

Word-level information influences phonetic learning in adults and infants

Naomi H. Feldman, University of Maryland

Emily B. Myers, University of Connecticut

Katherine S. White, University of Waterloo

Thomas L. Griffiths, University of California, Berkeley

James L. Morgan, Brown University

Running head: Word-level information influences phonetic learning

Address for correspondence:

Naomi Feldman

1401 Marie Mount Hall

College Park, MD 20742

nhf@umd.edu

301-405-5800

Abstract

Infants begin to segment words from fluent speech during the same time period that they learn phonetic categories. Segmented words can provide a potentially useful cue for phonetic learning, yet accounts of phonetic category acquisition typically ignore the contexts in which sounds appear. We present two experiments to show that, contrary to the assumption that phonetic learning occurs in isolation, learners are sensitive to the words in which sounds appear and can use this information to constrain their interpretation of phonetic variability. Experiment 1 shows that adults use word-level information in a phonetic category learning task, assigning acoustically similar vowels to different categories more often when those sounds consistently appear in different words. Experiment 2 demonstrates that eight-month-old infants similarly pay attention to word-level information and that this information affects how they treat phonetic contrasts. These findings suggest that phonetic category learning is a rich, interactive process that takes advantage of many different types of cues that are present in the input.

Keywords: language acquisition; phonetic category learning

One of the first challenges language learners face is discovering how speech sounds are organized in their language. Infants learn about the phonetic categories of their native language quite early. They can initially discriminate sound contrasts from many languages, but perception of non-native vowel contrasts declines by eight months (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994) and perception of non-native consonant contrasts follows soon afterwards by ten months (Werker & Tees, 1984). During the same time period, perception of at least some native language contrasts is enhanced (Kuhl et al., 2006; Narayan, Werker, & Beddor, 2010). Identifying the mechanisms by which sound category knowledge is acquired has been a central focus of research in language acquisition.

Distributional learning theories (Maye, Werker, & Gerken, 2002) propose that learners acquire phonetic categories by attending to the distributions of sounds in acoustic space. Learners hearing a bimodal distribution of sounds along an acoustic dimension may infer that sounds from the two “modes” belong to two different categories; conversely, a unimodal distribution may provide evidence for a single phonetic category. Adults and infants show sensitivity to this type of distributional information. After familiarization with a bimodal distribution of stimuli ranging from the voiceless unaspirated stop, [ta], to a voiced unaspirated stop, [da], adults assign endpoint [t] and [d] stimuli to distinct categories more often than after familiarization with the same sounds embedded in a unimodal distribution (Maye & Gerken, 2000). Similarly, 6-8-month-old infants show evidence of discriminating these endpoint stimuli after familiarization with a bimodal, but not a unimodal, distribution (Maye et al., 2002). Voiced and voiceless unaspirated stops are not contrastive in English, but are discriminable by 6-8-month-old infants (Pegg & Werker, 1997); thus, the difference between bimodal and unimodal familiarization conditions from Maye et al. (2002) is likely to reflect reduced sensitivity in the

unimodal condition. Distributional learning can also lead to increased sensitivity: Maye, Weiss, and Aslin (2008) showed that hearing a bimodal distribution enhances infants' sensitivity to a more difficult voicing contrast between prevoiced [da] and voiceless unaspirated [ta]. These findings indicate that young infants can compute and use distributional information, supporting a role for distributional learning in phonetic category acquisition.

Computational implementations of distributional learning have shown promising results in recovering well-separated phonetic categories (McMurray, Aslin, & Toscano, 2009; Vallabha, McClelland, Pons, Werker, & Amano, 2007). However, performance declines when categories have a higher degree of acoustic overlap (Dillon, Dunbar, & Idsardi, in press; Feldman, Griffiths, & Morgan, 2009). This poses a potential problem for the acquisition of certain natural language categories, particularly vowel categories. Vowel categories tend to overlap more in their formant values than do stop consonants in their voice onset time (e.g., see Hillenbrand, Getty, Clark, & Wheeler, 1995; Lisker & Abramson, 1964; Peterson & Barney, 1952, for production data from English). Discrimination of vowels has been shown to decline around 6-8 months (Bosch & Sebastián-Gallés, 2003; Kuhl et al., 1992; Polka & Werker, 1994), and it is still unknown which mechanisms guide this early vowel category acquisition. Demonstrations of distributional learning in infants have primarily tested consonant contrasts (Cristià, McGuire, Seidl, & Francis, 2011; Maye et al., 2002; Maye et al., 2008; Yoshida, Pons, Maye, & Werker, 2010). While evidence for distributional learning of vowels has been found in adults (Gulian, Escudero, & Boersma, 2007; Escudero, Benders, & Wanrooij, 2011), the high degree of overlap in natural language vowel categories suggests that learners' ability to acquire these categories might benefit from supplementary cues that can be used in parallel with distributional learning.

Recent work has begun to look at the influence of contextual cues, such as visual information from articulations and objects, on infants' phonetic learning (Teinonen, Aslin, Alku, & Csibra, 2008; Yeung & Werker, 2009). In each case, consistent pairings between sounds and contextual cues have been shown to enhance infants' sensitivity to contrasts among the sounds. Whereas this previous work has focused on visual context, here we examine an auditory contextual cue that nearly always co-occurs with sounds: the word context in which sounds appear. Previous work on distributional and contextual phonetic learning has implicitly assumed that infants analyze isolated sounds, without regard for the surrounding speech context, but infants show evidence of attending to word-sized units during the initial period of phonetic learning. 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). The word segmentation and recognition tasks used in these studies require infants to recognize similarities between words heard in isolation and new tokens of the same words heard in fluent sentences. Because isolated word forms differ acoustically from sentential forms (and even from each other), successful recognition requires categorizing together acoustically distinct tokens of the same word type. Infants can thus perform some categorization on word tokens before phonetic category learning is complete. The temporal overlap of sound and word categorization processes during development raises the possibility that knowledge at the word level may influence phonetic category acquisition.

Here we test whether word-level information can help separate overlapping vowel categories. This idea that distinctive word forms can serve as contextual cues for phonetic learning has been proposed several times in the literature (Feldman et al., 2009; Swingley, 2009; Swingley &

Aslin, 2007; Thiessen, 2007). For example, although the acoustic distributions for /ɛ/ and /æ/ overlap, learners might hear the two sounds in different word contexts, e.g., /ɛ/ sounds in the context of the word *egg* and /æ/ sounds in the context of the word *cat*. The acoustic forms of *egg* and *cat* are easily distinguishable. Categorization of these acoustic word tokens provides an additional word-level cue that can help distinguish the /ɛ/ and /æ/ phonetic categories. Word level information under this account is similar to a type of supervision signal, indicating to learners how they should categorize observed sounds. However, it differs from typical supervised learning accounts in that the supervision signal (the phonological forms of individual words) is not known in advance, but instead needs to be learned by observing and categorizing acoustic word tokens. This account is supported by data from 15-month-old infants (Thiessen, 2007, 2011), but it has not yet been demonstrated in young infants during the early stages of phonetic learning.

It is important to distinguish our account from previous ideas regarding the role of minimal pairs in phoneme acquisition. In theoretical linguistics, the discovery of *minimal pairs* - words that differ in only a single sound - is taken to be decisive evidence for positing phonemic categories. For example, if /bɛt/ (*bet*) and /bæt/ (*bat*) are words with different meanings, then /ɛ/ and /æ/ are distinct phonemes. This leads straightforwardly to a learning model: when infants notice that two words with different meanings differ by a single sound, this observation allows them to infer that the sounds are contrastive. Young infants are sensitive to the type of information available in minimal pairs (Yeung & Werker, 2009). However, as noted by Maye et al. (2002), learning from minimal pairs requires extensive knowledge of word meanings. Without knowledge of word meanings, members of a minimal pair might instead be interpreted as tokens of the same word. Analyses of children's early productive lexicons suggest that

although children do know some pairs of minimally different words, they know many fewer such pairs than adults (Charles-Luce & Luce, 1990, 1995; but see Coady & Aslin, 2003; Dollaghan, 1994). Thus, in early stages of development, minimal pairs might not help infants separate overlapping phonetic categories. Our account instead relies on the idea that non-minimal pairs, which provide distinct word contexts for similar sounds, are crucial for helping infants separate overlapping categories before semantic information is available.

Feldman, Griffiths, Goldwater, & Morgan (submitted; Feldman, 2011; Feldman et al., 2009) explored the potential benefit of interactive learning of sounds and words by formalizing a Bayesian model that recovered sound and word categories simultaneously from a corpus of isolated word tokens. The model was not given a lexicon a priori, but was allowed to learn a set of word types at the same time it was learning vowel categories from sequences of formant values. Feldman et al. compared the performance of this lexical-distributional learning model to that of previously proposed distributional models (Rasmussen, 2000; Vallabha et al., 2007), evaluating each on its ability to recover English vowel categories. They found that the lexical-distributional model outperformed distributional-only models, correctly recovering the complete set of English vowel categories despite their high degree of overlap. The success of the interactive lexical-distributional model in recovering English vowel categories indicates that the English lexicon used in their simulations contained a sufficiently low proportion of minimal pairs. Minimal pairs actually impair learning in this type of model, because the model can mistakenly interpret them as the same word and assign all vowels appearing in that word to a single category. Thus, the success of the model indicates both that attending to word contexts can help an ideal Bayesian learner identify categories that would be mistakenly merged by a

purely distributional learner, and that the early English lexicon contains sufficient disambiguating information to support interactive learning of English vowel categories.

If the interactive lexical-distributional hypothesis is correct, then we would predict that children should be able to use non-minimal pairs as contextual cues to constrain phonetic learning. In line with this, lexical contexts seem to influence children's use of phonemic contrasts in word-learning tasks. The first evidence for this was given by Thiessen (2007), who tested 15-month-old infants in the switch task (Stager & Werker, 1997). Infants in this task are habituated to an object-label pairing and then tested on a pairing of the same object with a novel label. Thiessen's first experiment 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, Thiessen introduced two additional object-label pairings during the habituation phase. Infants were thus habituated to three objects, labeled *daw*, *tawgoo*, and *dawbow*, respectively. As before, test trials contained the *daw* object paired with the label *taw*. Infants in this second group noticed the switch when the label changed from *daw* to *taw*. Intriguingly, when the additional objects were instead labeled *tawgoo* and *dawgoo* (i.e., were minimal pairs) during habituation, infants failed to notice the switch. These results suggest that the nature of the lexical context in which sounds are heard determines how those sounds are treated: 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. Thiessen and Yee (2010) replicated and extended these results from onsets to codas, showing that infants alter their treatment of minimally contrastive syllable-final consonants after hearing those syllable-final consonants in distinct word contexts.

The results from Thiessen (2007; Thiessen & Yee, 2010) are compatible with the idea that infants use word forms as contextual cues to constrain phonetic category acquisition. However, the non-minimal pairs in these experiments were presented in referential contexts, and it is possible that this combination of forms and meanings was necessary for infants to use the words as contextual cues. If true, this type of information may be of less use to infants in the early stages of phonetic category acquisition. Although infants do know the meanings of some familiar words at this stage of development (Benedict, 1979; Bergelson & Swingley, 2011; Mandel, Jusczyk, & Pisoni, 1995; Tincoff & Jusczyk, 1999), this semantic vocabulary is small, and they do not learn associations between words and objects in most laboratory settings until 12-15 months (Curtin, 2009; Schafer & Plunkett, 1998; Werker, Cohen, Lloyd, Casasola, & Stager, 1998; Woodward, Markman, & Fitzsimmons, 1994; but see Gogate & Bahrack, 1998; Shukla, White & Aslin, 2011). For this type of interactive learning to be viable in early development, then, it must be possible even in the absence of referential information. Thiessen (2011) tested whether distinct lexical contexts can facilitate older infants' use of sound contrasts even when word referents are not present. He familiarized 15-month-old infants with acoustic tokens of either *tawgoo* and *dawbow*, or *tawgoo* and *dawgoo*, in the absence of any objects. He then habituated infants to an object labeled *daw* and tested their reaction when the label was changed to *taw*, again using the switch task paradigm. Results showed that familiarization with non-minimal pairs, but not minimal pairs, facilitated infants' use of the *daw-taw* distinction in the switch task.

These previous results suggest that feedback from the developing lexicon plays a role in infants' use of phonetic contrasts during word learning. However, it is still an open question whether lexical context can guide infants' initial acquisition of phonetic contrasts. Here we

describe two experiments that probe the extent to which lexical information can influence phonetic learning. Experiment 1 asks whether the distribution of vowels across words affects adults' treatment of these vowels. Experiment 2 tests whether these same influences of word context on phonetic discrimination can be observed in eight-month-old infants, who are still in the early stages of phonetic learning. We find evidence that both adults and eight-month-old infants can use word-level information to constrain their treatment of phonetically similar vowels, suggesting that the developing lexicon may play an important role in infants' early acquisition of overlapping phonetic categories.

Our studies differ from Thiessen's (2011) in three important ways. First, whereas Thiessen used a stop consonant voicing distinction in his experiments (*taw* vs. *daw*), we instead use an [ɑ]-[ɔ] vowel contrast. Vowel categories typically exhibit more acoustic overlap than stop consonant categories, and they thus stand to benefit more from lexical information in phonetic category learning. Dialectal variation indicates that [ɑ] and [ɔ] can be treated as belonging to either one or two categories (Labov, 1998), suggesting that these sounds are neither too acoustically similar to be distinguished, nor too distinct to be assigned to a single category. Thus, this is an ideal example of a contrast for which learners may be able to switch between interpretations on the basis of specific cues in the input, such as word-level information. Using this vowel contrast allows us to directly investigate mechanisms involved in vowel acquisition and to assess whether word-level information can potentially solve the problem of overlapping vowel categories identified by Feldman et al. (2009).

Second, Thiessen measured children's use of a contrast in a word-learning context. Previous research has shown that infants can fail to use consonant contrasts in word-learning tasks but show sensitivity to those same contrasts in more standard discrimination tasks, suggesting that

word-learning tasks recruit resources beyond those required for simple phonetic learning (Stager & Werker, 1997). Here we are interested in measuring the influence of word-level information on the acquisition of a contrast, rather than its use. We adopt more direct measures of phonetic sensitivity, modeled on the distributional learning experiments of Maye and Gerken (2000) and Maye et al. (2002). Our tasks do not require participants to map words to referents. Instead, we ask for explicit same-different judgments of auditory stimuli from adults, and we look at infants' treatment of acoustic differences among auditory stimuli by measuring their looking times for alternating and non-alternating syllables.

Finally, and most importantly, Thiessen tested 15-month-old infants, who had presumably already acquired a good deal of knowledge about the phonetic categories of English. The plausibility of an interactive lexical-distributional approach to phonetic category acquisition critically depends on the extent to which infants are sensitive to word-level information during earlier stages of phonetic category learning. To get a measure of this earlier sensitivity, we test eight-month-old infants. Previous research indicates that infants at this age are actively acquiring knowledge of word forms (Bortfeld et al., 2005; Jusczyk & Aslin, 1995; Jusczyk et al., 1999). However, they have not yet finished learning the sound categories of their language: Perception of consonants is just beginning to change at this age (Werker & Tees, 1984), and although perception of vowels shows some language specificity as early as six months (Kuhl et al., 1992), it is likely that phonetic representations for vowels are still developing. The availability of word-level information, together with the fact that phonetic perception is still in flux, makes eight months an ideal age to investigate the influences of lexical context on phonetic category learning.

Experiment 1: Adults' sensitivity to word-level information

This experiment investigates whether adult learners are sensitive to the distribution of vowel sounds across words and can use this information to constrain phonetic category learning in a non-referential task. The experiment is modeled on the distributional learning experiment conducted by Maye and Gerken (2000), who tested adults' sensitivity to distributional information in a phonetic category learning task. In Maye and Gerken's experiment, participants listened to an artificial language consisting of monosyllables whose initial stop consonants were drawn from either a unimodal or a bimodal distribution along a continuum ranging from a voiced unaspirated [d] to a voiceless unaspirated [t]. In our experiment, rather than hearing unimodal or bimodal distributions of isolated syllables, all adult participants heard a *uniform* distribution of syllables with vowels from a vowel continuum ranging from [ɑ] (*ah*) to [ɔ] (*aw*). These vowels were embedded in the two-syllable words *gutah/aw* and *litah/aw* (see Stimuli section below for details about how the continua were constructed). The syllables *gu* and *li* preceding the target sounds provided word contexts that could potentially indicate to participants whether the *taw* and *tah* sounds belonged to two different categories or to a single category.

Participants were divided into two groups. Half the participants heard a NON-MINIMAL PAIR 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 MINIMAL PAIR corpus containing all four pseudowords. These participants therefore heard *tah* and *taw* interchangeably in the same set of lexical contexts. The interactive lexical-distributional hypothesis predicts that participants exposed to a NON-MINIMAL PAIR corpus should be more likely to separate the overlapping [ɑ] and [ɔ] categories because of their occurrence in distinct

lexical contexts. Appearance in distinct lexical contexts is predicted to influence phonetic categorization by indicating to listeners that there are multiple categories of sounds. Thus, participants in the NON-MINIMAL PAIR group should be more likely to respond that stimuli from the [ɑ] and [ɔ] categories are *different* than participants in the MINIMAL PAIR group, who hear the sounds used interchangeably in the same set of lexical contexts.

Method

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 eight-point vowel continuum in which vowels were preceded by aspirated [t^h], with stimuli ranging from *tah* ([t^hɑ]) to *taw* ([t^hɔ]), 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. For notational convenience, we refer to steps 1-4 of the continuum as *tah* and steps 5-8 of the continuum as *taw*.

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

The vowels in our natural recordings of *tah* and *taw* differed systematically only in 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 [a] 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 *gutah/aw*, *litah/aw*, *romu*, *pibu*, *komi*, and *nulo* were constructed through concatenation of these tokens. Thirty-two tokens each of *gutah/aw* and *litah/aw* were constructed by combining the four tokens of *gu* or *li* with each of the eight stimuli in the *tah* to *taw* 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.

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

Procedure. Each participant was assigned to one of two conditions, the MINIMAL PAIR condition or the NON-MINIMAL PAIR condition, and completed two identical familiarization-testing blocks. 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 (*gutah/aw* and *litah/aw*). Importantly, all participants heard each token from the *tah/taw* continuum eight times per block, but the lexical contexts in which they heard these syllables differed across conditions. Participants in the

MINIMAL PAIR condition heard each *gutah*, *gutaw*, *litah*, and *litaw* token once per block for a total of 64 experimental tokens. This was achieved by combining the 4 tokens of each context syllable with each of the 8 *tah/taw* tokens; with steps 1-4 of the continuum notated as *tah* and steps 5-8 referred to as *taw*. Participants in the NON-MINIMAL PAIR 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.

Instructions, which were given immediately prior to the first test phase, 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, two involving *tah* vs. *taw*: *tah*₁ vs. *taw*₈ (far contrast), *tah*₃ vs. *taw*₆ (near contrast), and a control contrast, *mi* vs. *mu*. 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 (32 total far and 32 total near) and 32 *different* and 32 *same* trials for the control contrast in each block. Responses and reaction times were recorded for each trial.

Results and Discussion

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

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

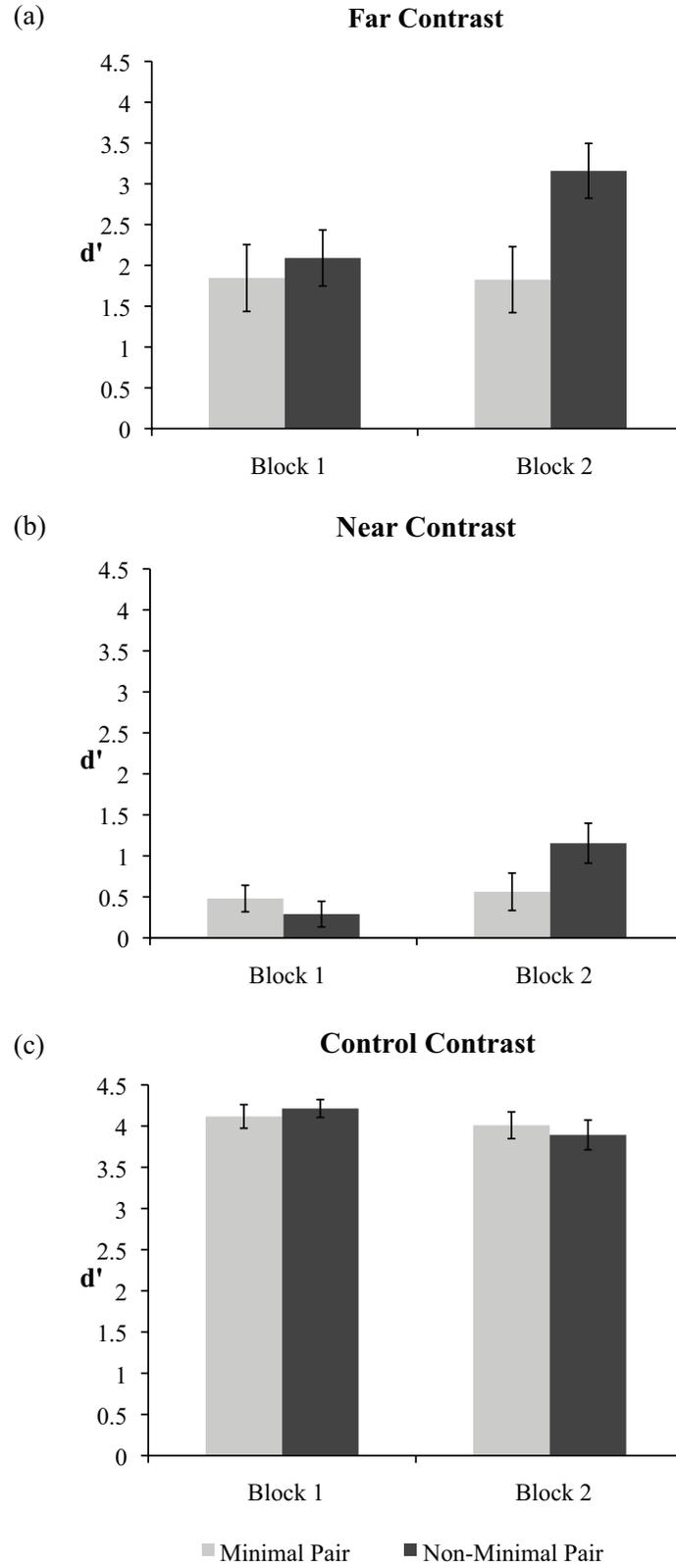


Figure 1. Sensitivity to category differences for the (a) far contrast, (b) near contrast, and (c) control contrast in Experiment 1.

in which a participant responded *same* on all trials. The d' scores for each contrast are shown in Figure 1.

A 2×2 condition (NON-MINIMAL PAIR vs. MINIMAL PAIR) \times block (1 vs. 2) 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). These interactions reflected the larger increases in d' scores from Block 1 to Block 2 for participants in the NON-MINIMAL PAIR condition as compared to participants in the MINIMAL PAIR condition, for both contrasts. 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 contrast, with participants in the NON-MINIMAL PAIR condition having higher d' scores than participants in the MINIMAL PAIR condition ($t(38)=2.54$, $p=.02$), but this comparison did not reach significance for the near contrast. A $2 \times 2 \times 2$ (condition \times block \times contrast) ANOVA² 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 \times 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 ($F(1,38)=5.90$, $p=0.02$), 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

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.

blocks on experimental trials and may reflect task-related effects, such as participants paying more attention to the difficult *tah/taw* sounds during the later portion of the experiment. There was no significant difference between groups and no interaction, showing that both groups did equally well on the control contrasts. 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.

The two groups' indistinguishable performance after the first training block provides strong evidence that the advantage for participants in the NON-MINIMAL PAIR condition after the second training block was the result of learning over the course of the experiment.³ This pattern of results is potentially compatible with two different learning processes that incorporate word-level information. One possibility is that the NON-MINIMAL PAIR 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 NON-MINIMAL PAIR group reflected perceptual learning that arose through simple exposure to the sounds, and that this perceptual learning was not apparent in the MINIMAL PAIR group because those participants learned to treat the experimental stimuli as the same based on their interchangeability in words (thus the net result of perceptual learning working against the collapse of categories was no change in performance). These results thus do not indicate whether the lexical information serves to separate overlapping categories or merge distinct acoustic categories that are used

³ A question then arises as to why participants did not show any effects of learning during the first block. One possibility is that because the *ah* and *aw* sounds were so similar, participants did not realize during the first training block that they should attend to these differences. As suggested by an anonymous reviewer, it is possible that learning performance during the first block might be improved by focusing participants' attention on vowel differences, e.g., through an alternate set of instructions.

interchangeably, but either possibility is consistent with the hypothesis that participants use information about the distribution of sounds in words to constrain their interpretation of phonetic variability. Given that distributional information can both decrease (Maye et al., 2002) and increase sensitivity to contrasts (Maye et al., 2008), both these possibilities may have contributed to the effect, such that the NON-MINIMAL PAIR corpus increased sensitivity and the MINIMAL PAIR corpus decreased sensitivity to the *tah/taw* contrast.

The results of Experiment 1 demonstrate that adults are sensitive to and can use contextual cues to constrain their interpretation of phonetic variability, assigning sounds to either one or two categories on the basis of their distribution in words. The patterns obtained here resemble those from Thiessen (2007; 2011), but show that word-level information affects listeners' treatment of a phonetic contrast even outside of a word-learning task. These findings suggest that the interactive learning strategy might be available to infants who are not yet learning the meanings of many words. Experiment 2 investigates this possibility directly by testing whether eight-month-old infants, who are learning phonetic categories in their own language, can use word context to help categorize speech sounds.

Experiment 2: Infants' sensitivity to word-level information

Infants' sensitivity to non-native phonetic contrasts decreases between six and twelve months, and they begin segmenting words from fluent speech during the same time period (Bortfeld et al., 2005; Jusczyk & Aslin, 1995; Jusczyk et al., 1999; Werker & Tees, 1984). Experiment 2 investigates whether these learning processes might interact by testing whether word-level information can affect eight-month-olds' treatment of phonetic contrasts.

Our experiment adapts the procedure used by Maye et al. (2002; Best & Jones, 1998). In their experiment, six- and eight-month-olds were familiarized with stimuli drawn from either a unimodal or a bimodal distribution. In our experiment, infants instead heard a uniform distribution of *tah* and *taw* syllables from a vowel continuum, but these syllables were embedded in the bisyllabic words *gutah/aw* and *litah/aw*. As with the adults, half the infants heard a NON-MINIMAL PAIR corpus containing either *gutah* and *litaw*, or *gutaw* and *litah*, but not both pairs of pseudowords, whereas the other half heard a MINIMAL PAIR corpus containing all four pseudowords. During test, infants were exposed to alternating trials, in which the continuum endpoints alternated, and non-alternating trials, which involved repetition of a single stimulus. If infants can use word-level information during phonetic category learning, then participants in the NON-MINIMAL PAIR group should be more likely to discriminate the continuum endpoints than participants in the MINIMAL PAIR group. Discrimination is indicated by a difference in looking times between alternating and non-alternating test trials.

Method

Participants. Forty full-term eight-month-old infants (241-277 days; mean=260 days) participated in this experiment. Participants had no known hearing deficits and were growing up in English-speaking households. An additional four infants were tested but their data were discarded due to fussiness (3) and equipment failure (1). Families were given a toy, book, or T-shirt as compensation for their participation.

Stimuli. This experiment used the *gutah/aw* and *litah/aw* tokens (during familiarization) and the isolated syllables from the *tah/taw* continuum (during test) from Experiment 1.

Apparatus. Infants were seated on a caregiver's lap on a chair in a sound-treated testing room. The testing booth consisted of three walls, one directly in front of the infant and one to each side.

A camera was positioned in the front wall to allow an experimenter in an adjacent control room to view the infant's looking behavior through a closed circuit television system. Three lights were located in the testing booth, one in each wall, at the infant's approximate eye level. A speaker, hidden by the testing booth wall, was located directly behind each of the side lights.

Stimulus presentation was controlled by customized experimental software. An experimenter in the control room, who was blind to experimental condition and test trial type, called for trials and coded looks by pressing a mouse button. Total looking times on each test trial were recorded using the same software that controlled stimulus presentation.

Procedure. The experiment used a modified version of the Headturn Preference Procedure (Kemler Nelson et al., 1995). At the beginning of each trial, a small light in front of the infant began to blink to attract the infant's attention. When the infant oriented to that light, it was extinguished, and a light on either the right or the left side of the infant began to blink. When the infant oriented to the blinking side light, stimuli began to play from a speaker located behind the wall that contained the blinking light. On familiarization trials, the light stopped blinking and remained on for the duration of the trial (21 seconds). On test trials, the light continued to blink until the infant looked away for two seconds or the maximum trial length (20 seconds) was reached. A test trial was repeated if the infant's total looking time on that trial was less than two seconds.

Participants were assigned to either the MINIMAL PAIR or the NON-MINIMAL PAIR condition. During familiarization, participants heard 128 experimental tokens (*gutah/aw* and *litah/aw*). Unlike Experiment 1, no filler syllables were included in familiarization. All participants heard each *tah/taw* token from the continuum sixteen times, but the lexical contexts in which they heard these syllables differed across conditions. Participants in the MINIMAL

PAIR condition heard each *gutah/aw* and *litah/aw* token twice. Participants in the NON-MINIMAL PAIR condition were divided into two subconditions. Participants in the *gutah-litaw* subcondition heard the 16 *gutah* tokens and the 16 *litaw* tokens four times each. They did not hear *gutaw* or *litah*. Conversely, participants in the *gutaw-litah* subcondition heard the 16 *gutaw* tokens and the 16 *litah* tokens four times each, but did not hear *gutah* or *litaw*.

Familiarization was divided into eight trials. On each trial infants heard eight *gutah/aw* and eight *litah/aw* tokens drawn from their respective familiarization condition, separated by a 500 ms interstimulus interval. Once these stimuli began playing, they continued to play for the remainder of the 21-second trial regardless of where the infant was looking.

Following familiarization, infants heard eight test trials consisting of isolated *tah/taw* stimuli. They heard two four-trial blocks each including two alternating trials ($tah_1-taw_8-tah_1-taw_8\dots$) and two non-alternating trials ($tah_3-tah_3-tah_3-tah_3\dots$ and $taw_6-taw_6-taw_6-taw_6\dots$). The interstimulus interval was 500 ms. On these trials, sound was contingent on the infant's looking behavior. As soon as the infant looked away, speech stopped playing but the light continued to blink for two additional seconds. If the infant looked back at the light during this two-second window, speech began to play again. Total looking times to each test trial were recorded.

Results and Discussion

As described earlier, discrimination is indicated by a difference between listening times for alternating and non-alternating trials. Therefore, we examined infants' preference scores, computed by subtracting looking times on alternating trials from looking times on non-alternating trials. The average preference score per condition, per block is shown in Figure 2. This preference score was 2.2 s during the first four trials and 0.1 s during the last four trials for the NON-MINIMAL PAIR condition, and -0.8 s during the first four trials and 0.4 s during the

last four trials in the MINIMAL PAIR condition. A 2×2×2 condition (NON-MINIMAL PAIR vs. MINIMAL PAIR) × trial type (alternating vs. non-alternating) × block (1 vs. 2) mixed ANOVA with condition as a between-subjects factor revealed a significant effect of block ($F(1,38)=21.8, p<0.001$), with looking times higher in the first four test trials than the last four test trials, and a marginally significant condition × trial type × block interaction ($F(1,38)=3.59, p=0.07$). This marginal three-way interaction suggests that the effect of interest, a condition × trial type interaction, differed across the two blocks of trials. We analyzed the first four and last four test trials separately to further explore this pattern of results.

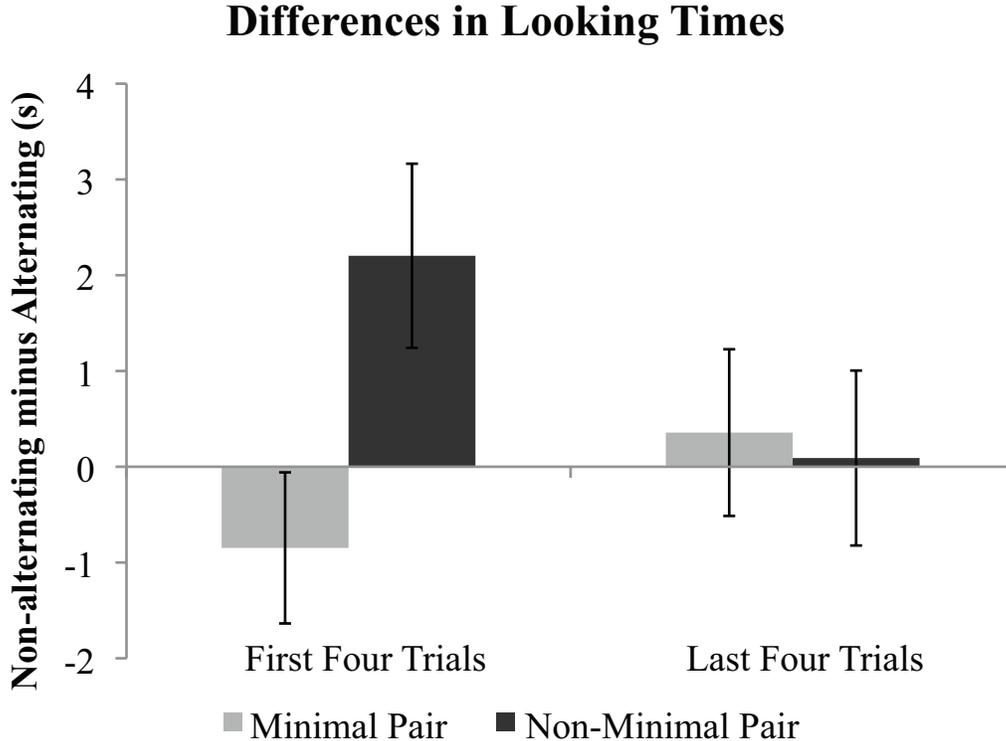


Figure 2. Differences in looking times to non-alternating and alternating test trials in Experiment 2.

Analysis of the first four test trials revealed a significant condition × trial type interaction ($F(1,38)=6.02, p=0.02$), reflecting the fact that infants’ listening preferences in the two

conditions differed at test. Infants in the NON-MINIMAL PAIR condition listened significantly longer to non-alternating stimuli ($t(19)=2.29$, $p=0.03$) whereas infants in the MINIMAL PAIR condition did not show a significant preference for alternating or non-alternating stimuli ($t(19)=1.07$, $p=0.30$). In principle the preference score near zero in the MINIMAL PAIR condition could be consistent with half the infants showing a large preference in one direction and the other half showing a large preference in the other direction; however, this did not seem to be the case, as individual infants' preference scores are also clustered near zero (Figure 3). Analysis of the last four test trials revealed no significant or marginally significant effects, suggesting that the effect of familiarization extended only to the first few test trials. Given the significant decrease in overall looking times between blocks, infants may simply have become fatigued and bored with the experiment by the second block of test trials. Note also that during test, infants heard significant numbers of *tah* and *taw* sounds in isolation, perhaps suggesting to infants that these sounds could be used interchangeably. Under an interactive account of phonetic learning, this interchangeability would actually suggest to infants that the sounds belong to the same category.

The difference between conditions in the first block of test trials provides evidence that eight-month-old infants are sensitive to word-level information and can use this information to constrain their interpretation of phonetic variability.⁴ These results provide strong support for an

⁴ As with Experiment 1, it is not known whether familiarization with the NON-MINIMAL PAIR corpus served to enhance infants' discrimination of an initially difficult contrast or whether exposure to the MINIMAL PAIR corpus blurred the distinction between sounds that were previously discriminable. To ascertain this, an additional 20 infants (242-276 days; mean=262 days; one additional infant was tested but discarded due to fussiness) were tested. During familiarization, these infants heard the sixteen tokens of pseudowords *komi* and *romu* from Experiment 1 four times each (128 tokens total), but did not hear any of the critical *gutah/aw* or *litah/aw* tokens with stimuli from vowel continuum. During the test phase, these infants heard the same alternating and non-alternating *tah/taw* trials as infants in Experiment 2. Because infants did not hear any of the *tah/taw* sounds during familiarization, their looking times during test should give a baseline measure of how easily infants discriminate these vowels. The mean preference scores were 1.2 s during the first four test trials and 0.7 s during the last four test trials. Preference scores for the first four test trials were numerically intermediate between the NON-MINIMAL PAIR and MINIMAL PAIR

interactive account of phonetic category learning. Previous research shows that infants at this age have begun to segment words from fluent speech (Bortfeld et al., 2005; Jusczyk & Aslin, 1995), and that infants' perception of non-native sounds begins to decline around this time (Polka & Werker, 1994; Werker & Tees, 1984). Our findings provide a link between these two contemporaneous learning processes by suggesting that the words infants segment from fluent speech contribute to their acquisition of phonetic categories.

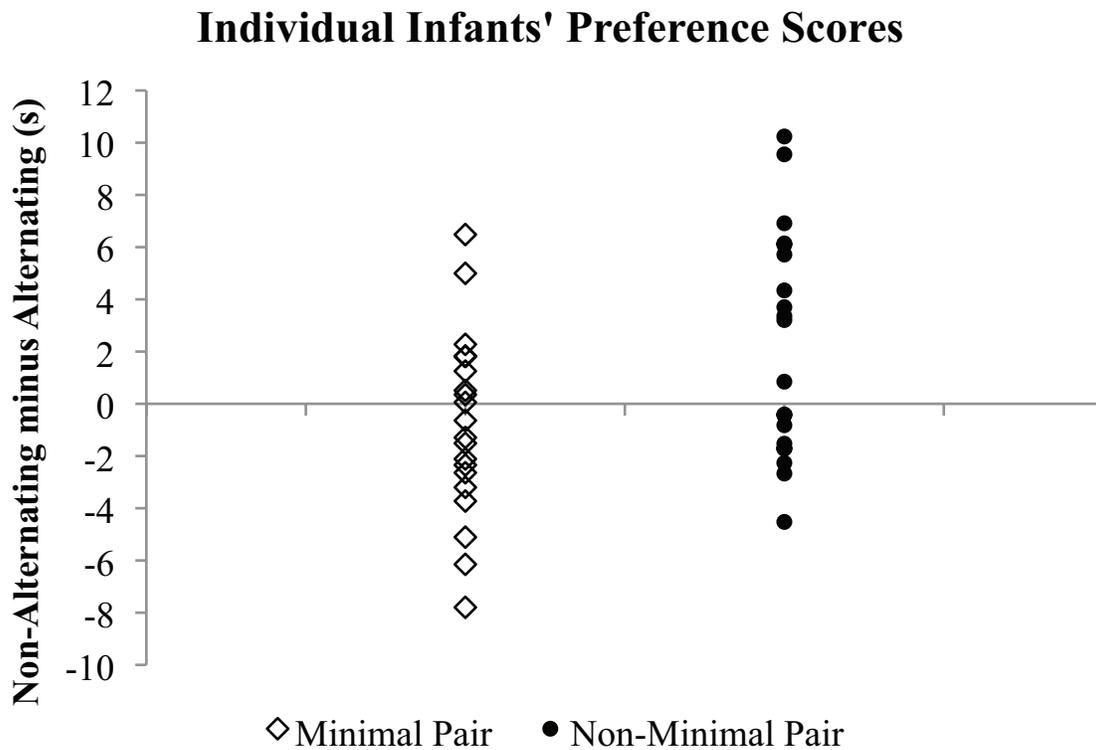


Figure 3. Individual infants' preference scores during the first block of test trials in Experiment 2.

conditions, but did not differ significantly from either of these two groups or from zero. This suggests that word-level information may either enhance or suppress phonetic perception, parallel to the bidirectional influences found in distributional learning studies (Maye et al., 2002; Maye et al., 2008).

General Discussion

Two experiments tested the hypothesis that learners use information from words when acquiring phonetic categories. The results support this interactive learning account for both adults and eight-month-old infants. In Experiment 1, adults assigned acoustically similar vowels to different categories more often when those vowels occurred consistently in different words than when they occurred interchangeably in the same set of words. Infants displayed similar behavior in Experiment 2, showing evidence of distinguishing similar vowels only after hearing those vowels occur consistently in distinct word environments. These results make three important extensions to the existing literature. They build on Thiessen's (2007, 2011) work by demonstrating that word-level information can affect learners' basic sensitivity to contrasts, in addition to affecting their use of these contrasts in word-learning tasks. They also provide empirical support for the computational analysis pursued by Feldman et al. (2009). Most crucially, they demonstrate that infants are sensitive to word-level cues while they are still in the process of learning phonetic categories, providing a plausible mechanism by which infants might acquire the vowel categories of their native language.

Infants are likely to use several types of cues when solving the phonetic category learning problem, including distributional information, semantic minimal pairs, and contextual information, including lexical context. Distributional learning (Maye et al., 2002; Maye et al., 2008) is appealing because it relies solely on acoustic characteristics of individual sounds, and acoustic information is always present when a sound is uttered. This means that infants receive distributional information for every sound they hear. Contextual information, such as visual information about articulation or likely referents (Teinonen et al., 2008; Yeung & Werker, 2009), can supplement distributional learning, but one potential problem with these types of contextual

cues is that the relevant visual information is only sometimes available. Information from the face primarily yields information about those contrasts that differ in place of articulation, and even place contrasts are not always easily observed visually. Referents are not always present when they are being spoken about. Thus, these types of visual contextual cues may not be highly reliable. Lexical contexts do not share this weakness of other contextual cues. Sounds nearly always occur as part of a word, and speech errors that substitute a sound into an incorrect word context are likely to be relatively rare. Because of this, pairings between sounds and words are likely to be more reliable than pairings between sounds and visual information.

The extent to which infants rely on each of these cues for acquiring different sound contrasts poses an interesting puzzle. For example, learners are sensitive to both distributional cues and lexical context for both consonant and vowel learning (e.g., Gulian et al., 2007; Escudero et al., 2011; Maye et al., 2002; Maye et al., 2008; Thiessen, 2007, 2011). However, the different cues may not be equally reliable across different sound contrasts. Distributional learning should be most effective for acquiring categories that have lower degrees of overlap, such as stop consonant contrasts. At the same time, computational simulations have indicated that lexical information can successfully distinguish English vowel categories, indicating that it provides a reliable cue for vowel category learning (Feldman, 2011; Feldman et al., submitted). Different cues to phonetic learning may also become available at different times in the developmental trajectory. For example, infants are likely to have access to more lexical information at later ages. Language specific influences have been observed earlier in vowel perception than in consonant perception (6-8 months for vowels vs. 10-12 months for consonants), suggesting that the availability of lexical information should be greater during the period when infants are acquiring consonant contrasts. Likewise, the availability of referential information for

distinguishing minimal pairs increases as infants learn the meanings of more words. Factors such as cue reliability and cue availability are likely to interact in complex ways to determining how and when each cue is used. Our work does not provide a solution to this puzzle, but instead underscores the importance of looking at how cues to phonetic learning interact throughout the time course of development and across different contrasts.

Infants' ability to use word contexts to distinguish acoustically similar phonetic categories also raises important questions about the way in which these word contexts interact with phonological contexts during phonological development. Phonological contexts are traditionally used to identify phonological alternations, in which the pronunciation of a sound changes depending on the identities of neighboring sounds. For example, /t/ surfaces as aspirated [t^h] when it occurs at the beginning of a stressed syllable as in *top*, but it is pronounced as unaspirated [t] when it follows a syllable-initial /s/ as in *stop*. Empirical and computational work has suggested that infants can learn phonological alternations by identifying patterns of *complementary distribution*, in which similar sounds appear consistently in distinct phonological contexts (Peperkamp, Le Calvez, Nadal, & Dupoux, 2006; White, Peperkamp, Kirk, & Morgan, 2008). Crucially, phonological and lexical contexts are confounded in linguistic input. Different phonemes appear consistently in different words, but phonological alternations also cause the *same* phoneme to consistently surface as different sounds in different phonological contexts (and thus in different words). This confound is present in natural language input and was also present in the NON-MINIMAL PAIR condition of our experiment. Participants hearing the NON-MINIMAL PAIR corpus did not have enough information to determine whether acoustic differences between *ah* and *aw* should be attributed to the differing lexical or phonological contexts. Under a lexical interpretation, the appearance of *ah* and *aw* in distinct lexical contexts

would allow learners to identify them as members of different categories. Under a phonological interpretation, however, the complementary distribution of *ah* and *aw* across phonological contexts would lead listeners to identify them as variants of a single phoneme *a*.

Learners need a strategy to distinguish between complementary distribution in phonological versus lexical contexts. One factor that infants might use to distinguish lexically driven alternations from phonological alternations is the consistency of an alternation across different words. In our experiments, all participants heard the vowel alternation in a single pair of lexical items. After hearing these lexical items, adults and infants in the NON-MINIMAL PAIR condition were more likely to assign the *tah* and *taw* sounds to different categories. Thus, it seems likely that participants in these conditions interpreted the alternations as being conditioned by lexical, rather than phonological, context. This contrasts with a previous experiment by White et al. (2008), in which infants were familiarized with phonetically motivated consonant alternations that were consistent across several novel words. Word pairs such as *bevi/pevi*, *bogu/pogu*, *dula/tula*, and *dizu/tizu* occurred either interchangeably or predictably following the words *na* and *rot*, which served as potential conditioning contexts for voicing assimilation. Infants treated voicing contrasts differently depending on whether they had heard these sounds interchangeably or only in restricted contexts, and White et al. argued that infants had chosen a phonological interpretation of these alternations. If correct, this suggests that infants' interpretation of alternations as being lexically or phonologically driven may depend on the number of lexical items that exhibit the critical alternation.

A second potential strategy for separating lexical from phonological alternations is to attend to phonetic naturalness. Cross-linguistically, some phonological alternations are more frequent than others. These frequent patterns of alternation can often be explained through factors such as

articulatory ease (e.g., Hayes & Steriade, 2004) and are often referred to as phonetically “natural” alternations. Previous research suggests that natural phonological alternations are easier to learn than arbitrary alternations (Peperkamp, Skoruppa, & Dupoux, 2006; 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 White et al. study described above, the alternations were indeed phonetically “natural”, involving voicing assimilation. In the present experiment, only one of the NON-MINIMAL PAIR subconditions presents participants with a natural alternation. In the *gutaw-litah* subcondition, the *gu* syllable with low F_2 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. The *gutah-litaw* subcondition 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. To determine whether participants were sensitive to the naturalness of the alternations they heard, we looked for differences between the *gutah-litaw* and *gutaw-litah* subconditions in each experiment. For adults in Experiment 1, 2×2 (subcondition \times block) ANOVAs showed no significant differences between the *gutah-litaw* and *gutaw-litah* subconditions and no interactions involving subcondition for any of the contrasts, suggesting that participants here were not sensitive to the naturalness of the alternation. Likewise, though we did observe somewhat larger preference scores for infants in the *gutah-litaw* subcondition, there were no significant interactions by subcondition for infants in Experiment 2. However, it remains possible that a more sensitive

design might show that infants are sensitive to phonetic naturalness when interpreting patterns of acoustic variability.

The present results demonstrate that learners are able to use word-form knowledge to constrain phonetic learning. Other work demonstrates the availability of word-form information by 6-8 months: In segmentation tasks, infants recognize similarities among different acoustic tokens of the same word (Jusczyk & Aslin, 1995) and they can also use highly familiar stored word-forms, such as their own name, to help find new words (Bortfeld et al., 2005; Mandel et al., 1995). At 11 months they appear to store frequent strings without associated meanings, as evidenced by their higher looking time toward high frequency words and non-words (Hallé & Boysson-Bardies, 1996; Ngon et al., in press; Swingley, 2005). Modeling results indicate that early form-based knowledge can potentially improve phonetic and phonological learning (Feldman et al., 2009; Martin, Dupoux, & Peperkamp, in press). Together, these findings suggest that early form-based knowledge of words has a substantial impact on phonetic learning processes. The extent to which meanings are associated with these early lexical items remains an interesting question for future research, as infants seem to associate at least some familiar words with their referents (Bergelson & Swingley, 2011; Tincoff & Jusczyk, 1999). Studying the nature of these early lexical representations can help us determine whether pure form-based knowledge is the critical factor in driving early top-down influences on phonetic learning, or whether meaning-based information might also make a significant contribution to this learning process.

We have argued that phonetic category acquisition is a richly interactive process in which information from higher-level structure, such as words, can be used to constrain the interpretation of phonetic variability. We have demonstrated that both adults and infants

interpret phonetic variation differently as a function of the word contexts in which speech sounds occur. More precisely, similar sounds that occur in distinct lexical contexts are more likely to be treated as belonging to separate phonetic categories, whereas sounds that occur interchangeably in the same contexts are more likely to be treated as belonging to the same category. This is opposite of what would be predicted if minimal pairs provide the primary cue to phonemic distinctions, but it reflects rational behavior for learners who have access to word forms but do not have access to meanings or referents. Crucially, lexical information appears to be available and used by infants early in development, supporting interactive accounts of early phonetic acquisition. Our results provide evidence that phonetic category learning can best be understood by examining the learning problem within the broader context of language acquisition.

Acknowledgments

We thank Lori Rolfe for assistance in running participants, Andy Wallace for assistance in making the vowel continuum, and Sheila Blumstein, members of the Metcalf Infant Research Lab, and two anonymous reviewers for helpful comments and suggestions. This research was supported by NSF grant BCS-0924821 and NIH grant HD032005. Portions of this work were presented at the 35th Boston University Conference on Language Development, the 3rd Northeast Computational Phonology workshop, and the 18th International Conference on Infant Studies.

References

- Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of Child Language*, 6, 183-200.
- Bergelson, E., & Swingle, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109, 3253-3258.
- Best, C. T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, 21 (Special Issue), 295.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341-345.
- Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16, 298-304.
- Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46, 217-243.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language*, 17, 205-215.
- Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighborhoods in young children's receptive vocabularies. *Journal of Child Language*, 22, 727-735.
- Coady, J. A., & Aslin, R. N. (2003). Phonological neighborhoods in the developing lexicon. *Journal of Child Language*, 30, 441-469.
- Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*, 39, 388-402.

- Curtin, S. (2009). Twelve-month-olds learn novel word-object pairings differing only in stress pattern. *Journal of Child Language*, *36*, 1157-1165.
- Dillon, B., Dunbar, E., & Idsardi, W. (in press). A single stage approach to learning phonological categories: insights from Inuktitut. *Cognitive Science*.
- Dollaghan, C. A. (1994). Children's phonological neighbourhoods: half empty or half full? *Journal of Child Language*, *21*, 257-271.
- Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *Journal of the Acoustical Society of America*, *130*, EL206-EL212.
- Feldman, N. H. (2011). *Interactions between word and speech sound categorization in language acquisition*. Unpublished doctoral dissertation, Brown University.
- Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (submitted). A role for the developing lexicon in phonetic category acquisition.
- Feldman, N. H., Griffiths, T. L., & Morgan, J. 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, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*, 103-138.
- Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*, 133-149.

- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Gulian, M., Escudero, P., & Boersma, P. (2007). Supervision hampers distributional learning of vowel contrasts. *Proceedings of the 16th International Conference on Phonetic Sciences*.
- Hallé, P. A., & Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development, 19*, 463-481.
- Hayes, B., & Steriade, D. (2004). Introduction: The phonetic bases of phonological markedness. In Hayes, B., Kirchner, R., & Steriade, D. (Eds.), *Phonetically Based Phonology*. Cambridge: Cambridge University Press.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099-3111.
- Jusczyk, P. W. & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159-207.
- Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*, 111-116.
- Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971-995.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*, F13-F21.

- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606-608.
- Labov, W. (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.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.
- Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*, 314-317.
- Martin, A., Peperkamp, S., & Dupoux, E. (in press). Learning phonemes with a proto-lexicon. *Cognitive Science*.
- Maye, J., & Gerken, L. (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, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, *11*, 122-134.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101-B111.
- McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*, 369-378.
- Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, *13*, 407-420.
- Ngon, C. A., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (in press).

- (Non)words, (non)words, (non)words: Evidence for a proto-lexicon during the first year of life. *Developmental Science*.
- Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America*, *102*, 3742-3753.
- Peperkamp, S., Le Calvez, R., Nadal, J.-P., & Dupoux, E. (2006). The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition*, *101*, B31-B41.
- Peperkamp, S., Skoruppa, K., & Dupoux, E. (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, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 421-435.
- Rasmussen, C. E. (2000). The infinite Gaussian mixture model. *Advances in Neural Information Processing Systems 12*, 554-560.
- Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, *69*, 309-320.
- Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*, 6038-6043.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception

- than in word-learning tasks. *Nature*, 388, 381-382.
- Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8, 432-443.
- Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B*, 364, 3617-3632.
- Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99-132.
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, 108, 850-855.
- Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, 56, 16-34.
- Thiessen, E. D. (2011). When variability matters more than meaning: The effect of lexical forms on use of phonemic contrasts. *Developmental Psychology*, 47, 1448-1458.
- Thiessen, E. D., & Yee, M. N. (2010). Dogs, bogs, labs, and lads: What phonemic generalizations indicate about the nature of children's early word-form representations. *Child Development*, 81, 1287-1303.
- Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10, 172-175.
- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, 104, 13273-13278.
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, 34, 1289-

1309.

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

White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, 107, 238-265.

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

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, 30, 553-566.

Yeung, H. H., & Werker, J. 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.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, 15, 420-433.