Learners Use Word-Level Statistics in Phonetic Category Acquisition

Naomi Feldman, Emily Myers, Katherine White,
Thomas Griffiths, and James Morgan

1. Introduction

One of the first challenges that language learners face is discovering which sounds make up their language. Evidence suggests that infants learn about the phonetic categories of their language between six and twelve months, as demonstrated by reduced discrimination of non-native contrasts and enhanced discrimination of native language phonetic contrasts (Werker & Tees, 1984; Narayan, Werker, & Beddor, 2010). The problem of how infants acquire phonetic categories is typically considered in isolation. In this paper, however, we consider the learning problem from a broader perspective, testing the hypothesis that word-level information can feed back to influence phonetic category acquisition.

Distributional learning accounts (Maye, Werker, & Gerken, 2002) propose that learners obtain information about which sounds are contrastive in their native language by attending to the distributions of speech sounds in acoustic space. If learners hear a bimodal distribution of sounds along a particular acoustic dimension, they can infer that the language contains two categories along that dimension; conversely, a unimodal distribution provides evidence for a single phonetic category. Distributional learning is supported by evidence that adults and infants are sensitive to these types of speech sound distributions. Maye and Gerken (2000) tested adults' sensitivity to distributional information in a phonetic category learning task. In their experiment, participants were told that they would be listening to a new language; the language consisted of monosyllables whose initial stop consonants were drawn from either a unimodal

[*] Naomi Feldman, University of Maryland, nhf@umd.edu; Emily Myers, University of Connecticut, emily.myers@uconn.edu; Katherine White, University of Waterloo, white@uwaterloo.ca; Thomas Griffiths, University of California, Berkeley, tom_griffiths@berkeley.edu; and James Morgan, Brown University, james_morgan@brown.edu. This work was supported by NSF grant BCS-0924821 and NIH grant HD032005. We thank Lori Rolfe for help recruiting subjects and Andy Wallace for assistance in creating the vowel continuum. The infant data that were included in the original conference presentation have been omitted from these proceedings following discovery of a technical error that affected the familiarization stimuli; we are currently re-running the experiment.

or a bimodal distribution. During test, they were asked to make explicit judgments about whether the endpoint stimuli belonged to the same category in the language they just heard. Collecting explicit judgments ensured that the results reflected inferences about category membership rather than low-level changes in discrimination. Participants in the bimodal condition responded *different* significantly more often to pairs of endpoint stimuli than participants in the unimodal condition, indicating that the former group treated the stimuli as belonging to two categories. Parallel results have been found with 6- and 8-month-old infants (Maye et al., 2002), suggesting that infants have access to distributional information at the earliest stages of phonetic category acquisition.

However, these demonstrations have been based on very simplified input languages in which categories are clearly separated. Recent computational modeling results call into question whether distributional learning is sufficient to recover phonetic categories from realistic speech data. Phonetic categories, particularly vowel categories, show substantial acoustic overlap (Hillenbrand, Clark, Getty, & Wheeler, 1995; Peterson & Barney, 1952). Overlapping categories can appear as a single unimodal distribution, leading a purely distributional learner to erroneously assign the sounds to a single category. Whereas computational models have shown good distributional learning results on well-separated categories (e.g., McMurray, Aslin, & Toscano, 2009; Vallabha, McClelland, Pons, Werker, & Amano, 2007), the models' performance deteriorates substantially when the categories have a higher degree of overlap (Feldman, Griffiths, & Morgan, 2009).

We propose that learners can overcome the problem of overlapping phonetic categories by using word-level contextual information to supplement distributional learning. Although phonetic category and word learning are often implicitly assumed to occur sequentially, findings reveal considerable temporal overlap between sound and word learning processes during development. Infants show a decline in sensitivity to non-native phonetic contrasts between six and twelve months (Werker & Tees, 1984). They begin to segment words from fluent speech as early as six months (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005), and this ability continues to develop over the next several months (Jusczyk & Aslin, 1995; Jusczyk, Houston, & Newsome, 1999). Word segmentation tasks require infants to map words heard in isolation onto words heard in fluent sentences. Because isolated word forms differ acoustically from sentential forms, successful segmentation indicates that infants are performing some sort of categorization on segmented words before phonetic category learning is complete. Thus, the temporal overlap of sound and word learning processes during development raises the possibility that knowledge at the word level may feed back to influence speech sound acquisition.

Using word-level information can potentially help learners separate overlapping speech sound categories if the categories occur in distinct lexical contexts. For example, whereas the acoustic distributions for /ɪ/ and /e/ overlap substantially, learners might hear the /ɪ/ sounds in the context of the word *milk* and the /e/ sounds in the context of the word *game*. Although young infants may not have access to meanings for these words, the acoustic forms of *milk* and *game* are easily distinguishable on the basis of distributional information. Categorization of these acoustic word tokens thus provides an additional word-level cue that can help distinguish the /ɪ/ and /e/ phonetic categories. Feldman et al. (2009) formalized this idea by building a computational "lexical-distributional" model of phonetic category acquisition that learned to categorize word tokens at the same time that it learned phonetic categories. They presented simulations demonstrating that word-level information improves performance by allowing learners to distinguish acoustically overlapping categories.

Direct evidence that word-level information can affect human learners' phonetic perception comes from a study by Thiessen (2007). In this study 15-month-olds were tested in the switch task, in which infants are habituated with an object-label pairing and then tested on a pairing of the same object with a novel label. He replicated Stager and Werker's (1997) finding that infants at this age fail to notice a switch between minimally different labels, in this case *daw* and *taw*, when tested in this paradigm. Testing a second group of infants, two additional object-label pairings were introduced during the habituation phase. Infants were habituated with three objects, labeled *daw*, *tawgoo*, and *dawbow*, respectively. As before, they were tested on the *daw* object paired with the label *taw*. In contrast to the first group, however, infants in the second group noticed the switch when the label changed from *daw* to *taw*. Infants' improvement in noticing the switch is compatible with the idea that they use words to constrain phonetic category acquisition. Intriguingly, when the additional objects were instead labeled *tawgoo* and *dawgoo*, infants failed to notice the switch. These results indicate that the nature of the lexical context in which sounds are heard plays a crucial role: hearing the target sounds in distinct lexical contexts facilitates attention to or use of the contrast, whereas hearing the sounds in minimal pair contexts does not. However, Thiessen's results may not provide direct evidence that learners can use word-level information in phonetic category acquisition, as the task used in this experiment involved mapping words to referents. It is not clear to what extent referents are available to young infants first acquiring phonetic categories (although cf. Yeung & Werker, 2009).

In this paper we investigate whether learners are sensitive to word-level information in a phonetic category learning task when referents are not available. If learners can use word-level cues to constrain their interpretation of phonetic variability, then these cues can potentially supplement distributional learning, leading to more robust learning of acoustically overlapping categories.

2. Design

This experiment was modeled on the distributional learning experiment from Maye and Gerken (2000). However, rather than hearing unimodal or bimodal distributions of isolated syllables, adult participants heard a uniform distribution of syllables from a vowel continuum ranging from /ta/ (*tah*) to /tɔ/ (*taw*), and these syllables were embedded in the two-syllable words *guta* and *lita*. (see Stimuli section below for details about how the continua were constructed). The syllables *gu* and *li* provided contexts that could help participants distinguish the *tah* and *taw* sounds. The /a/-/ɔ/ vowel contrast was selected because vowel categories typically exhibit more acoustic overlap than consonant categories, and thus stand to benefit more from lexical information in phonetic category learning. Dialectal variation indicates that the /a/ and /ɔ/ sounds can be treated as either one or two categories (Labov, 1998), suggesting that learners may be able to switch between these interpretations on the basis of specific cues in the input.

Participants were divided into two groups. Half the participants heard a LEXICAL corpus containing either *gutah* and *litaw*, or *gutaw* and *litah*, but not both pairs of pseudowords. These participants therefore heard *tah* and *taw* in distinct lexical contexts (the specific pairings were counterbalanced across participants). The other half heard a NON-LEXICAL corpus containing all four pseudowords. These participants therefore heard *tah* and *taw* interchangeably in the same set of lexical contexts. The lexical-distributional hypothesis predicts that participants exposed to a LEXICAL corpus should separate the overlapping *tah* and *taw* categories because of their occurrence in distinct lexical contexts. Appearance in distinct lexical contexts is predicted to influence phonetic categorization in a way similar to hearing a bimodal distribution of sounds. Participants in the LEXICAL group should be more likely to respond that stimuli from the *tah* and *taw* categories are *different* than participants who hear the sounds used interchangeably in the same set of lexical contexts.

3. Methods

Participants. Forty adult native English speakers with no known hearing deficits from the Brown University community participated in this study. Participants were paid at a rate of $8/hour.

Stimuli. Stimuli consisted of an 8-point vowel continuum ranging from *tah* (/ta/) to *taw* (/tɔ/) and ten filler syllables: *bu*, *gu*, *ko*, *li*, *lo*, *mi*, *mu*, *nu*, *ro*, and *pi*. Several tokens of each of these syllables were recorded by a female native speaker of American English.

Tokens of *tah* and *taw* differed systematically only by their second formant, F_2. An F_2 continuum was created based on formant values from these tokens, containing eight equally-spaced tokens along an ERB psychophysical scale (Glasberg & Moore, 1990). Steady state second formant values from this continuum are shown in Table 1. All tokens in the continuum had steady state values of F_1=818 Hz, F_3=2750 Hz, F_4=3500 Hz, and F_5=4500 Hz, where the first and third formant values were based on measurements from a recorded *taw* syllable. Bandwidths for the five formants were set to 130, 70, 160, 250, and 200, respectively, based on values given in Klatt (1980) for the /ɑ/ vowel.

To create tokens in the continuum, a source-filter separation was performed in Praat (Boersma, 2001) on a recorded *taw* syllable that had been resampled at

11000 Hz. The source was checked through careful listening and inspection of the spectrogram to ensure that no spectral cues remained to the original vowel. A 53.8 ms portion of aspiration was removed from the source token to improve its subjective naturalness as judged by the experimenter, shortening its voice onset time to approximately 50 ms.

Eight filters were created that contained formant transitions leading into steady-state portions. Formant values at the burst in the source token were F_1=750 Hz, F_2=1950 Hz, F_3=3000 Hz, F_4=3700 Hz, and F_5=4500 Hz. Formant transitions were constructed to move from these burst values to each of the steady-state values from Table 1 in ten equal 10 ms steps, then stay at steady-state values for the remainder of the token. These eight filters were applied to copies of the source file using the Matlab signal processing toolbox. The resulting vowels were then cross-spliced with the unmanipulated burst from the original token. The stimuli were edited by hand to remove clicks resulting from discontinuities in the waveform at formant transitions, resulting in the removal of 17.90 ms total, encompassing four pitch periods, from three distinct regions in the formant transition portion of each stimulus. An identical set of regions was removed from each stimulus in the continuum. After splicing, the duration of each token in the continuum was 416.45 ms.

Four tokens of each of the filler syllables were resampled at 11000 Hz to match the synthesized *tah/taw* tokens, and the durations of these filler syllables were modified to match the duration of the *tah/taw* tokens. The pitch of each token was set to a constant value of 220 Hz. RMS amplitude was normalized across tokens.

Bisyllabic pseudo-words *guta*, *lita*, *romu*, *pibu*, *komi*, and *nulo* were constructed through concatenation of these tokens. Thirty-two tokens each of *guta* and *lita* were constructed by combining the four tokens of *gu* or *li* with each of the eight stimuli in the /ta/ to /tɔ/ continuum. Sixteen tokens of each of the four bisyllabic filler words (*romu*, *pibu*, *komi*, and *nulo*) were created using all possible combinations of the four tokens of each syllable.

Table 1. Second formant values of stimuli in the *tah-taw* continuum.

Stimulus Number	Second Formant (Hz)
1	1517
2	1474
3	1432
4	1391
5	1351
6	1312
7	1274
8	1237

Apparatus. Participants were seated at a computer and heard stimuli through Bose QuietComfort 2 noise cancelling headphones at a comfortable listening level.

Procedure. Participants were assigned to one of two conditions, the NON-LEXICAL condition or the LEXICAL condition, and completed two identical blocks. Each block contained a familiarization period followed by test. Participants were told that they would hear two-syllable words in a language they had never heard before and that they would subsequently be asked questions about the sounds in the language.

During familiarization, each participant heard 128 pseudo-word tokens per block. Half of these consisted of one presentation of each of the 64 filler tokens (*romu*, *pibu*, *komi*, and *nulo*). The other half consisted of 64 experimental tokens (*guta* and *lita*). All participants heard each token from the continuum eight times per block, but the lexical contexts in which they heard these syllables differed across conditions. To describe the differences between conditions, we refer to steps 1-4 of the continuum as *tah* and steps 5-8 of the continuum as *taw*. Participants in the NON-LEXICAL condition heard each *gutah*, *gutaw*, *litah*, and *litaw* token once per block for a total of 64 experimental tokens (the 4 tokens of each context syllable combined with each of the 8 *ta* tokens). Participants in the LEXICAL condition were divided into two subconditions. Participants in the *gutah*-*litaw* subcondition heard the 16 *gutah* tokens (the 4 tokens of *gu* combined with the 4 tokens of *tah*) and the 16 *litaw* tokens (the 4 tokens of *li* combined with the 4 tokens of *taw*), each twice per block. They did not hear any *gutaw* or *litah* tokens. Conversely, participants in the *gutaw*-*litah* subcondition heard the 16 *gutaw* tokens and the 16 *litah* tokens twice per block, but did not hear any *gutah* or *litaw* tokens. The order of presentation of these 128 pseudowords was randomized, and there was a 750 ms interstimulus interval between tokens.

During test, participants heard two syllables, separated by 750 ms, and were asked to make explicit judgments as to whether the syllables belonged to the same category in the language. The instructions were as follows:

> Now you will listen to pairs of syllables and decide which sounds are the same. For example, in English, the syllables CAP and GAP have different sounds. If you hear two different syllables (e.g. CAP-GAP), you should answer DIFFERENT, because the syllables contain different sounds. If you hear two similar syllables (e.g. GAP-GAP), you should answer SAME, even if the two pronunciations of GAP are slightly different.
>
> The syllables you hear will not be in English. They will be in the language you just heard. You should answer based on which sounds you think are the same in that language. Even if

> you’re not sure, make a guess based on the words you heard before.

Participants were then asked to press specific buttons corresponding to *same* or *different* and to respond as quickly and accurately as possible.

The test phase examined three contrasts: *ta*$_1$ vs. *ta*$_8$ (far contrast), *ta*$_3$ vs. *ta*$_6$ (near contrast), and *mi* vs. *mu* (control). Half the trials were *different* trials containing one token of each stimulus type in the pair, and the other half were *same* trials containing two tokens of the same stimulus type. For *same* trials involving *tah/taw* stimuli, the two stimuli were identical tokens. For *same* trials involving *mi* and *mu*, the two stimuli were non-identical tokens of the same syllable, to ensure that participants were correctly following the instructions to make explicit category judgments rather than lower-level acoustic judgments. Participants heard 16 *different* and 16 *same* trials for each *tah/taw* contrast (far and near) and 32 *different* and 32 *same* trials for the control contrast in each block. Responses and reaction times were recorded for each trial.

4. Results

Responses were excluded from the analysis if the participant responded before hearing the second stimulus of the pair or if the reaction time was more than two standard deviations from a participant's mean reaction time for a particular response on a particular class of trial in a particular block. This resulted in an average of 5% of trials discarded from analysis.[1] The sensitivity measure d’ (Green & Swets, 1966) was computed from the remaining responses for each contrast in each block. A value of 0.99 was substituted for any trial type in which a participant responded *different* on all trials, and a value of 0.01 was substituted for any trial type in which a participant responded *same* on all trials. The d’ scores for each contrast are shown in Figure 1.

A 2×2 (condition × block) mixed ANOVA was conducted for each contrast. For the far contrast and the near contrast, the analysis yielded a main effect of block ($F(1,38)=10.42$, $p=0.003$, far contrast; $F(1,38)=17.99$, $p<0.001$, near contrast) and a significant condition by block interaction ($F(1,38)=11.25$, $p=0.002$, far contrast; $F(1,38)=12.30$, $p=0.001$, near contrast). This interaction reflected the larger increase in d’ scores from Block 1 to Block 2 of participants in the LEXICAL condition as compared to participants in the NON-LEXICAL condition. There was no significant main effect of condition for either contrast. Tests of simple effects showed no significant effect of condition in the first block; there was a significant effect of condition in the second block for the far

[1] Analyzing the data with these trials included yields similar results.

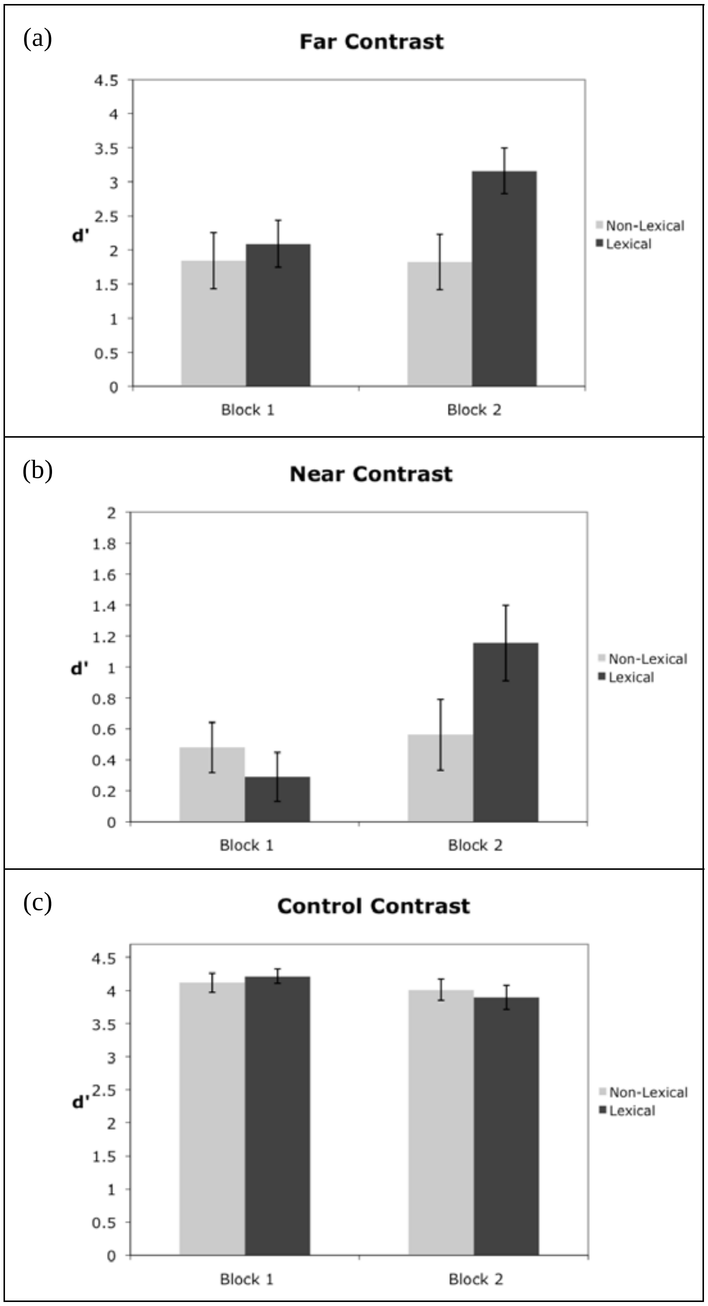

Figure 1. Sensitivity to category differences for the (a) far contrast, (b) near contrast, and (c) control contrast.

contrast (t(38)=2.54, p=0.03, Bonferroni corrected), but this comparison did not reach significance for the near contrast. A 2×2×2 (condition × block × contrast) ANOVA[2] confirmed that the near and far contrasts patterned similarly, showing main effects of block (F(1,38)=20.65, p<0.001) and contrast (F(1,38)=62.84, p<0.001) and a block×condition interaction (F(1,38)=18.18, p<0.001), but no interactions involving contrast.

On control trials, the analysis yielded a main effect of block (F(1,38)=5.90, p=0.019), reflecting the fact that d' scores were reliably lower in the second block. This decrease in d' scores between blocks was in the opposite direction from the increase in d' scores between blocks on experimental trials. There was no significant difference between groups and no interaction. Sensitivity to category differences was high, with an average d' measure of 4.17 on the first block and 3.95 on the second block, indicating that participants were performing the task.

5. Discussion

This experiment was designed to test whether human learners are sensitive to word-level cues to phonetic category membership in an artificial language learning task. Interactive learning predicts that participants in the LEXICAL group, who heard the *tah* and *taw* stimuli in distinct lexical contexts, should be more likely to treat the sounds as belonging to different categories as compared to participants in the NON-LEXICAL group, who heard *tah* and *taw* interchangeably in the same set of lexical contexts. This predicted pattern was obtained after the second block of training: Participants in the LEXICAL condition showed higher d' scores than participants in the NON-LEXICAL condition. These results show that adults alter their interpretation of acoustic variability on the basis of word-level information.

The two groups' indistinguishable performance after the first training block provides strong evidence that these differences in sensitivity were the result of learning over the course of the experiment. The direction of learning, however, cannot be inferred from these data. One possibility is that the LEXICAL group learned over the course of the experiment to treat the experimental stimuli as different. Under this interpretation, their increase in d' scores in the second block would reflect category learning that resulted from the specific word-level information they received about these sounds during familiarization. Another possibility is that the increase in d' scores in the LEXICAL group reflected perceptual learning that arose through simple exposure to the sounds, and that this perceptual learning was not apparent in the NON-LEXICAL group because those participants learned to treat the experimental stimuli as the same based on their interchangeability in words. These two possibilities cannot be

[2] The control contrast cannot be included in this direct comparison because the tokens in the *same* control trials were acoustically different, whereas the tokens in *same* experimental trials were acoustically identical.

distinguished without a baseline measure of categorization that is independent of familiarization condition. Thus, lexical information may have served to separate overlapping categories or to merge distinct acoustic categories that are used interchangeably. Regardless, either possibility is consistent with the hypothesis that participants used word-level information to constrain their interpretation of phonetic variability.

It is also possible, however, that participants' disambiguation of the *tah* and *taw* categories in the LEXICAL group arose from patterns of transitional probabilities between syllables rather than from word-level information. Transitional probabilities between specific *tah/taw* variants and context syllables were higher in the LEXICAL condition than in the NON-LEXICAL condition because the appearance of the *tah* and *taw* syllables was more constrained. In addition, transitional probabilities were different for the *tah* and *taw* syllables. If participants computed statistics separately for each step in the continuum, the statistical properties of the LEXICAL corpus may have enabled them to cluster sounds into distinct categories.

The confound between word-level information and transitional probabilities reflects an inherent ambiguity in language. Word contexts and phonological contexts are confounded in linguistic input; phonemic differences cause different sounds to appear consistently in different words, whereas phonological alternations cause different sounds to appear systematically in different phonological contexts (and thus in different words). Participants hearing the LEXICAL corpus do not have enough information to determine whether acoustic differences between *tah* and *taw* should be attributed to the differing lexical or phonological contexts.

Under a lexical interpretation, the pattern in the LEXICAL corpus arises because words in a language do not exhaust all possible phoneme sequences. Idiosyncratic differences across different lexical items occur because some lexical items contain one phoneme and some contain the other. If the input is interpreted in this manner, consistent acoustic differences across lexical contexts provide a source of disambiguating evidence that the sounds belong to different categories.

Under a phonological interpretation, however, the pattern in the LEXICAL corpus can arise due to a process like vowel-to-vowel coarticulation or vowel harmony. This type of phonological process results in complementary distribution of the target sounds. If participants hear the words *gutaw* and *litah*, it is possible to interpret *taw* and *tah* as allophones representing a single underlying phoneme conditioned by the preceding phonological contexts, /u/ and /i/. English-learning infants show evidence of sensitivity to vowel harmony patterns at seven months, despite lack of exposure to these patterns, suggesting that the phonological interpretation of acoustic variability may be available at the age when infants are first acquiring phonetic categories (Mintz, Walker, Welday, & Kidd, submitted). Under this interpretation, participants familiarized with a LEXICAL corpus might disregard acoustic variation that can be attributed to phonological factors and treat the sounds as different *less* often than participants familiarized with the NON-LEXICAL corpus.

The pattern of results obtained here suggests that participants adopted a word-level interpretation rather than a phonological interpretation of the alternations. To further rule out the possibility of a phonological interpretation, we tested for differences between the *gutah-litaw* and *gutaw-litah* subconditions. Evidence suggests that "natural" phonological alternations are easier to learn than arbitrary alternations (Peperkamp, Skoruppa, & Dupoux, 2006; Saffran & Thiessen, 2003; Wilson, 2006). Learners might therefore be more willing to attribute natural alternations to phonological factors, whereas unnatural alternations might be attributed more often to lexical factors. In the present experiment, only one of the LEXICAL subconditions presents participants with a natural alternation: In the *gutaw-litah* subcondition, the *gu* syllable with low F_2 and is paired with the *taw* syllable, which has lower F_2 than *tah*. Similarly, the *li* syllable with high F_2 is paired with the *tah* syllable with high F_2. This means the differences between *gu* and *li* are in the same direction as those between the *tah/taw* syllables, making this a natural alternation. In the *gutah-litaw* subcondition, however, the pattern represents a less natural alternation because the second formant in the *tah/taw* syllable is shifted in the opposite direction from what would be predicted on the basis of the context syllable. A 2×2 (subcondition × block) ANOVA showed no significant differences between the *gutah-litaw* and *gutaw-litah* subconditions and no interactions involving subcondition for any of the three contrasts, suggesting that participants were not sensitive to the naturalness of the alternation. Thus, participants appeared to interpret these patterns as reflecting phonemic differences. Taken together, these results suggest that listeners adopt a lexical rather than phonological interpretation of the *tah/taw* alternations evident in the LEXICAL condition.

This experiment provides support for the lexical-distributional account of phonetic category acquisition by showing that learners alter their interpretation of phonetic variability on the basis of word-level cues. Adults assigned sounds to different categories more often when they appeared in distinct lexical contexts. The patterns obtained here resemble the results from Thiessen (2007), but show that referents are not required for interactive learning. Taken together with previous results showing sensitivity to distributional cues (Maye & Gerken, 2000; Maye et al., 2002), these results indicate that human learners attend to the types of cues that are necessary to achieve lexical-distributional learning. Future work will examine whether infants are sensitive to these types of word-level cues between six and twelve months, at the age when they are first learning phonetic categories. If such cues are available to young infants, it would suggest that a more complete understanding of the phonetic category learning process can be obtained by taking into account interactions with contemporaneous learning processes.

References

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341-345.

Bortfeld, Heather, Morgan, James L., Golinkoff, Roberta M., & Rathbun, Karen (2005). *Mommy* and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*, 298-304.

Feldman, Naomi H., Griffiths, Thomas L., & Morgan, James L. (2009). Learning phonetic categories by learning a lexicon. In N. A. Taatgen & H. v. Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2208-2213). Austin, TX: Cognitive Science Society.

Glasberg, Brian R., & Moore, Brian C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*, 103-138.

Green, David M., & Swets, John A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hillenbrand, James, Getty, Laura A., Clark, Michael J., & Wheeler, Kimberlee (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099-3111.

Jusczyk, Peter W., & Aslin, Richard N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1-23.

Jusczyk, Peter W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*, 159-207.

Klatt, Dennis H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971-995.

Labov, William (1998). The three dialects of English. In M. D. Lin (Ed.), *Handbook of Dialects and Language Variation* (pp. 39-81). San Diego, CA: Academic Press.

Maye, Jessica, & Gerken, LouAnn (2000). Learning phonemes without minimal pairs. In S. C. Howell & S. A. Fish & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 522-533). Somerville, MA: Cascadilla Press.

Maye, Jessica, Werker, Janet F., & Gerken, LouAnn (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101-B111.

McMurray, Bob, Aslin, Richard N., & Toscano, Joseph C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*, 369-378.

Mintz, Toben, Walker, Rachel L., Welday, Ashlee, and Kidd, Celeste (submitted). Infants' universal sensitivity to vowel harmony and its role in speech segmentation.

Narayan, Chandan R., Werker, Janet F., & Beddor, Patrice S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, *13*, 407-420.

Peperkamp, Sharon, Skoruppa, Katrin, & Dupoux, Emmanuel (2006). The role of phonetic naturalness in phonological rule acquisition. In D. Bamman & T. Magnitskaia & C. Zaller (Eds.), *Proceedings of the 30th Boston University Conference on Language Development* (pp. 464-475). Somerville, MA: Cascadilla Press.

Peterson, Gordon E., & Barney, Harold L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.

Saffran, Jenny R., & Thiessen, Erik D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*, 484-494.

Stager, Christine L., & Werker, Janet F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381-382.

Thiessen, Erik D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*, 16-34.

Vallabha, Gautam K., McClelland, James L., Pons, Ferran, Werker, Janet F., & Amano, Shigeaki (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*, 13273-13278.

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

Wilson, Colin (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945-982.

Yeung, H. Henny, & Werker, Janet F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*, 234-243.