Infants exposed to fluent natural speech succeed at cross-gender word recognition

Marieke van Heugten and Elizabeth K. Johnson

University of Toronto

3359 Mississauga Road North, Mississauga, ON, L5L 1C6, Canada

marieke.vanheugten@utoronto.ca; elizabeth.johnson@utoronto.ca

Correspondence should be addressed to Elizabeth K. Johnson, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road North, Mississauga, Ontario, Canada, L5L 1C6. E-mail: elizabeth.johnson@utoronto.ca. Phone: 905-569-4785. Fax: 905-569-4326.

Abstract

Purpose: To examine the possibility that early signal-to-word form mapping capabilities are robust enough to handle substantial indexical variation in the realization of words.

Method: 7.5-month-olds were tested using the Headturn Preference Procedure. Half of the infants were exposed to words embedded in passages spoken by their mother and tested on lists of trained and novel isolated words spoken by their father. The other half of the infants were yoked pairs listening to unfamiliar speakers.

Results: In the test phase, infants listened longer to trained than to novel words, indicating that they successfully segmented the words from the passages. This result was not modulated by infants' familiarity with the speaker.

Conclusions: Under more naturalistic listening conditions, 7.5-month-olds exhibit the ability to recognize words in the face of substantial indexical variation regardless of whether or not speakers are familiar. This suggests that early word representations are, at least to some extent, independent of the speaker's gender and may reflect sophisticated abstraction capabilities on the part of the infants, which would render extreme episodic models of early speech perception untenable. Additional research using similarly ecologically valid testing methods is called for to elucidate the precise nature of early word representations.

Keywords: infant speech perception, word recognition, lack of invariance, indexical information, exemplar representations

Infants exposed to fluent natural speech succeed at cross-gender word recognition

Spoken word recognition is complicated by the lack of one-to-one mappings between acoustic forms and their corresponding lexical entries. Factors such as speaker gender, emotional affect, and accent can greatly alter the acoustic realization of a word. How do listeners cope with this lack of invariance in the speech signal? Adult listeners possess adaptable signal-to-word mapping processes enabling them to link phonetically distinct tokens of a word to the same underlying representation (see Cutler, 2008 for an overview). Infants, in contrast, neither have a lexicon full of possible word candidates stored in their memory nor do they possess adult-like signal-to-word mapping abilities. In fact, as a first step in building up their vocabulary, they must learn to extract reliable representations from the variable signal to start establishing the word forms to which they can later attach meaning.

Infants begin to exhibit some ability to recognize variable word forms from speech very early on. Word segmentation studies have shown that English-learning 7.5-month-olds exposed to isolated word tokens (e.g., *cup* and *dog*) later listen longer to passages containing different tokens of these words than to passages containing novel words (e.g., *feet* and *bike*). Similarly, 7.5-month-olds exposed to words embedded in passages listen longer to isolated tokens of these trained words as opposed to isolated tokens of novel words, indicating that previously presented words are subsequently recognized across new utterances (Jusczyk & Aslin, 1995; Seidl & Johnson, 2006). There is, however, substantial evidence that the ease with which infants map different tokens of a word onto the same underlying word representation is dependent on the degree of acoustic overlap between these items. More specifically, if two realizations of a word are not perceptually similar, infants' early word recognition abilities are severely compromised. Studies first presenting 7.5-month-olds with word tokens in a male voice, for example, show that infants subsequently do not recognize these words when they are produced in a perceptually distinct female voice (Houston & Jusczyk, 2000). Likewise, studies show that words in a happy affect are not recognized when they are later produced in a neutral affect (Singh, Morgan, & White, 2004) and that infants presented with words in Spanish-accented or even Canadian-accented English do not appear to recognize these words in a Midwestern American English accent (Schmale, Cristià, Seidl, & Johnson, 2010; Schmale & Seidl, 2009; see also Best, Tyler, Gooding, Orlando, & Quann, 2009 for related findings).

What does this imply for infants' speech processing abilities? Reports that 7.5-month-olds initially experience difficulty generalizing across acoustically distinct word form tokens have been taken as support for some of today's most influential models of early word recognition such as WRAPSA (Word Recognition and Phonemic Structure Acquisition; Jusczyk, 1993) and PRIMIR (Processing Rich Information from Multidimensional Interactive Representations; Werker & Curtin, 2005). These models assume that language learners store words in an instance-specific fashion. By considering early word forms to be exemplar based, these models can eloquently explain infants' apparent difficulty in coping with indexical variation. Previously encountered words presented in the same voice, affect, and accent share great acoustic overlap and hence are readily recognized as being the same word. Previously encountered words that differ in speaker gender, affect, or accent, in contrast, are perceptually less similar and therefore pose a greater challenge for word recognition. According to these models, only once a sufficient number of phonetically diverse tokens of a new word have been encountered and (individually) stored, a distinct word cluster will emerge and generalization can take place, allowing infants to more easily cope with indexical variation (e.g., Jusczyk, 1993). In line with this view, only infants exposed to highly variable tokens of a word appear to form robust representations and hence succeed in recognizing these same words when they are realized in an acoustically distinct fashion (Houston, 1999; Singh, 2008).

Episodic models indeed account for infants' observed difficulty mapping acoustically distinct tokens onto the same representation. Consistent with these models, research with adults has suggested that listeners do store some acoustic detail when listening to speech (e.g., Church & Schacter, 1994; Goldinger, 1998; Palmeri, Goldinger, & Pisoni, 1993). However, there is also evidence suggesting that an extreme exemplar view of speech perception is untenable (see Cutler, Eisner, McQueen, & Norris, 2010 for a review). If adult speech perception relies on abstract representations, then clearly at some point during language acquisition, infants must begin to develop word form representations that are at least partially abstract. It is thus very well possible that infants' emerging lexicon also contains some form of abstract representations. That is, even though indexical variation may hinder word recognition, 7.5-month-olds might still be able to cope with acoustic variability when tested under slightly more everyday-like listening conditions. Storage of abstract representations in the mental lexicon is supported by studies showing that infants are able to generalize across syllables or phonemes produced by different speakers (Dehaene-Lambertz & Pena, 2001; Jusczyk, Pisoni, & Mullennix, 1992), even when those speakers differ in gender (Kuhl, 1979).

In this study, we therefore examine the possibility that infants are able to map acoustically distinct tokens onto the same word representations. Past studies first presented infants with isolated words recorded by voice actors and subsequently tested their recognition of these words embedded in fluent speech recorded by another voice actor. In the current study, in contrast, infants are first presented with passages containing repeated tokens of a target word recorded by either their own mother or the mother of another similarly-aged infant. They are then tested on their recognition of these target words when produced in isolation by either their own father or the father of another infant. These listening conditions, more comparable to those in infants' daily life, potentially facilitate word form recognition in several ways. First, familiarizing infants with passages containing target words rather than lists of isolated words likely increases exposure to the speaker-specific idiosyncrasies in the passages and introduces infants to a wider range of the speaker's idiosyncratic realizations of speech segments and utterance-level prosody. Unlike isolated words, fluent speech may thus provide infants with the opportunity to adapt to the systematic variation of an unfamiliar voice, thereby allowing them to start building more robust and generalizeable representations of a word form (cf. Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras, Alexander, & Nygaard, 2009 for perceptual learning with adults after brief exposure). In addition, the increased variability in the realization of the target words available from the training passages (relative to the word lists in previous studies) may also facilitate word form extraction (cf. Rost & McMurray, 2009; Singh, 2008). Second, using speakers who, at the time of recording, were parents of 7.5-month-olds rather than voice actors may have resulted in utterances that are more representative of the input typically received by children. Parents, unlike the voice actors in previous studies, are speaking to infants of the tested age range on a daily basis. Because infant-directed speech differs from adult-directed speech in terms of pitch, accent, and vowel space (Burnham, Kitamura, & Vollmer-Conna, 2002; Kuhl et al., 1997), parental recordings may contain more naturalistic acoustic cues that enhance word recognition than the recordings made by actors used in past studies. Third, by presenting half of the infants with training and test materials produced by their own parents, we are able to test for the possibility that coping with variability in the realization of words may be facilitated by long-term exposure to the speakers.

Two groups of 7.5-month-olds are tested. The first group is exposed to passages recorded by their mothers. They are then tested on lists of isolated word tokens spoken by their fathers. Both speakers are thus highly familiar. The second group of infants hears the same stimuli as the first group. As the voices belong to another infant's parents, infants in this group have not had any prior access to the speakers' voices, apart from the exposure phase. If the use of more naturalistic stimuli in the exposure phase indeed yields more robust word encoding, infants should succeed in the cross-gender word recognition task. We thus hypothesize that infants will listen longer to trained than to novel words (cf. Jusczyk & Aslin, 1995; Seidl & Johnson, 2006) despite the change in gender of the speaker between the exposure and the test phase. Moreover, if voice familiarity plays a role, this effect should be more pronounced for those infants tested on their own parents' voices than for those tested on unknown voices. The predicted preference for trained over novel words may, in other words, be larger for the former as compared to the latter group.

Method

Participants

Forty-eight normally developing monolingual English-learning 7.5-month-olds with no reported hearing problems from the Greater Toronto Area were tested in this study (age range: 220 - 248 days; 25 girls). An additional eleven infants were tested, but excluded from the analysis due to failure to complete the study or extreme fussiness. Infants received a small gift and a certificate in appreciation of their participation.

Materials

Prior to test, twenty-four mothers and fathers whose infants were recruited to participate in the study were audio taped in a sound-attenuated booth. All parents were either native English speakers or had learned English before five years of age in an English-speaking country.[1]

Mothers recorded two of four six-sentence passages in infant-directed speech (see Appendix). Each passage contained a target word *(boat*, *cup*, *pear*, or *toque* – in Canadian English, 'toque' is the commonly used word for a knitted hat) that was repeated once in every sentence. Within a passage, this target word appeared twice in sentence-initial position following a function word (e.g., *Her boat had white sails*), twice sentence-medially (e.g., *That horn on the boat was really loud*), and twice in sentence-final position (e.g., *This girl will steer my big boat*). For each mother, an exposure video was created consisting of three alternating repetitions of each of the two recorded passages accompanied by a static photo of the mother. All exposure videos were between 110 and 120 seconds long and contained 18 tokens of each of the target words. Because infants were tested on each of the four target words (two being trained and two being novel), fathers recorded multiple isolated instances of *all four* target words, which were edited to create different test word lists. In each list, five tokens of a single target word were repeated three times, separated by an interstimulus interval of approximately 600 ms. Test lists were 17.31 seconds long. The average pitch level measured from the periodic portion of the vowel (or the vowel plus [ɹ] in the case of 'pear') was 187 Hz for male speakers and 249 Hz for female speakers.[2]

Procedure

Infants were seated on their parent's lap in front of a TV screen in a double-walled sound-attenuated booth. The experimenter, located outside the booth, started the exposure phase as soon as the infant oriented towards the TV screen. The movie continuously played until the end, after which the test phase started. In the test phase, an adapted version of the Headturn Preference Procedure (as used by Jusczyk & Aslin, 1995) tested infants' recognition of the trained words across different genders. First, a red light at the panel in front of the infant started flashing. Once the infant oriented towards this light, one of the two lights at the side panels started to flash. As soon as the child turned towards this flashing light, the word list started playing from the loudspeaker mounted underneath the blinking light. Trials either played until the end of the list or until the infant looked away for two seconds. Parents listened to masking music over closed headphones so that they could not bias the child's behavior.

Design

Half of the infants were presented with their own parents' voices, while the other half were yoked pairs listening to the same stimuli in unfamiliar voices. The target words used during exposure (*boat* and *toque*, *cup* and *boat*, *toque* and *pear*, or *pear* and *cup*) were counterbalanced across conditions, as was the order of presentation of the two passages in the exposure phase. Infants were tested on all four test lists. For each infant, two of these test lists contained trained and two contained novel words. Test lists were randomly presented once in each of three blocks (twelve test trials in total).

Results

First, orientation times were calculated for each infant in each trial. Orientation times for trials more than 2.5 standard deviations away from the infant's mean (3 out of 576 data points) were discarded. Mean orientation times to trained and novel words were then calculated for each infant separately. On average, infants tested on their parents' voices listened to lists with trained words for 10.18 s and to lists with novel words for 9.45 s, with 18 out of 24 infants listening longer to the trained words. Infants tested on unfamiliar voices listened to lists with trained words for 10.31 s on average and to lists with novel words for 9.51 s, with 16 out of 24 infants listening longer to the trained words (see Figure 1). A mixed 2×2 ANOVA with Word Familiarity (trained vs. novel words) as a within-subject factor and Voice Familiarity (own parents' voices vs. unknown voices) as a between-subject factor revealed a main effect of Word Familiarity ($F(1,46) = 4.224$; $p = .046$; $\eta_p^2 = 0.084$), showing that infants listened longer to word lists containing trained words as opposed to word lists containing novel words. No other significant main effects or interactions were found (all *F*s < 1), indicating that the performance of infants presented with their own parent's voices did not differ from those presented with the voices of another infant's parents.

(Figure 1 about here)

Discussion

These results demonstrate that infants as young as 7.5 months of age are capable of coping with naturally occurring surface variation in the realization of words. More specifically, this study is the first to show that 7.5-month-olds can map word tokens produced by a female and a male voice onto the same underlying representation in a word segmentation task. This is consistent with the view that early word representations are independent of the speaker's gender, such that infants are able to map perceptually dissimilar realizations of the same word onto the same linguistic representation.

An important difference between this study and earlier infant word segmentation studies is that we exposed infants to target words embedded in fluent speech and spoken by parents rather than presenting them with target words in isolation spoken by voice actors. The use of fluent speech instead of isolated word tokens may have assisted performance on this task in at least two ways. First, exposure to fluent speech provides infants with a wider range of segmental and prosodic information. This may better enable infants to adapt to the speaker. Better speaker adaptation may, in turn, allow for the generation of more robust word representations. Second, unlike words in isolation, fluent speech embedded in context, comparable to what infants typically experience in their everyday life, naturally contains some degree of variability that could be useful for extracting robust representations of word forms.

The use of parents' voices rather than voice actors may also have assisted performance in this study in at least two ways. First, conversational partners often display convergence in speaking rate (Webb, 1972), pitch (Gregory, 1990) and phonetic realizations (Pardo, 2006). For frequently interacting partners, such as the parents in this study, this type of accommodation could lead to long-term effects evidenced even in the absence of the partner and might have led to greater acoustic similarity between the word tokens produced in the male and female voice. Since infants were previously found to generalize across acoustically similar, but not acoustically distinct voices (Houston & Jusczyk, 2000), this explanation would be especially plausible if the difference between male and female speakers in this study would be smaller than the difference between male and female speakers in Houston and Jusczyk (2000). The relative increase in pitch from male to female speakers, however, is similar between the two studies (see Footnote 1, for a more elaborate explanation). It is nonetheless possible, of course, that male and female speakers in the current study are more acoustically similar based on properties not measured here (see Johnson, Westrek, Nazzi, & Cutler, in press; Remez et al., 2011 for a related discussion). Note, however, that under everyday listening conditions, infants typically receive a large proportion of their language input from speakers who spend a lot of time together and thus may exhibit this type of vocal convergence. Second, as discussed in the Introduction, speech recorded by infants' parents (as opposed to voice actors in previous studies) may better reflect speech encountered by infants in their everyday environment. This might decrease task demands such that more cognitive resources can be devoted to building up word representations. Although we cannot determine the exact factor(s) responsible for infants' early cross-gender word recognition success in the current study, this does not detract from the primary finding that when provided with more natural listening conditions, infants are competent in coping with variability in the speech signal.

In this study, we used speaker gender as a tool to examine infants' ability to deal with the lack of invariance in the speech signal. However, variability in the speech signal is, of course, not restricted to differences in speaker gender. Other factors, such as emotional affect and accent, also affect the acoustic realization of words. Recent findings, similar to findings on cross-gender word form generalization (Houston & Jusczyk, 2000), suggest that infants initially experience difficulty overcoming these forms of variability (Best et al., 2009; Schmale et al., 2010; Schmale & Seidl, 2009; Singh et al., 2004). These studies, however, have all presented infants with lists of isolated words. This could be potentially problematic, as isolated words do not contain the acoustic, segmental, and prosodic richness of fluent speech and may therefore reduce infants' opportunity to adapt to an unfamiliar speaker and subsequently build (Schmale et al., 2010; Schmale & Seidl, 2009; Singh et al., 2004) or recognize (Best et al., 2009) the underlying representation of a word form. Using isolated word tokens may, in other words, have failed to provide infants with the information they need to form generalizeable representations. In fact, even adult listeners' ability to apply newly acquired speaker idiosyncrasies is hindered when the type of linguistic material (e.g., sentences versus isolated words) changes (Nygaard & Pisoni, 1998). Albeit speculative, it is thus possible that infants would have been found successful at coping with acoustic variability in the speech signal due to affect or accent, had these previous studies used naturalistic fluent speech material similar to the current study instead.

The finding that infants generalize across acoustically distinct word tokens raises theoretically important questions regarding the nature of infants' early word representations. While our results argue against extreme episodic models lacking any form of abstraction, they do not rule out exemplar-based theories of early speech perception in general. That is, detailed memory traces including talker-specific information may be stored in infants' mental lexicons, attached to an emergent and more abstract prototypical representation of a word. Alternatively, word representations may be speaker-independent even at the earliest stages. For example, the developing lexicon may consist of abstract phonological words stripped off from lexically irrelevant information. In such an abstractionist view, indexical information would still play an important role in early word recognition, but only at the prelexical level (see Eisner & McQueen, 2005; Kraljic & Samuel, 2006, 2007; McQueen, Cutler, & Norris, 2006 for evidence that talker adaptation takes place at the prelexical level in adults). Future studies should disentangle these two possibilities.

Interestingly, infants' ability to cope with cross-gender variation in the realization of words held regardless of whether infants were presented with their own parents' or unfamiliar voices, suggesting that long-term exposure to speaker-specific idiosyncrasies might not have an effect over and above the effect of short-term exposure to a previously unknown speaker's voice. While previous work demonstrates that voice familiarity can enhance infants' speech processing when the familiar voice is presented concurrently with an unfamiliar speaker in the background (Barker & Newman, 2004), this finding may have been due to the more challenging listening conditions in that study. That is, if speaker adaptation in fluent speech is fast and at ceiling, and hearing a few sentences at most is sufficient for accommodating unfamiliar speakers in clear speech, it is plausible that long term voice familiarity only helps under more adverse listening conditions (e.g., speech in noise or multiple background talkers), when accommodation is more challenging. This interpretation nicely aligns with the observation that adult studies showing an advantage for processing speech produced by familiar over unfamiliar speakers typically involve the presentation of speech in noise (e.g., Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1998).

In short, while exemplar-based models might be accurate in their observation that recognizing words that vary greatly in their surface realization is more challenging than recognizing words that sound similar, our findings suggest that infants' early signal-to-word mapping skills under more naturalistic listening conditions are nonetheless sufficiently flexible to allow for non-trivial generalizations, such as across adult speakers' gender. Even at 7.5 months of age, infants may thus possess a readily available abstraction mechanism that allows them to generalize across phonologically irrelevant acoustic dimensions of the speech signal. Future models of early word recognition need to take into account these enhanced speech perception abilities when describing the transition from infant into adult listeners.

Acknowledgements

This research was supported by a Prins Bernhard Cultuurfondsbeurs awarded to MvH and CFI and NSERC grants awarded to EKJ. The authors thank the members of the ILSL for their assistance in running this study, Huiwen Goy for insightful comments on earlier manuscript drafts, and Derek Houston for the suggestion that partners' long-term vocal convergence may lead to greater acoustic similarity of the two partners' voices. Portions of this work were presented at the Annual Conference of the Canadian Acoustical Association (Niagara-on-the-Lake, 2009) and the 17th International Conference on Infant Studies (Baltimore, 2010).

References

Barker, B. A. & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition, 94*, B45-B54.

Best, C.T., Tyler, M.D., Gooding, T.N., Orlando, C.B., & Quann, C.A. (2009). Development of phonological constancy: toddlers' perception of native- and Jamaican-accented words. *Psychological Science, 20*, 539-542.

Bradlow, A.R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707-729.

Burnham, D., Kitamura, C. & Vollmer-Conna, U. (2002). What's new pussycat? On talking to babies and animals. *Science, 296*, 1435.

Church, B. A. & Schacter, D. L. (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental-frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 521-533.

Clarke, C. M. & Garrett, M. (2004). Rapid adaptation to foreign accented speech. *Journal of the Acoustical Society of America, 116*, 3647–3658.

Cutler, A. (2008). The abstract representations in speech processing. *Quarterly Journal of Experimental Psychology, 61*, 1601-1619.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 91-111). Berlin: de Gruyter.

Dehaene-Lambertz, G. & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport, 12*, 3155-3158.

Eisner, F. & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics, 67*, 224-238.

Goldinger, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partner's speech. *Journal of Nonverbal Behavior, 14*, 237-251 .

Houston, D. M. (1999). *The role of talker variability in infant word representations* (Unpublished doctoral dissertation). The Johns Hopkins University, Baltimore, MD.

Houston, D. M. & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1570–1582.

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (in press). Infant ability to tell voices apart rests on language experience. *Developmental Science.*

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics 21*, 3-28.

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

Jusczyk, P. W, Pisoni, D. W., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition, 43*, 253-291.

Kraljic, T. & Samuel, A.G. (2006). Generalization in perceptual learning for speech, *Psychonomic Bulletin and Review, 13*, 262-268.

Kraljic, T. & Samuel, A.G. (2007). Perceptual adjustments to multiple speakers, *Journal of Memory and Language, 56*, 1-15.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for perceptually dissimilar vowel categories. *Journal of the Acoustical Society of America, 66*, 1168-1679.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*, 684-686.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*, 1113–1126.

Nygaard, L. C. & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics, 60*, 355–376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Palmeri, T. J. Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of speaker's voice and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 19,* 309–328.

Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*, 2382-2393.

Remez, R. E., Dubowski, K. R., Broder, R. S., Davids, M. L., Grossman, Y. S., Moskalenko, M., Pardo, J. S., & Hasbun, S. M. (2011). Auditory-phonetic projection and lexical structure in the recognition of sine-wave words. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 968-977.

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

Schmale, R., Cristià, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy, 15,* 650-662.

Schmale, R. & Seidl, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science, 12,* 583-601.

Seidl, A. & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science. 9*, 566-574.

Sidaras, S. K., Alexander, J. E. D. & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America, 125*, 3306-3316.

Singh, L. (2008). Influences on high and low variability on infant word recognition. *Cognition 106*, 833-870.

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,* 173-189.

Webb, J. T. (1972). Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. In B. Pope & A. W. Seigman (Eds.), *Studies in dyadic communication* (pp. 115-133). New York: Pergamon.

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

Footnotes

[1] Not all parents spoke Canadian English. Some participating parents were born in regions of the world outside of North America where English is spoken as a first language, such as India, England, and Trinidad.

[2] The relative difference in average pitch level between the male and female recordings used in the current study was very similar to the relative difference in pitch level between the male and female recordings made by Houston and Jusczyk (2000). In Houston and Jusczyk (2000), the tokens recorded by male speakers were 34% higher in pitch than the tokens recorded by female speakers (average male pitch level: 227 Hz; average female pitch level: 306 Hz), comparable to the relative pitch level difference of 33% in the current study (average male pitch level: 187 Hz; average female pitch level: 249 Hz).

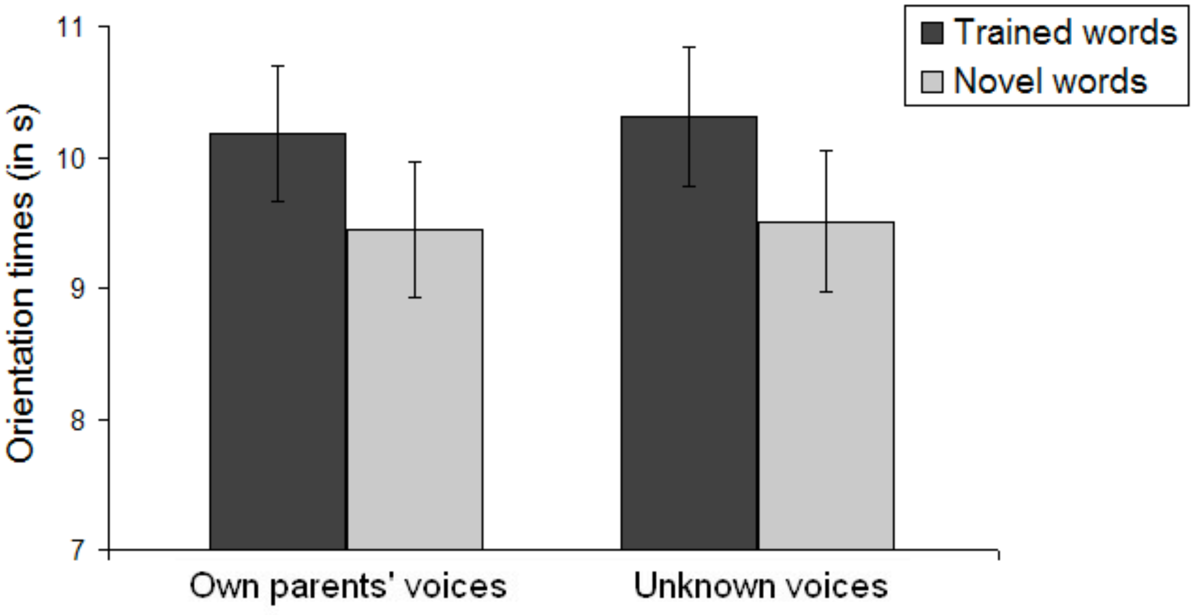

Figure 1. 7.5-month-old infants' mean orientation times in seconds (and standard errors of the difference scores) to trained and novel test words broken down by Voice Familiarity.

Appendix

Passages used in the experiment

Boat	Her boat had white sails. This girl will steer my big boat. That horn on the boat was really loud. He bought himself a new red boat. His boat could go quite fast. We always store your boat in our garage.
Toque	Your toque was soft and warm. She wore a red toque in the snow. Their brother had knitted this big toque. She liked how her toque covered my ears. Our friends also fancied the toque. His toque was blue and green.
Pear	Your pear came from my fridge. She washed her pear thoroughly. They wanted to eat a red pear. The pear in our basket looked good. Next to his pear was an apple. He enjoyed eating this big pear.
Cup	The cup was bright and shiny. A clown drank from that big cup. Some milk from his cup spilled on our rug. Your cup was filled with hot milk. They put her cup on their table. She then picked up a red cup.