Infants use phonetic detail in speech perception and word learning when detail is easy to perceive

Paola Escudero[1,2] and Marina Kalashnikova[3,1]

[1]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University

[2]ARC Centre of Excellence for the Dynamics of Language

[3]Basque Center for Cognition, Brain and Language

Abstract

Infants successfully discriminate speech sound contrasts that belong to their native language's phonemic inventory in auditory-only paradigms, but they encounter difficulties in distinguishing the same contrasts in the context of word learning. These difficulties are usually attributed to the fact that infants' attention to the phonetic detail in novel words is attenuated when they must allocate additional cognitive resources demanded by word-learning tasks. The present study investigated 15-month-olds' ability to distinguish novel words that differ by a single vowel in an auditory discrimination (Experiment 1) and a word-learning (Experiment 2) paradigm. These experiments aimed to tease apart whether infants' performance is dependent solely on the specific acoustic properties of the target vowels or on the context of the task. Experiment 1 showed that infants were only able to discriminate a contrast marked by a large difference along a static dimension (the vowels' second formant), while they were not able to discriminate a contrast with a small phonetic distance between its vowels, due to the dynamic nature of the vowels. In Experiment 2, infants did not succeed at learning words containing the same contrast they were able to discriminate in Experiment 1. The current findings demonstrate that both the specific acoustic properties of vowels in the infant's native language as well as the task presented continue to play a significant role in early speech perception well into the second year of life.

Infants use phonetic detail when detail is easy to perceive

During their first months of life, infants show the remarkable capacity to discriminate phonetic contrasts that do and do not belong to their language (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Infants become progressively more sensitive to the phonetic categories of their language, with a simultaneous decrease in sensitivity to most non-native phonetic categories (Aslin, Pisoni, Hennessy, & Perey, 1981; Best, 1984; Burnham, 1986; Kuhl, 2004; Kuhl & Iverson, 1995; Werker & Tees, 1984; 1999; Werker & Yeung, 2005). This is a gradual process that continues through childhood and even adolescence (McMurray, Danelz, Rigler, & Seedorff, 2018). As their linguistic skills develop, listeners become progressively more efficient in using the acoustic cues that signal lexical contrasts in their native language (Galle & McMurray, 2014) and to disregard irrelevant cues (e.g., amplitude, pitch, speaker-specific variation; e.g., Galle, Apfelbaum, & McMurray, 2015; Hay, Graf Estes, Wang, & Saffran, 2015; Rost & McMurray, 2009; Singh; Morgan, & White, 2004; Singh, White, & Morgan, 2008).

Infants' sensitivity to most native phonetic contrasts during their first year of life (Werker & Tees, 2005) runs counter to their difficulty in encoding subtle phonetic details when learning novel words at the beginning of their second year (Pater, Stager, & Werker, 2004; Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). Studies have shown that a number of factors determine early word learning performance including infants' lexical competence (Werker et al., 2002), the cognitive demands of the word-learning task (Yoshida, Fennell, Swingley, & Werker, 2009), the contextual information available in the task (Fennell & Waxman, 2010),

and the acoustic properties of the specific phonemes (Curtin, Fennell, & Escudero, 2009). The present study aims at teasing apart the effect of two such factors on 15-month-olds' ability to learn words that differ in a single phoneme, namely, phoneme acoustic properties and task demands.

Werker and colleagues (Stager & Werker, 1997; Werker et al., 1998) demonstrate the discrepancy in infants' sensitivity to native phonemes versus their word-learning skills. In Stager & Werker's first experiment, 14-month-olds were habituated to two word-object pairings with words that differed by a single consonant (known as minimal pairs; i.e., /bɪ/ + object A, /dɪ/ + object B). At test, infants were presented one of the word-object pairings from the habituation phase (Same trial; e.g., /bɪ/ + object A), and another that presented one of the words with the other object (Switch trial; e.g. /bɪ/ + object B). Fourteen-month-olds failed to learn the minimally different words, as measured by their inability to notice the switch in the word-object pairing during testing. Follow up experiments showed that infants were able to learn two words that differed in all their phonemes (/lɪf/ and /niːm/, Experiment 3), and they successfully discriminated the consonant minimal-pair in a version of their Switch task that did not require them to map the words to novel objects (a test of auditory discrimination, Experiment 4). Therefore, Stager and Werker suggested that the increased cognitive demands of a word-learning task prevent young infants from encoding consonant minimal pairs, and that this ability is developed later, at 17 months of age (Werker et al., 2002).

Later research challenged this conclusion with the suggestion that the task indeed imposed high cognitive demands not caused by word-learning per se, but by the use of the Switch task to test word learning (Fennell & Waxman, 2010; Yoshida et al., 2009). Specifically, the assumption of a surprise response manifested in longer

looking times, which is the dependent variable in the Switch task, presupposes that infants establish the mapping between the novel word and the novel object during habituation, retrieve this mapping during the test phase, and reject the switch word as a potential label for that object. That is, 14-month-olds may recognise that one of the words but not the other corresponds to the object in the habituation phase, but if their confidence in the initial mapping is low, they may accept the switch word as an admissible exemplar of the original one (i.e., a mispronunciation). To test this possibility, Yoshida et al. (2009) assessed 14-month-olds' ability to learn the same minimal pair /bɪ/-/dɪ/ using a visual choice task where infants identify the object that corresponds to the word without the need to reject an incorrect pairing. Further modifications to the Switch task such as inclusion of supporting linguistic or referential information also increase success at word-object association at 14 months (e.g., Fennell & Waxman, 2010; Fennell & Werker, 2003).

While it is evident that the Switch task imposes high cognitive and attentional demands on young infants, it is also possible that the abovementioned inconsistent findings reveal a more gradual development in native phoneme perception. That is, it is possible that while infants accurately discriminate native phoneme contrasts during their first year of life, they continue to fine-tune their competence about the cues in the acoustic signal that indicate lexical contrasts during their second year and further (Hay et al., 2015; Rost & McMurray, 2009). Accordingly, infants' difficulties in noticing a contrast between two words that differ in a single phoneme in a word-learning task may reflect their attention to acoustic cues that are used meaningfully in their language, but that are not used to differentiate phonemes. That is, in the Switch task, infants may fail to exhibit the expected surprise response because they recognise irrelevant cues in the Switch trial such as the coda of the word and other indexical

information about the speaker. In line with this account, infants' performance may be influenced by the type and amount of acoustic cues available to them in the task (Galle & McMurray, 2014).

In the present study, we examine 15-month-old's discrimination and learning of words that differ in a single vowel (i.e., vowel minimal pairs). This is because unlike the many early word-learning studies examining consonant minimal pairs, less is known about infants' ability to learn novel words that constitute vowel minimal pairs. Vowels represent an interesting test case for the controversy laid out above. On the one hand, phonological attunement to vowels is proposed to occur earlier than for consonants (Polka & Werker, 1994), likely because they are more salient in the speech signal due to greater loudness and duration (Repp, 1984). Also, some researchers have shown that vowel acoustic qualities tend to be exaggerated in infants' early linguistic input, i.e., infant-directed speech, which has been proposed to support the early acquisition of native phonetic categories (Burnham, Kitamura, & Vollmer-Conna, 2002; Kalashnikova, Carignan, & Burnham, 2017; Kuhl et al., 1997, but see McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013; Miyazaba, Martin, Kikuchi, & Mazuka, 2017 for alternative accounts). On the other hand, the ability to discriminate native vowel categories improves developmentally in infancy (see Tsuji & Cristia, 2014 for a review), and into childhood, as previous studies show non-adult like vowel perception in children between ages 4 and 7 (Gerrits et al. 2001; Giezen et al. 2010; Nittrouer et al. 2014). Unlike consonants, vowels carry less lexical but more prosodic and indexical information (Pisoni, 1973), to which infants are sensitive (Mulak et al. 2017). This sensitivity to phonetic information can result in young word learners' enhanced difficulty in encoding vowel changes in lexical tasks (Nespor, Peña, & Mehler, 2003).

Findings about infants' ability to distinguish vowel minimal pairs have been mixed. Monolingual English eighteen-month-olds failed to learn the vowel minimal pair /mʌn/-/mɪn/ in the Switch task (Singh, Fu, Tay, and Golinkoff, 2018). Conversely, monolingual English and Dutch eighteen-month-olds succeeded at the Switch task when tested with the vowel minimal pairs /tæm/-/tɛm/ and /tɑm/-/tɛm/ respectively (Dietrich, Swingley, & Werker, 2007). Given that both studies used the Switch task, it is evident that task properties alone do not account for these discrepant findings. Moreover, Nazzi (2005) showed that 20-month-old French-learning toddlers failed to learn vowel minimal pairs even when using an interactive object selection task. As well, Fikkert (2010) showed that infants' ability to distinguish vowel minimal pairs did not depend on the task but on the consonants that surrounded the vowels, as 14-month-olds in their study could detect a switch between /bɪn/ and /bɔn/ but not between /dɪn/ and /dɔn/. The author explained these findings with a phonological account according to which infants' ability to distinguish words depends on the phonological features they store in their lexicon (see also Escudero & Benders, 2010 for an in depth discussion of Fikkert's account and predictions).

Escudero and colleagues' findings support an explanation for infants' word-learning of vowel minimal pairs based on the acoustic properties of the specific phoneme contrasts. For instance, Curtin et al. (2009) found that 15-month-old infants acquiring Canadian English notice the switch between the novel words /diːt/ and /dɪt/ but not between /diːt/ and /duːt/, whereas Escudero, Best, Kitamura, & Mulak (2014) showed that 15-month-old infants acquiring Australian English noticed both switches when the tokens were produced in Canadian English but not when produced in their native Australian English variety. The authors used acoustic analyses of the stimuli to predict that infants would be most successful at discriminating the vowel contrasts

with the largest acoustic distance, regardless of whether they were produced in their native English variety.[1] Additionally, Australian 17-month-olds succeed at discriminating both Australian English /diːt/-/duːt/ and /diːt/-/dɪt/ in the Switch task, although learning was more robust for the former contrast (Escudero, Mulak, Elvin, & Traynor, 2018). The authors reported detailed acoustic analyses that showed that the vowels in the words /diːt/, /dɪt/, and /duːt/ differed significantly in their *dynamic properties*. This refers to how much the vowel changes from beginning to end in each produced vowel token, as measured by the trajectories of the vowels' first (F1) and second (F2) formants, which correspond to the vowels' height and backness. Infants struggle to distinguish /diːt/ and /dɪt/ in their native Australian English because these vowels overlap entirely in their F1 and F2 trajectories. Conversely, distinguishing /diːt/ and /duːt/ was less challenging because these words only overlap in their F1, allowing infants to use F2 for their distinction. Since adult Australian English listeners readily use dynamic vowel trajectories to distinguish vowels such as /i/ and /I/ (Williams, Escudero & Gafos, 2018), it seems that the ability to use dynamic properties in vowel perception emerges closer to the second year of life or later. These findings confirm the magnitude of the phonetic distance hypothesis advanced in Escudero et al. 2014, as dynamic properties tend to decrease the distance between the vowels in a contrast and increase difficulty in early word learning.

In sum, previous studies using the Switch task suggest that infants' age and the specific properties of the vowel contrast tested explain success at learning vowel minimal pairs. With respect to age, studies reveal a developmental progression

---

[1] Curtin et al. (2009) and Escudero and Benders (2010) show that these results did not conform with Fikkert's (2010) phonological approach, as the predictions did not account for the level of success shown in all contrasts. As well, this phonological approach focuses on universal phonological features without considering differences in acoustic properties resulting from accent variation, and would therefore not be able to explain Escudero et al.'s (2014) results. We discuss alternative theoretical approaches that explain these findings in the Discussion section.

whereby at 15 months infants' word-learning success is restricted to stimuli with the least amount of acoustic variation, but it becomes generalised to a wider range of stimuli by 17 months of age. This is in line with research suggesting that speech categories develop well beyond the first year of life and even into late childhood (see for example: Hazan & Barret, 2000 for consonants, and Nittrouer & Lowestein, 2014, for dynamic properties in vowels). However, it remains unclear whether the younger infants' (15-month-olds) selective difficulties with learning vowel minimal pairs (Curtin et al., 2009; Escudero et al., 2014) is due indeed to infants' developing sensitivity to vowel phonemes or to the high cognitive demands of the word-learning task used (i.e., the Switch task).

In the present study, we thus aim at disentangling the possible influence of acoustic properties and task demands by testing 15-month-olds on both discrimination (Experiment 1) and learning of vowel minimal pairs (Experiment 2). In Experiment 1, we tested infants' ability to discriminate vowel minimal pairs. The purpose of this experiment was to assess whether 15-month-olds perceive a difference between the target vowel phonemes, as was the case for the consonants tested in previous studies (Stager and Werker 1997). To assess the role of acoustic properties, Experiment 1 included the same two vowel contrasts (/diːt/-/duːt/ and /diːt/-/dɪt/) used in previous studies, which differ in their degree of acoustic overlap. We hypothesised that if the acoustic properties of specific vowel contrasts do not influence phoneme discrimination during the second year of life, vowel discrimination for /diːt/-/duːt/ and /diːt/-/dɪt/ should be equally successful. Alternatively, the acoustic properties of the vowels may impact phoneme discrimination, in that the phonetic variation in the production of /diːt/, which reduces the phonetic distance between /diːt/ and /dɪt/, may make this contrast more difficult than /diːt/-/duːt/ (Escudero et al., 2018).

In Experiment 2, infants were presented with /diːt/ and /duːt/ during a visual choice task. We hypothesised that if cognitive demands of the word-learning task determine infants' performance for both vowel and consonant minimal pairs (Yoshida et al., 2009), 15-month-old infants should demonstrate the ability to correctly map /diːt/ and /duːt/ to their respective referents. Alternatively, it may be that even a task with lower cognitive demands yields difficulty, which would suggest that further support from contextual cues (Waxman & Fennell 2010) may be needed for a contrast that only differs in a single acoustic dimension, namely F2 (Escudero et al., 2018).

Experiment 1

Experiment 1 assessed 15-month-old infants' performance in an auditory discrimination task. In this task, infants are not required to establish a word-referent mapping but instead they simply need to notice a difference between the habituated word and a word containing a different vowel, based solely on auditory information. This experiment used the two contrasts reported in Escudero et al. (2014; 2018), namely /diːt/-/dɪt/ and /diːt/-/duːt/, using the habituation word /diːt/. Recall that these two previous studies reported that 15-month-olds failed to notice a switch to either /dɪt/ or /duːt/ when the word was paired with a moving object. In this experiment, the habituation phase included a static picture of a bullseye, which according to Stager & Werker (1997) should not trigger word-referent association. If the task of associating a word to its object referent interferes with infants' ability to discriminate between minimally different words, an auditory discrimination task should show success at noticing a change in the habituated word regardless of the type of change: infants should succeed at noticing a change from /diːt/ to /dɪt/ and from /diːt/ to /duːt/ equally well. Alternatively, if the magnitude of the phonetic contrast matters for auditory

discrimination as it does for word learning (Escudero et al., 2018), infants should show better auditory discrimination of /diːt/- /duːt/ than /diːt/-/dɪt/.

Method

      Participants. Twenty-two 15-month-old infants (10 female, *M* age = 470 days, *SD* = 11.5, range 453 to 490) acquiring Australian English participated. During recruitment, parents completed a parental questionnaire, with their responses confirming that their infants were not at-risk for language or developmental disorders and were not exposed to a second language. An additional 12 infants participated but were excluded due to failure to meet the post-habituation criterion (*n* = 8, see Procedure below for details) and extreme fussiness during the task (*n* = 4). Infants were recruited through an infant laboratory database, and they received a small gift and a Baby Scientist degree for their participation, and their parents received $30 as a reimbursement for their travel expenses.

      Materials and apparatus. In this experiment we used the same auditory stimuli for the words /diːt/, /duːt/, and /dɪt/ that were used in Escudero et al. (2014, 2018) for both habituation and test phases. As described in Escudero et al. (2018), a female native speaker of Australian English produced the words in infant-directed speech using a variety of intonation contours. These recordings were used to create an audio string for each word, constructed by concatenating ten tokens of each word (the duration of each word token ranged from approximately 600 to 700 ms) into strings with 1300 to 1400 ms between tokens. This resulted in three 20-second strings, one for each word, which were identical across habituation and test phases. In addition to the three target words, an identical string with the same duration and number of tokens was concatenated using the word /laːd/ (this may be interpreted as a real English word, but it was unlikely to be familiar to infants of this age), and used in the

pre- and post-test trials of the task following previous studies that used a combination of visual and auditory stimuli to measure infant's task attention (e.g., Curtin et al., 2010; Escudero et al., 2018; Yoshida et al., 2009).

The two vowel contrasts chosen for this experiment differed in how much they overlapped in the F1/F2 acoustic plane, as measured and reported in Escudero et al. (2018). As shown in Figure 1, the vowel in /diːt/ has the most dynamic trajectory (i.e., the beginning and end of the vowels represented by the start and end of the arrow), exhibiting the largest change in both F1 and F2 values, which was confirmed with pairwise Levene's tests on the F1 and F2 values for the three vowels measured at 30 points of each vowel token (10 per vowel) (Escudero et al. 2018). As a result, the F1 and F2 values for /diːt/ and /dɪt/ overlap significantly, while /diːt/ and /duːt/ only overlap on F1 values and are well separated by F2 at every point of the vowel. If the magnitude of the phonetic distance between the members of a contrast, also expressed in the amount of overlap in their F1 and F2 values, plays a role in infants' auditory discrimination, they should find /diːt/- /duːt/ easier to discriminate than /diːt/-/dɪt/.
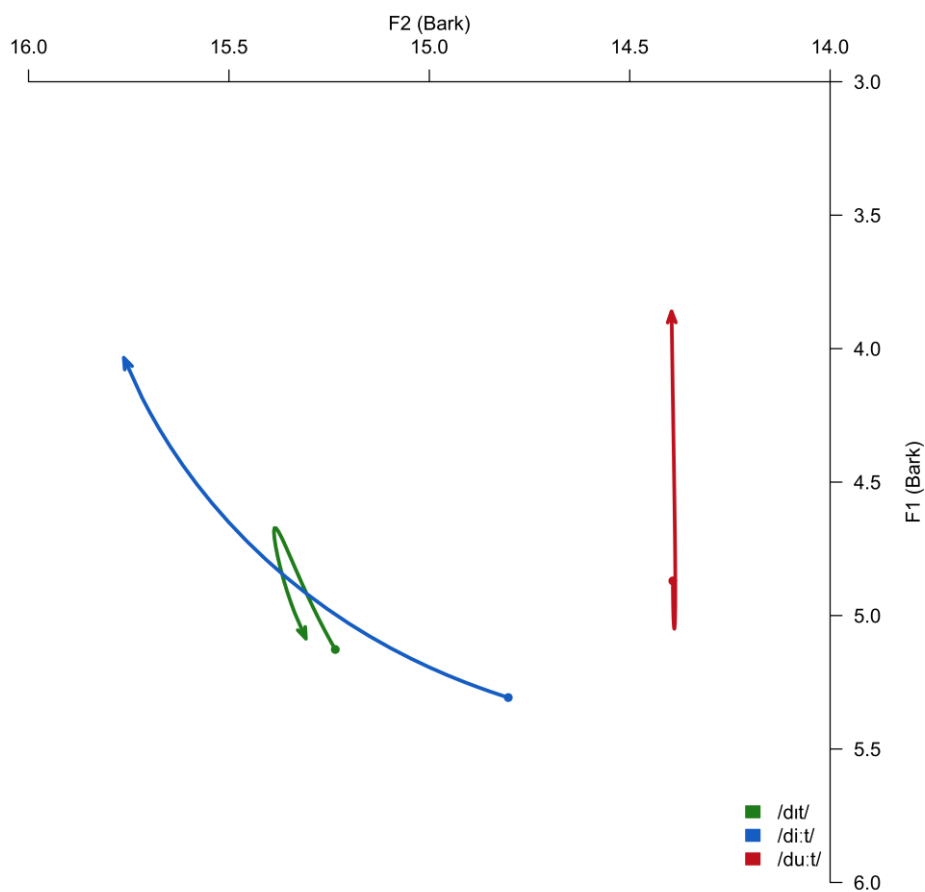
Figure 1. Average F1 and F2 trajectories (in Bark) for the tokens of the words /diːt/, /dɪt/, and /duːt/ used in Experiment 1.

A static image of a colourful bullseye was used as the visual stimulus for the habituation and test trials, and the video of a moving waterwheel was used as the visual stimulus for the pre- and post-test trials. The visual stimuli were presented on a single 17 inch screen, and auditory stimuli were played over loudspeakers located behind the screen. Stimuli were presented using the HABIT X 2.0 program (Cohen, Atkinson, & Chaput, 2004) using a Mac Book Pro.

Infants sat on their caregiver's lap approximately 60 cm away from the screen. Caregivers listened to masking sounds over headphones. An experimenter sat in an adjoining room and observed infants' gaze direction through a CCTV camera and

recorded whether the infant looked to the screen or away from the screen by pressing a key on the testing computer keyboard.

Procedure. We kept the same trial start and end procedure as in previous Switch studies (Escudero et al. 2014, 2018): both habituation and test trials started when the infant looked at a looming attention getter and ended when the sound file ended, which was fixed at 20 seconds. After infants fixated the screen, the habituation phase commenced. During habituation, infants saw the bullseye image on the screen and listened to repetitions of the word /diːt/. The habituation trials continued until the infant reached the habituation criterion (i.e., when looking time to a habituation trial was less than 65% of infant's average looking time to the first two habituation trials) or the maximum of 24 habituation trials. Next infants proceeded to the test phase in which they were presented with three test trials. One trial was identical to the habituation (/diːt/ trial) and two trials presented words that were different to habituation (/dɪt/ trials and /duːt/ trials). The trials in the test phase were administered in three fixed orders counterbalanced across participants. A third of the infants ($n = 8$) were presented with the trials in the order /diːt/, /dɪt/, /duːt/, a third ($n = 7$) with the order /duːt/, /diːt/, /dɪt/, and a third ($n = 7$) with the order /dɪt/, /duːt/, /diːt/. Infants also completed a pre-test trial at the start of the testing session before the first habituation trial, and an identical post-test trial at the end of the testing session after the last test trial.

Infants' looking duration to the screen on all trials was recorded online by the experimental software and used for analyses. Following the previous Switch studies that used identical auditory stimuli (Escudero et al. 2014, 2018), we included a post-habituation exclusion criterion whereby infants were excluded from the final sample (see Participants section) if they showed gaze recovery in the same trial (/diːt/). If

infants' gaze duration in the same trial was 65% or greater than in the last two habituation trials, it was considered that infants had not habituated (given that the same trial was identical to the habituation trials), so their data were excluded from analyses.

Results

First infants' performance in the pre- and post-test trials and the habituation phase was analysed to assess the possibility that decreases in looking times during test trials were due to an overall wane in attention throughout the task. Infants' looking duration to the pre- ($M = 17.05$, $SD = 4.32$) and post-test ($M = 17.55$, $SD = 2.71$) trials did not differ, $t(21) = .471$, $p = .642$, $d = .138$, indicating that infants were engaged throughout the task, and any decreases in looking time during the task were in response to the experimental stimuli and not an overall decrease in attention. On average, infants completed 7.27 ($SD = 4.21$, range 4 to 18) habituation trials before reaching the habituation criterion.

Infants' performance in the test phase was analysed using a mixed ANOVA with trial type (/diːt/, /dɪt/, /duːt/) as the within-subjects variable and order of test trial presentation as the between-subjects variable. The ANOVA was followed by planned within-subjects contrasts to compare looking times across trial types using the /diːt/ trial as the reference. The ANOVA yielded a main effect of trial type, $F(2, 18) = 4.585$, $p = .025$, $\eta^2 = .337$, no effect of trial order, $F(2, 19) = 2.155$, $p = .143$, $\eta^2 = .185$, and no trial type by trial order interaction, $F(4, 38) = .432$, $p = .784$, $\eta^2 = .044$. Mean looking duration at test trials in response to the three labels is shown in Figure 2. Infants looked significantly longer during the /duːt/ trials than the /diːt/ trials, $F(1, 19) = 9.538$, $p = .006$, $\eta^2 = .334$, but there were no significant differences between their looking duration in the /dɪt/ and /diːt/ trials, $F(1, 19) = .199$, $p = .661$, $\eta^2 = .010$.
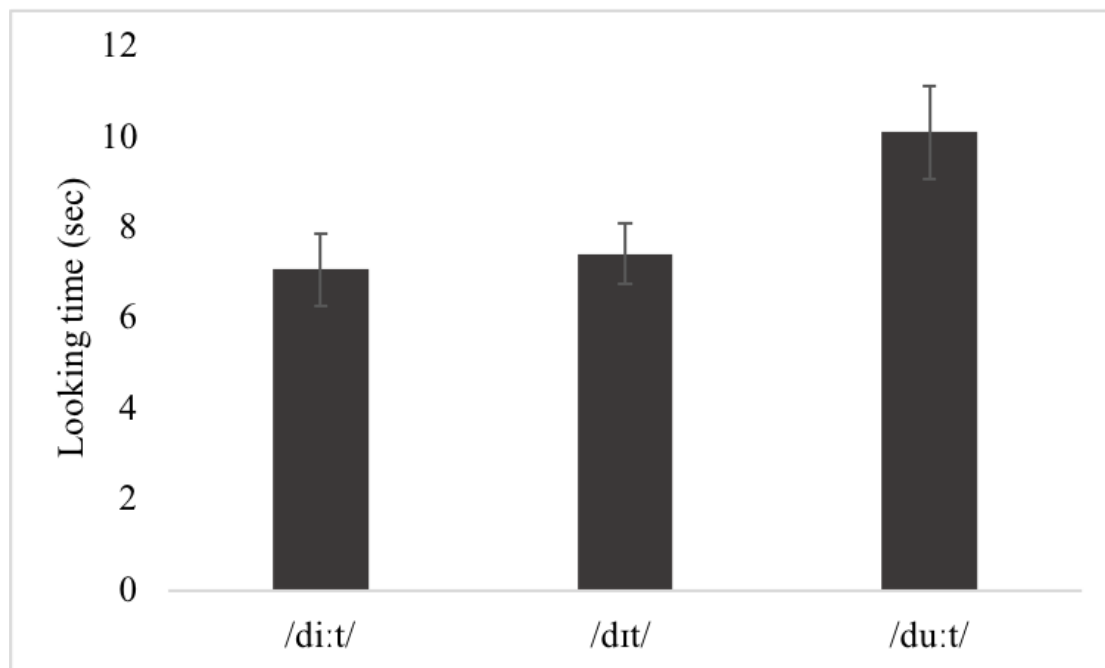
Figure 2. Mean looking time in response to /diːt/, /dɪt/, and /duːt/ test trials in Experiment 1 (error bars represent SEM).

Experiment 2

Experiment 1 showed that 15-month-old infants were sensitive to the acoustic properties of their native vowel categories in a task that involved phoneme discrimination, and not word learning. Specifically, they were only able to discriminate the contrast characterised by the least amount of spectral overlap (/diːt/-/duːt/) but not the contrast characterised by a larger amount of spectral overlap (/diːt/-/dɪt/). These findings support our prediction that infants' ability to discriminate native vowel contrast is constrained by the acoustic properties of the target vowels. This result also suggests that the previously reported difficulties detected in word-learning tasks may not have been entirely due to the additional cognitive demands of word-learning tasks per se. In order to test this possibility directly, Experiment 2 employed the simplified visual choice word-learning task introduced by Yoshida et al. (2009),

16

which was shown to be efficient in eliciting word-learning of consonant minimal pairs in 15-month-olds. Given that infants showed no successful discrimination of the /diːt/-/dɪt/, only word-learning of the /diːt/-/duːt/ contrast was assessed. We thus presented 15-month-old infants with a scenario that should result in successful word learning: a phonemic contrast that they were able to discriminate in Experiment 1 and a word-learning task characterised by low cognitive demands.

Participants. Twenty 15-month-old infants (11 female, *M* age = 469.25 days, *SD* = 12.9, range 452 to 494) acquiring Australian English participated. During recruitment, parents completed a parental questionnaire and confirmed that their infants were typically developing, were not exposed to a second language, and were not at risk for any developmental disorder. An additional group of 11 infants participated but were excluded for failing to complete the test trials (3) and for failing to fixate on both objects during the baseline phase of the task (8; see Procedure). Procedures for participant recruitment and reimbursement were identical to Experiment 1.

Materials and apparatus. The target words from Experiment 1 (also used by Escudero et al., 2014; 2018) were used in this experiment. The stimuli strings created for Experiment 1 were modified for use in this paradigm. The habituation audio strings were identical to those used in Experiment 1, which were 20 seconds long and were constructed by concatenating repetitions of 10 different tokens of each word (with a duration of 600 to 700 ms) into strings with 1300 to 1400 ms between tokens. The test audio strings (one for each word) were four seconds long and were constructed by concatenating four different tokens of each word that were previously presented in habituation with 500 ms between tokens. In addition to the habituation stimuli, the same speaker was also recorded producing four tokens of the word "*wow*"

and the phrases "*where is the ball*" and "*where is the clock*". These were concatenated into 4-second strings and used in baseline and filler trials respectively. Finally, a 20-second string of repetitions of the novel word "*pok*" was used in the pre- and post-test trials to make sure infants were still paying attention to the task.

Given that only 4 tokens were selected for the test phase compared to the 10 tokens used in habituation and test in Experiment 1, the acoustic properties of the three words were examined by re-plotting their vowel trajectories. As shown in Figure 3, the variability in the formant trajectories occurs along F1 and F2 for /diːt/, but only in F1 for /duːt/ in both the habituation and test strings.
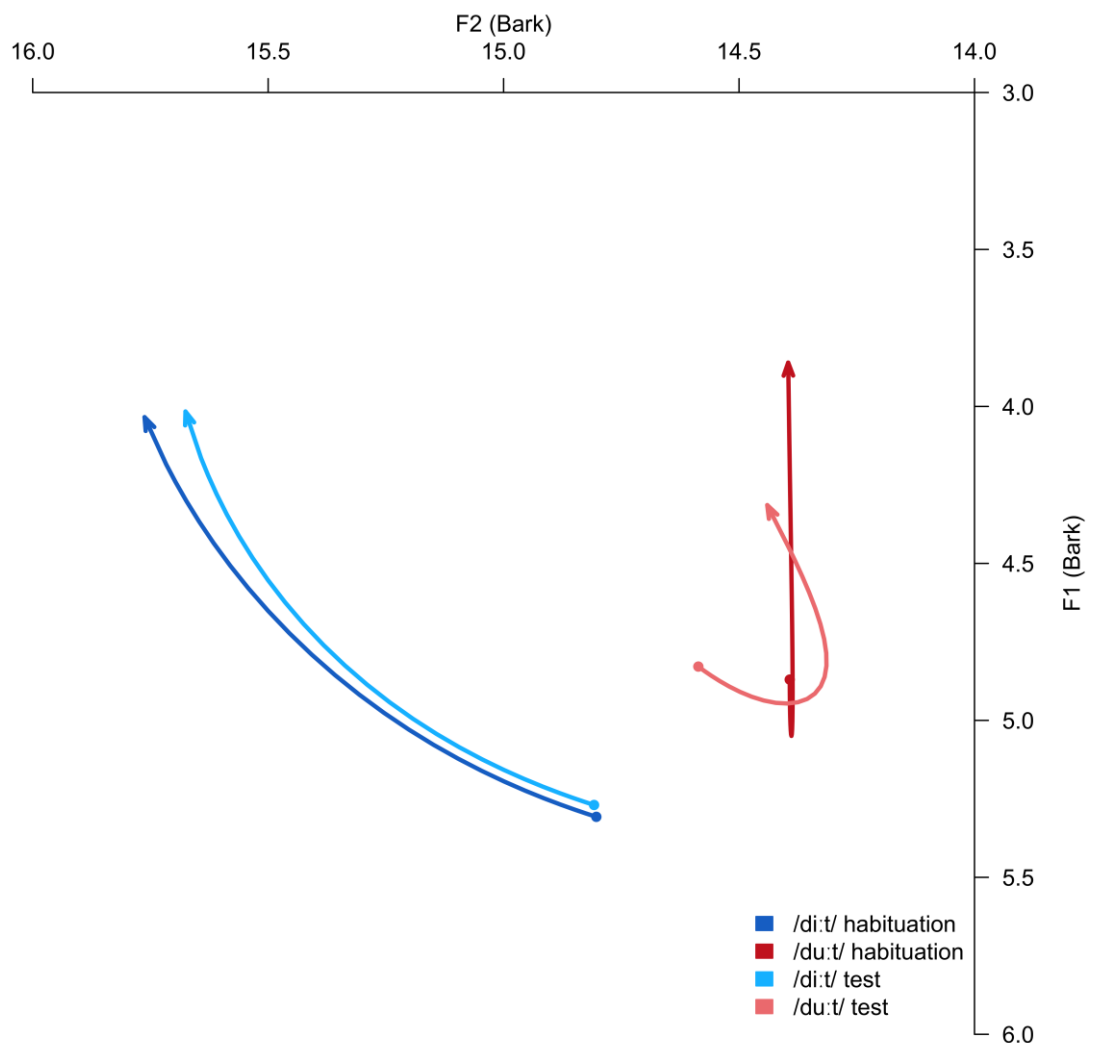
Figure 3. Average F1 and F2 trajectories (in Bark) for the habituation and test tokens of the words /diːt/ and /duːt/ used in Experiment 2.

The visual stimuli consisted of two objects selected from the Fribbles dataset (Yildirim & Jacobs, 2013) from the Tarr Object Databank, and which have been used successfully in previous word-learning paradigms (Kalashnikova & Burnham, 2016; Kalashnikova, Mattock, & Monaghan, 2015). The habituation stimuli consisted of videos of the two objects presented on a black background (Figure 4). In the videos, the objects moved slowly from left to right. In the test phase, the same objects were presented as static images. Each object was yoked with one of the novel labels (i.e., Object 1 – /diːt/; Object 2 – /duːt/). Infants also saw a moving waterwheel toy in the pre- and post-test trials of the task and static images of a ball and a clock in the filler trials.
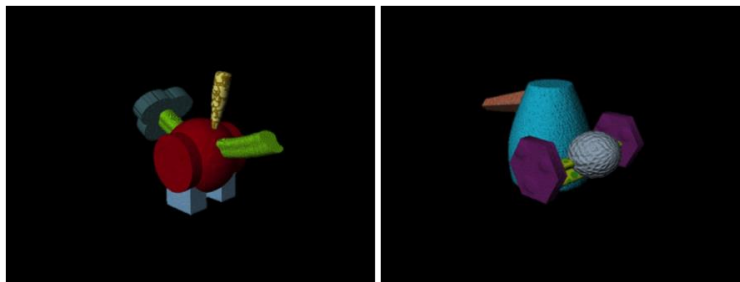

Figure 4. Static images of Object 1 (left) and Object 2 (right) used as visual stimuli in the word-learning task in Experiment 1.

Visual stimuli were presented on three 17 inch screens placed next to each other, and auditory stimuli were played over loudspeakers located behind the screens. The experiment was controlled using a MATLAB (Mathworks, Natick, USA) script presented on a computer running Windows XP. The pre-test, post-test, habituation, and attention getting stimuli were presented on the centre screen, and the test phase

stimuli were presented on the right and left screens. The habituation phase was infant-controlled, but the duration of trials in the test phase was fixed to 4 seconds. The remaining details of the experimental set up and procedures were identical to Experiment 1. For the recording of the infants' gaze, the experimenter recorded whether the infant looked at each screen (centre for habituation and left or right for baseline and test trials) by pressing a key on the testing computer keyboard.

Procedure. The present word-learning task consisted of a combination of a habituation paradigm and visual choice paradigm (Yoshida et al., 2009). First, infants were presented with a flashing red light accompanied by a beeping sound on the centre screen. After they fixated the screen, the habituation phase commenced. During habituation, infants were presented with each novel object-novel word pairing (Object 1-/diːt/; Object 2- /duːt/). Habituation trials for the two objects were presented in blocks, and the habituation phase continued until the infant has reached the habituation criterion (fixation duration to a given block decreased by 65% from the preceding block) or the maximum of 24 habituation trials.

Next, infants proceeded to the visual choice phase of the paradigm. This portion of the task included three types of trials: test, baseline, and filler. In the test trials, infants saw Objects 1 and 2 presented on the right and left screens and heard one of the target words (/diːt/ or /duːt/). In the baseline trials, the presentation of the objects was identical but infants heard the word "*wow*". In the filler trials, infants saw images of the familiar objects ball and clock presented on the right and left screens and heard either the phrase "*where is the ball*" or "*where is the clock*". The filler trials included a carrier phase to maintain infants' attention to the task and were not included in the analyses. Infants also completed a pre-test trial at the start of the

testing session before the first habituation trial, and an identical post-test trial at the end of the testing session after the last visual choice test trial.

The three trial types of the visual choice paradigm were presented in four blocks, each block containing two test trials (one with the word /diːt/ and one with the word /duːt/), one baseline, and one filler trial. The visual choice phase had a total of 16 trials (4 test trials for each word, 4 baseline, and 4 filler trials). The order of presentation of the trials within a block was randomised and the side of presentation for all the objects was counterbalanced within and across participants.

Infants' looking time to each screen (center in habituation, left and right in baseline and test trials) was recorded online by the experimental software. Following previous studies that have used a similar procedure (e.g., Swingley & Aslin, 2000; Yoshida et al., 2009), infants' looking durations to the right and left screens in the test and baseline trials were recorded from 0 to 2000 ms of the trial. While the total duration of the trial was 4 seconds to ensure that the progression from trial to trial was not too rapid to lose the attention of young infants, the window of interest for analyses was restricted to the first 2 seconds as fixations after this point cannot be reliably interpreted as a response to the auditory stimulus (Swingley & Fernald, 2002). Performance on the filler trials was not included in the analyses. For the test trials, proportion of looking duration at target out of the total looking duration at target and the distracter was calculated. As there was no target object in the baseline trials, we calculated the proportions of looking duration to both /diːt/ and /duːt/ objects. In order to ensure that infants had the opportunity to see the objects presented on both screens during the baseline and test trials, a trial was only included in the analyses if the infant had fixated both sides.

Results

First infants' performance in the pre- and post-test trials and the habituation phase was analysed to assess the possibility that decreases in looking times during test trials were due to an overall wane in attention throughout the task. Infants' looking duration was longer for the post-test (M = 19.95 sec, SD = 1.47) compared to the pre-test trials (M = 17.57 sec, SD = 4.74), $t(19)$ = -2.497, p = .022, d = .678, indicating that their attention was engaged throughout the task, and it was fully recovered after the habituation and the visual choice phases. On average infants completed 18.26 (SD = 6.48, range 8 to 24) habituation trials. As can be seen, overall infants required a large number of habituation trials, which suggests that establishing the mappings between the novel words and their referents was a challenging task for infants at this age. In fact, half of the infants in the study (n = 10) did not reach the habituation criterion. These infants proceeded to the test phase after reaching the maximum of 24 habituation trials and were included in the final analyses, following previous studies (e.g., Fennell & Waxman, 2010; Yoshida et al., 2009).

Infants' total looking duration in the habituation trials presenting the /diːt/ pairing (M = 95.69 sec, SD = 48.35) and the /duːt/ pairing (M = 93.08 sec, SD = 43.79) did not differ, $t(19)$ = .509, p = .617, d = .056, which demonstrates that infants did not direct greater attention or show a visual preference for one of the objects. Similarly, infants' looking duration to the object paired with /diːt/ (M = 6.31 sec, SD = 2.58) and the object paired with /duːt/ (M = 4.32 sec, SD = 2.81) did not differ in the baseline trials, $t(19)$ = 1.740, p = .098, d = .738, further suggesting neither image was more visually salient or attractive than the other.

To assess infants' performance in the visual choice portion of the task, their proportion of looking duration to each object was compared for the test and baseline

trials. The baseline trials allowed us to capture any initial visual preferences that the infants had for the two objects. Therefore, it was predicted that if infants selected the target object as the referent for the target word, their proportion of looking duration to that target object [target / (target+distracter)] would increase in the test trials compared to the baseline trials. A 2 (trial: baseline, test) x 2 (word: /diːt/, /duːt/) ANOVA yielded no main effect of trial, $F<1$, no main effect of word, $F(1, 19) =$ 1.239, $p = .280$, $\eta^2 = .061$, but a significant word by trial interaction, $F(1, 19) = 5.007$, $p = .037$, $\eta^2 = .209$ (Figure 5). The proportion of looking time that infants directed to the /duːt/ object increased in test trials compared to baseline but this difference did not reach statistical significance, $t(19) = 1.741$, $p = .098$, $d = .544$. However, the proportion of looking time to the /diːt/ object decreased significantly in test compared to baseline, $t(19) = -2.397$, $p = .027$, $d = .72$. That is, rather than directing their gaze to the object corresponding to the word /diːt/, infants looked longer at the incorrect object.

Finally, we tested infants' performance in the test trials by comparing looking proportion to the target object to chance (.5). While unlikely given the results presented above, it is possible that infants' looking to target did not differ in test compared to baseline, but they did show preference to the target object over the distracter after hearing its corresponding word. One-sample t-test analyses conducted for each label confirmed that this was not the case for /diːt/ trials, $t(19) = 1.270$, $p = .219$, $d = .288$, or for /duːt/ trials, $t(19) = .247$, $p = .808$, $d = .051$. Therefore, infants did not learn the object-word pairings in this task.
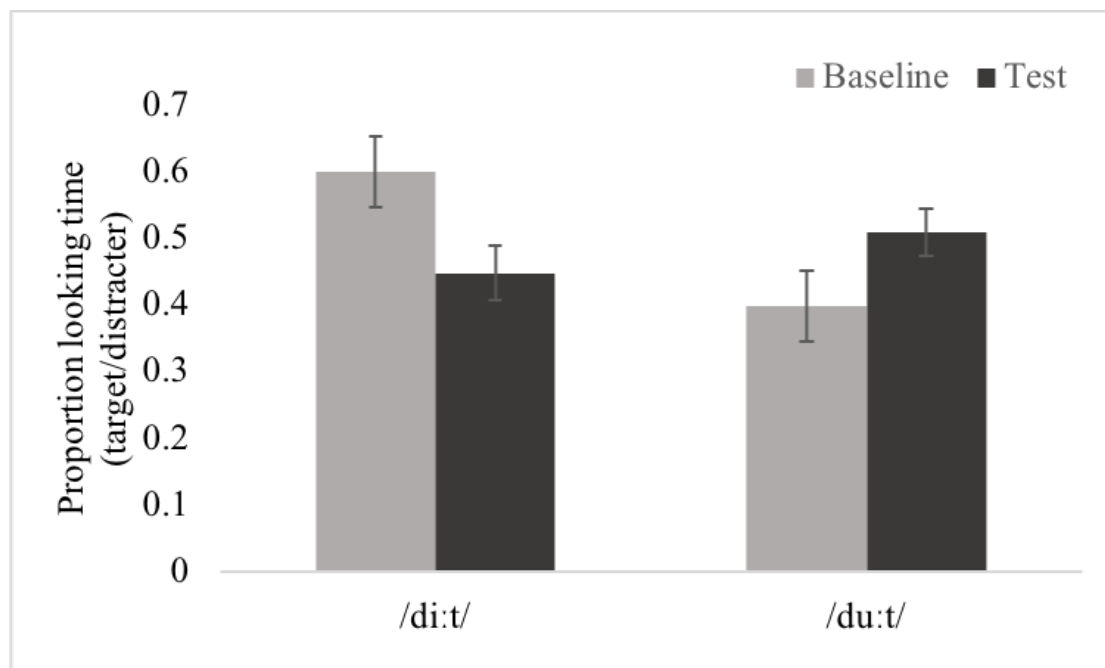
Figure 5. Proportion of looking duration at target in response to the words /diːt/ and /duːt/ in test and baseline trials in Experiment 2 (error bars represent SEM).

Discussion

This study assessed the ability to distinguish vowel minimal pairs in 15-month-old infants acquiring Australian English. We first tested whether vowel acoustic properties presented to the infants explained previous results, showing that infants at this age only successfully discriminate contrasts with the least phonetic overlap, namely /diːt/-/duːt/. Subsequently, we assesed word-learning of this vowel minimal pair using the visual choice task. Despite the simplicity of this word-learning task, infants were not successful. These results combined demonstrate that 15-month-olds' performance is affected by both the acoustic properties of the vowel categories comprising the minimal pair and whether the task is auditory discrimination or word learning.

Experiment 1 allowed us to directly test the effects of vowel acoustic properties on infants' phoneme discrimination. Following previous studies (Stager & Werker, 1997), it would be expected that if the failure at establishing mappings with vowel minimal pairs observed in previous studies (Escudero et al., 2014) is due solely to the cognitive demands involved in word learning, then infants should show successful discrimination of all contrasts presented in this type of auditory discrimination task. However, infants' performance was determined by the specific vowel contrast, whereby successful discrimination was only evident for the contrast characterised by the least amount of acoustic overlap (/diːt/-/duːt/). Unlike the results with consonant minimal pairs (Stager & Werker, 1997; Werker et al., 2002), for some vowel minimal pairs, word learning difficulty is connected to a developmental difficulty in phoneme discrimination.

Results of Experiment 1 are thus in line with studies showing that phoneme discrimination is not an 'all or none' phenomenon, and it does not take place at the same time for all contrasts of a child's native language. For instance, in the case of consonants, Tagalog infants can discriminate /ma/-/na/ at six to eight months but take another four months to discriminate /ɲa-/na/ (Narayan, Werker, & Beddor, 2010). Similarly, Canadian English infants take at least 12 months to discriminate /d/-/ð/ (Polka, Colantonio, & Sundara, 2001). For vowels, infants cannot discriminate Japanese vowel length contrasts until they are nine and a half months of age (Sato, Sogabe, & Mazuka, 2010), and Dutch infants can only discriminate /i/-/ɪ/ by 11 to 12 months (Liu & Kager, 2016). Thus, although seminal work (Polka & Werker, 1994; Werker & Tees, 1984) suggests that the time frame for language-specific tuning takes place around six and 10 months for vowels and consonants respectively, the specific acoustic salience of a contrast may play a role in infants' performance (Liu & Kager,

2016). Indeed, for vowel contrasts that involve dynamic properties, such as vowel trajectories, adult-like performance for the Australian English contrast /i/-/ɪ/ (Williams et al. 2018) seems to emerge at 17 months (Escudero et al. 2018).

The current findings demonstrate that specific properties of how individual phonemes are realised continue to play a role in phoneme discrimination beyond the first year of life. Models that take into account acoustic property distributions for phonological and lexical learning (e.g. Boersma, Escudero, & Hayes, 2003; van Leussen & Escudero, 2015 for infant learners; Kleischmidt & Jaeger, 2015 for adult learners) can easily explain why vowels with more variation in their realisations pose a large challenge. For instance, van Leussen and Escudero (2015) demonstrated that phonetic variability that leads to overlap in two neighbouring vowels, can in turn lead to only moderate categorisation accuracy, as simulated native adult listeners only achieved 80% accuracy when presented with vowels that had distributions that overlapped along the F1 continuum.

Despite the use of the contrast characterised by reduced acoustic overlap and by successful discrimination, Experiment 2 showed that infants were not successful at word learning, suggesting that they were not sensitive to the target vowel contrast in this context. This was despite the fact that our experiment used a paradigm proposed to reduce the cognitive demands of experimental word-learning tasks, and that has elicited successful word-learning of consonant minimal-pairs in infants as young as 14 months. The null results of this experiment must be interpreted with caution, but it can be argued that in a word-learning context, the great variability in the acoustic realisation of the vowels used here, even when the vowels had less overlap in their trajectory, prevented infants from establishing and learning a robust correspondence between the auditory and visual stimuli.

If acoustic overlap or the magnitude of the phonetic contrast, as proposed in Escudero et al. (2014; 2018), determine phoneme perception and word learning success in the second year of life, we predict that contrasts with large phonetic distinction, e.g. /i/-/a/, would be easier to learn. In general, the amount of phonetic distance between the vowels of a contrast would predict success and therefore better results are expected for the minimal pair (/dɪt/-/duːt/), as was shown in Escudero et al. (2018). This hypothesis could also explain why 14-month-olds learned /bɔn/-/bɪn/ in Fikkert (2010) but not /dɔn/-/dɪn/. These vowels seem to be acoustically closer together in the d-vowel-n context than in the b-vowel-n context (e.g. Escudero et al. 2009; Elvin et al. 2016), which would make the former context more difficult to learn. Although the magnitude of the phonetic contrast hypothesis has the potential of explaining previous findings, a detailed acoustic analysis of the stimuli used in previous studies would be required to support its predictions. This goes beyond the scope of the current paper but would be a great topic for a meta-analysis paper of all stimuli used in early word learning studies to test whether the sound contrast or task determines early word learning success.

Yoshida et al. (2009) proposed that transition to early word-learning success at 17 months results from development in general cognitive abilities and to gradual learning of phonological principles in the native language, such as the fact that phoneme changes signal lexical differences. They further suggested that these two possibilities should be disentangled for us to understand how the ability to use phonetic detail in word learning develops and strengthens. Our results for Experiment 1 combined with those reported in Escudero et al. (2018) suggest that the gradual learning of phonological distinctions depends on the phonetic properties of the sounds involved in a contrast. In other words, regardless of development of cognitive abilities

or the cognitive challenge for the infant, words containing phonetic variation that increases overlap between two sounds, such as contrasts involving dynamic formant trajectories, result in a slower learning trajectory.

However, it is noteworthy that our conclusions regarding infants' word-learning ability are based on the null findings of Experiment 2, so they must be considered with caution. Our task was an adaptation of the paradigm developed by Yoshida et al. (2009), which they proposed to be a successful method for assessing word-learning in young infants for whom the typically-used Switch task may be too taxing. However, our design did not include a control condition that could discard the possibility that infants in our study did not show evidence of learning due to the properties of this experimental task. Some alternative explanations for the null results in Experiment 2 based on the properties of the paradigm may be that our visual stimuli may have introduced unexpected challenges for the infants, although they led to successful learning in previous studies (Kalashnikova et al. 2015, Kalashnikova & Burnham, 2016). Additionally, infants' looking time was coded as a sum of looking duration to each object during the testing phase, which relied on the experimenter's ability to quickly code left and right looks online, and that importantly, did not provide us with a measure of infants' gaze patterns unfolding over time. The use of an eye-tracking measure as well as the inclusion of a control condition with labels that infants are expected to learn with ease (e.g., non-minimal pairs or consonant minimal pairs) in future designs should allow us to disentangle more clearly the effects of task demands and stimulus properties on infants' developing speech perception and word-learning skills.

Furthermore, while we used a simplified word-learning task following Yoshida et al. (2009), it is possible that other task adaptations would have yielded

successful learning, particularly for a vowel contrast that only differs in F2. Infants' word-learning performance could be fostered by statistical information to reinforce the word-object mappings (Escudero et al., 2016), by providing additional referential cues in the task (Fennell & Waxman, 2010), or by presenting labels produced by multiple speakers (Rost & McMurray, 2009). It is also possible that young infants are able to navigate these challenging acoustic contrasts outside the laboratory environment by relying on additional cues that support word-learning available in their language input. While the speaker who produced the stimuli was asked to use infant-directed speech and acoustic properties compared well to those of female adult-directed speech (Elvin et al. 2016), it is possible that infants' linguistic input is less challenging than the auditory stimuli we presented here. Our current research investigates this possibility.

Given that 17-month-olds are able to overcome both the challenges presented by the specific task and the acoustic variability in the target vowels (Escudero et al. 2018), it can be concluded that success at early word learning indeed depends on both the development of general cognitive abilities and increased experience not only with the native language phonology but with specific phonemes and their phonetic properties. Additionally, infants must learn to selectively attend to the abundant acoustic cues present in their linguistic input to differentiate the cues that are meaningful but not lexically relevant in their language. Infants must not disregard these cues entirely as attention to low-level acoustic detail can assist processing by allowing them to contend with the vast variation in the linguistic input (e.g., Galle et al., 2015; Rost & McMurray, 2009; 2010). They should attend to communicatively-relevant segmental and supra-segmental information but they must develop competence about what cues to consider to support native phoneme perception (see

Galle & McMurray, 2014 for a related view). This proposition is consistent with the framework for Processing Rich Information from Multidimensional Interactive Representations (PRIMIR), which states that while infants' reliance on phonetic dimensions decreases over time as phonemes emerge, task demands and developmental level ultimately determine learning and perceptual biases only to a lesser extent (Curtin, Byers-Heinlein, & Werker, 2011; Werker & Curtin, 2005).

In sum, the present study demonstrates that the magnitude of the phonetic distance between the vowels of a contrast has an impact on phoneme discrimination and on word learning abilities beyond the first year of life. When predicting developmental trajectories in early childhood, models of phonological and lexical development should take into account that factors such as acoustic salience, phonetic variation and frequency distributions interact with and may sometimes overrule the role of task demands.

References

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52*(4), 1135.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words, 167*(224), 233-277.

Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In *Proceedings of the 15th International Congress of Phonetic Sciences* (Vol. 1013). Barcelona.

Burnham, D. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7*(3), 207–239.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435. https://doi.org/10.1126/science.1069587

Cohen, L. B., Atkinson, H, H., & 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.

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913-934.

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development : PRIMIR in focus. *Journal of Phonetics, 39*(4), 492–504. https://doi.org/10.1016/j.wocn.2010.12.002

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

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences, 104*(41), 16027–16031.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303-306.

Elvin, J., Williams, D., & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *Journal of the Acoustical Society of America*, *140*, 576–581.

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America*, *126*(3), 1379-1393.

Escudero, P., & Benders, T. (2010). Phonetic and phonological approaches to early word recognition: empirical findings, methodological issues, and theoretical implications. In M. Everaert, T. Lentz, H. De Mulder, Ø. Nilsen, & A. Zondervan (Eds.), *The Linguistics Enterprise: From Knowledge of Language to Knowledge in Linguistics*. Amsterdam: John Benjamins.

Escudero, P., Best, C. T., Kitamura, C., & Mulak, K. E. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents, *Frontiers in Psychology, 5*, 1059.

Escudero, P., Mulak, K. E., & Vlach, H. A. (2016). Infants encode phonetic detail during cross-situational word learning. *Frontiers in Psychology*, *7*, 1419.

Escudero, P., Mulak, K. E., Elvin, J., & Traynor, N. M. (2018). "Mummy, keep it steady": phonetic variation shapes word learning at 15 and 17 months. *Developmental Science*, e12640.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*(2/3), 245–264. https://doi.org/10.1177/00238309030460020901

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. https://doi.org/10.1111/j.1467-8624.2010.01479.x.

Fikkert, P. (2010). Developing representations and the emergence of phonology: Evidence from perception and production. *Laboratory Phonology*, *10*(4), 227-255.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and speech*, *5*(4), 171-189.

Galle, M. E., & McMurray, B. (2014). The development of voicing categories: A quantitative review of over 40 years of infant speech perception research. *Psychonomic bulletin & review*, *21*(4), 884-906.

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2015). The role of single talker acoustic variation in early word learning. *Language Learning and Development*, *11*(1), 66-79.

Gerrits, E. (2001). *The categorisation of speech sounds by adults and children: a study of the categorical perception hypothesis and the developmental weighting of acoustic speech cues*. Utrecht, The Netherlands: LOT Dissertation Series 42.

Giezen, M. R., Escudero, P., & Baker, A. E. (2010). Use of acoustic cues by children with cochlear implants. *Journal of Speech, Language and Hearing Research*, *53*(6), 1440–1457. doi:10.1044/1092-4388(2010/09-0252)

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. doi:10.1111/cdev.12269

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28*(4), 377-396.

Kalashnikova, M., & Burnham, D. (2016). Novel word learning, reading difficulties, and phonological processing skills. *Dyslexia*, *22*(2), 101-119.

Kalashnikova, M., Mattock, K., & Monaghan, P. (2015). The effects of linguistic experience on the flexible use of mutual exclusivity in word learning. *Bilingualism: Language and Cognition*, *18*, 626–638.

Kalashnikova, M., Carignan, C., & Burnham, D. (2017). The origins of babytalk : smiling , teaching or social convergence. *Royal Society Open Science, 4*, 170306.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*, 148-203.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V,

    Ryskina, V. L., … Lacerda, F. (1997). Cross-Language Analysis of Phonetic

    Units in Language Addressed to Infants. *Science, 277*(5326), 684–686.

    https://doi.org/10.1126/science.277.5326.684

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech

    using signal detection theory and multidimensional scaling. *The Journal of the

    Acoustical Society of America*, *97*(1), 553-562.

Liu, L., & Kager, R. (2016). Perception of a native vowel contrast by Dutch

    monolingual and bilingual infants: A bilingual perceptual lead. *International

    Journal of Bilingualism*, *20*(3), 335-345.

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in

    infant-directed speech: More breathy and more variable, but not clearer.

    Cognition, *166*, 84-93.

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization

    develops slowly through adolescence. *Developmental Psychology*, *54*(8), 1472.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant

    directed speech and the development of speech perception: Enhancing

    development or an unintended consequence?. *Cognition*, *129*(2), 362-378.

Mulak, K. E., Bonn, C. D., Chládková, K., Aslin, R. N., & Escudero, P. (2017).

    Indexical and linguistic processing by 12-month-olds: Discrimination of

    speaker, accent and vowel differences. *PloS one*, *12*(5), e0176762.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between

    acoustic salience and language experience in developmental speech perception :

    Evidence from nasal place discrimination. *Developmental Science, 13*(3), 407–

    420. https://doi.org/10.1111/j.1467-7687.2009.00898.x

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition, 98*(1), 13-30.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio, 2*(2), 203-230.

Nittrouer, S., & Lowenstein, J.H. (2014). Dynamic spectral structure specifies vowels for adults and children. *Language and Speech*, *57*, 487–512.

Object DataBank [Computer Software] (1996). Providence. RI: TarrLab, Brown University.

Pater, J., Stager, C., Werker, J. F., & District, T. V. (2004). The Perceptual Acquisition of Phonological Contrasts. *Language, 80*(3), 384–402.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253-260.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of/d/– /ð/perception: evidence for a new developmental pattern. *The Journal of the Acoustical Society of America, 109*(5), 2190-2201.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology Human Perception & Performance, 20*(2), 421–435.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 10). (pp. 243-335). New York: Academic Press.

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

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608-635.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology*, *46*(1), 106-119. https://doi.org/10.1037/a0016718.

Singh, L., Fu, C. S., Tay, Z. W., & Golinkoff, R. M. (2018). Novel word learning in bilingual and monolingual infants: evidence for a bilingual advantage. *Child development*, *89*(3), e183-e198.

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*(2), 173-189.

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, *4*(2), 157-178.

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. https://doi.org/10.1038/41102

Swingley, D., & Aslin, R. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166.

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language*, *46*(1), 39-56.

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis.
Developmental Psychobiology, 56(2), 179–191.
https://doi.org/10.1002/dev.21179

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a
second language: the L2LP model revised. *Frontiers in Psychology*, *6*.
https://doi.org/10.3389/fpsyg.2015.01000

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. A. (2005). PRIMIR: A developmental framework of infant
speech processing. *Language Learning and Development, 1*(2), 197–234.
https://doi.org/10.1080/15475441.2005.9684216

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants ʼ
Ability to Learn Phonetically Similar Words : Effects of Age and Vocabulary
Size. *Infancy, 3*(1), 1–30.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech-perception: Evidence for
perceptual reorganization during the 1st year of life. *Infant Behavior and
Development, 7*(1), 49–63. https://doi.org/Doi 10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a
new synthesis. *Annual Review of Psychology*, *50*(1), 509-535.

Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word
learning. *Trends in Cognitive Sciences, 9*(11), 519–527.
https://doi.org/10.1016/j.tics.2005.09.003

Williams, D., Escudero, P., & Gafos, A. (2018). Spectral change and duration as cues in Australian English listeners' front vowel categorization. *The Journal of the Acoustical Society of America*, *144*(3), EL215-EL221.

Yildirim, I., & Jacobs, R. A. (2013). Transfer of object category knowledge across visual and haptic modalities: experimental and computational studies. *Cognition, 126*, 135–148.

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. https://doi.org/10.1111/j.1467-7687.2008.00789.x