Mapping non-native pitch contours to meaning: Perceptual and experiential factors

Jessica F. Hay[a,*], Ryan A. Cannistraci[a], Qian Zhao[1]

[a] University of Tennessee, Knoxville, Department of Psychology, United States

ABSTRACT

Infants show interesting patterns of flexibility and constraint early in word learning. Here, we explore perceptual and experiential factors that drive associative learning of labels that differ in pitch contour. Contrary to the salience hypothesis proposed in Experiment 1, English-learning 14-month-olds failed to map acoustically distinctive level and dipping labels to novel referents, even though they discriminated the labels when no potential referents were present. Conversely, infants readily mapped the less distinctive rising and dipping labels. In Experiment 2, we found that the degree of pitch variation in labels also does not account for learning. Instead, English-learning infants only learned if one of the labels had a rising pitch contour. We argue that experience with hearing and/or producing native language prosody may lead infants to initially over-interpret the role rising pitch plays in differentiating words. Together, our findings suggest that multiple factors contribute to whether specific acoustic forms will function as candidate object labels.

At its most basic level, word learning involves mapping sounds to meaning. There are many factors that affect infants' ability to associate sounds with referents, including, but not limited to, infant characteristics (e.g., age & vocabulary size; Werker, Fennell, Corcoran, & Stager, 2002) and task characteristics (e.g., referential support; Fennell & Waxman, 2010; experimental design; Yoshida, Fennell, Swingley, & Werker, 2009). Further, even when researchers test infants of the same age and use the same basic methodologies, different patterns of results can emerge depending on characteristics of the labels used. For example, at 12 months English-learning infants reject distinct communicative vocal sounds (e.g. *ooh*, *ssh*) and sound sequences that do not conform to native-language phonotactic patterns (e.g., the Czech word *ptak*) as labels for novel objects, although, they continue to map phonotactically legal words even if they come from a different language (e.g., the Japanese words *sika* & *hashi*) (MacKenzie, Curtin, & Graham, 2012; MacKenzie, Graham, & Curtin, 2011). Further, at 14 months English-learning infants appear to have difficulty mapping labels that differ by a single consonant (e.g., minimal pairs *bih* and *dih*; Stager & Werker, 1997) but succeed in mapping labels that minimally differ only in pitch contour (e.g., rising /ku/ and falling /ku/; Hay, Graf Estes, Wang, & Saffran, 2015). Thus, it is not immediately apparent why certain sounds are mapped to referents more easily than others. In the current work, we aim to uncover why certain stimulus characteristics,

specifically pitch contour information, support the mapping of labels to meaning by young learners.

Seminal work by Stager and Werker (1997) has suggested that even when sounds are lexically contrastive (i.e., they are used to differentiate word meaning in a given language) and are easily discriminated in an object-free task (i.e., when the words are presented with a checkerboard instead of an object), 14-month-old infants have a difficult time attending to the fine acoustic features in minimal pair object labels (e.g., *bih* and *dih*). Follow-up studies have replicated this pattern of failure by 14-month-olds with numerous consonant-based minimal pairs (e.g., *bin/din, bin/pin, pin/din*; Pater, Stager, & Werker, 2004; *buk/puk*; Rost & McMurray, 2009; *daw/taw*; Thiessen, 2007) and some vowel-based minimal pairs (e.g., *deet/doot* and *dit/doot*; Curtin, Fennell, & Escudero, 2009). One account of the difficulty 14-month-olds have mapping these types of minimal pairs is that infants of this age may not have strong, rapid access to minimally-distinctive phonemic contrasts during cognitively demanding word-learning tasks (Werker & Curtin, 2005; for related evidence see Fennell, 2012; Fennell & Waxman, 2010; Yoshida et al., 2009). Consistent with this hypothesis, as infants gain language experience, they become more sensitive to phonetic differences in minimal pair words and show success in mapping them by 17–20 months of age (Werker et al., 2002).

Work examining infants' ability to map minimal-pairs that differ in

---

suprasegmental features has revealed a different pattern of performance than the work on consonant-based minimal pairs. For example, at 12 months English-learning infants are able to map words that differ in lexical stress (e.g., BEdoka vs deDOka) to novel objects (Curtin, 2009). Further, similar success in minimal pair leaning has been observed for lexical tones (e.g., Graf Estes & Hay, 2015; Hay et al., 2015; Singh, Hui, Chan, & Golinkoff, 2014). Much like changing the consonant or vowel in non-tonal languages functions to change the meaning of a word, in tonal languages (e.g., Mandarin Chinese, Thai), tones are also lexically contrastive. Lexical tones are realized somewhat differently from one tonal language to the next, but at their most basic level they are characterized by their pitch height and contour. For example, in Mandarin Chinese there are four lexical tones: Tone 1 (high-level), Tone 2 (high-rising), Tone 3 (low-dipping), and Tone 4 (high-falling) (Chao, 1948; Howie, 1976), and the same consonant-vowel (CV) sequence has a different meaning based how its pitch is realized. For example, in Mandarin Mā (Tone 1/T1) means mother, Má (Tone 2/T2) means hemp, Mǎ (Tone 3/T3) means horse, and Mà (Tone 4/T4) mean to scold.

Recent work by Hay et al. (2015) has demonstrated that at 14 months – an age at which mapping consonant-based minimal pairs is difficult – English-learning infants readily map minimal pair words that vary in pitch contour to novel objects. Using a modified version of the Switch paradigm (Werker, Cohen, Lloyd, Casasola, & Stager, 1998), monolingual English-learning infants were habituated to two novel label-object pairings. The labels were the syllable /kʊ/ produce with a rising pitch contour and /kʊ/ produced with a falling pitch contour, similar to mandarin lexical Tone 2 and Tone 4, respectively. Following habituation infants were presented with Same trials, in which the label-object pairings from habituation were maintained, and Switch trials, in which the label-object pairings from habituation were switched or violated (e.g., label A was paired with object B and vice versa). If infants notice the violation in the pairing they should look longer on Switch as compared to Same trials. In the first experiment, infants readily noticed the label-object violations, suggesting that they had learned the label-object mappings. In a subsequent study, bilingual 14-month-olds, who were not learning a tone language, also treated rising vs. falling pitch contours as lexically contrastive (Graf Estes & Hay, 2015).

Across the second half of the 2nd year, monolingual and bilingual infants who are not exposed to a tone language appear to go through a period of interpretive narrowing, whereby pitch contour differences cease to differentiate meaning. Specifically, Hay et al. (2015) demonstrated that by 17–19 months, English-learning monolinguals cease to treat rising vs. falling pitch contours as lexically contrastive in these types of associative learning tasks. Further, Graf Estes and Hay (2015) have found that bilingual infants who were not exposed to a tone language show a similar pattern of interpretive narrowing – although this interpretive change occurs 3–5 months later for bilinguals than for their monolingual peers. Thus, in contrast to the consonant-based minimal pair word learning studies where infant performance improves across development, English-learning and non-tonal language learning bilingual infants appear to become worse at mapping pitch contour-based minimal pairs to objects across the second year of life. Both of these developmental trajectories are adaptive and demonstrate that infants are homing in on the relevant features of their native language(s), while narrowing their interpretation of less relevant features, across development. One critical question that arises from these findings, however, is why do 14-month-olds, who are not learning a tone language, treat pitch contours as lexically contrastive in the first place?

Hay et al. (2015) suggest that the acoustic salience of the pitch contours may drive their early contrastive use. Specifically, the differences between pitch contour-based minimal pairs unfold over hundreds of milliseconds and, thus, may be more perceptually salient than acoustic differences between consonant-based minimal pairs (e.g., bin/pin, bin/din), which unfold over milliseconds of voice-onset-time or formant transitions. However, evidence to support the hypothesis that acoustic salience drives label-object association, early in development, is mixed. In support of the role of acoustic salience, 14-month-olds demonstrate better learning of labels with non-overlapping phonological features, such as lif and neem than less distinctive minimal pairs such as bih and dih (Stager & Werker, 1997). Further, work by Curtin et al. (2009) has demonstrated that English-learning 15-month-olds successfully mapped vowel-based minimal pairs that differed in vowel height (i.e., deet and dit), but failed to map two other vowel contrasts that did not share this distinctive feature (i.e., deet/doot and dit/doot). As previously mentioned, English-learning 12-month-olds can map minimal-pairs that differ in lexical stress, which is a suprasegmental feature that unfolds over a longer time frame and contains salient pitch information (Curtin, 2009). Additionally, a recent study by Archer and Curtin (2018), suggests that English-learning 14-month-olds rely on salient coarticulatory cues to learn minimal pairs such as bleet and breet. Finally, the salience of infant-directed speech (IDS) appears to facilitate label-object mapping compared to adult-directed speech (ADS), which contains significantly less acoustic variation (Graf Estes & Hurley, 2013; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; but see Robertson, von Hapsburg, & Hay, 2017). Thus, there is evidence that acoustically distinct labels may support early word learning.

Conversely, work by MacKenzie et al. (2011) suggest that acoustic salience is not sufficient to drive early label-object mapping – at 12 months infants fail to map both consonant-based (mmm and shhh) and vowel-based (oooh and aaah) communicative sounds to novel objects even though these labels are acoustically distinctive. They also fail to learn non-communicative, yet acoustically distinctive English consonant sounds, / l / and / ʒ /, even though they readily map novel CVC words, wug and fep. Finally, older infants fail to map a whole host of acoustically distinctive non-speech sounds to objects, even when they are provided with referential support (Hirsh-Pasek, Golinkoff, & Hollich, 2000; Namy, 2001; Woodward & Hoyne, 1999). Given these seeming contradictory findings, we do not yet have a clear understanding of the role acoustic salience plays in determining how labels are mapped to objects.

In Experiment 1, we test the hypothesis that the acoustic salience of pitch contour labels drives learning in the associative mapping task used by Hay et al. (2015), Graf Estes and Hay (2015). To test this, we took advantage of the fact that lexical tones have varying degrees of acoustic overlap (see Fig. 1) and varied the acoustic distinctiveness between the labels in our task accordingly. For example, the Mandarin rising (Tone 2) and falling (Tone 4) tones used by Hay et al. (2015), Graf Estes and Hay (2015) both begin at very a different fundamental frequency (F0; Tone 2 begins with a low F0, Tone 4 begins with a high F0) and exhibit highly dissimilar F0 trajectories. Thus, acoustically, Tones 2 and 4 are highly distinctive. Tone 1 (level) and Tone 3 (dipping) also have acoustically distinctive pitch contours – they have very dissimilar F0s both in their height (Tone 1 has a high F0, Tone 3 has a much lower F0) and trajectory (the F0 of Tone 1 remains relatively constant across the tone, whereas the F0 of Tone 3 falls and then rises). Conversely, some lexical tones are much less distinctive. For example, Tone 2 (rising) and Tone 3 (dipping) both begin in the mid frequency range and also display a similar fundamental frequency (F0) trajectory, first falling and then rising. The major difference between Tones 2 and 3 arises at the turning point of the contour, where it changes from falling to rising. Thus, of all of the lexical tones, Tone 2 (rising) and Tone 3 (dipping) are most acoustically similar.

Consistent with the idea that acoustic distinctiveness may drive perceptual salience, Tsao (2008) showed that 10- to 12-month-old Mandarin-learning infants were better at discriminating Mandarin Tone 1 (level) vs. Tone 3 (dipping), but were less accurate in discriminating acoustically similar contrasts Tone 2 (rising) vs. Tone 3 (dipping). This result was partly confirmed by So and Best (2010) across a number of different language groups (i.e., Cantonese, Japanese, and Canadian English). They found that Tone 1 (level) vs. Tone 3 (dipping), Tone 2
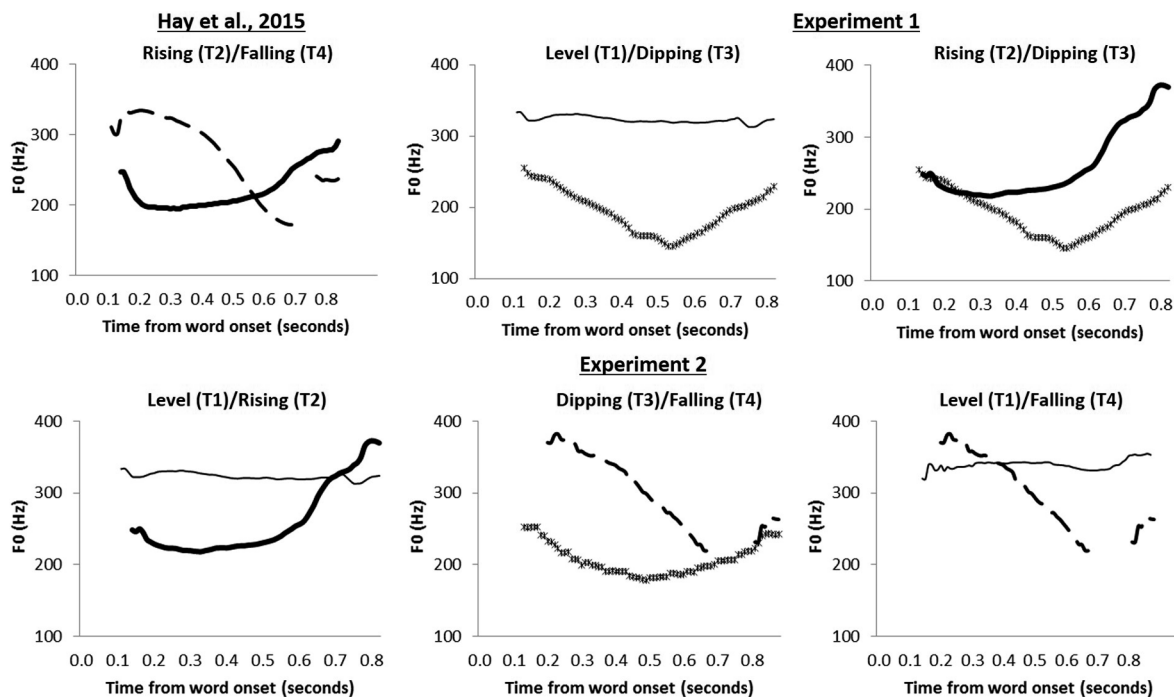
Fig. 1. F0 contours of lexical tones used in Experiments 1 and 2, plotted by condition. F0 contours used by Hay et al. (2015) are also included for comparison purposes.

(rising) vs. Tone 4 (falling), and Tone 3 (dipping) vs. Tone 4 (falling) are all more easily discriminated than Tone 1 (level) vs. Tone 2 (rising), Tone 1 (level) vs. Tone 4 (falling) and Tone 2 (rising) vs. Tone 3 (dipping). Mandarin-learning infants are also more likely to mispronounce and less likely to notice a mispronunciation in tone pairs that are less acoustically distinctive (Li & Thompson, 1977; Singh, Tan, & Wewalaarachchi, 2017).

Based on the characteristic pitch contours of Mandarin lexical tones, in Experiment 1, we used a modified version of the Switch Paradigm (Werker et al., 1998) to train one group of English-learning 14-month-olds to map acoustically salient level vs. dipping pitch contours to two novel objects and a second group to map less acoustically salient rising vs. dipping pitch contours. We expected that if acoustic salience is a driving factor in early label-object mapping, that infants in the salient pitch contour contrast condition should outperform infants in the non-salient contrast condition, as evidenced by longer looking on Switch test trials where the original label-object mapping is violated.

Experiment 1

Methods

Participants

Thirty-two 14-month-old (mean = 13.9 months, range = 13.4–14.6 months; 15 female) monolingual English-learning infants from the greater Knoxville area participated in Experiment 1. Sample sized was determined based on prior studies using the Switch Paradigm (e.g., Stager & Werker, 1997; Archer & Curtin, 2018). All infants were born full-term, had fewer than four previous ear infections, and had no history of hearing or vision problems according to parental report. Infants were recruited from the Child Development Research Group database maintained at the University of Tennessee. Data from 10 infants were excluded due to fussiness or crying (6), inattentiveness (1), experimental error (1), looking for the total duration on four or more test trials (1), and a current ear infection (1).

Auditory stimuli

A female whose native language was Mandarin produced all of the speech tokens. In order to ensure that our findings were not limited to a single CV sequence, we selected two different CV nonsense words to serve as object labels: the CV sequence /kʊ/ because it was used in the previous studies by Graf Estes and Hay (2015), Hay et al. (2015) and the CV sequence /di/ because it is phonotactically legal in both English and Mandarin. The particular CV sequences (i.e., /kʊ/ or /di/) that served as the object label was counterbalanced across participants. Both CV sequences were produced with each of the three pitch contours used in Experiment 1: level, rising, and dipping (see Fig. 1). Another nonsense word, /mi/, was produced in a neutral tone and was used as a pre-habituation and post-test stimulus. The labels were recorded in a soundproof booth at a sampling rate of 4400 Hz. For a given CV sequence, tokens were selected to have similar overall durations: 838–903 ms for /kʊ/ (with 683–749 ms of voicing) and 620–664 ms for /di/ (see Fig. 1). During referent training and testing a single token of the target pitch contour was repeated with an interstimulus interval (ISI) of 750 ms. Labels were modified to have similar overall durations using Praat (Boersma & Weenink, 2001) and were RMS matched for equal loudness in Adobe Audition 3.0™.

Visual stimuli

Novel objects were multicolored, two-dimensional images (see Fig. 2). In order to ensure that our findings would be generalizable beyond a single set of objects, two different object pairs were also used. Approximately half of the infants saw Object Pair 1, and half saw Object Pair 2, counterbalanced across label (/kʊ/ vs. /di/) and condition (level/dipping and rising/dipping). A fifth object was used for the pre-habituation and post-test trials. On each trial, objects were presented against a grey background and bounced continuously across the screen to help maintain infants' attention. Movement of the objects was not temporally synced with the presentation of the auditory stimuli.

Procedure

Infants sat on their caregivers' lap in a 2.3 m × 2.3 m soundproof booth, approximately 1 m from a 42-inch flat screen television, which
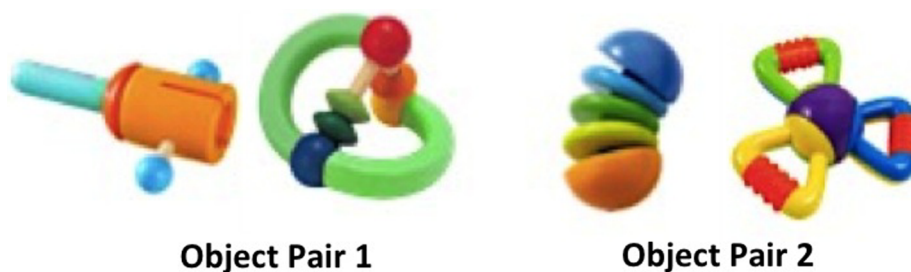
Fig. 2. Objects used in Experiments 1 and 2.

was used to present the visual stimuli. Auditory stimuli were presented from two ORB audio speakers that were hidden behind the left and right corners of the center screen. Habit X 1.0 (Cohen, Atkinson, & Chaput, 2004), installed on a Mac, was used to control the experimental procedure and record infant looking time. A Security Labs 22× Optical Power Zoom digital video camera was used to relay the visual image of infants looking behavior to the control room, via iMovie™ installed on a MacMini. An experimenter, blind to the stimulus being presented, recorded the infants' looking behaviors by pressing keys on the computer running Habit. In order to avoid any other sources of bias, caregivers wore headphones that played masking music throughout the experiment.

We assessed label-object association using a modified version of the Switch paradigm (Werker et al., 1998). We used a more stripped down associative learning task, instead of one that provided more referential support (i.e., through the use of carrier phrases or having a live speaker) for two reasons: (1) we wanted to replicate the methods of Hay et al. (2015) as closely as possible, and (2) associative learning tasks are sometimes able to reveal underlying processing differences that can be masked in more referential tasks. In order to familiarize infants with the task and to help mitigate against artificially short or long first looks that could lead to erroneous habituation times, the first trial was always an unrelated pre-test stimulus (neutral \mi\ paired with a novel object). Next, infants were habituated to two novel label-object pairs presented one at a time, in randomized order. In the Salient Condition the labels were maximally acoustically distinctive: /di/ level vs. /di/ dipping or /kʊ/ level vs. /kʊ/ dipping. In the Non-salient Condition the labels were minimally acoustically distinctive: /di/ rising vs. /di/ dipping or /kʊ/ rising vs. /kʊ/ dipping. On each trial a label-object pair was presented until the infant looked away from the screen for 1 s or after 20 s had elapsed. The habituation phase ended and the test phase began when the habituation criterion was met (i.e., looking on the last three trials decreased to 65% of the looking on the first three trials) or after 25 habituation trails. During the Test phase infants were presented with two types of test trials: Same trials and Switch trials (see Fig. 3). On the Same trials, the label-object pairings from the Habituation phase were maintained. On the Switch trails, the labels of the objects were switched, such that label A was presented with object B, and vice versa. If infants learn the label-object pairing they should look longer on Switch than on Same trials. Infants' looking was directed back to the screen between trials with a spinning pinwheel that served as the attention getter. The experimenter began each trial only after the infant looked at the attention getter. There were 8 test trials; four Switch trials and four Same trials counterbalanced in 8 different testing orders. Finally, to verify that infants maintained attention through the experimental procedure, the pre-test stimulus was presented a second time as a final post-test trial.

Results and discussion

There were no differences between Conditions in the number of trials to habituate (Salient level/dipping = 10.5, SD = 3.95; Non-

salient rising/dipping = 10.0, SD = 5.40), $F < 1$, $p > .7$, or in the total time to habituate (Salient level/dipping = 118.6 s, SD = 49.5; Non-salient rising/dipping = 114.2 s, SD = 76.0), $F < 1$, $p > .8$. This suggests that infants in both conditions showed similar levels of interest in the task.

There were no significant main effects or interactions involving infant sex, CV label (/kʊ/ vs. /di/), or object pair (1 vs. 2) used in training; thus, we collapsed across these factors in all subsequent analyses. To examine the effects of acoustic distinctiveness on label-object mapping we performed a between Condition (Salient level/dipping vs. Non-salient rising/dipping) × within Trial Type (Switch vs. Same) repeated measures ANOVA. There were no significant main effects of Condition, $F(1,30) = .047$, $p = .83$, $\eta_p^2 = .002$, or Trial Type, $F(1,30) = .236$, $p = .630$, $\eta_p^2 = .008$, however, there was a significant Condition X Trial Type interaction, $F(1,30) = 10.199$, $p = .003$, $\eta_p^2 = .254$. Planned comparisons, using paired t tests, revealed that infants in the Non-salient rising/dipping Condition looked significantly longer to Switch ($M = 8.93$, $SD = 3.83$) than Same trials ($M = 6.98$, $SD = 1.99$), $t(15) = 3.00$, $p = .009$, $d = 1.037$ (all t tests are two-tailed; effect sizes reported for t tests are Cohen's d), indicating that they learned to map the rising and dipping pitch contours to separate objects (see Fig. 4). Thirteen out of 16 participants looked longer on the Switch trials. There were no significant differences in looking between Switch ($M = 7.47$, $SD = 3.12$) and Same ($M = 8.90$, $SD = 4.11$) trials for infants in the Salient level/dipping Condition, $t(15) = -1.714$, $p = .107$, $d = .433$, suggesting that they did not learn the mapping between the labels and the objects. Four of the 16 infants looked longer on the Switch trials.

Although previous research suggests that lexical tones with level versus dipping pitch contours are some of the most acoustically distinctive and perceptually salient tone pairs, it is possible that infants in our task failed to map them onto novel objects because they lacked the ability to discriminate them. Although in natural speech, Tone 1 (level) is typically shorter than Tone 3 (dipping), our tokens were matched for length so that we could compare the performance across different pitch contours and thus we may have reduced their distinctiveness. If infants are unable to discriminate a given contrast, then they would necessarily fail to map them to distinct objects. In order to determine whether performance in the Salient level/dipping Condition from Experiment 1 was the result of a failure to discriminate the contrast, in Experiment 1B we tested infants' discrimination of level vs. dipping pitch contours using an object-free habituation/dishabituation task.

Experiment 1B

Methods

Participants

Ten 14-month-old (mean = 13.8 months, range = 13.6–14.1 months; 4 female) monolingual English-learning infants participated in Experiment 1B. Here, like in our previous work (e.g., Hay et al., 2015), we used a smaller sample size for the tone discrimination

Fig. 3. Schematic of experimental design for level and dipping /kʊ/.

task. Based on an effect size of .98 (from Hay et al., 2015), a sample size of 10 should yield statistical power of 87%, well above the best practice threshold of 80% power. Exclusion criteria and recruitment procedures were identical to Experiment 1. Data from 4 additional infants was excluded due to fussiness (3) and failure to habituate (1).

Stimuli

The auditory stimuli were identical to those used in Experiment 1. The visual stimulus consisted of a static black and multicolored 9 × 14 in. checkerboard.

Procedure

Like Experiment 1, Experiment 1B consisted of 2 phases – habituation and test – plus one pre-test and one post-test trial. In Experiment 1B, infants were habituated to either level /kʊ/ (or /di/) or dipping /kʊ/ (or /di/), paired with the checkerboard. The assignment of habituation stimulus was counterbalanced across participants. Trial

duration and habituation criteria were the same as in Experiment 1. Following habituation, infants were presented with two test trials, one where the auditory stimulus remained the same and one where the auditory stimulus was changed to the other pitch contour (e.g., habituate: level /kʊ/ →test: dipping /kʊ/). Again, the order of the test trials was counterbalanced across participants such that half of the infants heard the same trial first, and half heard the change trial first.

Results and discussion

There were no significant main effects or interactions involving infant sex, CV label (/kʊ/ vs. /di/), or Pitch Contour (level vs. dipping), used during habituation; thus, we collapsed across these factors in all subsequent analyses. To examine whether English-learning 14-month-olds can discriminate level versus dipping pitch contours, we performed a paired samples t test on the looking time to the Same versus the Change trial. Infants looked significantly longer on the Change trial
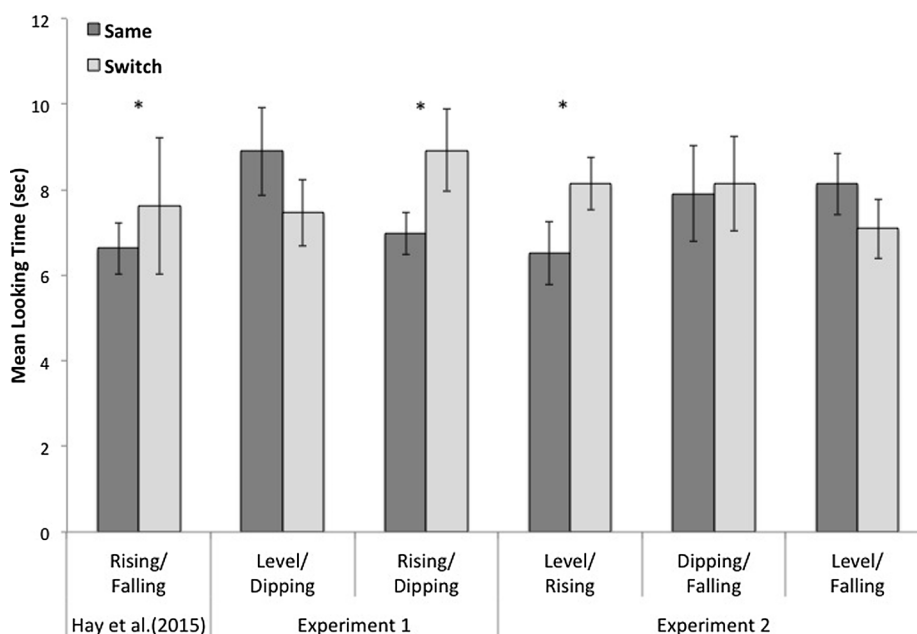

Fig. 4. Results of Experiments 1 and 2: Mean looking times (± 1SE) for Same and Switch trials. Results from 14-month-olds from Hay et al. (2015) are also included for comparison purposes. *$p < .05$.
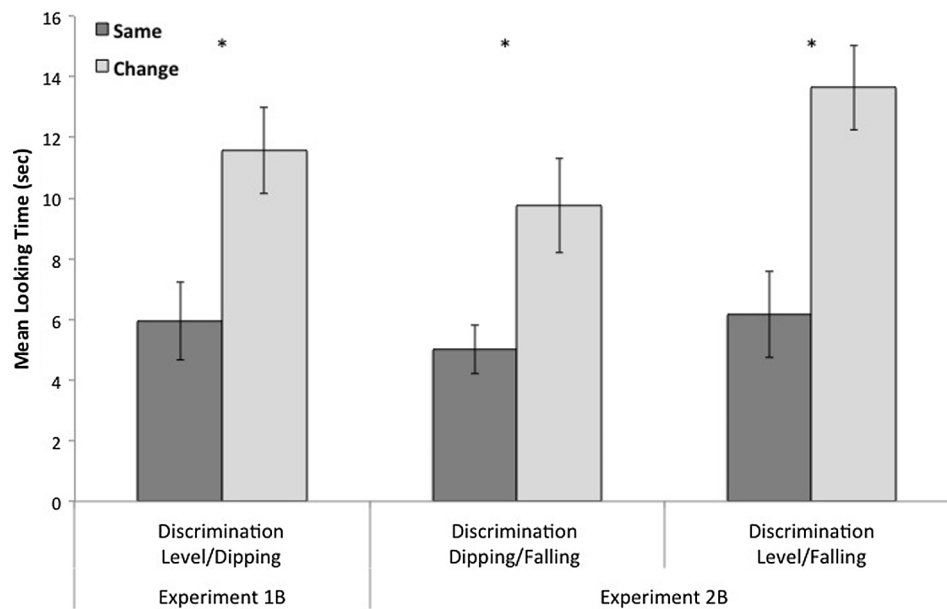
Fig. 5. Results of discrimination Experiments 1B and 2B: Mean looking times (± 1SE) for Same and Change trials. *p < .05.

(change: $M = 11.58$, $SD = 4.50$; same: $M = 5.95$, $SD = 4.06$), $t(9) = 2.55$, $p = .031$, $d = .806$ (see Fig. 5). All 10 participants showed this same pattern of results, readily discriminating level vs. dipping pitch contours. These results suggest that failure to map level versus dipping pitch contours in Experiment 1 was not due to a lack of acoustic distinctiveness between the labels. Thus, some factor other than the acoustic salience of the pitch contour minimal pairs must be driving early label-object mapping.

Experiment 2

In Experiment 2, we consider two competing hypotheses for why infants show early interpretive flexibility when mapping rising/falling (Graf Estes & Hay, 2015; Hay et al., 2015) and rising/dipping (Experiment 1) pitch contours to novel objects, but fail map level/dipping pitch contours. The acoustic distinctiveness of the labels cannot account for these findings. Instead, infants may map rising, dipping, and falling pitch contours to novel objects because the labels themselves contain more variability in F0 than level tones, which show minimal variation in pitch across the label. Labels with level pitch contour are low in entropy, in that there is relatively little change in F0 across time. According to Shannon Information Theory (Shannon, 1948, see also Kluender, Stilp, & Kiefte, 2013), when there is no variability in a signal, there is total predictability, and hence no information is transmitted. Thus, labels with a level pitch contour, which contain low variability in F0, likely transmit comparatively less information and may therefore be more difficult to map to meaning. Learning would also be difficult if the pitch contour of a label were to vary randomly (i.e., high entropy), because it would be impossible to predict F0 across time. The rising, falling, and dipping labels, may provide a sweet spot or Goldilocks effect (Kidd et al., 2012; Kidd, Piantadosi, & Aslin, 2014) in pitch variation, where learning is optimal. The degree of F0 variability in the contour labels may add to the richness of information contained in the labels and thus might help the infants encode more details about the novel label-object associations. This in turn may make the word mapping easier to accomplish. By this account, contour-contour distinctions may be more privileged during label learning than level-contour distinctions.

A second possibility is that rising pitch contours are driving learning in the rising/falling (Graf Estes & Hay, 2015; Hay et al., 2015) and

rising/dipping (Experiment 1) conditions. There are a number of reasons that rising pitch may have a privileged status for young English-learning infants. First, data from Snow (2006) and Kent and Murray (1982) suggest that while rising pitch contours appear to be under-represented in English-learning infants' earliest productions, they are the only class of pitch contours that appears to increase significantly across the first year of life. This suggests that English-learning infants may still be in the process of mastering the production of rising pitch, which may lead to heighted perceptual sensitivity to rising pitch. A link between production experience and speech perception sensitivities has been well documented in the consonant babbling literature (e.g., DePaolis, Vihman, & Nakai, 2013; Majorano, Vihman, & DePaolis, 2014), and may extend to the prosodic domain.

A second reason that rising pitch contours may be driving learning is that English-learning infants have a considerable amount of experience with rising pitch signaling meaningful information. For example, rising pitch contours can differentiate yes/no questions from statements (Bolinger & Bolinger, 1989; Hadding-Koch & Studdert-Kennedy, 1964) and are used extensively in infant-directed speech (Fernald & Kuhl, 1987; Fernald, 1992). Although variation in pitch does not differentiate word meaning in English, 14-month-old infants may still treat rising pitch contours as though they are functionally relevant. Thus, early in development English-learning infants may seize on to the meaningfulness of rising pitch to support word learning in an otherwise unclear associative-learning task. By both the production and meaningfulness accounts, minimal pair distinctions that contain a rising pitch should be privileged over minimal pairs that do not contain a rising pitch contour label.

To test these hypotheses in Experiment 2, we trained 48 additional English-learning 14-month-olds to map lexical tones to novel objects in 1 of 3 conditions: (1) level vs. rising, (2) dipping vs. falling, and (3) level vs. falling. If early interpretive flexibility is being driven by the informational richness of the contour-contour distinction, then we would expect infants to only learn in the dipping/falling condition. However, if interpretive flexibility is being driven by the rising pitch contours then we would expect to see learning only in the level/rising condition. Although both hypotheses would predict no learning in the level/falling condition, we included it here to have data on all of the Mandarin pitch contour contrasts, and as a validation of our predictions.

Methods

Participants

Forty-eight 14-month-old (mean = 14.1 months, range = 13.5–15.2 months; 23 female) monolingual English-learning infants participated in Experiment 2. Exclusion criteria and recruitment procedures were identical to Experiment 1. Data from 20 additional infants were excluded due to fussiness or crying (13), inattentiveness (3), experimental error (3), or parental interference (1).

Stimuli

The visual stimuli were identical to those used in Experiment 1. For the level/rising condition we used the level and rising /kʊ/ and /di/ stimuli from Experiment 1. Since we did not use the falling lexical tone in Experiment 1, we recorded new exemplars of level, dipping, and falling stimuli for the level/falling and dipping/falling conditions. As seen in Fig. 1, the pitch contours of the new level and dipping recordings were very similar to those used in Experiment 1.

Procedure

The procedure was identical to that of Experiment 1. One third of the infants (i.e., 16) participated in each of the three word learning conditions.

Results and discussion

There were no differences between Conditions in the number of trials to habituate (level/rising = 12.0, SD = 4.65; level/falling = 10.66, SD = 5.90; dipping/falling = 9.19, SD = 4.02), F(2,45) = 1.31, p = .28, or in the total time to habituate (level/rising = 123.5 s, SD = 71.3; level/falling = 118.5 s, SD = 81.5; dipping/falling = 94.6 s, SD = 52.3), F < 1, p > .7. This suggests that infants in the three conditions showed similar levels of interest in the task.

There were no significant main effects or interactions involving infant sex, CV label (/kʊ/ vs. /di/), or object pair (1 vs. 2) used in training; thus, we collapsed across these factors in all subsequent analyses. A between subjects Condition (level/rising vs. dipping/falling vs. level/falling) × within subjects Trial Type (Switch vs. Same) repeated measures ANOVA revealed no significant main effects of Condition, F(2,45) = .201, p = .818, η_p^2 = .009, or Trial Type, F(1,45) = .502, p = .482, η_p^2 = .011. As predicted, there was a significant Condition X Trial Type interaction, F(2,45) = 4.044, p = .024, η_p^2 = .152. Planned comparisons, using paired t tests, revealed that infants in the level/rising Condition looked significantly longer to Switch (M = 8.14, SD = 2.43) than Same trials (M = 6.51, SD = 2.94), t(15) = 2.401, p = .029, d = .608, indicating that they learned the label-object mappings (see Fig. 4). Twelve of the 16 infants looked longer on the Switch trials. There were no significant differences in looking between Switch and Same trials for infants in either the level/falling Condition (Switch M = 7.09, SD = 2.80; Same M = 8.14, SD = 2.89), t(15) = −1.433, p = .172, d = −.358, or the dipping/falling Condition (Switch M = 8.14, SD = 4.42; Same M = 7.91, SD = 4.45), t(15) = .399, p = .695, d = .099, indicating that they did not show evidence of learning the mapping between the level/falling or the dipping/falling labels and the objects. Five of the 16 infants in the level/falling Condition and 10 of the 16 infants in the dipping/falling Condition looked longer on the Switch trials.

Before we provide an interpretation of our results, we wanted to verify that infants of this age are able to successfully discriminate the level/falling and dipping/falling pitch contour contrasts. Thus, in Experiment 2B we replicated the discrimination experiment laid out in 1B with the level/falling and dipping/falling pitch contour contrasts.

Experiment 2B

Methods

Participants

Twenty 14-month-old (mean = 14.0 months, range = 13.5–14.9 months; 10 female) monolingual English-learning infants participated in Experiment 2B. Exclusion criteria and recruitment procedures were identical to Experiments 1 and 2. Data from 6 additional infants were excluded due to fussiness or crying (4), inattentiveness (1), or failure to habituate (1).

Stimuli

The visual stimuli were identical to those used in Experiments 1B. The auditory stimuli were identical to those used in Experiment 2.

Procedure

The procedure was identical to that of Experiment 1B. Half of the infants participated in the level/falling pitch contour discrimination task and half participated in the dipping/falling discrimination task.

Results and discussion

There were no significant main effects or interactions involving infant sex, habituated pitch contour, or CV label (/kʊ/ vs. /di/) for either the dipping/falling or the level/falling discrimination tests; thus, we collapsed across these factors in all subsequent analyses. To examine whether English-learning 14-month-olds can discriminate our dipping/falling and level/falling pitch contours, we performed paired t tests on the looking time to the Same versus the Change trials in each condition. Infants looked significantly longer on the Change versus Same trials in both the dipping/falling discrimination task (Change M = 9.75, SD = 4.88; Same M = 5.02, SD = 2.57), t(9) = 3.01, p = .015, d = 1.022, and in the level/falling discrimination task (Change M = 13.64, SD = 4.37; Same M = 6.16, SD = 4.50), t(9) = 3.46, p = .007, d = 1.093, (see Fig. 5). In both conditions, 8 of the 10 participants looked longer on the trial in which the pitch contour was changed. Again, these results suggest that failure to demonstrate learning of the label-object mappings in Experiment 2 was not due to infants' inability to differentiate the pitch contours in the labels.

Together our findings suggest that although the amount of pitch variation in the labels may be a relevant factor in determining what makes a good object label, it is not the driving factor in the early interpretive flexibility we see in English-learning 14-month-olds' associative learning of non-native pitch contours. Across 6 pitch contour contrasts, infants learned the label-object mappings in the 3 conditions where one of the labels had a rising pitch contour: rising/falling (Hay et al., 2015), rising/dipping (Experiment 1), and level/rising (Experiment 2). Infants failed to learn in the 3 conditions where none of the labels had a rising pitch contour: level/dipping (Experiment 1), level/falling (Experiment 2), and dipping/falling (Experiment 2). This pattern of results supports the hypothesis that rising pitch contours may be driving learning in our label-object mapping task. We suggest that rising pitch contours may be especially salient for young English-learning infants, and that infants may generalize what they know about the relevance of rising pitch contours during a task that indicates that pitch contours may also be lexically relevant.

General discussion

At the outset of these studies, our goal was to explain the finding that 14-month-old English-learning infants are able to map two non-native lexical tones, such as rising and falling /kʊ/, to novel objects with relative ease (Graf Estes & Hay, 2015; Hay et al., 2015). These findings contrast the long-observed phenomenon that infants at this age typically have a difficult time mapping consonant-based minimal pairs

(e.g., bin/din) in associative learning tasks (e.g., Stager & Werker, 1997). In Experiment 1, we tested the hypothesis that acoustic salience may be driving the learning of these pitch contour-based minimal pairs. Although infants could readily discriminate acoustically salient level vs. dipping pitch contours, they failed to show evidence of mapping them to novel objects. Instead, infants mapped the less acoustically distinctive rising and dipping pitch contours.

In Experiment 2, we considered two hypotheses: (1) Labels that contain pitch movement (i.e., rising, dipping, falling) transmit more information and thus will be mapped to meaning more readily than labels with minimal pitch variation (i.e., level pitch). On this account, infants should map contour-contour labels to meaning more readily than level-contour labels. (2) Rising pitch contours are privileged early in development, and thus, are driving pitch contour-based minimal pair learning. On this account, infants should learn to map the labels if one of the labels has a rising pitch contour, but should fail if none of the labels have a rising pitch contour. Our data support the latter hypothesis. Infants learned to map labels with rising vs. dipping (Experiment 1), rising vs. level (Experiment 2) and rising vs. falling pitch contours (Hay et al., 2015), but failed to learn when the label pairs did not include a rising pitch contour; they failed to map level/dipping (Experiment 1), level/falling (Experiment 2), and falling/dipping (Experiment 2), even though all of the contrasts were readily discriminable at this age (i.e., 14 months).

One potential explanation for our results is that children's own experience mastering the production of native-language prosody may impact their sensitivity to pitch contour information during early word learning. Previous work on the link between consonant babbling and speech perception suggests that speech production experience impacts speech perception sensitivities (e.g., DePaolis et al., 2013; Majorano et al., 2014). While, to our knowledge, there have been no explicit tests of the relationship between English-learning infants' pitch contour production and perception, there are data to suggest a developmental trend in prosodic production capabilities. Kent and Murray (1982) suggest that while falling pitch contours appear to be most pervasive in English-learning infants' earliest vocalizations, the rate of vocalizations with rising pitch appears to increase across the first year. Further, data from Snow (2006) suggests that by 12–14 months English-learning infants appear to be producing rising and falling pitch contours in equal proportions (see also Whalen, Levitt, & Wang, 1991 for evidence that language background influences the distribution of pitch information in children's early vocalizations). Thus, given that 14-month-old English-learning infants are likely to still be in the process of mastering the production of rising pitch contours, rising pitch may be particularly salient to infants of this age. This emergent feature of English-learning infants' productions across the first year does not appear to extend to other pitch contours (see Kent & Murray, 1982), and thus other pitch contours may not be as salient. Future longitudinal explorations of pitch contour productions with English-learning infants, as well as cross-linguistic comparisons (e.g., Hallé, De Boysson-Bardies, & Vihman, 1991; Whalen et al., 1991) will be necessary to further probe how perception-production links are realized in the prosodic domain.

An alternative explanation that may function in parallel with infants' emergent production abilities is that rising pitch contours may be especially meaningful for the young English-language learner, and this meaningfulness may drive associative learning. There is a great deal of evidence suggesting that not only are rising pitch contours prevalent in infants' auditory environment, but that infants show sensitivity to rising pitch. For example, rising pitch contours are prevalent in infant-directed speech (Fernald & Kuhl, 1987; Fernald, 1992) and are broadly used to elicit infant attention (Fernald & Mazzie, 1991; Sullivan & Horowitz, 1983). Further, there is evidence that adults' productions of yes/no questions and statements are differentiated based on the presence or absence of rising pitch contours (Bolinger & Bolinger, 1989; Hadding-Koch & Studdert-Kennedy, 1964). Indeed, in a recent corpus analysis (Hay, Cannistraci, Moore, & Graf Estes, 2017) we probed

whether rising pitch contours in utterance final words are more predictive of the types of sentences that infants are hearing than are other pitch contours. Our acoustic analyses of the speech of 12 mothers from the Brent corpus (Brent & Siskind, 2001) revealed that although a number of different sentence types have certain pitch contours that are over-represented, only rising pitch contours are both over-represented in Yes/No questions while being under-represented in all other sentence types. Thus, rising pitch contour may be a meaningful, frequent, and pragmatic cue to interpreting the linguistic exchange. Consistent with these findings, we suggest that rising pitch contours provide infants with meaningful information, which may in turn push infants to treat rising pitch as a potential lexical cue, in a task that provides little other relevant information. Other pitch contours do not provide infants with as much meaningful information; thus, infants may fail to interpret them as lexically relevant in our task, as is appropriate.

The idea that infants may be able to glean information from pitch patterns in speech is not new. For example, English-learning 12-month-olds can use a constellation of features including intensity, duration, and pitch information to map labels that differ in lexical stress to novel objects (Curtin, 2009). Infants of the same age can also use prosodic patterns in speech to infer the communicative intent of the speaker (Fernald, 1989). This sensitivity to pitch information appears early in development. Indeed, infants are sensitive to rising pitch contours from as early as 2 months (e.g., Sullivan & Horowitz, 1983), and they are able to use intonational cues, and specifically phrase final rising pitch, to differentiate questions from statements in both English and European Portuguese within the first year of life (Frota, Butler, & Vigário, 2014; Geffen, 2014; Kaplan, 1975; Soderstrom, Ko, & Nevzorova, 2011; Sullivan & Horowitz, 1983). Given their early sensitivity to pitch information, the meaningful nature of rising pitch contours in their everyday environments, and the emergent nature of rising pitch in children's own vocal productions, we suggest that English-learning infants may initially over-interpret rising pitch contours during early word learning.

Further support for this claim comes from recent work with non-speech stimuli. Graf Estes, Antovich, and Hay (2018) took the original rising and falling /kʊ/ used by Hay et al. (2015) and filtered out the harmonic information, preserving only the original pitch contours. Using a very similar experimental paradigm to Hay et al. (2015) and to the one used here, they found that English-learning 14-month-olds learned to map the non-speech rising vs. falling pitch contours to novel objects. These findings are particularly remarkable as similarly aged English-learning infants fail to map much more speech-like sounds (e.g., mmmm, shhhh; MacKenzie et al., 2011). Further, they are consistent with the hypothesis that rising pitch contours may be especially meaningful for young English-learning infants.

Our findings are also consistent with Werker and Curtin (2005) PRIMIR model (see also Curtin & Werker, 2018), which is a developmental framework for Processing Rich Information from Multi-dimensional Interactive Representations. PRIMIR posits that performance on associative learning tasks involves the integration of 3 dynamic filters; (1) the initial perceptual biases of the infant, (2) the developmental stage of the infant, and (3) the task demands. Thus, one account of the difficulty 14-month-olds have mapping consonant-based minimal pair labels to objects is that they may not have strong, rapid access to phonetic details during cognitively demanding word-learning tasks (for related evidence see Fennell, 2012; Fennell & Waxman, 2010; Yoshida et al., 2009). Here, we suggest that pitch contours provide an acoustically salient signal that infants are sensitive to from an early age. Further, infants appear to have access to the acoustic structure of pitch contour information, but only use it during associative learning when they have prior experience that indicates that pitch, and specifically rising pitch, is a relevant feature in their language.

Consistent with the predictions of PRIMIR (Werker & Curtin, 2005), across the second half of the second year, infants appear to become more resistant to treating pitch contours contrastively in these types of

associative learning tasks, even if one of the labels has a rising pitch (Graf Estes & Hay, 2015; Hay et al., 2015). Although 14-month-olds readily map labels with rising vs. falling pitch contours to novel objects, 17- to 19-month-olds do not (Hay et al., 2015). Interestingly, similar to the developmental trajectory observed by Hay et al. (2015), Graf Estes et al. (2018) found that although 14-month-olds mapped the non-speech analogues of the rising and falling lexical tones to novel objects, 19-month-olds did not. As infants gain native language knowledge, the linguistic role of rising pitch contours may become better defined. Infants may learn that differences in pitch are informative about the types of sentences they are hearing, but that pitch itself does not differentiate words. Thus, although older infants may continue to derive a processing benefit when they hear rising pitch contours, we suggest that their meaningfulness is no longer broadly applied in the context of associative learning tasks.

A final consideration of the PRMIR model (Werker & Curtin, 2005) involves the task demands. Studies that have provided infants with referential support typically find that this interpretive narrowing for lexical tones occurs somewhat later in the second year (Singh et al., 2014). Because previous studies of lexical tone mapping have almost exclusively used rising versus falling pitch contours, it is unclear whether infants may be pushed to learn to map contours that do not have a rising pitch contour if they are provided with the types of cues that are generally found to support word learning. Some recent work suggests that the addition of referential support, through exposure to familiar label-object pairs prior to referent training, supports the mapping of non-speech sounds that are not otherwise learned (Graf Estes et al., 2018). Future research can address whether other pitch contours may be mapped if the learning is aided by referential support.

Given that we find that not all pitch contours are treated equally during associative learning tasks, researchers need to be cautious in making broad interpretations based on data from only a subset of lexical tones. For example, to our knowledge, virtually every study conducted on lexical tone learning by non-native listeners has used some variation of rising and falling pitch contours (e.g., Graf Estes & Hay, 2015, Hay et al., 2015; Singh & Foong, 2012; Singh et al., 2014). Testing infants' interpretation of various lexical tones, as we have done here, will be necessary in order for us to refine our understanding, claims, and theories about the developmental trajectories of infants' sensitivities to lexical tones and about factors that drive early word learning, more broadly.

In sum, our findings suggest that infants can take advantage of multiple sources of information when learning associative links between sounds and visual referents. When infants have little information about what to do in a task, they may use whatever information is available to them and apply it to the task at hand. Although acoustic distinctiveness between labels is a necessary condition for early associative mapping, our findings suggest that it is not the primary driver of learning. Further, the amount of information in the acoustic signal available to be linked to meaning, while likely an important feature of labels, also does not appear to be the sole driver of learning. Here, we suggest that meaningfulness may work in concert with these other acoustic cues to facilitate associative learning. Early in development, rising pitch contours provide a salient and meaningful source of information, which may lead infants to initially over-interpret its role in differentiating words.

Acknowledgments

References

Boersma, P., & Weenink, D. (2001). Praat speech processing software. *Institute of Phonetics Sciences of the University of Amsterdam*.

Bolinger, D., & Bolinger, D. L. M. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford University Press.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*(2), B33–B44.

Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken Chinese.* Harvard University Press.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0)*. Austin: University of Texas.

Curtin, S. (2009). Twelve-month-olds learn word-object associations differing only in stress patterns. *Journal of Child Language, 36*, 1157–1165.

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word–object associative learning. *Developmental Science, 12*(5), 725–731.

Curtin, S., & Werker, J. F. (2018). PRIMIR on Tone. *Frontiers in Psychology, 9*.

DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development, 36*(4), 642–649.

Fennell, C. T. (2012). Habituation procedures. *Research Methods in Child Language: A Practical Guide,* 1–16.

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development, 81*(5), 1376–1383.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development,* 1497–1510.

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. *Nonverbal Vocal Communication: Comparative and Developmental Approaches,* 262–279.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279–293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209–221.

Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy, 19*(2), 194–213.

Geffen, S. (2014). *When and how infants discriminate between declaratives and interrogatives.* University of Southern California.

Graf Estes, K., Antovich, D., & Hay, J. F. (2018). Intersecting constraints on label learning: Effects of age, label properties, and referential context. *Journal of Cognition and Development,* 1–20.

Graf Estes, K., & Hay, J. F. (2015). Flexibility in bilingual infants' word learning. *Child Development, 86*(5), 1371–1385.

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy, 18*(5), 797–824.

Hadding-Koch, K., & Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. *Phonetica, 11*(3–4), 175–185.

Hallé, P. A., De Boysson-Bardies, B., & Vihman, M. M. (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech, 34*(4), 299–318.

Hay, J. F., Cannistraci, R., Moore, D., & Graf Estes, K. (2017). Mapping lexical tones to meaning: The role of native language prosody. In Paper presented at the 2017 Biennial Meeting of the Society for Research in Child Development, Austin, TX.

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child Development, 86*(1), 10–22.

Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. (2000). An emergentist coalition model for word learning. Becoming a Word Learner: A Debate on Lexical Acquisition (pp. 136–164).

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones.* Cambridge University Press.

Kaplan, R. M. (1975). On process models for sentence analysis. *Explorations in Cognition,* 117–135.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America, 72*(2), 353–365.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development, 85*(5), 1795–1804.

Kluender, K. R., Stilp, C. E., & Kiefte, M. (2013). Perception of vowel sounds within a biologically realistic model of efficient coding. *Vowel inherent spectral change* (pp. 117–151). Berlin: Springer.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*(2), 185–199.

Ma, W., Golinkoff, R. M., Houston, D., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development, 7*, 209–225.

MacKenzie, H., Curtin, S., & Graham, S. A. (2012). 12-Month-Olds' phonotactic knowledge guides their word-object mappings. *Child Development, 83*(4), 1129–1136.

MacKenzie, H., Graham, S. A., & Curtin, S. (2011). Twelve-month-olds privilege words over other linguistic sounds in an associative learning task. *Developmental Science, 14*(2), 249–255.

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development, 10*(2), 179–204.

Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy, 2*(1), 73–86.

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language,* 384–402.

Robertson, S., von Hapsburg, D., & Hay, J. F. (2017). The effect of hearing loss on novel

word learning in infant- and adult-directed speech. *Ear & Hearing, 38*(6), 701–713.

Rost, G., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*(2), 339–349.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423.

Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition, 124*(2), 128–142.

Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science, 17*(1), 94–109.

Singh, L., Tan, A., & Wewalaarachchi, T. D. (2017). Lexical tone variation and spoken word recognition in preschool children: Effects of perceptual salience. *Journal of Child Language, 44*(4), 924–942.

Snow, D. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development, 77*(2), 281–296.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*(2), 273–293.

Soderstrom, M., Ko, E. S., & Nevzorova, U. (2011). It's a question? Infants attend differently to yes/no questions and declaratives. *Infant Behavior and Development, 34*(1), 107–110.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*(6640), 381–382.

Sullivan, J. W., & Horowitz, F. D. (1983). The effects of intonation on infant attention: The role of the rising intonation contour. *Journal of Child Language, 10*(3), 521–534.

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

Tsao, F. M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology, 50*(2), 111–124.

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*(6), 1289.

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

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*(1), 1–30.

Whalen, D., Levitt, A., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language, 67*, 297–319.

Woodward, A., & Hoyne, K. (1999). Infants' learning about words and sounds in relation to objects. *Child Development, 70*(1), 65–77.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12*(3), 412–418.