Learning Allophones from the Input

Christine Shea and Suzanne Curtin
University of Calgary

1. Introduction

Perceptual contrast and its consequences play a large role in both first and second language acquisition. Infants acquiring their native language must learn to divide the speech stream into the vowels and consonants that perform a contrastive role in the inventory of their language. Researchers have proposed that infants initially engage in a process of phonetic category learning and subsequently, after exposure to their native language contrasts, begin to unify these phonetic exemplars into more abstract phoneme categories (Pierrehumbert 2003). The objects of this phonetic learning process are the language's allophones, or contextualized surface variants of the more abstract phoneme categories. The infant learner must determine whether the variation perceived in the speech stream is attributable to allophonic or phonemic contrasts in the language's sound categories.

The mechanism used by children to carry out this task is assumed to involve a statistical analysis of the acoustic space and the grouping together of input tokens to form areas of high density exemplars, which eventually abstract away and become phoneme categories (Pierrehumbert 2003, Werker and Curtin 2005). The precise characterization of this mechanism has yet to be determined, although there is growing evidence in favour of a role for distributional frequencies in the input (Maye & Gerken, 2000; Maye, Werker, Gerken, 2002). While individual tokens of a sound category may differ considerably, all tokens of the same type will be more similar amongst themselves than across different sound categories (the phenomenon of categorical perception) and more similar along specific acoustic dimensions, such as VOT in the case of consonants or F1/F2 values in the case of vowels. Acoustic dimensions that are important for the differentiation of a language's phonetic categories will serve to organize tokens into a bimodal distribution along this dimension; dimensions which are not informative will lead to the formation of a token cluster at the center of the distribution (Maye, Werker & Gerken 2002) In a series of experiments, Maye, Werker and Gerken (2002) showed that the perception of allophonic contrasts by six and eight month-old infants can be modified after exposure to either bimodal or monomodal distributions of the allophones [d] and the unaspirated voiceless stop [t] in English. Infant participants were exposed to a continuum of eight tokens ranging from [d] – [t] in eight equal VOT steps. These particular contrasts were selected because they do not constitute a phonemic contrast in English and are both perceived as members of the /d/ category when in word initial position. After exposure, infants in the bimodal group performed significantly better (20%) than the monomodal group on a discrimination task, indicating that they had begun to separate the sounds into two categories. There was no effect for age. Their finding indicates that infants at these age groups are in fact sensitive to statistical distributions in the input, which is what one would suppose if in fact they are using this mechanism during the first year of life when native language categories are developing.

One of the major advantages of Maye, Werker, & Gerken's distribution-based learning algorithm is that it precludes the need for a lexicon. A considerable body of research has demonstrated that infants as young as 8-months are able to track statistical regularities in the speech stream after as little as two minutes of exposure (Saffran, Aslin, Newport, 1996), results which indicate that a lexicon is not necessary for the acquisition of sound categories to begin. As Peperkamp, Pettinato & Dupoux (2003) argue a distributional analysis of the input exploits the complementary distribution between phonemes and their allophones whereby pairs of segments whose lists of contexts have an empty intersection are in complementary distribution and occur as allophonic contrasts. These studies suggest that infants have available to them a statistical learning mechanism that can track distributions in the absence of a lexicon.

Maye & Gerken (2000) used the same distribution-based training cited above with adult learners, employing the identical allophonic contrasts of [d] and [t]. They exposed adult native English speaker listeners to either the bimodal or monomodal distributions. Their results showed a significant effect for distribution, where the bimodal group scored significantly better than the monomodal group (20%) on the contrasts that included the segments of interest.

Is the same statistical mechanism available to adults, and can it track distributions without relying on a lexicon to derive meaningful contrasts? Presumably, adults can make use of lexical contrasts and feedback in the acquisition of minimal pairs, a fact which has been observed in Japanese learners of the /r/ vs. /l/ contrast. McCandliss, Fiez, Protopapas, Conway &McClelland (2002) showed that when accompanied by stimuli exaggerated along the F3 dimension (the acoustic dimension used by native English speakers to distinguish between these two phonemes), the presentation of minimal pairs and the use of feedback did increase accuracy on identification tasks. However, both feedback and lexical contrasts are forms of explicit information that learners take advantage of in order to direct their perceptual learning strategies. The present experiments sought to determine a role for implicit learning in adult speech perception and what effect contrast and allophony might have upon this process.

Investigators have shown that adults can distinguish between the allophones of a phoneme when engaged in a phonetic perception task. Research by Pegg and Werker (1997) demonstrated that adults performed worse on an AX task contrasting the allophones [d] and the voiceless unaspirated [t] than on a phonemic contrast. Similarly, Peperkamp et al. (2003) used an AX task to tested native French speakers on the uvular voiced fricative [ʁ] and the voiceless uvular fricative [χ] contrast, where the latter is the allophonic variant of the first and is found only adjacent to voiceless consonants. The first in their series of experiments involved presenting subjects with contextualized VC sequences using the [ʁ] and [χ], but without the following segment that would normally condition the allophonic alternation. Subsequently, the researchers carried out another AX task but with phonologically contextualized stimuli of the shape VC.CV, where the second consonant either agreed in voicing with the first or did not. Peperkamp et al. found two results that are relevant for the present investigation: first, there was an effect for context, where discrimination was significantly better on the phonetic contrast than on the phonologically contextualized contrast; second, participants performed significantly better on discriminating the phoneme sequence than on the allophone sequence in the phonologically contextualized task. Their results show that in isolation, the allophone and phoneme segments are sufficiently different to be distinguished on a discrimination task but once contextualized, the phoneme took the advantage. In their second set experiments, Peperkamp et al tested whether the Maye and Gerken (2000) distribution-based learning effect could be replicated when the allophonic contrast is presented within context. They created a continuum between [χ] - [ʁ] along the dimension of frication and followed the VC sequences with a context syllable. They created another set of stimuli in which the context syllable agreed in voicing with the coda syllable of the first word. Peperkamp et al hypothesized that the voicing agreement would bias the participants into creating a bimodal distribution if in fact they were relying upon context to determine allophones in this artificial language. Participants completed an AX pre-training task and subsequent to training, a second AX task on the same stimuli. Pre and post training AX scores were compared and the results from this experiment showed minimal statistical learning, without significant pre-post exposure improvement. The researchers attributed these results to the very nature of speech perception itself: If the subjects already have a phonological category for these allophones, then they may be operating at the level of phonological perception, rendering it more difficult to create distinct distributions solely by means of implicit training. On this view, allophones in context are more difficult to perceive than allophones out of context, as in the case of Maye and Gerken (2000), whose studies yielded much higher learning percentages.

While previous studies have used adults learning allophonic contrasts without phonological context and allophonic contrasts with phonological contexts, all participants have been trained on contrasts that exist in their native language. These studies provide evidence for a degree of continuity between infant and adult distribution-based learning in the native language. In the present study, we asked whether adults learning a second language exploit the same statistical frequency-distribution based learning mechanism used by infants acquiring the phonemes and allophones of their native language. In the case of adults learning a second language, the way in which the target language sound system contrasts with the native system will play a large role in the acquisition of the new sounds. We know that the perceived relationship between phonetic segments in an L2 and the L1 sound system will play a crucial role in how those phonetic segments will be discriminated and classified (Aoyama, Flege, Guion, Akahane-Yamada & Yamada 2004) Based on results collected over the years from experiments of Japanese speakers learning the /r/-/l/ contrast in English, we also know that explicit teaching of non-native sound categories using natural tokens with little phonetic variability merely serves to further entrench the incorrect perception of these new categories (McCandliss et al 2002). The research presented here seeks to discover whether the distributional mechanism active in infant phonetic acquisition is available for adult second language (L2) learners and if so, whether the L1 phonemic status of the categories being implicitly learned plays a role in their "acquisition".

Based upon the results obtained by Maye and Gerken (2000) and Peperkamp et al (2003), we hypothesized that distribution-based learning would have differential effects upon distinct native speaker groups. To this end, we selected as our target for acquisition a phonological context that conditioned specific allophonic variants. Our second step was to find languages whose inventory categorized these allophones in different ways and whose speakers would thus exhibit differential effects for contextualized distribution-based learning. Our goal was to tease apart the effects of phonological contextualization and phonological categorization. We found such a contrast in the Arabic emphatic/non-emphatic consonant contrast and the allophonic variation in the low vowels which are triggered by the emphatic secondary pharyngealized articulation. Specifically, phonemic emphasis in Arabic corresponds to a primary articulation in the dental/alveolar region and a secondary pharyngeal articulation in the upper pharynx:

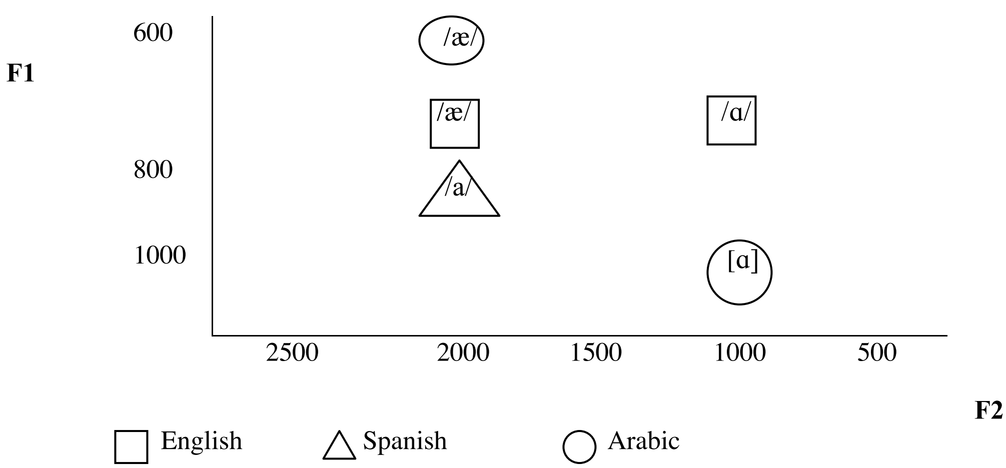
(1) [tæb´t] ‘she repented’ vs. [tˤɑb´t] ‘she recovered’
(Al-Masri and Jongman, 2003)

Acoustically, emphasis is consistently manifested by a lowering of the second formant frequency of the vowel following the emphatic consonant (Al-Masri & Jongman 2003). The alveolar stops /d dˤ t tˤ/ were chosen as the conditioning consonants and were paired with the three short vowels that are underlying in the Arabic vocalic system: /æ i u/. As described below, the allophonic variation in the contiguous low vowels triggered by the pharyngealized alveolar stops leads to segments that have a different status in the sound systems of Spanish and English (Table 1).

In this study, we report on the differential effects of distribution-based learning of allophonic contrasts across different native speaker groups. Following Peperkamp et al (2003), we considered the effects of phonological context and how cues to the context might affect the assimilation of non-native sounds into native language sound categories in a distribution-based learning paradigm.

2. Present Study

Our participants were native speakers of Spanish and English. Neither of these languages has emphatic consonants in their native language inventory and more importantly, while Arabic, English and Spanish all have low vowels, the way in which this part of the vowel space is divided up in terms of the allophone-phoneme relationship is distinct for each language:



Values taken from Shahin 2003 (Arabic); Stevens 1997 (English); Escudero and Boersma 2003 (Spanish)

Figure 1: Formant values for low vowels in Arabic, Spanish and English

As can be seen in Table 1, Arabic has a surface contrast between [æ] and [ɑ] while in English this contrast is phonemic. Spanish has only one low vowel, /a/ and Spanish speakers tend to perceive the two low vowels of English as members of this category (Escudero and Boersma 2003). We hypothesized that the native language vowel space would affect sensitivities to the frequency distribution in distinct ways across the native English speaker and native Spanish speaker groups.

We further supposed that the hypothesis put forth by Peperkamp et al (2003) stating that context will affect the learning of allophonic alternations merited further consideration. If allophones are conditioned alternations, there must be something in the speech stream that can indicate to the learner where they must use one or the other variant. For example, as described above, the pharyngealization of the alveolar stops in Arabic triggers a lowering in the F2 formant and conditions the alternation between [æ] and [ɑ] in that language. In English, these segments are phonemes and not conditioned by any context and in Spanish, they are neither phonemes nor allophones. Therefore, in order to acquire the allophonic alternations, our participants had to implicitly acquire the acoustic cues that indicate the allophonic alternation in Arabic. As stated, in English, the two Arabic allophones correspond to phonemes and in Spanish the two Arabic allophones correspond to a single low vowel category. This led us to hypothesize that each native speaker group would pay attention to different aspects of the input in acquiring this alternation. For example, native speakers of English might pay more attention to the vowel as indicating the emphatic-non-emphatic distinction because the contrast is phonemic in their native language and not indicated by any specific phonological context. In contrast, for the native Spanish speakers, the consonant might play a larger role in acquiring the contrast because these listeners must learn what conditions the alternation. Because Spanish listeners do not have a low vowel contrast as part of their native language inventory, they may learn to distinguish between the two Arabic allophones based upon the consonant that triggers the alternation.

3.0 Method

The goal of this study is to explore whether adults can learn an allophonic variant using a distributional analysis. Moreover, this study explores how the native language effects which cue is utilized when learning a new sound alternation.

3.1 Participants

Sixteen monolingual native English speakers, none of whom had any reported hearing difficulties, were tested. Eight were native speakers of Western Canadian English and eight were native speakers of Western Pennsylvania English. They were students at the University of Calgary (eight) and the University of Pittsburgh (eight).

We also tested sixteen native Spanish speakers who were living either in Calgary, Alberta or Pittsburgh, Pennsylvania at the time of the experiment. Of the sixteen speakers, four had been in the U.S. less than three months, four had been in the US for around nine months, three had been in Canada for 3 months and four had been in Canada for an average of fifteen months. All spoke Spanish at home and with friends.

Participants were paid five dollars.

3.2 Stimuli

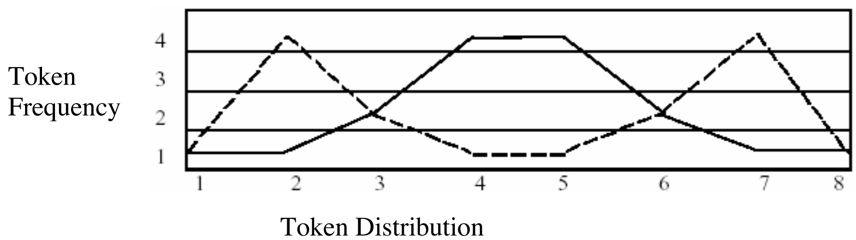
The stimuli were recorded by a male native speaker of Kuwaiti Arabic and digitized on a Macintosh computer at 16 kHz and 16 bits. The CV sequences were taken from CVCV units where the speaker produced the target CV syllable first followed by [tæ]. For example, if the target CV was [dˤɑ], then the speaker produced [dˤɑtæ] and the second CV was spliced away. There were a total of four CV sequences that involved the contrast of interest, the low front vowel-back vowel alternation: [dæ dˤɑ tæ tˤɑ]. There were eight filler sequences, using the same coronal consonants and the other two vowels found in Arabic: [dˤi dˤu di du tˤi tˤu ti tu]. The stimuli were recorded four times, and tokens from all four recordings were used in the creation of the artificial language.

In order to explore the role of attentional mechanisms, we created two different continua. The first continuum consisted of vowel tokens where naturalistic [æ] and [ɑ] served as anchors and the other tokens varied along the F2 continuum in eight equal steps. These eight exemplars were paired with a steady consonant onset [D] to create our "vowel variation" continuum: [Dæ] – [Dɑ]. The F2 locus was the same for all exemplars in order to give the acoustic impression of similar place of articulation.

We then created another set of tokens that varied along a VOT continuum in the consonant and for which the vowel was held steady: [dA] - [dˤA]. The F2 locus for these tokens varied. The vowel was held steady at an average F2 value of 1300 Hz. These averages were taken from naturalistic [æ] - [ɑ] and was our "consonant variation" continuum.

The stimuli for the continua were synthesized from natural speech samples using the Kay Elemetrics

ASL Model 5104. The vowel variation continuum was created from naturalistic tokens where the vowel was spliced away from the original [dæ] and [dˤɑ] syllables. The F2 steady state values were 1124 Hz and 1562 Hz respectively. Eight equal steps of 62 Hz were devised along the F2 continuum. The [D] onset was created from averaging the VOT and intensity of [d] and [dˤ] and an average locus was taken at 1483 Hz. The consonant variation continuum was created by splicing away the [d] and [dˤ] portions of the CV syllables and creating a continuum of values for the VOT for each consonant. The F2 locus varied from 1896 Hz for [d] to 1070 Hz for [dˤ]. The F2 of the steady state [A] vowel was held constant at 1300 Hz.



CONDITION (1) [dA] [dˤA]
CONDITION (2) [Dæ] [Dɑ]
________ monomodal _ _ _ _ _ _ _ bimodal
(1) consonant variation (2) vowel variation

Figure 2: Distribution of stimuli used for training (adapted from Maye and Gerken 2000)

3.3 Procedure

All experiments were carried out on a Macintosh computer running PsyScope (Cohen, MacWhinney, Flatt, and Provost 1993). The stimuli were presented through a set of headphones and the subjects were tested sitting individually in a quiet room.

In order to test our hypotheses, we designed a pre-training ABX task using the two alveolar emphatic/non-emphatic segments and the three Arabic vowels. Training on one of the four continua followed: monomodal + vowel variation, bimodal + vowel variation, monomodal + consonant variation, bimodal + consonant variation. We followed the training session with another ABX task to determine if any change in perception had occurred post-training.

The experiment began with a brief training session during which subjects were told they would be hearing short words from an unfamiliar language. They were to indicate on their keyboard whether the third "word" was the same as the first or the second word they heard by pressing "first" or "second" keys on the keyboard, where "first" or "second" appeared on the lower left and lower right corners respectively. Five ABX stimuli were presented during the initial training session and feedback was provided.

The second task involved 132 ABX trials created orthogonally from the four consonant/three vowel combinations produced in naturalistic tokens of CV or CˤV syllables. These syllables were spliced from disyllabic nonwords (see below). Each CV sequence lasted an average of 479ms, with 500 ms between each syllable and an ISI of 1000ms, for a total of 7.7 minutes. The CV syllables were presented randomly.

Following this, the participants participated in a training session where they were exposed to twelve minutes of recorded stimuli from one of the two continua: (1) variation in the vowel F2, from [æ] - [ɑ] with the F2 locus held steady at the voiced coronal consonant [D] or (2) variation in the consonant F2 transition locus from the pharyngeal [dˤ] - [d] to a steady state vowel at F2 1300 Hz. There were four lists, differing along two dimensions: distribution (monomodal vs. bimodal) and variation (consonant vs. vowel). The stimuli at the end of the distributions were natural tokens. In the bimodal distribution, the subjects heard stimuli 2 and 7 four times more than the others while in the monomodal distribution, stimuli four and five were heard four times more than the others (see Figure 2, above). In order to distract attention from the varying dimensions, we used four filler stimuli: [ti tˤi tu tˤu] (Maye, Werker and Gerken 2002). Fillers were presented two times each per sequence of training stimuli, resulting in 16 stimuli per sequence, with an ISI of 1000 ms. The training continua stimuli were presented 504 times. Subjects heard 3 Blocks of four minutes each, for a total listening time of 12 minutes and were given a rest period of two minutes between each block during which time they could take off their earphones and leave the computer.

Finally, the participants completed another ABX task using the second half of the randomly selected emphatic/non-emphatic syllables presented in the pre-training task to determine if any changes in their sensitivity had occurred. The entire experiment took approximately thirty-five minutes to complete.

Our hypotheses were as follows:

Predictions for pre-post-training ABX test		**Dæ - Dɑ**	**dA-dˤA**
	English Speakers	Post training SAME than pre-training	Post training WORSE than pre-training
	Spanish Speakers	Post training the SAME as pre-training.	Post training SAME as pre-training
	English Speakers	Post training the SAME as pre-training.	Post training BETTER than pre-training
	Spanish Speakers	Post training BETTER than pre-training.	Post training BETTER than pre-training

Table 1: Predictions for change in pre-post training ABX task scores

We predicted that the two native speaker groups would have very low error rates on the contrasts that did not involve the two low front vowel contrasts. The vowels /i/ and /u/ are very distinguishable in their emphatic/nonemphatic contexts. Therefore, our predictions will only concern the contrasts that involved the low front vowel /æ/.

If native English speakers are relying more upon the vowel to distinguish between the emphatic/non-emphatic sequences, we hypothesize that they will achieve the same results on pre and post training ABX task in the monomodal training with vowel variation condition because they will be reinforcing native vowel categories and the steady state consonant [D] will not draw their attention to the consonant as playing a role in triggering the allophonic variation. There is no context available to assist with learning the allophonic contrast in this situation. The consonant variation–monomodal condition with native English speakers should result in worse post-training results because the vowel is steady and the consonant that triggers the allophonic shift is drawn from closely related exemplars. Therefore, these listeners will not have much variation in the place of articulation information form the consonant onset.

In the case of the Spanish speakers and the monomodal vowel variation continuum, we predict that there will be no change in their performance from the pre to post training ABX test. Native Spanish speakers will be reinforcing the unitary category that exists in their native language under the monomodal condition. After being trained on the consonant variation continuum, native Spanish speakers should also have the same post training results because they will be reinforcing native language unitary low vowel category and the existence of one consonant. These two facts are consistent with the Spanish sound inventory.

In the case of the bimodal vowel variation continuum, the English speakers should perform the same as they did on the pre-training task after exposure to the vowel variation continuum because they are essentially reinforcing the native language vowel categories. Holding the consonant steady will not draw their attention towards it as a conditioning factor in the variation. If they do exhibit improvement, it may be attributable to a practice effect. On the bimodal distribution-consonant variation continuum training, the native English speakers should perform better because they are beginning to pay more attention to the consonant and less to the vowel.

The Spanish speaking participants are hypothesized to perform better on the vowel variation continuum in the bimodal condition because they are starting to become sensitized to the existence of two different vowel sounds. On the consonant variation continuum in the bimodal condition, the native Spanish speakers are predicted to perform best of all eight groups. This is because these listeners are having their unitary vowel category reinforced by the steady vowel and all their attention can be drawn towards the consonant which triggers this alternation.

In summary, we predict the following order in terms of the magnitude of change between the pre-training and posttraining blocks:

Vowel Variation
Bimodal Spanish > Bimodal English, Monomodal Spanish, Monomodal English
Consonant Variation
Bimodal Spanish > Bimodal English > Monomodal Spanish > Monomodal English

4.0 Results

To eliminate possible between group differences prior to training, we carried out a t-test to compare the means of the native English speakers and the native Spanish speakers on the pretest results for all the items and also for the sequences containing the contrasts of interest, the emphatic/non-emphatic low vowel alternations. There were no significant differences found (t (30) = .553, n.s. and 0.1, n.s., respectively). We therefore assumed that any significant differences found between language groups would be due to training effects.

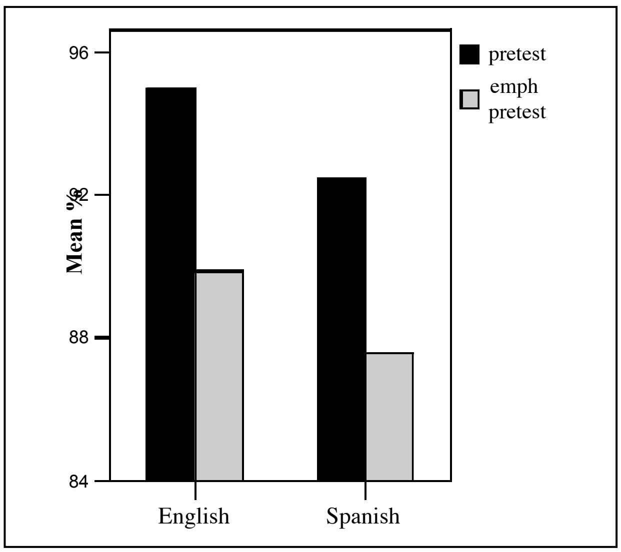


Figure 3: Mean percent correct on pretests/posttests

A global 2X2X2 ANOVA was run on the error percentage, with pretest-posttest as a repeated measure and language, distribution and variation as between. We ran the tests on the all the sequences, including the fillers and subsequently only on the contrasts of interest, those that involved the emphatic-nonemphatic sequences [tæ tˤɑ dæ dˤɑ]. We did not obtain significant results for the filler items, whether filler + emphatic/nonemphatic or only filler. This was attributed to extremely high ceiling effects, due to the ease of discriminating among the three vowels. We only analyzed trials where the contrast involved two of the four syllables including a low front vowel, in either of its surface realizations and with either of the onset consonants. Therefore, the following discussion will only treat the emphatic/nonemphatic contrasts.

Mean scores on the pretest and posttest for the trials with emphatic/nonemphatic low vowel contrasts are provided in Figure 3. A mixed-design ANOVA (2 Language X 2 Distributions X 2 Variation X Pretraining/ Posttraining) revealed a main effect of test $F(1,24)=20.5;p>0.01$). Distribution main effects almost reached significance ($F(1,31)=0.057$, n,s,). No significant main effects were detected for language or variation.

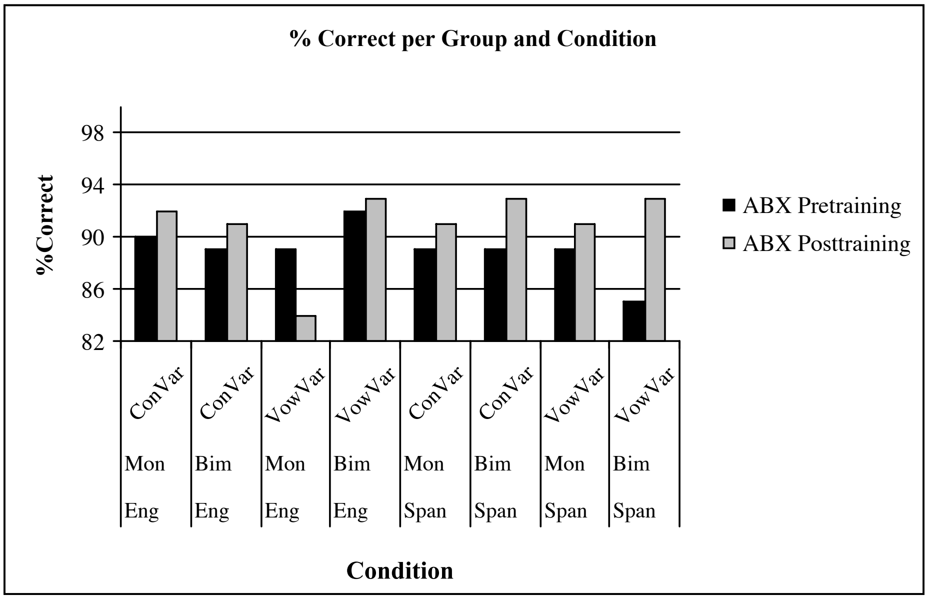


Figure 4

We found an interaction effect for test and language (F(2, 30)=20.5;p>0.01), test and distribution (F(2,30)=13.81; p>0.01) and finally, test and variation (F(2,30)=13.8;p>0.01). There was also a three-way interaction between test, variation and language (2,29)= 11.88;p>0.01) as well as test, variation and distribution (F(2, 29)= 23.77;p>0.01).

We found an interaction between test and consonant variation (F(2,14)=15.9;P>0.01) but there was no interaction between consonant variation*language nor consonant variation*distribution. In the vowel variation condition, however, we found a significant interaction with test (F(2,14)=6.7;p>0.01), distribution and language.

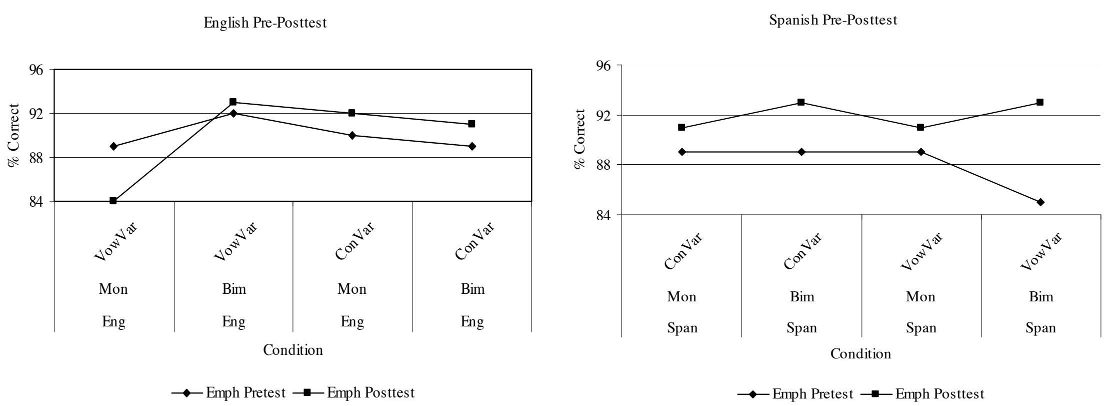


Figure 5

Turning to results by language group, we conducted a series of planned comparisons to further examine the interaction between pre-post training scores of each language group and distribution and variation. The results are depicted in Figures 4 and 5. We obtained a significant effect for test*variation interaction with the Spanish bimodal group (F(1,2)=22.6;p=0.01 but results did not reach significance in the Spanish monomodal condition. The English monomodal condition reached significance at p>0.01 (F(1,2)=11.143) while in the bimodal distribution, the English speaker group's ABX test score differences did not reach significance.

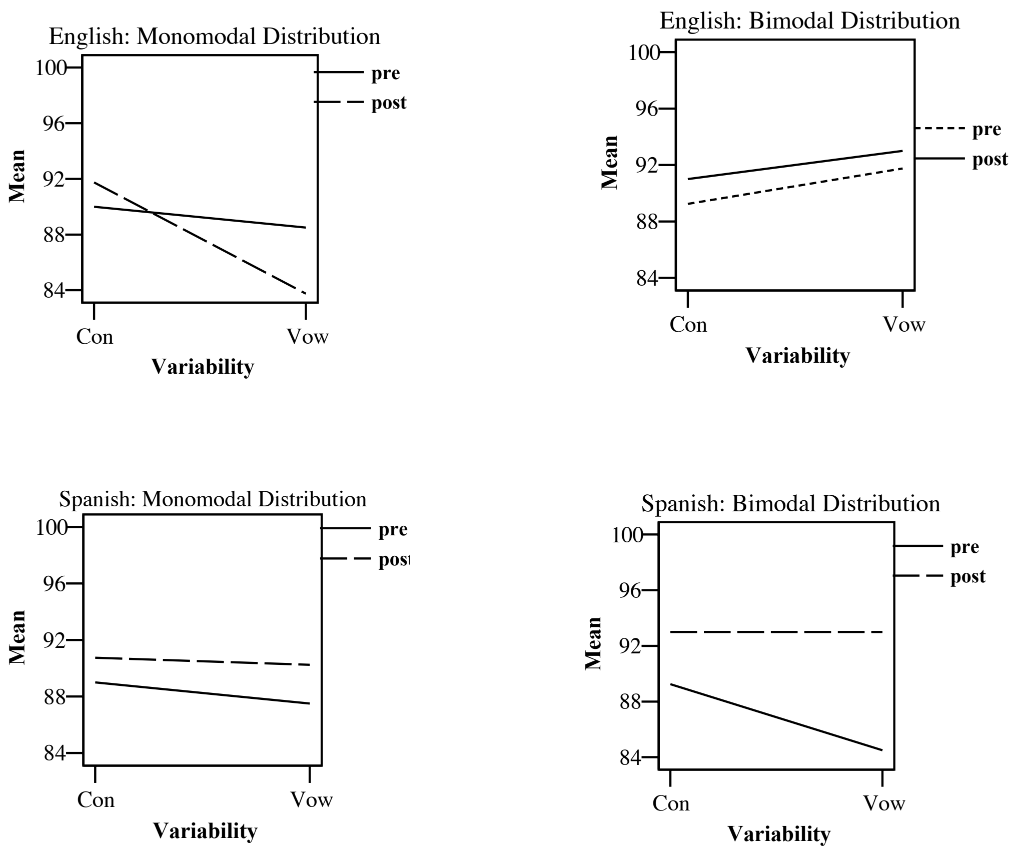


Figure 6

In Figure 6 we show the magnitude of change for each group in each condition. In terms of variation, as predicted, the vowel variation group showed the greatest magnitude in change scores, with the Spanish bimodal group highest, followed by the Spanish monomodal, English bimodal and finally the English monomodal group which demonstrated a negative change.

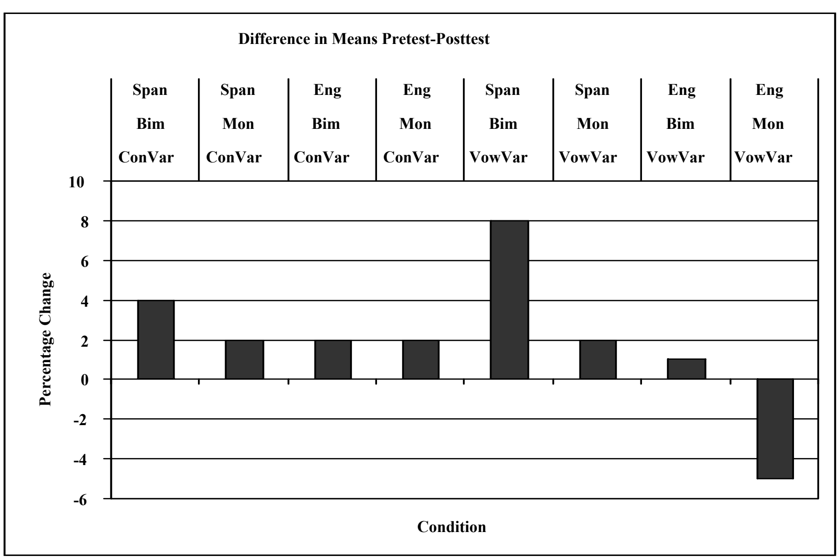


Figure 7

While part of the change in the Spanish bimodal group may be attributable to their low scores on the pretraining emphatics ABX test (see Figure 6), they also had one of the highest scores on the post training task. For the consonant variation continuum, magnitude of change was less, and all changes were positive in nature. The greatest change occurred in the Spanish bimodal group, followed by the other three groups.

4. General Discussion

Studies such as those carried out by Maye et al (2000; 2001; 2003) looked at statistical learning primarily as a function of frequency in the input. Peperkamp et al (2003) carried this notion further and looked at how contextual factors influence the acquisition of phonemes in adults exposed to their native language. In the present study, we hypothesized that frequency, context and native language categories would have an effect upon the way in which the statistical learning mechanism operates in a specific learning context. We considered how different native language groups would be affected by exposure to distinct statistical frequencies in the input and how this might interact with contextualization of the contrast being learned. The exposure to different statistical distributions in the input was operationalized by means of either a monomodal or bimodal frequency training condition. Contextualization of the contrast during training was varied along two continua: subjects were exposed to eight exemplars varying in steps along a consonant continuum or a vowel continuum. The results we obtained bear upon three areas of adult second language speech perception, which roughly correspond to the independent variables used in our study. First, our results suggest that the distribution-based mechanism shown to operate in infant (Maye, Werker and Gerken 2002) and adult (Maye and Gerken 2000) native language sound category acquisition is also available for adult learners of a second language. Second, we showed that the effects of statistical learning in an adult L2 context depend upon the contrast being learned and the way in which it fits into the native language sound inventory. Finally, and related to the second point, our results demonstrate that the acquisition of foreign language allophones is context-dependent and proceeds from an implicit awareness of the contextual cues that signal their occurrence. We will address each of these points in turn.

Part of our initial hypothesis was that adults learning a second language make use of a statistical learning mechanism for acquiring non-native contrasts in a similar way as infants do when acquiring their native language. Our results showed a significant interaction effect for test and distribution, across both language groups, indicating that significant learning had taken place. Therefore, we can conclude that under these experimental conditions, adults can take benefit from implicit input in the speech signal to initiate new category formation. These findings also suggest that adults do not rely exclusively upon lexical contrasts when acquiring the sound system of a new language.

Related to the differential effect of training distribution and continua upon our learners, our results support previous research in demonstrating that the categorization of non-native sounds will depend upon the relationship of these new sounds to existing native language categories (Best 1995; Flege 1995). Flege's Speech Learning Model characterizes long-term sound representations as phonetic categories and the target language sounds that fit into these native language categories will be distinguished in a native-like manner. However, if the native language sounds differ on a sub-segmental level the non-native category will be assimilated into the native sound category. Only the perception of sub-phonemic features will make L2 learners perceive a new category. The results we obtained from out study showed that adult listeners are capable of shifting subphonemic boundaries after a relatively short exposure to continua of input varying along one of two dimensions, either consonant variation or vowel variation and that the native language categories interact with the distribution to which participants were exposed. Specifically, we did not find a main effect for language but we did find an interaction for language with the test variable and the other two independent variables of distribution and variation. We interpret this to mean that the native language may not play the deciding role in new sound categorization but rather is one of a number of factors that interact in second language speech perception. The native language categories will interact with the task at hand (test variable), the statistics found in the input (distribution variable) and the type of input (variation). Based upon this result, our next step will be to look at more language groups and different types of contrasts.

Our learners were exposed to one of two continua, composed of eight exemplars varying in the consonant or the vowel. The sounds were allophones of the low front vowel found in Arabic. Allophones are contextualized surface variants of a language's phoneme inventory. "Context" can be understood as statistical dependencies between two features where one reliably indicates the presence of the other and vice versa. In this study we looked at how variability in one part of the context indicating allophonic variation affects the acquisition of the contrast. Specifically, we looked at how varying the consonant that cues the back/ front low vowel alternation in Arabic affects discrimination by non-native speakers.

Cue weighting refers to how listeners use the multiple acoustic dimensions that exist in the speech stream to identify speech sound categories (Holt and Lotto 2005); some acoustic dimensions tend to be more strongly correlated with certain categories than others and variance in the specific cue, or acoustic dimension, will affect auditory cue weighting. The relationship is not a simple one, however, and depends upon the nature of the variance in the signal. Within-category variation will decrease the informativeness of a cue and even indicate that it might not be a reliable indicator of the category itself. On the other hand, across category variability will tend to indicate robust category boundaries. Our experiment examined two aspects of variability in cue weighting: within/across category variability by means of our consonant and vowel continua and statistical variability in the input by means of our two distribution conditions. Both of these variables proved to have significant interaction effects across both language groups. The variability in the training block worked to either reinforce the native language categories or weaken native language categories, depending upon the condition. Listeners will weigh acoustic cues according to their use and distribution in the native language and when confronted by different cue weighting requirements when acquiring a foreign language, learners are presumed to use a statistical learning mechanism such as the one discussed here. While this may seem to go against the concept of categorical perception, recent research by McMurray, Tannenhaus, Aslin and Spivey (2003) demonstrates that the perception of within category variability by listeners assists with the lexical recognition process because it aids the listener in overcoming the inherent variability that exists in the speech signal.

5. Conclusions

The results of this study point towards the availability of a statistical learning mechanism for adults acquiring a foreign language. In order to refine the nature of the statistical learning mechanism used by our participants, we require further studies using larger groups and different contrasts. We must also consider the duration of the effects from statistical learning. Are they lasting or very short term in nature? Another interesting research question is whether biasing learning by means of lexical items may affect the degree of statistical learning. In conclusion, we propose that adult second language learners are capable of taking advantage of implicit distributions in the speech signal in the acquisition of foreign language allophones.

References

Alwan, A. (1989). Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants. *JASA* 86 (2), 549-556.

Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32, 233-250.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues.* Baltimore: York Press, 171–204.

Boomershine, A. Currie Hall, K., Hume, E., & Johnson, K. (2005). The impact of allophony versus contrast on speech perception. Handout from the Hispanic Linguistics Symposium. November 15. Penn State University.

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

Cohen, J.D., MacWhinney, B., Flatt, M., & Provostm J. (1993). Psyscope: A new graphic environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers* 25 (2), 257-271.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 38 (4).

Flege, J.E. (1995). Second language speech learning: Theory, .ndings and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues.* Baltimore: York Press. 233–277.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. *Proceedings of the 24th Annual Boston University Conference on Language Development Volume 2.* Somerville, MA: Cascadilla Press.

Maye, J., & Gerken, L.A., (2001). Learning phonemes: How far can the input take us? In Anna H.J. Do et al (eds.), *Proceedings of the 25th Annual Boston University Conference in Language Development* Volume 1. Somerville, MA: Cascadilla Press, 480-490.

Maye, J., Werker, J., Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101-B111.

McCandliss, B.D., Fiez, J.A., Protopapas, A., Conway, M., & McClelland, J.L. Success and failure in teaching the [r] – [l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective and Behavioral Neuroscience* 2(2), 89-108.

McMurray, B., Tannenhous, M., ASlin, R., & Spivey, M. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research* 32 (1), 77-97.

Peperkamp, S., Pettinato, M., & Dupoux, E., (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown and F. Conlin (eds.) *Proceedings of the 27th Annual Boston University Conference in Language Development* Volume 2. Sommerville, MA: Cascadilla Press, 650-661.

Pierrehumbert, J.B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46, 115-147.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926-1928.

Shahin, K. (1999). *Postvelar Harmony*. Boston: John Benjamins Publishers.

Stevens, K. (1998). *Acoustic Phonetics*. Boston: MIT Press.

Werker, J.F., & Curtin, S. (2005). PRIMIR: A developmental model of speech processing. *Language Learning and Development*, 1(2), 197-234.

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