attested in any of the world's languages. Given an equal amount of exposure to stimuli crossing both natural and unnatural category boundaries, the theories outlined above make different predictions. If perception of phonemic contrasts is facilitated by general learning principles and is only dependent on statistical distribution in the input, then adults should not differ in their ability to learn both natural and unnatural categories. Alternatively, if humans are prewired with sensitivity to linguistically relevant information in the input, then natural categories should be easier to learn than unnatural categories. The current study tests these predictions.

In the experiments described below, adult native speakers of English were exposed to non-native categories using a bi-modal statistical frequency distribution modeled after Maye and Gerken (2000). Half of the participants were trained to perceive the natural category boundary representing the Jordanian Arabic uvular – pharyngeal distinction, while the other participants were trained to perceive an unnatural boundary located within the pharyngeal side of the continuum. Exemplars in both training groups were presented with the same bi-modal statistical distribution. Immediately after training, participants completed an A-X delayed comparison task, where they were presented with stimuli that crossed category boundaries. If adults use general learning mechanisms dependent on statistical distribution information to perceive category boundaries, then the participants should not differ in their ability to distinguish across-category pairs. Therefore, the null hypothesis is the following: there will be no difference between the unnatural and natural training groups in their ability to distinguish across-category pairs.

3. Experiment 1 method

3.1. *Experiment 1 Stimulus Materials*

Experiment 1 used three types of stimuli, described in detail below. The experimental stimuli were CV syllables whose consonants varied along a pharyngeal – uvular (/Ξi/ – /□i/) continuum, the filler stimuli were various tokens of the syllables /bi/ and /di/ and the practice stimuli were natural recordings of English words.

3.1.1. *The Experimental Stimuli*

The experimental and filler stimuli were synthesized speech tokens created using the Synthworks digital terminal analog speech synthesizer (Tehrani 2000). The experimental stimuli consisted of a voiceless uvular/pharyngeal fricative continuum, modeled after the synthesized stimuli used in El-Halees (1985). In a study designed to investigate the role of F_1 in the perception of uvular and pharyngeal fricatives, El-Halees created a 12-point uvular – pharyngeal /Ξid/ – /□id/ continuum. The CV transition for the first stimulus contained an F_1 that was set at 350 Hz. For all subsequent stimuli, the F_1 was increased in 50-Hz steps up to 900 Hz, holding all else constant. According to El-Halees, native speakers of Jordanian Arabic perceived a category boundary at about 550 Hz, such that stimuli with a CV transition containing an F_1 below 550 Hz were generally identified as voiceless uvulars, and stimuli with an F_1 above 550 Hz were generally identified as voiceless pharyngeals.[2]

Following El Halees (1985), the experimental stimuli in the current study consisted of a 17-point voiceless uvular/pharyngeal fricative continuum, in which all parameters were held constant except the F_1. All experimental CV stimuli contained a high front vowel and a syllable-initial voiceless fricative ranging in steps from /Ξi/ — /□i/. The duration of the initial fricative was 120 ms, and the duration of the following vowel was 280 ms. F_0 was set at 180 Hz at the beginning of the vowel, linearly increased up to 190 Hz during the first 115 ms of the vowel, and remained constant for the final 65 ms of the stimulus. F_1 for the entire first stimulus was 350 Hz. For each subsequent stimulus along the continuum, F_1 was increased by steps of 40 at the onset of the transition, and decreased linearly down to the steady state of 350 Hz at the end of the first 55 ms of the vowel. Thus, the initial F1 frequencies in the CV transition ranged from 350 Hz to 990 Hz.[3] For all stimuli, F_2 began at 1750 Hz, increased linearly up to 2200 Hz at the end of the first 75 ms of the vowel, and was held constant for the final 100 ms of the vowel. F_2, F_3 and F_4 were excited during the fricative, with amplitudes of 40 db each. Voicing amplitude was set at 20 db at the onset of the vowel, rose linearly to 50 db at the end of the first 35 ms of the vowel and remained at a steady state for the duration of the stimulus. Frication amplitude was initially set at 40 db, increased linearly up to 55 db at the end of the first 20 ms of the stimulus and held constant throughout the fricative. F_3 – F_6 frequencies were identical in all stimuli, with values of F_3 = 2750 Hz, F_4 = 4000 Hz, F_5 = 4500 Hz and F_6 = 5500 Hz. Bandwidths for F_1 – F_6 were 50, 70, 110, 250, 200, and 100 Hz, respectively.[4]

[2]There was some variation among participants for tokens closest to the 550 Hz boundary. For example, only 70% of the participants identified stimuli containing a CV transition of 500 Hz as uvulars, and 70% of participants identified stimuli containing a CV transition of 600 Hz as pharyngeals.

[3]Although the F1 continuum used in El-Halees (1985) ranged from 350 Hz – 900 Hz, the continuum in the current study ranges from 350 Hz – 990 Hz, covering a slightly broader range. I assume that the human vocal tract is capable of producing first formant frequency up to 990 Hz in pharyngeals based on Al-Ani (1970) and Ghazeli (1977). Al-Ani (1970) provides spectograms of natural Arabic utterances, one of which is in the /□i/ environment. This spectrogram shows an F1 transition frequency around 800 Hz. Given the fact that this measurements is from a male speaker, I assumed that female speakers or children, who generally produce higher frequencies in all formants, can produce pharyngeals containing first formant frequencies as high as 990 Hz. Further, as indicated in Alwan (1986), Ghazeli (1977) reports F_1 frequencies in voiced pharyngeals around 900 Hz (it is unclear whether this is based on data from male or female speakers).

[4]In terms of synthesis parameters, the experimental stimuli used in the current study were identical to the stimuli in El-Halees (1985) with two exceptions. First, while the previous stimuli consisted of CVC tokens that were real words of Arabic, the stimuli in

3.1.2. *The Filler Stimuli*

The 10 filler stimuli consisted of five different variants of the two CV syllable types /bi/ and /di/. All filler stimuli were 245-275 ms in duration. The five variants of both syllable types differed slightly with respect to consonant length, vowel length, F_2, F_3, and/or transition duration.

3.1.3. *The Practice Stimuli*

Ten naturally recorded words of English were used as practice stimuli. The practice stimuli consisted of the following five pairs, which differed minimally by the word-initial consonant: *coat*, *goat*; *moon*, *noon*; *ram*, *lamb*; *sack*, *Zack*; *fat*, *vat*.

The experiment was created and implemented on a Macintosh using PsyScope (Cohen, MacWhinney, Flatt and Provost 1993), a computer software program designed for experimental research.

3.2. *Experiment 1 Participants*

Ten adult UCLA undergraduates participated in Experiment 1. All participants were paid five dollars.

Each participant was randomly assigned to a natural training group or an unnatural training group. Tokens in the natural training condition were distributed around a natural category boundary (F_1 = 550), while tokens in the unnatural training condition were distributed around an unnatural category boundary (F_1 = 790) (See figure 2 below for a diagram of the uvular/pharyngeal continuum used in this study.) Stimuli on the uvular side of the continuum included a CV transition containing an F_1 ranging from 350 – 510 Hz, and will be referred to as U1 for ease of exposition. Stimuli on the other side of the continuum were all considered to be pharyngeal fricatives, which were further split into two groups. The first group on the pharyngeal side of the continuum included a CV transition containing an F_1 ranging from 590

---

the current study were simple CV syllables. Since the participants in the current study were native speakers of English, there was no need to use real words of Arabic. Thus, the stimuli were shortened to maximize the amount of attention allotted to the first consonant. Second, the previous study used a 12-point continuum in which the F_1 was increased by 50 Hz steps, while the current study used a 17-point continuum in which the F_1 was increased by 40 Hz steps. Because two category boundaries were created from the same continuum (one category boundary for each group), it was necessary to construct a longer continuum.

– 750 Hz, and will be referred to as P1. The second group of pharyngeals included a CV transition containing an F_1 ranging from 830 – 990, and will be referred to as P2. Participants in the natural training group heard stimuli from U1 and P1, while participants in the unnatural training group heard stimuli from P1 and P2. As noted in figure 2, the natural category boundary was at F_1 = 550 Hz (between U1 and P1), and the unnatural category boundary was at F_1 = 790 Hz (between P1 and P2).[5] Essentially, we might think of U1, P1 and P2 as phonemic categories (both natural and unnatural) containing different variants of distinct non-native phonemes.

[5]Although phonemic distinctions within the larynx are quite rare, some languages of the Caucasus and the Pacific Northwest contrast pharyngeal and epiglottal fricatives. However, Esling (1999) performed a laryngoscopic analysis of pharyngeals and epiglottals, and found that the two phonemes were distinguished by manner and larynx position rather than location of the constriction. They report that epiglottal fricatives are trilled and have a raised larynx while pharyngeals are not trilled and have a relatively lower larynx. Based on the fact that the epiglottal – pharyngeal distinction is one of manner and larynx position rather than place, it was assumed that no languages of the world make a phonemic distinction within the pharyngeal side of the continuum. Thus, a category boundary located at 790 Hz was considered to be unnatural.

<<INSERT FIGURE 2 HERE>>

3.3. *Experiment 1 Procedure*

Participants were told that they would be listening to non-English words so that we could learn more about how native speakers of English perceive sounds that they have never heard before. They first completed a practice session, followed by the training session and testing phase, described below.

Practice session. First, participants completed a practice session in which they listened to English word pairs and were asked to indicate whether or not they heard the same word repeated twice, or two different words. Ten word pairs were presented, half of which contained different words (e.g., *moon*, *noon*) and half of which contained the same word repeated twice (e.g., *moon*, *moon*). Participants responded by pressing a button on a button box labeled "same" or "different". Right-handed participants pressed a button labeled "same" with their right index finger and a button labeled "different" with their left index finger. For left-handed participants, the labels were reversed; left-handed participants pressed the button labeled "same" with their left index finger and the button labeled "different" with their right index finger. The participants were asked to respond as quickly as possible, but not so fast that they would make mistakes. The practice stimuli pairs were separated by a 500 ms inter-stimulus interval, and there was a 2 second pause between each of the 10 trials.

Training session. Participants were told that they would hear a long list of non-English words during the listening portion of the experiment, and all they needed to do was simply sit and listen to the words.

During the training, all tokens were presented using a bi-modal statistical distribution. As indicated by the vertical lines in figure 2, stimuli in the center of U1 were presented four times as often as stimuli at the edges of U1. Likewise, stimuli in the center of P1 and stimuli in the center of P2 were presented four times as often as stimuli at the edges of P1 and P2. For example, participants in the natural training group heard CV syllables with transitions containing an F_1 of 430 Hz four times as often as stimuli with transitions containing an F_1 of 350 or 510 Hz, and twice as often as stimuli with transitions containing an F_1 of 390 or 470 Hz. Following Maye and Gerken (2000), this statistical distribution was designed to facilitate a categorical perception effect, such that tokens closest to the most frequently presented stimuli would be perceived as part of a single category.

Participants in each group heard 20 experimental and 20 filler stimuli per block of training. For example, participants in the natural training group heard 10 tokens from U1, 10 tokens from P1, and five tokens of /bi/ and /di/ per block. Likewise, participants in the unnatural training group heard 10 tokens from P1, 10 tokens from P2, and five tokens of /bi/ and /di/ per block.

Each token was separated by a 1 second inter-stimulus interval. Participants from each group heard 20 experimental and 20 filler stimuli per block of training. The entire training session consisted of 5 training blocks, for a total listening time of 4.3 minutes.[6]

Test session. Participants were told that during the last phase of the experiment they would be presented with pairs of non-English words and would be asked to indicate whether they thought the pairs were the same word repeated twice, or two different words. They were asked to respond as quickly as possible, but not so fast that they would make mistakes. Participants were reminded that this task was similar to the task in the practice session. The stimulus pairs were separated by an inter-stimulus interval of 500 ms, and there was a 2 second lag between trials.

The testing phase consisted of 68 test trials containing seven different test pair types. The significant items were the across-category pairs (e.g., /Ξi/$_{350\text{ Hz}}$ – /□i/$_{750\text{ Hz}}$ for the natural category group) and within-category pairs (e.g., /Ξi/$_{350\text{ Hz}}$ – /Ξi/$_{510}$ Hz for the natural group). Each across-category and within-category pair was presented twice in reverse order (e.g., /Ξi/$_{350\text{ Hz}}$ – /□i/$_{750\text{ Hz}}$, /□i/$_{750\text{ Hz}}$ – /Ξi/$_{350\text{ Hz}}$, /Ξi/$_{350\text{ Hz}}$ – /Ξi/$_{510\text{ Hz}}$, /Ξi/$_{510\text{ Hz}}$ – /Ξi/$_{350\text{ Hz}}$). Across-category pairs were a distance of 400, 240, and 80 Hz from each other along the continuum for both groups. The within-category pairs were all 150 Hz apart. Experimental stimuli were also paired with filler stimuli (e.g., /Ξi/$_{350\text{ Hz}}$ – /di/). Additionally, filler stimuli were paired with different filler word types (e.g. /di/ – /bi/) and variants of the same word type (e.g.,

[6] The participants in the current study were exposed to the experimental stimuli with a slightly different frequency than those in Maye and Gerken (2000). In the previous study, the experimental stimuli were presented 64 times, while in the current study the experimental tokens were presented a total of 100 times. As will be discussed in section *5*. Maye and Gerken actually used three different word-types representing the /t/ — /d/ continuum. Each of the three word types were presented a total of 64 times, so the participants actually heard 192 repetitions from the /t/ — /d/ continuum during the entire training session in Maye and Gerken (2000). This difference will become important in Experiment 2.

/di/$_1$ – /di/$_2$). Finally, all filler and experimental stimuli were presented in identical pairs (e.g., /Ξi/$_{350\ Hz}$ – /Ξi/$_{350\ Hz}$, /di/$_1$ – /di/$_1$). Since participants were expected to respond "same" to within-category pairs, identical experimental pairs, identical filler pairs and similar filler pairs (e.g., /di/$_1$ – /di/$_2$), all "same" responses to these pairs were counted as correct and all "different" responses were counted as incorrect. Likewise, since participants were expected to respond "different" to across-category pairs, filler – experimental pairs, and different filler pairs (e.g., /bi/ – /di/), all "different" responses to such pairs were counted as correct and all "same" responses were counted as incorrect. The entire test session contained 34 "same" pairs and 34 "different" pairs. Reaction times for each response were also recorded.

The test trials were randomized and presented in two different fixed orders, labeled list A and list B. Each list consisted of two blocks, and each block contained half of the items for each test pair type, including 17 "same" and 17 "different" items. Thus, each block contained 4 across-category pairs, 2 within-category pairs, 5 experimental identical pairs, 5 filler identical pairs, 8 experimental – filler pairs, 5 filler similar pairs and 5 filler different pairs. The first block of list A was the last block of list B, and the last block of list A was the first block of list B. Participants were given the option to take a break after completing the first testing block.

4. Experiment 1 Results

Responses to within-category pairs and across-category pairs are included in the analysis. Recall that across-category pairs for participants in the natural training condition contained one stimulus from U1 and one stimulus from P1, while across-category pairs for participants in the unnatural training condition contained one stimulus from P1 and one stimulus from P2 (see figure 2). Within-category pairs for participants in the natural training condition contained two stimuli from within U1, or two stimuli from within P1. Within-category pairs for participants in the unnatural training condition contained two stimuli from within P1, or two stimuli from within P2. Because stimuli varied along an F_1 continuum, the CV stimuli will be identified by the F_1 in the CV transition for ease of exposition (for example, the stimulus /□i/$_{510}$ has a CV transition containing an F_1 of 510 Hz, so it will simply be referred to as 510). The within- and between-category pairs included in the analysis are listed in Table 1.

Table 1. Within- and between-category pairs for participants in the natural and unnatural training groups

	Training Group		Difference in hertz
	Natural (N=5)	Unnatural (N=5)	
Within-category pairs	350—510	590—750	160 Hz
	590—750	830—990	160 Hz
Between-category pairs	350—590	590—830	240 Hz
	510—750	750—990	240 Hz
	350—750	590—990	400 Hz
	510—590	750—830	80 Hz

Because participants were expected to learn phonemic contrasts, responses were counted as correct if they indicated knowledge of the category distinction. Thus, "different" responses to across-category pairs were counted as correct, as were "same" responses to within-category pairs.[7]

Recall that each pair in table 1 was presented twice in reverse order. Thus, each within- and across-category pair in the analysis contains data from a total of 10 responses (2 per person in each group). Table 2 shows the number of correct "different" responses to across-category pairs for participants in the natural and unnatural training groups.

[7] It should be noted that all within- and across-category pairs were technically different sounds. For example, a within-category pair from U1 included a stimulus with a CV transition containing an F_1 of 350 Hz and a stimulus with a CV transition containing an F_1 of 510 Hz. Strictly speaking, the stimuli in this pair are different, so a "different" response would not technically be incorrect. However, since this experiment was designed to investigate whether participants could learn phonemic distinctions (which involves grouping phonetically different sounds into single categories and ignoring uninformative phonetic differences), only "same" responses to within-category pairs were counted as correct.

Table 2. Number of correct "different" responses to across-category pairs for each training group (Experiment 1)

Training Group					
Natural (N = 5)			Unnatural (N = 5)		
Stimulus pair	"different" responses		Stimulus pair	"different" responses	
	# / 10	%		# / 10	%
350-590	2	20%	590-830	(3/10)	30%
510-750	1	10%	750-990	(0/10)	0%
510-590	1	10%	750-830	(2/10)	20%
350-750	2	20%	590-990	(1/10)	10%
Total	6/40	15%	Total	(6/40)	15%

As illustrated in table 2, there is very little difference between responses for participants in the natural and unnatural training groups, both groups responded "different" to across-category pairs about 15% of the time. In order to obtain a normal distribution, an arcsine transformation was applied to the proportion of "different" responses for each participant. A one-way ANOVA indicates that there is not a significant difference in proportion of correct responses to across-category pairs between participants in the natural and unnatural training groups ($F(1,8) = 0.00$, $p = 1.00$).

Table 3 shows the number of correct "same" responses to within-category pairs for participants in each group.

Table 3. Number of correct “same” responses to within-category pairs
for each training group (Experiment 1)

Training Group					
Natural (N = 5)			Unnatural (N = 5)		
Stimulus pair	“same” responses		Stimulus pair	“same” responses	
	# / 10	%		# / 10	%
350-510	8	80%	590-750	9	90%
590-750	9	90%	830-990	10	100%
Total	17/20	85%	Total	19/20	95%

As table 3 indicates, “same” responses to within-category pairs are generally high, with little difference between the two groups. Again, an arcsine transformation was applied to the proportion of “same” responses to within-category pairs. There is not a significant difference between participants in the natural and unnatural training groups in proportion of correct “same” responses to the within-category pairs, as indicated by a one-way ANOVA (F(1,8) = .76, p = .41).

5. Experiment 1 Discussion

The results for Experiment 1 indicate no difference in performance between participants in the natural and unnatural training groups. Participants in both groups responded “different” to across-category pairs and “same” to within-category pairs with virtually the same frequency.

These results suggest that after 4.3 minutes of exposure to non-native phonemes, adults fail to demonstrate sensitivity to the “naturalness” of a phonemic category boundary. That is to say, although adults in the natural training condition were exposed to sounds that crossed a natural category boundary (attested in many natural languages), it was no easier for them to distinguish the natural phonemic boundary than it was for participants exposed to sounds that crossed a contrived, unnatural category boundary (attested in no natural language). On the face of it, these results suggests that the nature of the linguistic input does not affect acquisition of non-native contrasts and supports the notion that language learning might be driven by general learning principles not specific to language. An alternative explanation is that participants were not given enough exposure to the experimental stimuli to distinguish across-category pairs, and increasing the amount of exposure to the experimental stimuli would result in increased sensitivity to category boundaries.

Recall that the frequency distribution of the stimuli in Experiment 1 was slightly different from the distribution used in Maye and Gerken (2000). Maye and Gerken trained participants to categorically perceive voiceless unaspirated /t/ and voiced /d/ using three different word types. The experimental stimuli for each of the three word types were presented 64 times, for a total of 192 repetitions throughout their 9-minute experiment. The experimental stimuli used in Experiment 1 of the current study were presented a total of 100 times over a 4.3-minute time span. The purpose of the next experiment was to see whether doubling the amount of exposure from 100 to 200 repetitions (more similar to the 192 exposures used in Maye and Gerken (2000)) would facilitate categorical perception of the non-native contrasts. Experiment 2 tested this prediction by doubling the number of presentations of experimental stimuli during the natural and unnatural training sessions.

6. Experiment 2 method

The experimental method for Experiment 2 was identical to the method for Experiment 1 with one exception: the number of blocks in the training session was increased from 5 to 10 in Experiment 2. Thus, the amount of exposure to all stimuli in the training session was doubled, for a total listening time of 8.6 minutes. Importantly, exposure to the experimental stimuli was increased to 200 repetitions in Experiment 2.

7. Experiment 2 Results

Recall that we were interested in responses that would provide information about knowledge of category boundaries. Thus, only "different" responses to within-category pairs and "same" responses to across-category pairs were counted as correct. Table 4 shows the number of correct "different" responses to across-category pairs for participants in each group.

Table 4. Number of correct "different" responses to across-category pairs

for each group (Experiment 2)

Training Group					
Natural (N = 5)			Unnatural (N = 5)		
Stimulus pair	"different" responses		Stimulus pair	"different" responses	
	# / 10	%		# / 10	%
350-590	4	40	590-830	2	20
510-750	3	30	750-990	1	10
510-590	4	40	750-830	0	0
350-750	8	80	590-990	3	30
Total	19/40	48	Total	6/40	15

As Table 4 indicates, participants in the natural training group responded "different" to across-category pairs more often than participants in the unnatural training group. In fact, participants in the unnatural training group were able to distinguish across-category pairs in 19/40 responses (48%), while participants in the unnatural training group only distinguished across-category pairs in 6/40 responses (15%). A one-way ANOVA using an arcsine transformation of proportion correct indicates that there were significantly more "different" responses to across-category pairs for participants in the natural training group than for participants in the unnatural training group ($F(1,8) = 5.09$ $p = .05$).

Table 5 shows the number of "same" response to within-category pairs for each training group.

Table 5. Number of correct "same" responses to within-category pairs

for participants each group (Experiment 2)

Training Group					
Natural (N = 5)			Unnatural (N = 5)		
Stimulus pair	"same" responses		Stimulus pair	"same" responses	
	# / 10	%		# / 10	%
350-510	7	70	590-750	8	80
590-750	6	60	830-990	9	90
Total	13	65	Total	17	85

As table 5 indicates, participants in the natural training group responded "same" to within-category pairs 65% of the time, while participants in the unnatural training group responded "same" 85% of the time. A one-way ANOVA using an arcsine transformation of proportion correct indicates that there is not a significant difference in the number of "same" responses to within-category pairs between the two groups ($F(1,8) = 3.24$, $p = .11$).

Comparing table 2 (from Experiment 1) to table 4 (from Experiment 2), we can see that the proportion of "different" responses to across-category pairs changed when the amount of exposure to experimental stimuli (100 repetitions in the Experiment 1 vs. 200 repetitions in the Experiment 2) was increased. For participants in the natural training condition, "different" responses to across-category pairs increased from 15% to 48% when the frequency of exposure was increased from 100 repetitions to 200 repetitions. Figure 3 shows the proportion of "different" responses to across-category pairs for participants in Experiment 1 and Experiment 2.

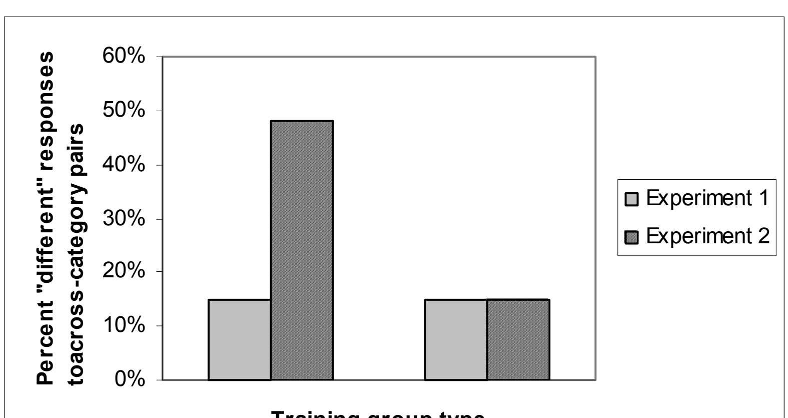

Figure 3. Percent of different responses to across-category pairs as a function of training group type.

Figure 3 suggests that participants in the natural training group in Experiment 2 responded "different" to across-category pairs a larger proportion of the time than all other participants. However, a two-way ANOVA indicates that there is no interaction between the type of training (natural vs. unnatural) and amount of exposure to the experimental stimuli in the training phase (100 repetitions vs. 200 repetitions) ($F(1,16) = 2.09$, $p = .17$).

8. Discussion

Comparing results from Experiment 1 and Experiment 2, we see a striking difference in responses to across-category pairs. Participants who were exposed to the experimental stimuli only 100 times (Experiment 1) demonstrated a low proportion of "different" responses to across-category stimuli in both groups. However, the results from Experiment 2 indicate that when the amount of exposure to the experimental stimuli was increased, participants in the natural training condition were able to distinguish across-category pairs significantly more often than participants in the unnatural training condition. Clearly, the results from Experiment 2 refute the null hypothesis, which predicted no difference between the two groups in ability to distinguish across-category pairs.

The results fail to support the general claim that language acquisition is solely driven by general learning principles not specific to language, and are more in line with the notion that humans are genetically endowed with a learning mechanism specifically designed for language acquisition. While theories of innateness and the general learning mechanism approach to language acquisition both predict normal language acquisition given exposure to natural language input, they crucially differ in their predictions for language learning in response to unnatural language input. If language learning is a reflex of general learning principles coupled with exposure, then the quality (or naturalness) of the input should be of little relevance as long as the frequency of exposure is high enough. On the other hand, if language acquisition is facilitated by mechanisms designed specifically for the acquisition of natural language, then acquisition ability should be affected by the input: unnatural input should be more difficult (if not impossible) to acquire than natural input. The latter prediction is borne out in Experiment 2. Despite the fact that participants in the both training groups were exposed to the experimental stimuli with identical frequency, participants in the natural training group were better able to distinguish across-category pairs. Simply put, the natural category boundary was easier to acquire than the unnatural boundary. If language learning—or more specifically categorical perception—were dependent only on general learning principles, then there should have been no difference between participants trained to perceive the natural category distinction vs. the unnatural category distinction, since frequency in the input was held constant. Thus, the results suggests that language learning is not solely dependent only on general learning principles but instead may be facilitated by principles specific to natural language.

There are several alternative explanations for these results. First, one might claim that the results are not necessarily indicative of phonemic learning, but instead simply reflect the participants' ability to hear phonetic differences across categories. In order to show that participants grouped the sounds into phonemic categories, we would ideally compare responses to equidistant within- and across-category pairs. The ability to distinguish across-category pairs coupled with the inability to distinguish equidistant within-category pairs would certainly be indicative of phonemic knowledge. Unfortunately, this comparison was impossible given the experimental design. The within-category pairs were 180 hz apart, and the across-category pairs spanned a distance of 400, 240 and 80 Hz. While a high number of "same" responses to within category pairs (distance = 180 Hz) and a low number of "same" responses to across-category pairs (distance = 80 Hz) could certainly provide evidence for phonemic learning, this was not the result. Tables 4 and 5 show that participants in the natural training group responded "same" to within category pairs (distance = 180 Hz) between 60% – 70% of the time, and they responded "same" to the 510-590 across-category pairs (distance = 80 Hz) 60% of the time. However, this result is somewhat uninteresting given the fact that the distance between the across category pair is much closer than the within-category pair, making it harder to hear the distinction despite any phonemic learning. Further, the natural category boundary located at 550 Hz is based on El Halees (1985), who did not find a sharp categorical distinction for stimuli close to this point. That is, only 70% of the participants in the El Halees study labeled stimuli containing an F1 at 500 Hz as uvular, and only 70% of them labeled stimuli at 590 Hz as pharyngeal. Since there is obviously not a sharp boundary at 550 Hz, it is not surprising that participants often responded "same" to the 510-590 pair.

It is interesting, however, to compare the natural training group's responses to the 510-590 across-category pair (distance = 80 hz) with the unnatural training group's responses to the 710-850 across-category pair (distance = 80 Hz). Although the distance was exactly the same and relatively short, participants in the unnatural group responded "same" 100% of the time, while participants in the natural training group responded "same" only 60% of the time. Clearly, participants in the natural training group more successfully distinguished across-categories pairs that spanned a short distance, providing further evidence for the claim that natural category distinctions are easier to acquire than unnatural category distinctions.

One might argue that acoustic differences are simply easier to perceive toward the left side of the continuum (when the F1 is lower) than toward the right side of the continuum (when the F1 is higher). However, Kewley-Port and Watson (1994) have demonstrated that people can detect just noticeable differences as low as 14.5 Hz in vowels containing first formant frequencies up to 800 Hz, showing that it is indeed possible for humans to hear slight differences between tokens with high first formant frequencies (albeit their stimuli did not reach 990 Hz). I know of no phonetic studies that have investigated whether differences are easier to detect in stimuli containing lower first formant frequencies. This suggestion could be tested using the current paradigm. Werker and Logan (1985) showed that decreasing the ISI from 500 ms to 250 ms in an A-X discrimination task caused participants to use acoustic knowledge rather than phonemic knowledge to distinguish word pairs.[8] In order to see if stimuli at the

[8] Werker and Logan discuss a three-factor model of speech perception, in which stimuli pairs separated by 1500 ms are processed "phonemically", stimuli separated by 500 ms are processed "phonetically" and stimuli separated by 250 ms are processed "acoustically". They define "phonetic" processing as the ability to distinguish pairs that are not phonemic in one's own language, but are phonemic in some natural language.

left side of the continuum are acoustically easier to distinguish than stimuli at the right side of the continuum, the current study could be replicated using as ISI of 250 ms, thus tapping acoustic knowledge rather than phonemic knowledge. However, if Werker and Logan are correct in assuming that discrimination tasks using 500 ms ISIs tap phonemic knowledge, then a replication of the current study with a shorter ISI would be unnecessary. Since the current study used a 500 ms ISI, one might arguably assume that the participants were demonstrating phonemic awareness in their responses.

The results of the current study show that when frequency of exposure is held constant, participants are better able to learn natural category distinctions than unnatural category distinctions, refuting the hypothesis that categorical perception is dependent only on the statistical frequency distribution of speech elements in the input. Frequency in the input was, however, an important factor. It was not until the frequency of exposure to the experimental stimuli was doubled from 100 repetitions in Experiment 1 to 200 repetitions in Experiment 2 that participants in the natural training group could reliably distinguish across-category pairs. From a learnability perspective, it seems natural that increased exposure would result in improved acquisition ability, but it is not clear why such a slight change in frequency should have such a large effect, or why 200 exposures, rather than 100 exposures, is needed. Future research might investigate this matter in more detail.

Figure 3 suggests that participants in the natural training condition in Experiment 2 were able to distinguish across-category pairs more often than participants in all other groups. This difference, however, was not statistically significant, as there was not an interaction between type of training and amount of exposure to experimental stimuli across the two experiments. Since there were only 5 participants in each group, it is possible that the lack of statistical significance might be attributed to the small sample size. The trend illustrated in figure 3 should be tested using a larger sample size.

It is not surprising that the number of "same" responses to within-category pairs was generally higher than the number of "different" responses to across-category pairs in both experiments. Since uvular and pharyngeal consonants are not evidenced in English, participants would be expected to respond "same" to within-category pairs regardless of whether they had knowledge of the category boundary. This being the case, "same" responses to within-category pairs are

---

Since this definition fits our description of phonemic knowledge, it is referred to as "phonemic" above.

somewhat uninformative. Further, participants in the natural and unnatural training groups did not different in responses to within-category pairs in either experiment, so this measure proved to be less informative than responses to across-category pairs.

9. CONCLUSION

The current study has shown that adults demonstrate sensitivity to non-native phonemic contrasts with very little exposure, replicating the results of Maye and Gerken (2000). Further, the results showed that it is easier for adults to distinguish non-native phonemes that cross a natural category boundary than phonemes that cross an unnatural category boundary, suggesting that categorical perception is not solely dependent on general learning principles, but instead may be driven by a system specifically designed to facilitate language acquisition. The results of the current study also indicate that adults' ability to demonstrate sensitivity to non-native phonemic contrasts is dependent on the amount of exposure to the contrast. More research with a larger sample size is needed to further investigate the role of exposure in the acquisition of non-native contrasts.

REFERENCES

BATES, ELIZABETH., and JEFFREY ELMAN. 1996. Learning rediscovered. *Science*, *274*, 1849-1850.

BRADLOW, ANN, REIKO AKAHANE-YAMADA, DAVID PISONI, and YOH'ICHI TOHKURA. 1999. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech perception. *Journal of the Acoustic Society of America*, *101*(4), 2299-2310.

EL-HALLES, YOUSEF. 1985. The role of F1 in the place-of-articulation distinction in Arabic. *Journal of Phonetics*, *13*, 287-298.

LIVELY, SCOTT, DAVID PISONI, REIKO YAMADA, YOH'ICHI TOHKURA, and TSUNEO YAMADA. 1994. Training Japanese listeners to identify English /r/ and /l/. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076-2087.

MACKAIN, KRISTINE, CATHERINE BEST, and WINIFRED STRANGE. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369-390.

MAYE, JESSICA and LOUANN GERKEN. 2000. Learning phonemes without minimal pairs. In S. Catherine Howell, Sarah A. Fish and Thea Keith-Lucas (Eds.), *Proceedings of the 24th Annual BUCLD* (pp. 522-533). Somerville, MA: Cascadilla.

RUMELHART, DAVID and JAMES MCCLELLAND. 1994. On Learning the Past Tenses of English Verbs. In P. Bloom (Ed.), *Language*

Acquisition: Core Readings (pp. 423-471). Cambridge. MA: MIT Press.

TREHUB, SANDRA. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*, 466-472.

WERKER, JANET, JOHN GILBERT, KEITH HUMPHREY, and RICHARD TEES. 1981. Developmental aspects of cross-language speech perception. *Child Development*, *52*, 349-355.

WERKER, JANET, and RICHARD TEES. 1983. Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, *37*(2), 278-286.

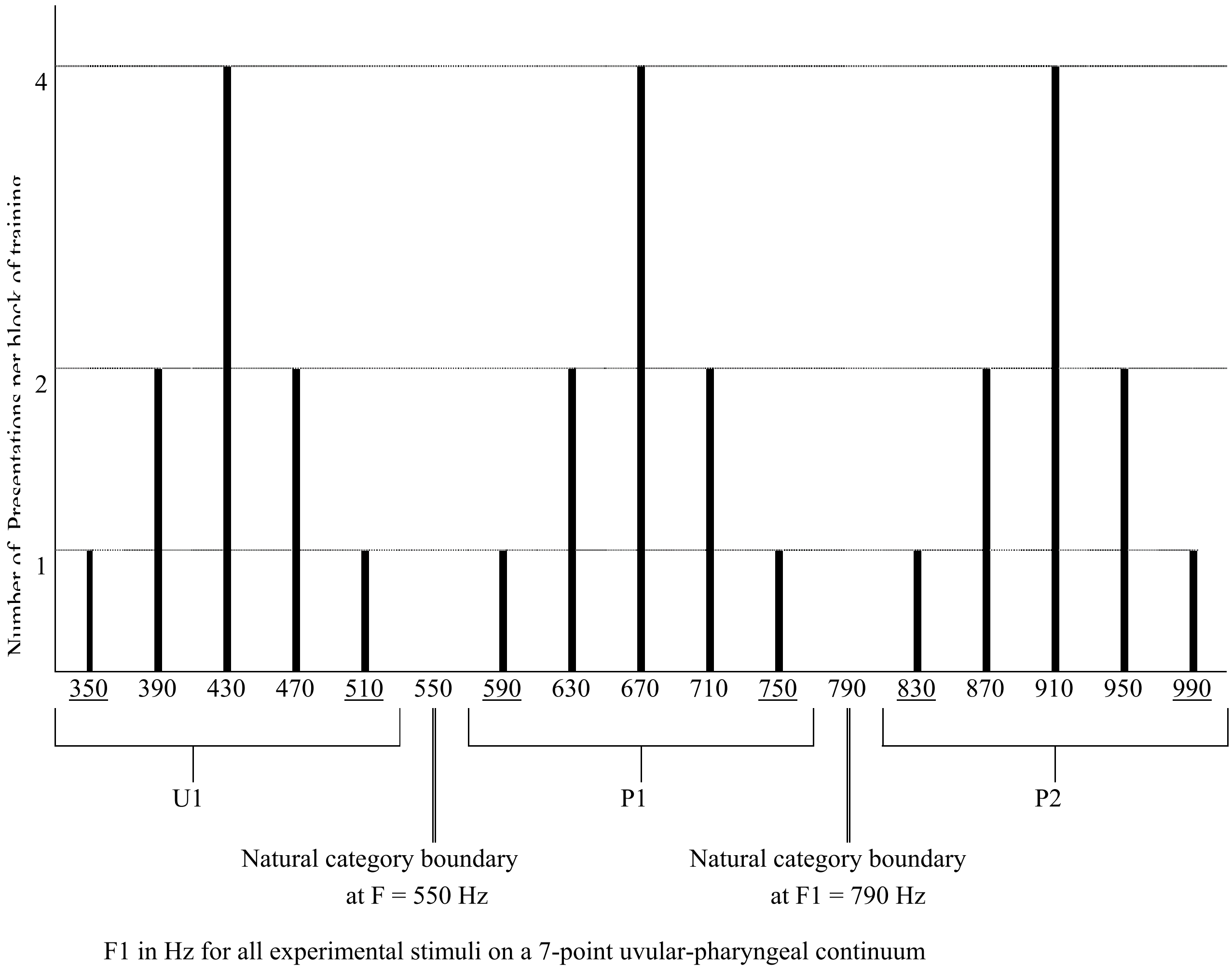

Figure 2. Stimuli presentation frequency during training phase