Referential Labeling Can Facilitate Phonetic Learning in Infancy

H. Henny Yeung, Lawrence M. Chen, and Janet F. Werker
University of British Columbia

All languages employ certain phonetic contrasts when distinguishing words. Infant speech perception is rapidly attuned to these contrasts before many words are learned, thus phonetic attunement is thought to proceed independently of lexical and referential knowledge. Here, evidence to the contrary is provided. Ninety-eight 9-month-old English-learning infants were trained to perceive a non-native Cantonese tone contrast. Two object–tone audiovisual pairings were consistently presented, which highlighted the target contrast (Object A with Tone X; Object B with Tone Y). Tone discrimination was then assessed. Results showed improved tone discrimination if object–tone pairings were perceived as being referential word labels, although this effect was modulated by vocabulary size. Results suggest how lexical and referential knowledge could play a role in phonetic attunement.

For many types of perceptual stimuli (i.e., speech, faces, musical rhythms, etc.), human infants quickly attune to relevant perceptual differences, and learn to ignore irrelevant ones (Scott, Pascalis, & Nelson, 2007). Phonetic perception offers a classic example of attunement, as experience with the native language begins affecting the perception of phonetic contrasts from the 1st year of life (e.g., Best, McRoberts, & Goodell, 2001; Narayan, Werker, & Beddor, 2010; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005; Sato, Kato, & Mazuka, 2012). A classic developmental pattern is maintenance for the discrimination of native contrasts, and a decline for the discrimination of non-native contrasts (Werker & Tees, 1984).

One reason that phonetic attunement is interesting is because its acquisition describes a learning problem. Early thinkers (e.g., Jakobson, 1942/1968) suggested that a language’s unique set of phonetic contrasts could be learned by comparing minimal pairs of words (e.g., the pair of words bat and pat minimally differ in the first consonant). However, such a mechanism is seen as implausible in infancy, as learners would require sufficient numbers of phonological neighbors—words with similar phonological forms—to compare with one another to deduce native phonetic categories, and this has not yet occurred by the end of the 1st year (e.g., Coady & Aslin, 2003). How then can phonetic attunement occur if infants do not have a sufficiently sizable lexicon?

Mechanisms of Phonetic Attunement

One approach considers how native language sound patterns can be deduced from the statistical frequency of phonetic features in acoustic space (Gauthier, Shi, & Xu, 2007; Jusczyk, 1993; Kuhl, 1993; McMurray, Aslin, & Toscano, 2009; Salminen, Tiitinen, & May, 2009; Vallabha, McClelland, Pons, Werker, & Amano, 2007; Werker et al., 2007). Maye and colleagues tested this hypothesis in the laboratory by presenting 6- to 8-month-olds with phonetic tokens that varied along a single acoustic dimension for just a few minutes. Tokens were distributed around either a single mode (the unimodal group), or two separate modes (the bimodal group), simulating exposure to languages that had one or two phonetic categories defined by this dimension. Only the bimodal group discriminated the relevant contrast, demonstrating distributional learning from small amounts of speech exposure (Maye, Werker, & Gerken, 2002), an effect now replicated several times (Cristia, 2011; Cristia, McGuire, Seidl, &
Recent work has suggested, however, that statistical information from frequency distributions alone is not sufficient to explain phonetic attunement (e.g., McMurray et al., 2009; Werker, Yeung, & Yoshida, 2012). Supplementary mechanisms may conspire with statistical information. For example, phonetic learning is enhanced when speech input comes from “social” scenarios involving live experimenters (Kuhl, Tsao, & Liu, 2003). Distributional learning may also be supplemented by redundant visual cues about lip movements (Teinonen, Aslin, Alku, & Csibra, 2008). Moreover, other kinds of co-occurring cues, like visual objects (Hayes-Harb, 2007), may conspire with statistical information in supplementary ways, perhaps helping infants recognize two categories when statistical cues are muddled. For example, infants may frequently hear speech of one form (e.g., doll) co-occurring with one kind of object, and speech of another form (e.g., ball) co-occurring with another. Phonetic learning from this kind of supplementary visual cue is the focus of this study.

A recent report demonstrated infant learning from these sorts of visual co-occurrences: Nine-month-olds’ discrimination of a non-native speech contrast was improved when one type of speech sound was consistently paired with one object, and the other type was consistently paired with a different object (Yeung & Werker, 2009). This phenomenon was attributed to a learning principle, termed acquired distinctiveness (Kluender, Lotto, Holt, & Bloedel, 1998; Lawrence, 1949; Thiessen, 2011), such that when Token A is consistently heard in Context B (e.g., co-occurring with one object), and Token A’ is consistently heard in Context C (e.g., co-occurring with another object), then the perceptual distance between Tokens A and A’ is increased.

Phonetic learning from acquired distinctiveness has some similarities to the earlier described minimal pair theory of phonetic learning (i.e., maintaining a /b/-/p/ distinction because bat and pat are contrastive words). However, two important features of acquired distinctiveness circumvent the learning problem inherent to the minimal pair theory. First, acquired distinctiveness is a basic, domain-general learning principle that is likely available to infants from early in life. While young infants may not have words for concepts like bat (the animal) or pat (the gesture), they may still hear these speech sounds in different environmental contexts. Second, acquired distinctiveness draws from a potentially richer data set than just the number of known words. Indeed, distinctive contexts come not only from co-occurring visual cues (Hayes-Harb, 2007; Yeung & Werker, 2009), but also from the auditory contexts (i.e., lexical forms that infants may have already segmented from the speech stream) in which phonetic tokens occur (Feldman, Myers, White, Griffiths, & Morgan, 2013; Swingley, 2009).

The Role of the Lexicon

The mechanisms of infant phonetic learning that are described above suggest that the lexicon may indeed be important for phonetic category acquisition, but only insofar as it structures the statistical aspects of phonetic input, and the contexts in which input is given. Nevertheless, some researchers still hypothesize ways in which the actual lexical knowledge of an infant can act upon phonetic category learning. Specifically, certain kinds of lexical or referential cues may increase the efficacy of phonetic learning (Heitner, 2004; Werker et al., 2012).

Consider a closely related perceptual categorization study, where adults learned a visual distinction between two kinds of “Greeble,” an artificially created set of alien-like figures (Lupyan, Rakison, & McClelland, 2007). The two kinds were learned more easily when cued by a lexical contrast (i.e., knowing a different label for each kind of Greeble), than when cued by a nonlexical contrast (i.e., knowing the two locations where each kind of Greeble lived). Top-down information about the status of a “label” may have deepened perceptual learning of the Greeble categories through some feedback process not available in the nonlexical condition (the “label feedback hypothesis,” Lupyan, 2012).

Visual categorization in infancy is similarly affected, as hearing linguistic stimuli promotes object categorization in ways that hearing nonlinguistic stimuli does not (e.g., Ferry, Hespos, & Waxman, 2010). Here, we ask whether a similar kind of “label feedback” can influence phonetic learning as well. That is, does the lexical status of a label affect categorization, not just of visual object categories but also of phonetic information in the labels themselves (see also Heitner, 2004)? Acquired distinctiveness contexts offer an elegant way to test this question. In such contexts, are infants better at differentiating two kinds of speech sounds (paired with two objects) when these sounds are known to be lexically contrastive (i.e., when the sounds...
represent two labels) compared to when they are nonlexically contrastive?

Two experiments are reported, both of which adapt the acquired distinctiveness training procedure from Yeung and Werker (2009). Nine-month-old infants in that previous report were trained to discriminate a non-native phonetic contrast using two consistent object–sound pairings. Phonetic learning (i.e., improved discrimination of the target contrast) was seen when consistent pairings were trained, compared to when no training or inconsistent training was given. The present Experiment 1 used a similar procedure to again train infants, but this time on an even harder non-native contrast. Experiment 2 then asked whether phonetic learning could be facilitated when object–sound pairings were presented in a more referential manner (i.e., when pairings were interpreted as word labels, rather than just co-occurrences).

Experiment 1

Experiment 1 probed the role of acquired distinctiveness contexts (i.e., the co-occurrence of two different speech sounds with two visually distinct objects) in the establishment of the native language phoneme repertoire. Two differences from the original Yeung and Werker (2009) study were introduced.

The first difference was that a new phonetic contrast was tested. Yeung and Werker (2009) tested English-learning infants’ discrimination of two /d/–like phones: a dental [ɻ] and a retroflex [ɖ], which differ along a familiar phonetic dimension (i.e., place of articulation, or where in the vocal tract a consonant is produced). English normally has a single intermediate category (alveolar /d/) in this part of the dimension, and thus the dental–retroflex contrast was one that partitioned the native /d/ into a slightly more fronted /d/ category (a dental place of articulation), or a slightly more back /d/ category (a retroflex place of articulation). Thus, acquired distinctiveness in Yeung and Werker (2009) was shown to create new categories along a familiar phonetic dimension (rather than bring attention to a contrast differentiated along a novel phonetic dimension). Moreover, the perception of this dental–retroflex contrast in English-learning infants is in the midst of decline at 9 months of age (Werker & Tees, 1984), and it was not clear if acquired distinctiveness could do more than simply enhance an already weakly discriminable contrast. For these two reasons, the target phonetic contrast was changed.

Here, we chose to use Cantonese lexical tones, which are cued primarily by pitch variations within a single syllable (Khouw & Ciocca, 2007). Tone represents a stronger test of perceptual learning for the two reasons above. Tone distinctions are based on a phonetic dimension (pitch) that is not, by itself, lexically relevant in English. Moreover, tone perception is more entrenched than dental–retroflex contrasts at this age: English-learning 9-month-olds consistently fail to discriminate tones (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008; Yeung, Chen, & Werker, 2013), and further ignore pitch variation in the recognition of segmented words (Singh, White, & Morgan, 2008). Thus, evidence of a newly learned tone contrast would constitute a clearer, more unambiguous case of learning (not just enhancement) in phonetic perception.

The second difference from the acquired distinctiveness procedure used by Yeung and Werker (2009) was a slight, but theoretically important modification. In the earlier work, the co-occurrence of speech sounds and objects was purposely controlled: The presentation of the syllables was synchronized with the onset of object movement. This tight coupling of movement and speech sounds can occur in actual speech input from parents (Gogate, Bahrick, & Watson, 2000), and facilitates associative learning (Gogate & Bahrick, 2001). However, real word labels do not always occur in perfect temporal synchrony with their referents, and such synchrony can often promote a different interpretation of the speech sound (as an “amodal” feature of the object). To eliminate these complications in the present work, the co-occurrence of the tones and the objects was maintained, but synchrony in the timing of presentation was eliminated.

Method

Participants

Infants were healthy, full-term infants around 9 months of age recruited from a database of parents who expressed interest in research studies (primarily recruited at local maternity wards). Infants were exposed to English at least 80% of the time in their environment, and heard no other tone languages. Twenty infants were tested (eight boys, M = 270 days, range = 260–282 days). Nine additional infants were not included for the following reasons: excessive fussiness (n = 6) and experimenter or equipment error (n = 3).
Stimuli

Target auditory stimuli were identical to those used in another report, in which 9-month-old English-learning infants failed to discriminate a rising versus level tone contrast, even though English-learning 4-month-olds, and Chinese-learning 4- and 9-month-olds succeeded (Yeung et al., 2013). The tones used are native to Cantonese, and it is common to notate these tones with numeric sequences describing the starting and ending pitch on a subjective scale from 1 to 5 (higher numbers = higher pitch). Adult-directed consonant–vowel syllables, “chee” (/ʨʰi/), were spoken by a female native Cantonese speaker with either a high-rising “Tone 25” (此 this; 始 start), or a midlevel “Tone 33” (次 next; 刺 thorn). Four tokens of each type were recorded and well matched in f0 range, duration, and amplitude (Figure 1a; see Yeung et al., 2013, for further details).

Visual stimuli during the training phase consisted of a set of animated cartoon objects (Figure 1b), which were displayed while infants heard one of the tone stimuli. The visual stimulus during the test phase was a static checkerboard (Figure 1b).

Procedure

Infants sat on their parent’s lap approximately 36 in. from a 27-in. screen placed in a sound-attenuated and dimly lit room. Parents listened to music over headphones, and were instructed not to speak or point at the screen. A video camera allowed the experimenter to see the infant from a separate room. Sound was presented at about 65 dB, and each trial consisted of stimuli presentation until the infant looked away for 2 s. An attractive silent animation was then presented, and the next trial began when the infants refixated the screen. Stimuli presentation was controlled using Habit X software (http://habit.cmb.ucdavis.edu/) running on OS X on a Mac Pro (Apple Inc., Cupertino, CA).

Training trials consisted of an animation displaying a rotating object moving horizontally on the screen for 7 s. Five phonetic tokens from one tone category (four unique) were also presented with a variable interstimulus interval (ISI) of around 850 ms, which prevented audiovisual synchronization. The next trial started immediately, featured a different object moving in the same manner, and presented with five tokens (four unique) from the other tone category. Trials continued until 2 min of looking was accumulated.

The test phase followed training, and consisted of a modified alternating/nonalternating discrimination procedure. The first trial contained a single checkerboard as a visual stimulus, presented in silence to give infants a chance to process the new visual stimulus after the training phase. Eight trials followed, all containing the same visual checkerboard. Nonalternating (Non-Alt) trials contained all the unique tokens of one tone type, whereas alternating (Alt) trials also contained four unique tokens, ordered such that one tone type always followed a contrasting type for the first four stimulus presentations, and then both tone types occurred randomly. Maximum looking to each trial was set to 30 s, and all auditory stimuli were presented with a 1-s interstimulus interval.

The logic of this procedure is that looking times will be different for Alt and Non-Alt trials if the phonetic contrast is discriminable, and previous work suggested that longer looking to the Alt trial
type is seen when discriminating this (or similar) tone contrasts (Mattock et al., 2008; Yeung et al., 2013). Alt trials were always followed by Non-Alt trials, and vice versa. Altogether, four Alt trials and four Non-Alt trials were presented, counterbalancing which trial type occurred first, as well as what order the two kinds of Non-Alt trials occurred (see Yeung et al., 2013, for details). Trained observers blind to the auditory stimuli coded looking offline for statistical analysis.

**Results**

Results showed no evidence of tone discrimination, as looking in the test phase to Alt trials was not different from looking to Non-Alt trials, \( t(19) = .42, p = .82 \) (Figure 2).

**Discussion**

Infants in Experiment 1 failed to show tone discrimination, even after being trained on consistent object–tone pairings that highlighted the contrast. This differs from the results reported by Yeung and Werker (2009), in which a dental–retroflex consonant contrast was tested. It is not clear if this failure to discriminate stems from the procedural changes made in this study—particularly the desynchronization of object and sound—or from the use of a tonal contrast rather than a consonant distinction. Nevertheless, this result allowed us to directly test whether phonetic learning could be improved if object–tone pairings were presented in a more referential context. Experiment 2 used the same procedure as Experiment 1, but manipulated the referential interpretation of the training.

Previous work from the word mapping literature provided inspiration for the experimental manipulation (Fennell & Waxman, 2010). Although infants at 14 months of age can discriminate many phonetic contrasts, they have difficulty using this perceptual knowledge in word mapping. For example, even though infants successfully map two phonetically distinct forms (e.g., *lif* with Object A; *neem* with Object B), they have difficulty mapping phonetically similar forms (e.g., *bin* with Object A; *din* with Object B; see Stager & Werker, 1997). Fennell and Waxman (2010) found that they could make this task more “referential,” and improve mapping abilities. They did this by first presenting infants with three trials containing familiar object–label pairings (e.g., *car* with an image of a car). They argued that this provided a referential context because infants knew that they would be seeing images of objects and hearing their correct, referential labels. The recognition of this referential context was considered helpful to infants, as they better accessed phonetic detail when learning novel word mappings.

![Figure 2](image_url)
In Experiment 2, a similar manipulation was used to test 9-month-old infants’ sensitivity to referential contexts, but unlike Fennell and Waxman (2010), the focus was not on word mapping. Instead, this study assessed whether the referential manipulation introduced by Fennell and Waxman could enhance phonetic learning of a non-native contrast. If infants were indeed sensitive to the referential value of the manipulation, it should facilitate phonetic learning.

**Experiment 2**

Two conditions were tested. In the nonreferential condition, a group of infants participated in the same procedure as Experiment 1. In the referential condition, another group of infants did the same task, except three trials containing familiar object labels preceded the training phase (see Fennell & Waxman, 2010). If this referential distinction was meaningful to infants, it was predicted that improved phonetic discrimination—as measured in the test phase—would be seen in the referential condition (where infants are trained on two-word labels) compared to the nonreferential condition (where infants are trained on two co-occurring sounds).

Because our hypotheses were related to the lexical knowledge of infants, and because 9-month-olds in our sample were younger (and likely had smaller vocabularies) than the 14-month-olds tested in Fennell and Waxman (2010), a measure of vocabulary size was also recorded. The short form of the MacArthur Communicative Development Inventories (CDI) was used, as it offers an easy-to-complete checklist of common words that infants may comprehend or produce (Fenson et al., 2000). Here, we modified the short form to include the familiar object labels used in the experimental procedure.

**Method**

**Participants**

Infants were recruited as in Experiment 1. The nonreferential condition included 40 infants (19 boys, $M = 273$ days, range = 255–302 days). Twelve additional infants were not included for the following reasons: hearing a tone language ($n = 2$), excessive fussiness ($n = 7$), and experimenter/equipment error ($n = 3$). The referential condition included 38 infants (20 boys, $M = 270$ days, range = 256–294 days). Seventeen additional infants were not included for the following reasons: excessive fussiness ($n = 10$), experimenter or equipment error ($n = 5$), and being out of camera range ($n = 2$).

**Stimuli**

Auditory and visual stimuli were identical to those in Experiment 1, except that four auditory tokens each of *banana*, *car*, and *keys* were recorded from the same speaker who produced the tone stimuli. These words were chosen because they are highly frequent for infants, and because animated cartoon stimuli similar to the novel objects in Experiment 1 were also available for these items (Figure 1b).

**Procedure**

The procedure was identical to Experiment 1 in both conditions, except that in the referential condition, three pretraining trials were shown. Here, animations of the familiar objects (*banana*, *car*, and *keys*) were paired with five tokens (four unique) of their corresponding labels. Pretraining trials were presented for 14 s each, but not in an infant-controlled manner, ensuring that these trials were not inadvertently cut short by looks away. All parents also completed our modified version of the MacArthur CDI (Fenson et al., 2000), which consisted of the standard short form, altered such that *banana*, *car*, and *keys* all appeared.

**Results**

A preliminary analysis tested for differences in attention during the training phase between conditions. We measured how many infant-controlled trials were needed to accumulate this preset looking criterion. This is similar to examining total looking time during habituation to control for attentional differences across conditions (Cohen, Deloache, & Rissman, 1975), but because the total amount of looking was fixed to 2 min, the only freely varying parameter was the number of trials needed to reach the criterion. An analysis of this variable showed no difference between experimental conditions, $M_{\text{nonreferential}} = 21.10$, $SD_{\text{nonreferential}} = 2.60$, $M_{\text{referential}} = 20.26$, $SD_{\text{referential}} = 2.31$, $t(76) = 1.50$, $p = .14$.

**Overall Patterns**

Was phonetic discrimination of the tone contrast (as measured in the test phase) improved in the ref-
ential condition relative to the nonreferential one? A mixed analysis of variance (ANOVA) with a between-subjects factor of condition (nonreferential or referential) and a within-subjects factor of trial type (Alt or Non-Alt trials) revealed no significant main effects or interactions (as with all ANOVA comparisons here, alpha = .05). This suggested neither an overall effect of phonetic discrimination nor any differences across conditions.

To test our hypothesis that the lexical knowledge of young infants might interact with phonetic learning, two additional analyses were conducted. First, parental reports of their infants’ knowledge of the pretraining items (i.e., banana, car, or keys) from the referential condition) were examined. It was hypothesized that infants might only benefit from the referential manipulation if they knew at least one of the words presented in pretraining phase (see also Fennell & Waxman, 2010). Second, the overall vocabulary scores of infants were examined. It was hypothesized that better word learners should also be better at understanding referential contexts in general, irrespective of knowing the specific pretraining words. One possibility is that simply seeing a familiar object on the screen and hearing a label may have cued certain individuals (i.e., lexically savvy infants) to the fact that they were in a “word learning” context. Both analyses are described separately below.

Comprehension of Pretraining Words

A preliminary analysis examined the number of pretraining words that infants were reported to have known from the CDI. Slightly less than half of the parents in our total sample reported knowledge of banana, car, or keys (n = 31 reported that their infant knew one word, n = 1 knew two words, and n = 2 knew all three words). The proportion of infants knowing at least one pretraining word did not differ between groups, M_{nonreferential} = .38, SD_{nonreferential} = .49, M_{referential} = .50, SD_{referential} = .51, t(76) = 1.12, p = .27.

An ANOVA with between-subjects factors of pretraining (understood no pretraining words, or understood at least one pretraining word) and condition, as well as a within-subjects factor of trial type showed no significant effects. Results thus suggested that parental reports of whether their infants understood any of the pretraining words did not bear on tone discrimination, and this did not significantly change according to the condition in which infants were tested.

Overall Vocabulary

A preliminary analysis examined overall CDI scores. These scores did not significantly differ between groups, M_{nonreferential} = 8.10, SD_{nonreferential} = 6.61, M_{referential} = 12.31, SD_{referential} = 15.07, t(77) = 1.61, p = .11, but differences in means and variances between conditions were sizable. A further analysis showed that these differences stemmed from a highly non-normal, and positively skewed, distribution of raw scores (common with checklists; see also Fenson et al., 2000), and in neither condition was there a normal distribution, S-W_{nonreferential}(40) = .91, p = .004, S-W_{referential}(38) = .73, p < .001. Three different analyses were conducted to correct for this problem.

A categorical analysis. The first method was a categorical analysis, which assigned infants to groups based on the overall median score (seven words), which was the same median within both experimental conditions. Infants with CDI scores below the median were counted as having low vocabularies, whereas the others were counted as having high vocabularies. A preliminary ANOVA on the number of trials needed to complete the training phase with between-subjects factors of vocabulary (low or high vocabulary) and condition revealed no significant effects.

The critical ANOVA examined looking times during the test phase with between-subjects factors of vocabulary and condition, and a within-subjects factor of trial type. This showed only a significant three-way interaction, F(1, 74) = 12.22, p = .029, η² = .063, such that tone discrimination differed as a function of vocabulary group and condition. Follow-up two-way ANOVAs were conducted within each condition.

In the nonreferential condition, no significant effects or interactions were found. In the referential condition, only the interaction between vocabulary and trial type was significant, F(1, 36) = 5.56, p = .024, η² = .13, indicating that looking times to Alt and Non-Alt trials differed in low- versus high-vocabulary groups. Planned comparisons within each level of vocabulary (Bonferroni-corrected, pooled variance) showed that looking times to the two types of test trials were not different in the low-vocabulary group, t(37) = .64, p = .53, but that infants looked significantly longer at Alt compared to Non-Alt trials in the high-vocabulary group, t(37) = 2.55, p = .015, d = .41. Results thus showed that only high-vocabulary infants succeeded at tone discrimination after referential training, but low-vocabulary infants did not (Figure 2).
Removing outliers. The second method was to remove outliers. A cutoff of 20 on the CDI was chosen because it was the lowest threshold at which no outliers were present (i.e., all scores were within the 1.5 interquartile ranges). As shown in Figures 3a and 3b, outlier-removed CDI scores did not differ between groups, \( M_{\text{nonreferential}} = 7.54, \) \( SD_{\text{nonreferential}} = 5.65, \) \( M_{\text{referential}} = 6.73, \) \( SD_{\text{referential}} = 4.60, \) \( t(70) = .66, \) \( p = .51. \)

A hierarchical multiple regression was conducted on “discrimination scores” from the test phase (i.e., looking to Alt trials minus looking to Non-Alt trials), using outlier vocabulary (outlier-removed CDI scores), condition, and their interaction as predictors. A reduced model including just outlier vocabulary and condition was marginally improved by addition of their interaction, \( F(1, 68) = 3.16, \) \( p = .080, \) change in \( R^2 = .043, \) although the complete model remained nonsignificant, \( F(3, 68) = 1.76, \) \( p = .16. \) These results are weak, but suggest that the interaction between vocabulary scores and referential context could explain some of the variance in discrimination scores beyond the portion explained by vocabulary or referential context alone. As seen in Figures 3a and 3b, discrimination scores were not correlated with outlier-removed CDI scores in the nonreferential condition, \( r(39) = .001, \) \( p = .99, \) but these variables were marginally correlated in the referential condition, \( r(33) = .30, \) \( p = .088. \)

Normalizing vocabulary scores. The third method was to transform CDI scores. In this analysis, a \( \ln(x + c) \) transform was applied: \( \ln \) was the natural log, \( x \) was the raw CDI score, and \( c \) was a constant. As several infants had CDI scores of zero, a constant had to be added (the constant 2 was the lowest integer that resulted in a normal distribution of scores, \( S-W(78) = .98, \) \( p = .15, \) but identical statistical patterns are found with either 1 or 2). As shown in Figures 3c and 3d, log-transformed scores did
not differ between groups, \( M_{\text{nonreferential}} = 2.08, SD_{\text{nonreferential}} = .73, M_{\text{referential}} = 2.23, SD_{\text{referential}} = .77, t(76) = .90, p = .37. \)

Another hierarchical multiple regression was conducted on discrimination scores from the test phase, using log vocabulary (log-transformed CDI scores), condition, and their interaction as predictors. A reduced model including just log vocabulary and condition was significantly improved by addition of their interaction, \( F(1, 74) = 4.95, p = .029, \) change in \( R^2 = .060, \) and the complete model was significant, \( F(3, 74) = 2.22, p = .038, R^2 = .11. \) This analysis more clearly shows that the interaction between vocabulary size and referential context explains a unique portion of the variance in discrimination scores beyond the portion explained by vocabulary or referential context alone. As seen in Figures 3c and 3d, discrimination scores were not correlated with log-transformed CDI scores in the nonreferential condition, \( r(40) = -.051, p = .75, \) but were significantly correlated in the referential condition, \( r(38) = .37, p = .022. \)

**Discussion**

Results suggest that phonetic learning (as assessed by tone discrimination) was affected by both the referential context of training and infants’ overall vocabulary scores. This was shown in three analyses, correcting for a non-normal and positively skewed distribution of vocabulary scores. First, a categorical analysis showed that infants receiving the referential training had improved tone discrimination—but only if they also had high vocabularies (i.e., scores at the median or higher)—whereas infants in the nonreferential condition showed no learning, regardless of vocabulary size (Figure 2). The second and third analyses used regression to show that an individual infants’ success at tone discrimination was predicted by vocabulary size (using either raw or normalized scores), but only in the referential condition (Figure 3). Together, these results provide convincing evidence that tone discrimination is preferentially improved (a) when a referential context helps to indicate that two object labels are distinguished by tones, and (b) when infants have greater overall vocabulary sizes. The effects of referential context and vocabulary are discussed separately below.

**Nonreferential Versus Referential Contexts**

Previous findings show that the perceptual distinctiveness of speech sounds is enhanced in acquired distinctiveness contexts (Yeung & Werker, 2009). The current results indicate that the nature of these contexts matters. Specifically, 9-month-old infants—at least those with high vocabularies—paid special attention to the acoustic properties that distinguished two (linguistically important) sounds when those sounds were potential word labels. Infants did not differentiate these same acoustic properties when they simply co-occurred with the two objects. Theoretically, this finding is closely related to the notion that words play a special role in both adult (Lupyan, 2012) and infant (Ferry et al., 2010) categorization. In that literature, visual features of object categories are learned more easily when referential cues (i.e., labels) indicate category membership, compared to when nonreferential cues (i.e., tones, spatial positions) indicate the same. Here, we similarly show that the acoustic features of the labels themselves are learned more easily when referential cues (i.e., contrasting word labels) indicate category membership, compared to when nonreferential cues (i.e., co-occurring visual contexts) indicate the same.

It remains important to note that the current work is different from—although perhaps related to—the literature on word learning. In that literature, infants’ ability to map phonetic forms onto novel object referents is difficult before 13–14 months of age (e.g., Hollich et al., 2000; Werker, Cohen, Lloyd, Casasola, & Stager, 1998), particularly when the phonetic forms to be mapped are similar (Stager & Werker, 1997; although see Fennell & Waxman, 2010). In this study, it is therefore unlikely that 9-month-old infants are learning new lexical entries (successfully mapping novel word forms onto their referents). Results suggest instead that some infants—those with high vocabularies—are better able to identify and compare the phonetic content of speech sounds when those sounds are possible word labels. These infants then used this newly acquired phonetic knowledge to discriminate a perceptual pattern normally considered to be non-native in their language environment.

**Effects of Overall Vocabulary Size**

The above discussion raises another question: Why did vocabulary size interact with the referential nature of the training? One hypothesis is that having a larger vocabulary size increased the chance that infants had banana, keys, and/or car in their lexicon, which might have been a prerequisite for recognizing the referential context. This hypothesis
was not supported because the reported comprehension of these individual items did not predict discrimination performance. However, even though CDIs are quite accurate at the item level by at least 20 months of age (Ring & Fenson, 2000), it remains possible that parental reports do not accurately reflect infants’ specific knowledge of individual words at 9 months of age. Without any independent means of verifying the comprehension of these three particular words, it is thus difficult to definitively exclude the possibility that overall vocabulary was actually a better estimate of whether an infant really understood banana, keys, or car than CDI reports of those specific items.

We consider another hypothesis to be more likely, however. It may be the case that specific knowledge of the three pretraining words was not the true source of infants’ referential understanding in the critical condition. Rather, simply seeing familiar objects and hearing an unknown label bears striking similarity to other ostensive scenarios that infants may have previously experienced. Consider book-reading scenarios, where infants are commonly shown familiar objects whose labels they may not yet have learned. Here, they are repeatedly presented with the corresponding word label (e.g., books showing familiar objects beginning with letters A–Z). Familiarity with activities like these may just be enough to help linguistically savvy 9-month-olds recognize the referential condition as a word learning “game,” where (phonetically contrastive) word forms are heard. This may be where the crucial difference between the referential and non-referential conditions lies because in the latter condition the referential nature of the object–tone pairings was not as transparent.

This latter hypothesis entails two distinct possibilities that further explain the interaction between the experimental condition and vocabulary size. The first possibility is that two independent factors acted to improve phonetic learning, pushing the recognition of distinct tone categories just above threshold only in the referential condition. One factor is the top-down benefit from being trained with contrasting word forms in a referential context, as discussed earlier (e.g., Lupyan, 2012). The other factor is that infants who are better phonetic learners may also have had higher vocabularies, a proposal related to studies prospectively correlating language-specific phonetic sensitivity at younger ages with larger vocabularies at older ages (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Tsao, Liu, & Kuhl, 2004). As learning tone distinctions is quite difficult for English-learning infants at the age tested here, discrimination may only have been possible when both of these factors conspired to aid learning. That is, only good phonetic learners (i.e., infants with high vocabularies), who were also trained in a referential context, succeeded at the task.

The second possibility is that vocabulary size is related to infants’ abilities at recognizing the referential context. For example, high-vocabulary infants might be exposed more frequently to ostensive labeling situations like book reading, a reasonable hypothesis given correlations between vocabulary size and these sorts of infant–parent interactions (e.g., Raikes et al., 2006). Provocatively, this predicts that referential understanding (as measured by one’s overall vocabulary size) is one cause of variance in phonetic sensitivity, or at least of the variance in infants’ ability to learn new phonetic contrasts. That is, those infants with precocious lexical knowledge may be better at identifying referential contexts, and thus be better at recognizing relevant perceptual contrasts between distinct word labels. Future work will need to further explore the relation between referential abilities, vocabulary, and phonetic sensitivity, but one prediction is that some measure of lexical or referential knowledge (one more appropriate for infants under 12 months of age than parent checklists) can prospectively predict infants who are better at native phonetic contrasts (and worse at non-native phonetic contrasts). Preliminary evidence for this hypothesis comes from recent work, showing that the pace of phonetic development can be predicted by the style of mother–child interactions (Elsabbagh et al., 2013), which is then likely correlated with vocabulary size.

**General Discussion**

Young infants acquire native language phonetic patterns defined by the lexicon, but this happens long before infants have learned more than a handful of words. Statistical (i.e., distributional) mechanisms of phonetic learning have been proposed (e.g., Cristia et al., 2011; Maye et al., 2002; McMurray et al., 2009), which do not rely on an infant’s lexical knowledge. Infants may also use information from the unique (acquired distinctiveness) contexts in which phonetic tokens are heard, which could supplement statistical learning (Feldman et al., 2013; Swingley, 2009; Yeung & Werker, 2009).

A weakness of the acquired distinctiveness approach is that the notion of informative visual or
auditory contexts is unconstrained. For example, many phonetically irrelevant acoustic properties consistently appear in unique contexts that are not lexically distinct (see also the discussion of allophones in Feldman et al., 2013). In the visual domain, for example, recent work reports adult phonetic learning when talker identity distinguishes a phonetic contrast (Mani & Schneider, 2013). Other work has also shown that infants are consistently exposed to unique pitch deviations in certain caregiving contexts, which have pragmatic meanings, rather than lexical ones (e.g., Stern, Spieker, & MacKain, 1982). An unsophisticated computational model might mistakenly consider all such contexts to be distinctive, yet human infants successfully attune to phonetic contrasts that generalize over lexically irrelevant acoustic information such as talker identity (e.g., Kuhl, 1983) and pitch (e.g., Singh et al., 2008). From a theoretical point of view, acquired distinctiveness can thus only be effective if infants also recognize the uniquely referential contexts where relevant phonetic contrasts are made.

This study suggests one kind of constraint on learning from acquired distinctiveness: infants’ understanding of the way that words are used (Heitner, 2004), as observed in their referential understanding of labeling contexts. A procedure similar to that reported by Yeung and Werker (2009) was run here, where consistent pairings between distinct objects and speech sounds highlighted a difficult tone contrast. In Experiment 1, infants failed to learn the contrast from this training, showing no evidence of tone discrimination. In Experiment 2, infants succeeded at tone discrimination, but only when trained in a more referential context (i.e., it was made clearer that the object–tone pairings were labels, and not simply co-occurrences), although this effect was seen only in infants with higher vocabularies.

Two possibilities describe how referential context and vocabulary size might interact. The first is that a referential context boosts tone discrimination, as does having a high vocabulary (i.e., those infants are better phonetic learners). Together, these independent factors pushed only those high-vocabulary infants in the referential condition above threshold. The second possibility is that having a high vocabulary allows infants to more easily understand the referential manipulation. That is, having a high vocabulary is likely associated with a better understanding of referential contexts (i.e., knowing that word labels are heard when seeing objects and hearing speech). Future research will need to distinguish these possibilities, but both implicate lexical knowledge (vocabulary size) and referential knowledge (sensitivity to labeling contexts) in phonetic learning.

Word learning and phonetic attunement are rarely considered mutually influential, as it is widely assumed that the latter influences the former, and not the other way around. This assumption stems from the fact that the most active period of attunement occurs in the 1st year of life, before infants have learned many words. In conjunction with other work, however, there is increasing evidence that—in addition to statistical learning—lexical and referential factors can affect the learning of perceptual categories from early in development. For example, word labels appear to preferentially aid infants’ detection of category features (for objects) from at least 3 to 4 months of age (Ferry et al., 2010), and this study reports a similar phenomenon in the phonetic domain at 9 months of age. Moreover, infants from 6 to 12 months of age have already begun to accumulate several lexical items (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012), an indication that infants in this age range already have some knowledge of how referential labels work.

All proposed mechanisms of phonetic development, including statistical, bottom-up approaches, depend crucially on the structure of the lexicon. However, only a few previous accounts of phonetic learning (Heitner, 2004; Werker et al., 2012) had also proposed that phonetic learning might also be influenced by lexical and referential knowledge on the part of the infants themselves. The current results suggest that these proposals have some traction. However, the present findings also highlight the need for broader research that tackles two further issues.

One issue is a computational one. Is acquired distinctiveness a mechanism that depends strictly on minimal pairs, which are thought to be relatively rare in parental input? Previous work using an acquired distinctiveness approach (Feldman et al., 2013; Thiessen, 2011) has shown that infants can use a more probabilistic learning mechanism, learning phonetic contrasts (e.g., between ah [a] and aw [ɔ] vowels) after be presented with nonminimal pairs of word labels (e.g., gutah and litaw). It remains to be seen whether this kind of phonetic learning from nonminimal pairs is seen in object labeling situations, and if so, if it is similarly subject to the kinds of referential constraints observed here. A second future research topic concerns the precise relation between lexical knowledge (i.e., the number of lexical entries) and referential abilities (i.e., the
capacity to understand social and ostensive cues when using words), particularly when phonetic attunement is most active in the 1st year of life. A better understanding of this process may help interpret the mediating role that lexical knowledge had in the currently reported studies where phonetic learning was modulated by labeling objects.

References


