Published in final edited form as:
Child Dev. 2013 ; 84(6): 2064–2078. doi:10.1111/cdev.12087.

Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent

Karen E. Mulak[1,2], **Catherine T. Best**[1,2], **Michael D. Tyler**[3,1], **Christine Kitamura**[3,1], and **Julia R. Irwin**[2,4]

[1]MARCS Institute, University of Western Sydney, Sydney NSW, Australia

[2]Haskins Laboratories, New Haven, CT, United States

[3]School of Social Sciences and Psychology, University of Western Sydney, Sydney NSW, Australia

[4]Southern Connecticut State University, New Haven, CT, United States

Abstract

By 12 months, children grasp that a phonetic change to a word can change its identity (*phonological distinctiveness*). However, they must also grasp that some phonetic changes do *not* (*phonological constancy*). To test development of phonological constancy, 16 15-month-olds and 16 19-month-olds completed an eye-tracking task that tracked their gaze to named versus unnamed images for familiar words spoken in their native (Australian) and an unfamiliar non-native (Jamaican) regional accent of English. Both groups looked longer at named than unnamed images for Australian pronunciations, but only 19-month-olds did so for Jamaican pronunciations, indicating that phonological constancy emerges by 19 months. Vocabulary size predicted 15-month-olds' identifications for the Jamaican pronunciations, suggesting vocabulary growth is a viable predictor for phonological constancy development.

Keywords

early word recognition; phonetic specificity; phonological constancy; regional accents; phonetic variability

The pronunciation of a given word can display notable phonetic variation across utterances, which adult native perceivers are able to “hear through” in order to accurately and rapidly recognize the word. The sources of variation that confront speech perceivers range from between-speaker differences in vocal tract characteristics, to within-speaker differences in emotional state and speech style, through to pronunciation patterns that differ systematically between regional accents of the same language but are shared among speakers within each accent. To recognize spoken words despite this range of phonetic variation, perceivers must identify the underlying phonological form of the word that remains constant across the variations.

Correspondence concerning this article should be addressed to Karen Mulak (k.mulak@uws.edu.au) or Cathi Best (c.best@uws.edu.au), MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith NSW 2751, Australia.

To discover the more abstract phonological form of a word, perceivers must determine whether a given phonetic difference alters the word's underlying identity, or leaves its identity intact. For example, an adult American English (AmE) perceiver who recognizes the Australian English (AusE) pronunciation of the word NICE [nɑ$^{\mathrm{e}}$s] (low back vowel with mid front offglide) as equivalent to the AmE pronunciation [na$^{\mathrm{I}}$s] (low central vowel with high front offglide) has grasped the *phonological constancy* (Best, Tyler, Gooding, Orlando, & Quann, 2009) of the word, which is not altered by the phonetic differences between AusE [ɑ$^{\mathrm{e}}$] and AmE [a$^{\mathrm{I}}$]. Conversely, perceivers must also recognize the complementary type of phonetic variation, that is, that which *does* signal a lexical distinction. For example, if the same American English perceiver recognizes that the AusE pronunciation [n^{ə}i:s] (high front vowel with centralized onglide) is not NICE but the contrasting word NIECE (pronounced [ni:s] in AmE, without an onglide), they have grasped the concept of *phonological distinctiveness* (Best et al., 2009) between the words in each accent, thus generalizing the phonological contrast across the two accents. Adults are experts at spoken word recognition because they make efficient use of phonological constancy and distinctiveness, the complementary principles that relate surface phonetic variations to more abstract phonological forms. Expert word recognition is thus flexible enough to accept most speaker and accent variation, but is usually exacting enough to accept only those variations that preserve the identity of the word.

By these definitions, the phonological form of a word is not simply a sequence of individual phonemic categories. In grasping phonological constancy, the perceiver recognizes that phonetic variation in a word can violate native-accent phonemic boundaries without changing the identity of the word, provided that it does so systematically. That is, perceivers can adapt to new phonemic boundaries in a non-native accent that are systematic in their relation to the native regional accent. As well, it is not necessary that systematic variation maintain the same distinctions across regional accents, that is, they do not always have a one-to-one equivalence across regional accents. This can be seen in cross-accent variation that contains phonemic mergers in some regional accents that are still maintained as distinct contrasting phonemes in the perceiver's native regional accent. While this would not stop perceivers from adapting to the variation, it may cause difficulties in their recognition of novel words and lexical ambiguities.

Unlike adults, young children are word recognition novices. The ability to discern the abstract phonological form of words takes time to develop based on experience with the specific phonetic patterns of the native language as spoken in their environment. Until they discover phonologically specified word forms, they must recognize words as phonetic patterns, specifically those of the pronunciations they have previously encountered. Phonetic patterns refers to the fine-grained detail of the phones as pronounced in a given accent regardless of how they would be categorized phonemically in the native accent, or when put in the context of a phonological word. This includes any subset of the panoply of nuances that can be incorporated into definitions of phonemes, including the results of coarticulation (segmental variation where articulation of a given phoneme is influenced by the articulation of phonemes before and after) and allophony (permissible variation in how a phoneme is realized, which can be governed by linguistic factors such as word position, or may be free in variation). This is contrasted with phonologically specified word forms, which are word forms that are represented without respect to concrete phonetic detail or to native accent phonetic categories, but instead to more abstract phonological structure that allows for systematic phonetic variation within phonemic categories (i.e., in a regional accent), as described above. Thus, it can be seen that phonologically specified, rather than phonetically detailed, word forms are required for cross regional-accent perception.

Many theories on the development of word learning and recognition propose that as young children learn their native language, their recognition of spoken word forms does not become based on phonological principles until 18–20 months (e.g., Swingley, 2008; Thiessen, 2007; Werker & Curtin, 2005). That premise, however, is largely based on discrimination tests of minimal pair phonetic differences. At 14 months, discrimination of minimal pairs is unreliable, with some studies showing successful discrimination (Swingley & Aslin, 2002), and others showing a failure to discriminate (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). By 18–20 months discrimination of minimal pairs has become reliable (Swingley, 2003; Swingley & Aslin, 2000). However, if task demands are reduced relative to other studies in the literature, either by testing children's discrimination of familiar rather than newly learned words (Fennell & Werker, 2003; Swingley & Aslin, 2002), or by increasing contextual support in studies using training with novel words (Fennell & Waxman, 2010; Yoshida, Fennell, Swingley, & Werker, 2009), it appears that even 14-month-olds show some sensitivity to minimal pair word differences. From these results it might be concluded that phonologically specified word forms may be emerging yet fragile at 14 months, and have become robust by 18–20 months.

Studies of minimal pair word discrimination, however, can provide only a partial picture of the nature of early word form recognition. Such results can suggest but cannot confirm whether the older or especially the younger children have even achieved an understanding of phonological distinctiveness alone. To fully address the proposition that 18-to 20-month-olds have phonologically specified word representations, it is necessary to also examine the complementary principle of phonological constancy.

Efficient phonologically based word representations cannot be overly stringent and detailed, that is, phonetically *overspecified*, as this would to lead to rejection of phonologically constant phonetic variations (e.g., an AmE-perceiver mustn't reject the AusE pronunciation of NICE as an acceptable variant of that word). On the other hand, they also cannot be too flexible and lacking in sufficient differentiating detail, that is, phonetically *underspecified*, as this would lead to acceptance of phonetic variations that actually signify phonologically distinct words (e.g., the AmE perceiver must not accept the AusE pronunciation of NIECE as an acceptable pronunciation of NICE). Minimal pair discrimination tests can tell us only whether or not children's early word forms are *under*specified; they cannot tell us whether they are *over*specified. Critically, this means that minimal pair tests alone cannot pinpoint whether the children who succeed are attending to phonetic details or to phonological structure, as both types of difference are involved in the distinction, for example between BABY and the non-word VABY. And as for the younger children who do not reliably succeed, we again can conclude only that they fail to reliably detect *either* the phonetic difference or the phonological difference.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Therefore, the possibility remains that children’s early word forms may be *over*specified, comprised of very specific and even richly detailed phonetic patterns built up through experience with the native accent input via statistical learning or exemplar registration mechanisms. This is supported by studies showing that 7.5-month-old children are unable to recognize previously familiarized words when uttered by a different speaker of the same regional accent and same gender, but can do so if the talker is held constant (Houston & Jusczyk, 2003), and do not segment words from passages when familiarized to phonetically similar foils (e.g., failing to segment BIKE from test sentences when familiarized to GIKE: Jusczyk & Aslin, 1995). That is, where reduced task demands have permitted 14-month-olds to discriminate minimal pairs (Fennell & Waxman, 2010; Fennell & Werker, 2003; Swingley & Aslin, 2002; Yoshida et al., 2009), this may have simply facilitated the children's access to less efficient, phonetically (over)specified word forms, rather than uncovered an emerging grasp of phonological distinctiveness.

To resolve these issues, the complementary skill of phonological constancy must be examined. As phonological constancy is reflected in phonetic variation that does not alter a word's underlying phonological form, testing phonological constancy calls for word pronunciation differences that contain phonetic variation, but not phonological changes (that is, stimuli that do not violate the phonological form of the word). Using regional accent variations in pronunciations of words, which offer a natural way of meeting those experimental requirements, a recent study pioneered the investigation of phonological constancy in young toddlers (Best et al., 2009). Pronunciations of a given word in two different regional accents of the same language generally share the same abstract phonological structure, but still contain perceptible phonetic differences, as in the example of NICE in AmE [na^{I}s] versus AusE [nɑ^{e}s]. Because of this, phonetic variations can be separated from phonological distinctions by careful selection of words according to the naturally occurring pronunciation differences between the native versus some other unfamiliar regional accent, providing an ideal tool for examining recognition of phonological constancy.

Best et al. (2009) examined 15- and 19-month-olds' listening preference for high frequency toddler vocabulary words versus low frequency adult words spoken either in their native accent (AmE) or in a phonetically quite different regional accent that the participants had not experienced previously (Jamaican Mesolect English: JaME). They found that both age groups preferred listening to frequent toddler words over unfamiliar low frequency adult words in AmE, indicating recognition and preference for familiar words. However, only the 19-month-olds also displayed this preference for the JaME-accented words, suggesting they had gained some command of phonological constancy that the 15-month-olds had not yet achieved.

To explain these results in relation to the prior developmental findings on discrimination of minimal pair word modifications, Best et al. (2009) proposed a perceptual attunement account, in which word forms have begun to be phonologically specified by 19 months, but not yet at 15 months, when they are still phonetically defined and specific to the child’s native accent input. The younger children exploit their perceptual attunement to the specific phonetic patterns in their language environment as a means to recognize their first words, but this hinders their ability to generalize recognition to known words spoken in other regional accents. As a result they are able to recognize words across speakers, affects, and the phonetic variations permissible within the phonemic definitions of their native accent, but not the phonetic variations of other accents they haven't experienced. However, increasing exposure to between- and within-speaker variation soon fosters the emergence of more abstract phonological knowledge of the word structure by around the time of the vocabulary spurt (around 18 months). The older children's shift of focus to the abstract phonological structure of words allows them to accept a wider range of phonetic variations in word pronunciation across accents that nonetheless preserve the invariant phonological structure of the words. Thus, according to the Perceptual Attunement account, as they mature children shift their attention from environment-specific phonetic patterns to the higher-order, more abstract phonological structure of words. This is what allows them to understand other regional-accent pronunciations they have never before encountered (Best, 1994, 1995).

However, that may not be the only possible explanation of Best and colleagues' findings. Alternatively, as noted earlier, phonological knowledge may emerge as early as 14 months if the cognitive demands of the experimental task are reduced (as in Fennell & Waxman, 2010; Fennell & Werker, 2003; Yoshida et al., 2009). If that interpretation is correct, then the developmental shift observed by Best and colleagues (2009) could simply reflect a developmental increase in the cognitive resources children can bring to bear on recognizing phonological structure in non-native pronunciations of words they know.

But that alternative account was based solely on children's discrimination of minimal pairs in their native accent (e.g., a change from /b/ to /v/ in BABY and the non-word VABY). As we have argued, success on minimal pair word discrimination tasks can be achieved through a focus on *either* phonological structure *or* detailed phonetic patterns. Thus, Fennell and colleagues interpret these findings as indicating simply that reduced cognitive demands allow 14-month-olds to more easily access the full *phonetic* details of words in their own accent. Importantly, tasks that examine phonological constancy cannot be successfully performed with phonetically detailed, input-specific word representations, as the phonetic differences between regional accents would false alarm that the word forms are different. Therefore, even if task demands were greatly reduced, 15-month-olds would fail on a phonological constancy task if, as the Perceptual Attunement account predicts, they retain a focus on specific phonetic details (overspecified word forms, with regard to the native accent) and do not yet attend to phonological structure in words. However if they can detect phonological information under reduced task demands, then they should display knowledge of phonological constancy and recognize familiar words even when spoken in an unfamiliar regional accent. This would suggest that 15-month-olds do have phonologically specified word forms, which are fragile and easily masked by demands on their more limited cognitive resources than 19-month-olds are able to access for the task (see also Best, Tyler, Kitamura, & Bundgaard-Nielsen, 2010; Best, Tyler, Kitamura, Notley, & Bundgaard-Nielsen, 2008).

To tease these possibilities apart, the present study examined whether 15- and 19-month-olds can identify familiar spoken words with their meaningful real-world referents in a task with relatively low cognitive demands. Specifically, we assessed whether they can match words spoken in the native accent versus an unfamiliar regional accent to their visual referents, as reflected in their direction of gaze between an image of the named word (target image) versus an unnamed distractor image (of a different known word). To reduce cognitive demands, we tested children’s identification of familiar words rather than newly learned words (Fennell & Werker, 2003). Further, as research shows that the more contextual information available to a child, the more they are able to access word forms (Fennell & Waxman, 2010), words were presented in a sentence context, which had the added benefit of increasing the child’s exposure to the systematic variation of the non-native regional accent. As well, a reward stimulus that played at the end of each trial regardless of performance reinforced the objective of gazing at the named picture, further reducing cognitive demands and helping to maintain the children's interest in the task. We also assessed each child’s expressive vocabulary, as development of phonological distinctiveness in children's spoken word recognition has been linked with vocabulary size (e.g., Werker et al., 2002), as has their development of phonological constancy in the word-preferences task (Best et al., 2010).

If 15-month-olds do have a nascent grasp of the phonological structure of known words, but it is masked by their difficulties with cognitively taxing tasks, then the increased contextual support of the current lower-demand word recognition task should facilitate their access to phonological word structures, resulting in identification of the non-native pronunciations. Such findings would suggest that in Best et al. (2009) the 15-month-olds’ failure to recognize the accented pronunciations of familiar words was due to the demands of the listening preference task with sets of numerous known words versus sets of phonetically similar low frequency adult words, rather than being due to a lack of phonological knowledge.

On the other hand, if reduced task demands instead only facilitate young children's access to their more detailed *phonetically* specified word forms, then we would expect 15-month-olds to show above chance identification of native-accented words, but to continue to fail to identify the unfamiliar non-native accented pronunciations. By contrast, the 19-month-olds should be above chance across both regional accents as we expect them to have become attuned to the phonological form of the words, that is, we expect them to show phonological constancy across the accents.

As past results show a link between increasing vocabulary size and increasing phonological knowledge (Best et al., 2010; Werker et al., 2002), we further predicted that vocabulary size would be positively correlated with the ability of the 15-month-olds to identify words in the non-native regional accent. A correlation with the native regional accent is not predicted, as phonological constancy is not required for word recognition in the native accent, where identification can be achieved through recognition of the familiar phonetic patterns. No correlations between vocabulary size and word recognition in the native or non-native regional accent are predicted for the 19-month-olds, on the other hand, as we postulate they have already gained a competent grasp of phonological constancy, which is no longer changing in tandem with the growth of their lexicon.

Method

Participants

Participants were 16 18.5- to 19.5-month-olds, (*M* = 19 months, 3 days; 8 per gender), and 16 14.5- to 15.5-month-olds (*M* = 15 months, 4 days; 8 per gender) from AusE-speaking households in Sydney, Australia. Their amount of exposure to non-native languages or non-AusE accents ranged from 0 to no more than 4 hours per week as indicated by parental report on the OZI vocabulary inventory (see below). Of the 32 participants, only 2 15-month-olds and 4 19-month-olds were reported as having any regular exposure to any languages or accents besides AusE (*M* = 1.83 hrs/week), but no participants had any exposure to the non-native accent for this study, JaME. Participants were primarily Caucasian, from middle- to upper-middle-class households, and were recruited via advertisements in a regional parents' magazine. Data from an additional 19 15-month-olds and 18 19-month-olds were collected but not included in the analysis due to fussiness or inattentiveness resulting in < 45% eyetracking for either test (*n* = 17), failure to complete both the native and non-native accent tests (*n* = 12), technical problems (*n* = 7), or parental interference (*n* = 1).

Stimuli and Materials

Audio target words and sentences—Stimuli were produced by a 39-year-old male native speaker of the children's native regional accent (AusE) and a 50-year-old male native speaker of the non-native regional accent (JaME), who were instructed to produce the tokens as if they were speaking to a toddler. JaME was selected as the non-native regional accent as it differs markedly from AusE in phonetic realizations of its vowels, consonants, and prosody (Patrick, 1999; Wassink, 2006), and because it is quite unfamiliar to families living in the testing region. Vowels in JaME appear to retain a length distinction not present in Australian English, and many consonants differ from Australian English consonants either in place or manner of articulation, or both. Furthermore, using JaME in this study allows direct comparison to previous findings on development of phonological constancy, which used the same non-native accent but different word sets and recordings by different female speakers (Best et al., 2010, 2008) and a different native accent (Best et al., 2009). Eighteen target words were selected on the basis that they are easily depicted and appear in most toddlers' early vocabularies, having a mean frequency of 71% in 15-month-olds' receptive vocabularies (Dale & Fenson, 1996). While receptive frequencies are not available at 19 months, the target words would necessarily have at least the same, though more likely a much higher, mean frequency in receptive vocabularies at that age. The target words, their receptive frequencies, and their phonetic realizations in AusE and JaME are presented in Table 1. Phonetic realizations are given here as narrow IPA transcriptions, based on judgments of three phonetically trained listeners. Target words were presented in four carrier sentences (Can you see the ____?; Where is the ____?; Let's find the ____.; Look at the ____.). Four reward sentences (That's the one!; There it is!; Yeah that's right!; You got it!) were selected to serve as task reinforcement at the end of each trial. Multiple recordings of each carrier sentence, reward sentence, and target word were produced by the same AusE and JaME speakers. The final tokens of each target word (two per speaker per word) were selected based on similarity across accents in voice quality and infant-directed speech quality, by consensus among the authors and informal verification by other lab personnel.

In the original recordings, gathered for a separate experiment, many of the carrier sentence and target word combinations required for this task had not been recorded. In order to maintain consistency across stimuli, all carrier sentence and target combinations for the present study were created by splicing a token of the target word into the final position in a carrier sentence. Care was taken to ensure that targets were spliced into carrier sentences that originally contained a word having a phonetically similar or identical onset (e.g., the target word BABY was spliced into a sentence originally ending with the word BALL). An additional token of each target word was selected to serve as the second repetition of the word in isolation. Due to differences in speaker rate, the resulting JaME sentences consisting of the carrier sentences and spliced target words were substantially shorter in duration (*M* = 1224 ms; *SD* = 150 ms) than the AusE sentences (*M* = 1668 ms; *SD* = 188 ms). Therefore, to ensure that the mean sentence duration did not differ significantly across the stimuli sets, the JaME sentences were lengthened (*M* = 1456 ms; *SD* = 187 ms), and the AusE sentences were shortened (*M* = 1439 ms; *SD* = 158 ms) using the duration manipulation function in Praat version 5.0.11 (Boersma & Weenink, 2008). The durations of the tokens used for the second repetition of the target words did not differ significantly across the AusE (*M* = 696 ms; *SD* = 150 ms) and JaME (*M* = 658 ms; *SD* = 131 ms) stimulus sets, and therefore were not altered. Naturalness of the resulting modified phrases and words was verified auditorily by three phonetically trained judges (co-authors CTB, MDT, and CK). A given item was accepted for the final stimulus set only if at least two of these judges agreed that it sounded natural; in most cases all three agreed.

Target and distractor images—For each word, two color photographic images were selected for visual depiction of the word. The two images depicting a given word were matched for size, clarity, and impressionistically for visual salience. Individual images were then paired with another image depicting a different word from the target familiar-words list to create target-distractor image pairs (e.g., the first CAR image was paired with the second SPOON image, and the second CAR image was paired with the first BOAT image). Each target-distractor image pair was matched as best as possible on visual (e.g., color, complexity) and semantic (e.g., animals, inanimate objects) characteristics. Further, monosyllabic item names were paired only with monosyllabic distractors, and disyllables were paired only with disyllabic distractors. Each pairing was placed on an 800 × 600 pixel 10% gray background, placing one image on the left (center pixel: x = 160, y = 300), and one on the right (x = 640, y = 300; see Figure 1). Two such arrangements were created for each pairing so that an image appeared on the left and right side of the display equally often, and with equal designation of each item as the target and distractor image in a given condition (see *Procedure and counterbalancing*, below). All images measured 280 × 274 pixels, apart from the toothbrush and spoon, which were oriented diagonally and therefore were scaled to 200 × 196 pixels so that the length of the diagonal equaled the horizontal width of the other images. From a 95 cm viewing range, the full display subtended a 20.17° × 16.18° visual angle. The visual stimuli (apart from the toothbrush and spoon images) subtended a 5.89° × 5.82° visual angle, and the diagonally oriented toothbrush and spoon images, a 4.22° × 4.16° visual angle.

Trial videos—Audio and visual stimuli were combined into audiovisual videos, as detailed in the *Procedure and counterbalancing* section below.

The Australian English vocabulary inventory (OZI)—To measure the participants' expressive vocabularies, the Australian English vocabulary inventory (OZI) was completed by the parent who brought the child in for the test session. The OZI is an adaptation of the MacArthur-Bates Communicative Development Inventory (MCDI): Words and Sentences form that is designed for use with 16- to 30-month-old children (Fenson et al., 2007). The OZI was developed at the Marcs Institute at the University of Western Sydney, and comprises two sections: an expressive vocabulary measure, and a measure of developing grammar skills (Schwarz & Burnham, 2006). While the MCDI measures the vocabulary of AmE-learning infants and toddlers, the OZI was adjusted to better reflect contemporary AusE, as well as to shorten the overall administration of the measure by including only nouns, verbs, and descriptives in the inventory. Although normative data are still being collected for children in the age range tested here, preliminary results show that Australian children's scores on the MCDI and OZI are highly correlated for 24- and 30-month-olds (Schwarz & Burnham, 2006).

Apparatus and Setup

Participants' gaze during the experimental task was measured using a Tobii X120 eye tracker (Tobii Technology AB) sampling at 120 Hz. This eye tracker is accurate within 0.5° and has a 0.2° compensation error for head movements, and has a 100 ms recovery time when tracking is lost. It implements both dark-pupil and bright-pupil technology to minimize data loss, and tracks both eyes simultaneously, which allows for data collection even when one eye is not being tracked. This binocular tracking also allows for correction of drift through continuous averaging of drift effects between the two eyes.

Two hand-drawn areas of interest (AOIs) were defined that coincided with the left and right images on the video monitor the infants viewed. Each AOI measured 323 × 468 pixels, and the pairs were positioned such that the left AOI had its center at 161.5 × 300 pixels, and the right at 638.5 × 300 pixels (the centers of the two side-by-side pictures -- see Figure 1). The same AOIs were used for all trials.

The testing room was set up with a 17" Diamond Digital LCD monitor (Mitsubishi Electric) 25 cm behind the back of the eye tracker, and with its lower edge positioned 26.5 cm above the table on which the eye tracker sat. The monitor was angled at approximately 5° backward tilt (top tilting away from the child). Two Edirol MA-15D speakers were centered 41 cm below the tabletop. A Logitech QuickCam Orbit AF camera was placed 15 cm to the right of the eye tracker, allowing the experimenter to view participants from the adjoining room, and verify that participants' gaze was being tracked when they were oriented toward the screen. Stimuli were presented using Tobii Studio 2.0.2 software (Tobii Technology). So that the lag between presentation of the stimuli would not be dependent on the computer's speed, Tobii Studio was set to load one video in advance.

Procedure and counterbalancing

Participants were seated on their caregiver's lap so that their eyes were 70 cm from the front of the eye tracker. For the duration of the study, caregivers wore Macally MTUNE headphones that played a mixture of music and speech, and were instructed to look down or to the side during the experiment. This kept caregivers blind to the experimental conditions, and also served to assure that it was the children's and not the caregivers' gaze that was tracked. Before testing began, the participant's gaze was calibrated to a moving object presented once at the top left and once at the bottom right corner of the screen. For this purpose, a dynamic cartoon measuring 6.35 cm × 6.35 cm was paired with sound to attract and maintain the children's interest. The experimenter determined participants to be looking at the calibration stimuli when their gaze was fixed at a point on the screen at or in close proximity to the calibration object.

Following calibration, each participant completed two tests, one per regional accent, with 18 trials (one for each target word) per test. The entire procedure lasted about 8 minutes. A trial began with a central looming bull's-eye video clip that played until the child fixated on it, at which point they saw a four-second silent video depicting the target-distractor image pair, which served to familiarize them to the pictures. Immediately following that, the test video began with the looming bull's-eye. Once the child fixated on the bull's-eye, the carrier sentence began. The looming bull's-eye continued to play during the sentence, until the onset of the target word at the end of the sentence, at which point the target-distractor image pair replaced the bull's-eye. In order that the onset of the target word and the corresponding image pair would occur at 1160 ms after fixation of the bull's-eye for all trials, silence was inserted at the beginning of each carrier sentence as needed to compensate for the duration variations among the carrier sentences. Onset of the second target word repetition occurred 2000 ms after the onset of the word within sentence frame (at 3160 ms from the start of the test video). This created two time windows of 2000 ms for fixation analyses, one per word repetition. Animation (e.g., spinning, blinking) of the target image began 2000 ms after the onset of the second word repetition (5160 ms into the test video), which was on average 1302 ms (SD = 144 ms) after the word had ended (the words had a mean duration of 696 ms in AusE, and 658 ms in JaME). The reward sentence began playing 480 ms after onset of the animation (at 5640 ms); the animation continued until the reward sentence finished, at which point the trial ended and the looming bull's-eye beginning the next trial began. As reward sentences varied in length, the total durations of the test videos ranged from 6600 ms to 7520 ms (M = 7001 ms; SD = 285 ms). A schematic of each trial appears in Figure 2.

Trial order and regional accent order were counterbalanced across participants. Images were counterbalanced such that in each test the participant saw all 18 target-distractor image pairings, arranged so that one of the two images corresponding to a target word (e.g., one of the two CAR images) appeared on the left side, and the other appeared once on the right side, with equal designation of each image as the target or distractor image across the two tests in a given condition. Raw gaze points to the computer screen were recorded throughout the trial by the Tobii X120 eye tracker.

Results

For the duration of the experiment, participants' gaze was tracked on average 72.5% of the time (*SD* = 9.8%). The time participants' gaze was not tracked was due primarily to looking away from the screen, rather than to failure of the eyetracker. Raw gaze points were converted to fixations by applying a fixation filter that compiled groups of gaze points occurring within 50 pixels and within at most 200 ms of each other. This is the recommended fixation filter for use with participants viewing still images (as opposed to text or videos; “Tobii Studio 1.2 user manual,” 2008). This removed gaze points associated with saccades, so that only gaze points associated with actual fixations were measured. The fixation filter also served to compensate for missing data points over a short time period (e.g., when the participant was blinking) by replacing missing points with the mean point of the fixation.

The percentage of total fixation time that fell within the AOI of the target image was calculated by examining the durations of fixations from word onset to 2000 ms after word onset for each of the two word repetitions per trial, creating a 4000 ms window of analysis. The total duration of fixations to the target image in this interval was then divided by the total fixation duration to the target and distractor image during the interval. Some analyses of infant gaze data for this type of task exclude the first 367 ms post word onset from analysis, as that is the presumed time it takes young children to carry out a saccade (Swingley & Aslin, 2000). Since we instead filtered the raw gaze data to only count fixation times, this approach simultaneously excluded gaze points belonging to saccades at any point during the analysis interval. As well, trials began once children fixated on a central looming bull’s-eye that persisted through the carrier sentence until the simultaneous presentation of the spoken target word and target and distractor images. Thus, children were not already fixated on either the target or distractor image at the onset of the first target word repetition, and always had to shift their gaze from the center of the screen to view either of the images during the first spoken word presentation.

Percent fixation time to the target image was examined in 2 × 2 ANOVA comparing regional accent (AusE, JaME) as a within-subject factor and age group (15 months, 19 months) as a between-subjects factor. A main effect of regional accent was found, $F(1, 30) = 13.72$, $p < .001$, reflecting that overall there was a higher percentage of looking to the target image in the native regional accent, $M = 60.62$, compared to the non-native regional accent, $M = 54.78$. No other effects were found. While no main effect of age was found, this is not entirely surprising, as it was predicted *a priori* that 15- and 19-month-olds would perform in the same manner in the native accent condition, and differently only in the non-native accent condition. To test our prediction that looking to the target for the 19-month-olds would be above chance in both regional accent conditions, but only above chance in the native accent condition for 15-month-olds, percentage of fixation time to the target image was then compared against chance (50%) for each age group and accent condition using one-sample *t*-tests, and percentage of fixation time to the target image across regional accent conditions for each age group was compared using paired-samples *t*-tests. Fifteen-month-olds' percentage of fixation time on the target image was above chance for the native regional accent, $M = 59.31$, $t(15) = 3.89$, $p = .001$, but was at chance for the non-native regional accent, $M = 51.21$, $t(15) = 0.45$, $p = .658$ (Figure 3). Percentage of target fixation time differed across regional accents, $t(15) = 3.55$, $p = .003$. In contrast, 19-month-olds' percentage of fixation time on the target image exceeded chance in both the native, $M = 61.94$, $t(15) = 5.35$, $p < .001$, and non-native regional accent, $M = 58.35$, $t(15) = 4.47$, $p < .001$, with the percentage of fixations to the target image failing to differ significantly across regional accents, $t(15) = 1.65$, $p = .120$.

Nineteen-month-olds' performance was not affected by condition order. That is, percent fixation to the target image was above chance in the native regional accent condition regardless of whether participants were first exposed to the native or non-native regional accent (native regional accent first: $M = 59.86$, $t[7] = 3.23$, $p = .014$; non-native regional accent first: $M = 64.02$, $t[7] = 4.27$, $p = .004$). The same held true for their performance in the non-native regional accent condition (native regional accent first: $M = 58.37$, $t[7] = 2.42$, $p = .046$; non-native regional accent first: $M = 58.32$, $t[7] = 4.89$, $p = .002$). Fifteen-month-old participants who first completed the non-native regional accent condition showed some improvement in the native regional condition relative to those participants who were first exposed to the native regional accent condition (native regional accent first: $M = 58.67$, $t[7] = 2.09$, $p = .075$; non-native regional accent first: $M = 59.94$, $t[7] = 3.71$, $p = .008$), perhaps due to increased familiarity with the objective of the task. There was no difference in performance in the non-native regional accent (native regional accent first: $M = 48.60$, $t[7] = −0.34$, $p = .744$; non-native regional accent first: $M = 53.82$, $t[7] = 1.10$, $p = .308$).

To examine the roles of vocabulary and age in predicting children's performance on the task, two stepwise regressions were carried out comparing log-transformed vocabulary and age (in days) against percentage of fixation time to the target in the native and the non-native regional accents, with $p < .1$ as the entry criterion. Vocabulary scores were log transformed to minimize any distorting effect of the rapid expansion of expressive vocabulary around the time of the vocabulary spurt. For the non-native regional accent, the log-transformed vocabulary score reached significance across the two age groups $b = .388$, $t(31) = 2.31$, $p = .028$, explaining 15% of the variance in performance, $F(1, 31) = 5.32$, $p = .028$, while age did not add any predictive power to the model. Neither factor predicted children's performance in the native regional accent. Separate regression analyses were then performed for each age group, to examine vocabulary scores as a predictor of performance in the non-native regional accent. There was an effect of log-transformed expressive vocabulary size against 15-month-olds' performance in the non-native regional accent, $b = .561$, $t(15) = 2.54$, $p = .024$, explaining 32% of variance in performance, $F(1, 15) = 6.43$, $p = .024$ (Figure 4). Vocabulary size did not predict the 19-month-olds' performance in the non-native regional accent (Figure 5).

Following a similar approach as was carried out by Richmond and Nelson (2009), fine-grained analyses were also conducted on looking patterns for the two accents, to determine whether there were more subtle differences in the looking patterns of the two age groups. In particular, if 15-month-olds did identify the words in the non-native regional accent, but at a slower rate than the 19-month-olds, their correct identifications could have been masked by examining total fixation time over the whole trial. Because application of a fixation filter to data disrupts the relation between time and spatial location of participants' gaze, we analysed the raw gaze data and normalized the participants' gaze so that a value of 400 (pixels) represented the vertical midpoint of the screen, and looks with horizontal gaze location values above 400 corresponded to looking to the target image side of the screen, and those with values below 400 corresponded to looking at the distractor image side of the screen. These values were then averaged across all trials for each participant, and collected into bins with separate analyses conducted for bins of length 100 ms, 200 ms, 400 ms, and 500 ms. Bin values were then compared to chance via *t*-test analysis across age group and accent to determine at which points in time participants' gaze was directed at the side of the screen containing the target or distractor image. No systematic differences between age groups emerged from these analyses, primarily due to the fact that the 15-month-olds' raw gaze data never differed from chance. Additional analyses compared the time point at which the maximum horizontal gaze towards the target image was achieved. A 2 × 2 ANOVA comparing time of maximum value for the factors of age group (15 and 19 months) and accent (AusE and JaME) found only a main effect of age group, $F(1, 30) = 4.33$, $p = .046$, reflecting that 19-month-olds reached their maximum gaze point sooner than 15-month-olds ($M_{19\ mos}$ = 2228.13 ms, SD = 910.24 ms; $M_{15\ mos}$ = 2737.50, SD = 1138.97 ms). This suggests that 19-month-olds identify the target word sooner than 15-month-olds, but this interpretation is mitigated by the fact that 15-month-olds' raw gaze data never differed from chance. Gaze point data are presented in Figures 6 and 7. The patterns in the non-native regional accent condition seem to suggest that 15-month-olds show a similar looking pattern to 19-month-olds, but that this pattern is both delayed and weaker in comparison, likely reflecting 15-month-olds' inability to identify words in the non-native regional accent. However, the younger age displays too much variance in gaze to draw clear conclusions from these gaze measures over time.

Discussion

When compared against chance, 19-month-olds identified words spoken in both the native and non-native regional accent, with total fixation time on the target images exceeding chance in both conditions. Fifteen-month-olds, on the other hand, identified words spoken only in the native regional accent; their recognition of words in the non-native regional accent did not exceed chance. Word recognition did not differ across regional accents for the 19-month-olds, but did for the 15-month-olds. While these results do not demonstrate a clear effect of age, they are suggestive of a possible link between performance and age. This link may have been obscured to an extent by individual differences in phonological development within each age group. However, such differences may be linked to expressive vocabulary, and expressive vocabulary size was indeed correlated with the 15-month-olds' recognition of words in the non-native regional accent. It was predictive of their performance on words in the non-native accent, but not with the native accent nor with 19-month-olds' performance on either accent.

These findings are inconsistent with the hypothesis that phonological knowledge emerges as early as 14 months. Given the simple nature and relatively low cognitive demands of the current task, the current findings support the notion offered by some researchers that 14-month-olds' ability to discriminate minimal word pairs in previous studies suggests that low cognitive demands permit them to access phonetically detailed word representations, and not that they already possess phonologically specified word forms (e.g., Fennell & Waxman, 2010; Fennell & Werker, 2003; Yoshida et al., 2009). If they had discriminated minimal pairs via phonologically specified word forms, then they should have identified the non-native regionally accented words in the present study. Thus, the reduced demands simply appear to allow them better access to inefficient, detailed phonetic specifications of known words that they have experienced in their native accent.

The disparity between the 15- and 19-month-olds' recognition of words in the non-native regional accent in our procedure is instead compatible with the perceptual attunement hypothesis (Best et al., 2009) that children first become attuned to the phonetic information in native-accented words, but that further experience with natural variations shifts their attention to the more abstract phonological structure of words by around 18 months, a time at which most children’s expressive vocabulary has begun to grow at a rapid rate. Early phonetically specific native accent lexical representations are too rigid to accept unfamiliar, phonetically different non-native regional accent pronunciations, whereas later phonologically specified lexical representations can accommodate the phonetically varied but phonologically invariant non-native pronunciations. Thus the 19-month-olds have made a further leap of abstraction where they can consider unfamiliar phonetic deviations as still belonging to a given phonemic class in the context of word recognition, in theory due to the systematic nature of the variation. By this account, of course, word recognition requires more than an understanding of native phonemic categories, as recognition is still possible across variation that violates native phonemic boundaries.

The present findings echo the challenge to phonological underspecification accounts (Brown & Matthews, 1997; Metsala & Walley, 1998) set forth by previous studies on phonological distinctiveness. The underspecification accounts propose that children’s early lexical representations begin as global and underspecified, with more detailed phonetic specification arising as needed to resolve ambiguities in the expanding lexicon. Earlier findings that 14-month-olds can discriminate pairs of words based on contrasts that are not required to separate minimal pairs in their existing lexicon demonstrates that their lexical representations are more richly phonetically specified than the phonological underspecification account requires. Similarly, if 15-month-olds had underspecified lexical representations, they should be more likely than 19-month-olds, who have more fully specified lexical representations, to accept the non-native regional accent pronunciations. However, exactly the converse was found here and in Best et al. (2009).

In addition to being in line with the Perceptual Attunement approach, the current results are also consistent with other contemporary theories that posit phonological knowledge emerges by 18–20 rather than by 14–15 months (Swingley, 2003; Thiessen, 2007; Werker & Curtin, 2005). Further, while the emergence of phonological knowledge is unrelated to the presence of phonological neighbors in the child’s lexicon (as proposed in Brown & Matthews, 1997; Metsala & Walley, 1998), such knowledge is thought to emerge once children have learned a sufficient number of words from which they can draw phonological generalizations (Best et al., 2009; Swingley, 2003; Thiessen, 2007; Werker & Curtin, 2005). This knowledge is posited to derive either from the abstraction of phonemes from experienced native accent exemplars (Werker & Curtin, 2005), or from the distribution of phonemes in diverse lexical contexts (Thiessen, 2007), or from exposure to systematic variation in speech that allows children to derive the underlying phonological structure of the words (Best et al., 2009; Mulak & Best, in press). This relation with vocabulary size is compatible with our finding that although 15-month-olds overall did not identify target words in the non-native regional accent overall, their performance with the non-native accent was positively correlated with their expressive vocabulary size. It was not correlated with the 15-month-olds' performance in the native regional accent, which may be interpreted as evidence that phonological knowledge is not required for them to identify *native* pronunciations. This is important, as it suggests the emergence of a specific phonological skill – phonological constancy – rather than a simple improvement in general cognitive skills. The lack of correlation between vocabulary and performance in either regional accent for the 19-month-olds suggests that they had already achieved phonological constancy, such that their recognition of the non-native pronunciations was no different than that of the native pronunciations. This is consistent with recent findings linking lexical development to the presence of phonological constancy via children's ability to recognize non-native pronunciations of familiar words (Best et al., 2010). A similar correlation between expressive vocabulary size and the presence of phonological distinctiveness, via children's ability to learn new words containing minimal pair differences (Werker et al., 2002), demonstrates that vocabulary growth is linked with overall development of phonological knowledge, and is consistent as well with recent evidence that expressive vocabulary size is linked to the growth of phonological knowledge even in adult second-language learning (Bundgaard-Nielsen, Best, & Tyler, 2011a, 2011b).

An issue related to the development of phonological specification is how children learn to handle allophonic variation within native phoneme categories. In one study, Canadian English (CanE)-learning and Dutch-learning 18-month-olds’ awareness of language-specific linguistically relevant versus irrelevant phonetic detail was examined in a discrimination test using newly learned nonsense words that were minimal pairs by a vowel length difference that is contrastive in Dutch but not CanE (Dietrich, Swingley, & Werker, 2007). Short [ɑ] and long [a:] are lexically contrastive in Dutch, for example, in the words /stat/ (*city*) versus /sta:t/ (*stands*), whereas they are both variants of a single vowel /a/ in CanE. In Dutch, therefore, the target stimuli used in the toddler study were minimal pairs, but in CanE they were simply phonetically varied tokens of the same phonological words. Children of each language group were habituated to word-object pairings for each item of the minimal pair (e.g., /stat/ paired with object A, and /sta:t/ paired with object B), and then tested on a switched pairing (/stat/ paired with object B). A change in looking time indicated discrimination of the minimal pair. The Dutch-learning toddlers discriminated the minimal pairs, but the CanE-learning children did not, suggesting the former but not the latter group perceived them as phonologically distinct. This study was not a direct test of phonological constancy, in part due to the use of foreign words and a foreign speaker. Furthermore, the distinction involved may be considered comparable to that between allophonic variants. This raises the question of how children come to master perception of allophonic variation, and how this is related to phonological perception. It is our contention that phonemic perception includes the variation allowable within a category both in terms of its canonical production and in terms of variations that follow from phonotactics and coarticulation. The crucial distinction between allophonic variation and cross-accent variation is that allophonic variation is experienced regularly and systematically as part of quotidian exposure to the native regional accent, and is thus part of what is learned in the development of a native-accented phonemic category. In contrast, phonological constancy allows successful perception of a word even in cases where the variation has never been experienced in the native regional accent.

Considering the results of the current study together with other research on cross-accent word perception gives view to a more complete account of the development of phonological constancy. Recent research on cross-accent speech perception in infants has revealed that 12-month-olds, but not 9-month-olds, are able to segment words familiarized in one regional accent from sentences presented in another regional accent (Schmale, Cristia, Seidl, & Johnson, 2010). This suggests that a grasp of phonological constancy appears to have begun to emerge at this age. However, this skill is likely fragile at this point, and limited to cognitively undemanding tasks like the recognition of word forms, which does not require any semantic processing. This notion is supported by recently completed research that shows a less reliable ability of 12-month-olds to segment word forms across accents when the stimulus load is doubled in comparison to prior studies that have used a similar procedure (Mulak, Best, Tyler, & Kitamura, in preparation). The current study demonstrates that by 19 months, however, children's grasp of phonological constancy has strengthened to the point that they are able to recognize known words across accents, in a paradigm that requires both semantic and word form processing. This development of phonological constancy continues, however. In an even more cognitively demanding task that required participants to learn novel word-object associations, in which only 30-month-olds, and not 24-month-olds, were able to generalize newly learned words across accents (Schmale, Hollich, & Seidl, 2011). This study tested children's ability to generalize across a *foreign* accent, which may contain more non-systematic variations and thus be more cognitively demanding to resolve. Thus, children's developing grasp of phonological constancy can be seen in their performance on increasingly demanding tasks. This sensitivity to task difficulty is consistent with results showing that pre-exposure to a specific accent facilitates cross-accent word recognition (White & Aslin, 2010; see also Schmale et al., 2011), and is a characteristic mirrored in the literature outlining children's developing ability to access phonetic detail in speech (see Fennell & Werker, 2003; Werker & Curtin, 2005).

In this experiment we examined variation to vowels, consonants, and suprasegmental features. There is reason to believe that there may be differences in how each of these, and in particular vowels and consonants may emerge in the specificity of children's word recognition. For instance, vowels are perceived less categorically than consonants, with less of a clear divide between the perception of one vowel and another compared to consonant perception (Eimas, 1963), and consonants and vowels appear to serve different roles in word perception by young children (e.g., Bonatti, Peña, Nespor, & Mehler, 2007; Hochmann, Benavides-Varela, Nespor, & Mehler, 2011), with consonants being tied to word identity, and vowels linked more to structural features. Recent results suggest that in terms of developing phonological specificity, vowels and consonants may follow different developmental paths (Mulak et al., in preparation; Mulak, Best, Tyler, & Kitamura, 2012).

Overall, the current study together with other recent research on early development of cross-accent word perception (Schmale et al., 2010; Schmale & Seidl, 2009; White & Aslin, 2010) underline the importance of examining phonological constancy as the complement to phonological distinctiveness in order evaluate overall level of phonological knowledge. Both of these phonological abilities need to be considered in current theories of language development. The concept of phonological constancy appears to have particular utility for separating the child’s use of phonologically versus phonetically specified word forms. The Perceptual Attunement account (Best et al., 2009) specifically addresses the complementary relation between these two aspects of phonological knowledge. The results of the present study support a developmental transition in perceptual attunement from attention specifically to phonetic detail, toward greater reliance on phonological structure between 15 and 19 months. That the latter point in development coincides with the vocabulary spurt, and that expressive vocabulary size correlates with recognition of non-native accented words during the transitional period, taken together, are highly suggestive of a link between an increasing lexicon and developing phonological knowledge. This relation may be a complex one, in which the need to rapidly expand one's lexicon encourages attunement to the systematic phonological properties of words, which in turn reduces the cognitive load required to access a known word. Attunement to the more efficient phonological properties might then bootstrap further vocabulary growth, allowing children in the midst of the vocabulary spurt to go on to increase their lexicon at a previously unmatched rate as they form more abstract and versatile lexical representations.

Acknowledgments

This research was funded by ARC Discovery grant DP0772441 (Chief Investigators: C. Best & C. Kitamura) and NIH grant DC000403 (PI: C. Best), with contributions from the Dean of the College of Arts at the University of Western Sydney, and the Marcs Institute via Research Training Scheme funds. Acknowledgement is also given to the first author's support from an EIPRS scholarship from the Australian Department of Education, Employment & Workplace Relations.

We express our gratitude to our toddler participants and their parents for their invaluable contributions and interest in the research. We also thank Rikke Bundgaard-Nielsen, Lidija Krebs-Lazendic, Anna Notley, and Bethany Wootton for their work in stimulus development, Alizee Foucher, Carly Hayman, and Michelle Pal for their assistance in running the study, and Lauren Tornatore for her help with earlier pilot work. Finally, we thank Damien Smith for feedback on earlier versions of this manuscript.

References

Best, CT. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, J.; Nusbaum, HC., editors. Development of speech perception: The transition from speech sounds to spoken words. Cambridge: MIT Press; 1994. p. 167-224.

Best, CT. A direct realist perspective on cross-language speech perception. In: Strange, W.; Jenkins, JJ., editors. Speech perception and linguistic experience: Issues in cross-language research. Timonium, MD: York Press; 1995. p. 171-204.

Best CT, Tyler MD, Gooding TN, Orlando CB, Quann CA. Development of phonological constancy: Toddlers’ perception of native- and Jamaican-accented words. Psychological Science. 2009; 20:539–542. [PubMed: 19368700]

Best, CT.; Tyler, MD.; Kitamura, C.; Bundgaard-Nielsen, RL. Presented at the International Conference on Infant Studies. Baltimore, Maryland: 2010. Vocabulary size at 17 months and the emergence of phonological constancy in word recognition across native and nonnative dialects.

Best, CT.; Tyler, MD.; Kitamura, C.; Notley, A.; Bundgaard-Nielsen, RL. Presented at the International Conference on Infant Studies. Vancouver, Canada: 2008. Phonetic specificity of early words? Australian toddlers' perception of Australian versus Jamaican English pronunciations.

Boersma, P.; Weenink, D. Praat: Doing phonetics by computer. 2008. Retrieved from http://www.praat.org

Bonatti LL, Peña M, Nespor M, Mehler J. On consonants, vowels, chickens, and eggs. Psychological Science. 2007; 18(10):924–925. [PubMed: 17894612]

Brown, CA.; Matthews, J. The role of feature geometry in the development of phonemic contrasts. In: Hannahs, SJ.; Yount-Scholten, M., editors. Focus on phonological acquisition. Amsterdam: John Benjamins; 1997. p. 67-112.

Bundgaard-Nielsen RL, Best CT, Tyler MD. Vocabulary size is associated with second language vowel perception performance in adult learners. Studies in Second Language Acquisition. 2011a; 33:433–461.

Bundgaard-Nielsen RL, Best CT, Tyler MD. Vocabulary size matters: The assimilation of L2 Australian English vowels to L1 Japanese vowel categories. Applied Psycholinguistics. 2011b; 32:51–67.

Dale PS, Fenson L. Lexical development norms for young children. Behavior Research Methods, Instruments, & Computers. 1996; 28:125–127.

Dietrich C, Swingley D, Werker JF. Native language governs interpretation of salient speech sound differences at 18 months. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:16027–16031. [PubMed: 17911262]

Eimas PD. The relation between identification and discrimination along speech and non-speech continua. Language and Speech. 1963; 6:206–217.

Fennell CT, Waxman SR. What paradox? Referential cues allow for infant use of phonetic detail in word learning. Child Development. 2010; 81:1376–1383. [PubMed: 20840228]

Fennell CT, Werker JF. Early word learners' ability to access phonetic detail in well-known words. Language and Speech. 2003; 46:245–264. [PubMed: 14748446]

Fenson, L.; Marchman, V.; Thal, DJ.; Dale, PS.; Reznick, JS.; Bates, E. MacArthur-Bates Communicative Development Inventories. Users guide and technical manual. 2nd. Baltimore: Paul H. Brookes Publishing Co.; 2007.

Hochmann J-R, Benavides-Varela S, Nespor M, Mehler J. Consonants and vowels: different roles in early language acquisition. Developmental Science. 2011; 14:1445–1458. [PubMed: 22010902]

Houston DM, Jusczyk PW. Infants' long-term memory for the sound patterns of words and voices. Journal of Experimental Psychology: Human Perception and Performance. 2003; 29:1143–1154. [PubMed: 14640835]

Jusczyk PW, Aslin RN. Infants' detection of sound patterns of words in fluent speech. Cognitive Psychology. 1995; 29:1–23. [PubMed: 7641524]

Metsala, JL.; Walley, AC. Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In: Metsala, JL.; Ehri, LC., editors. Word recognition in beginning literacy. Mahwah, NJ: Erlbaum; 1998. p. 89-120.

Mulak, KE.; Best, CT. Development of word recognition across speakers and accents. In: Gogate, LJ.; Hollich, G., editors. Theoretical and computational models of word learning: Trends in psychology and artificial intelligence. IGI Global: Robotics Division; in press

Mulak, KE.; Best, CT.; Tyler, MD.; Kitamura, C. Development of phonological specification of word forms: 15-month-olds' sensitivity to vowel variation, and 19-month-olds' sensitivity to consonant variation from the native accent. in preparation

Mulak, KE.; Best, CT.; Tyler, MD.; Kitamura, C. Twelve-month-olds may segment words from continuous speech in the face of non-native suprasegmental features; Presented at the International Conference on Infant Studies; Minneapolis, USA. 2012.

Patrick, PL. Urban Jamaican Creole: Variation in the Mesolect. Amsterdam: Benjamins; 1999.

Richmond J, Nelson CA. Relational memory during infancy: Evidence from eye tracking. Developmental Science. 2009; 12:549–556. [PubMed: 19635082]

Schmale R, Cristia A, Seidl A, Johnson EK. Developmental changes in infants' ability to cope with dialect variation in word recognition. Infancy. 2010; 15:1–13.

Schmale R, Hollich G, Seidl A. Contending with foreign accent in early word learning. Journal of Child Language. 2011:1–13.

Schmale R, Seidl A. Accommodating variability in voice and foreign accent: Flexibility of early word representations. Developmental Science. 2009; 12:583–601. [PubMed: 19635085]

Schwarz, IC.; Burnham, DK. Stern, R. Proceedings of INTERSPEECH-2006, Pittsburgh (DVD). Adelaide: Causal Productions; 2006. New measures to chart toddlers' speech perception and language development: A test of the lexical restructuring hypothesis.

Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature. 1997; 388:381–382. [PubMed: 9237755]

Swingley D. Phonetic detail in the developing lexicon. Language and Speech. 2003; 46:265–294. [PubMed: 14748447]

Swingley D. The roots of the early vocabulary in infants' learning from speech. Current Directions in Psychological Science. 2008; 17:308–312. [PubMed: 20523916]

Swingley D, Aslin RN. Spoken word recognition and lexical representation in very young children. Cognition. 2000; 76:147–166. [PubMed: 10856741]

Swingley D, Aslin RN. Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological Science. 2002; 13:480–484. [PubMed: 12219818]

Thiessen ED. The effect of distributional information on children's use of phonemic contrasts. Journal of Memory and Language. 2007; 56:16–34.

Tobii Studio 1.2 user manual. Tobii Technology AB. 2008. Retrieved from www.tobii.com

Wassink AB. A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties. Journal of the Acoustical Society of America. 2006; 119(4):2334–2350. [PubMed: 16642847]

Werker JF, Curtin SA. PRIMIR: A developmental framework of infant speech processing. Language Learning and Development. 2005; 1:197–234.

Werker JF, Fennell CT, Corcoran KM, Stager CL. Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. Infancy. 2002; 3:1–30.

White KS, Aslin RN. Adaptation to novel accents by toddlers. Developmental Science. 2010; 14:372–384. [PubMed: 21479106]

Yoshida KA, Fennell CT, Swingley D, Werker JF. Fourteen-month-old infants learn similar-sounding words. Developmental Science. 2009; 12:412–418. [PubMed: 19371365]

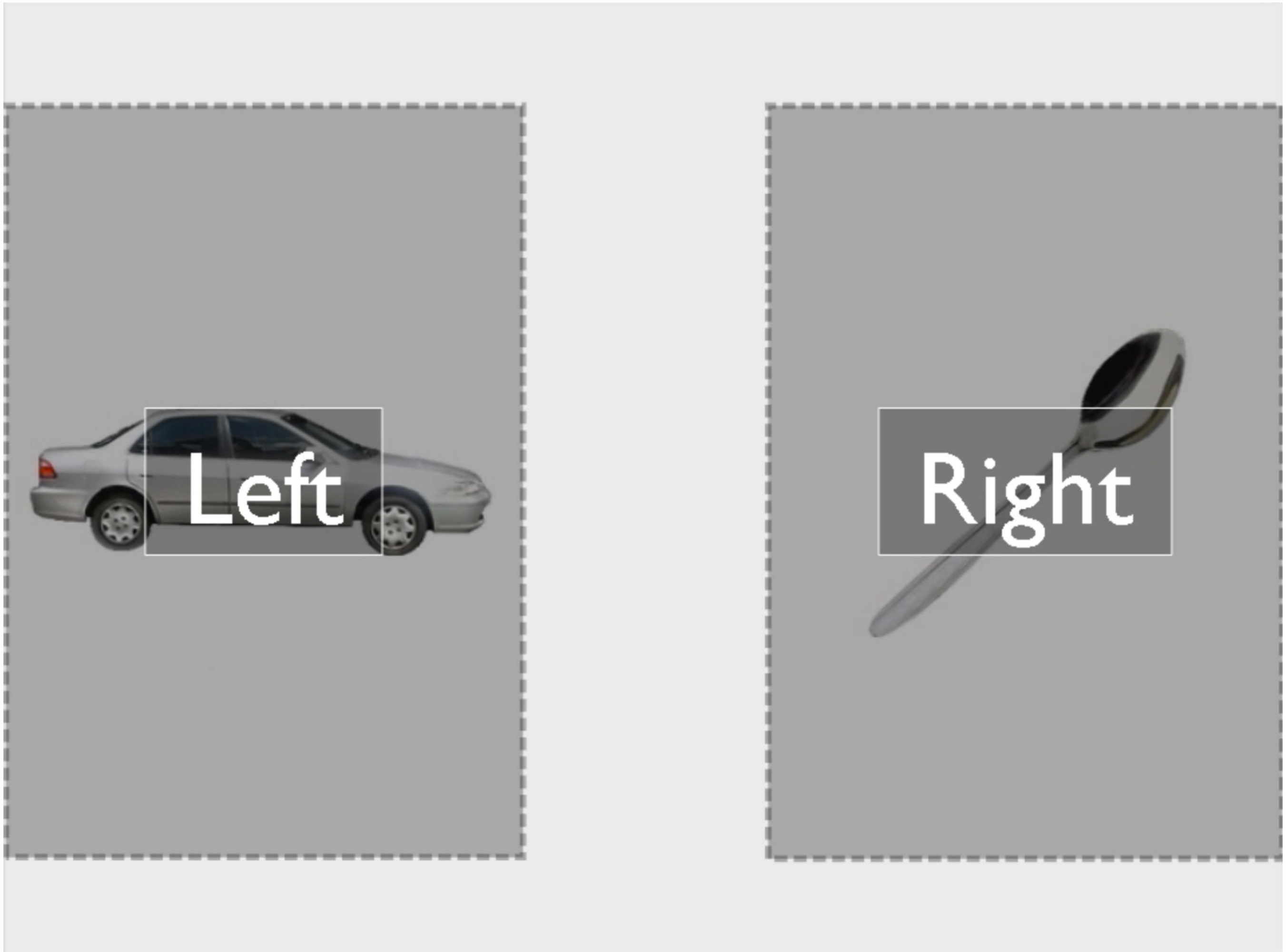

Figure 1.
Placement of images and areas of interest. Left and right images were centered vertically, and situated horizontally at 20% and 80% of the screen's width from the left border for the left and right images, respectively. Areas of interest (in dark gray) were placed around the left and right images. Fixations within those areas of interest indexed looking at either the left or right image.

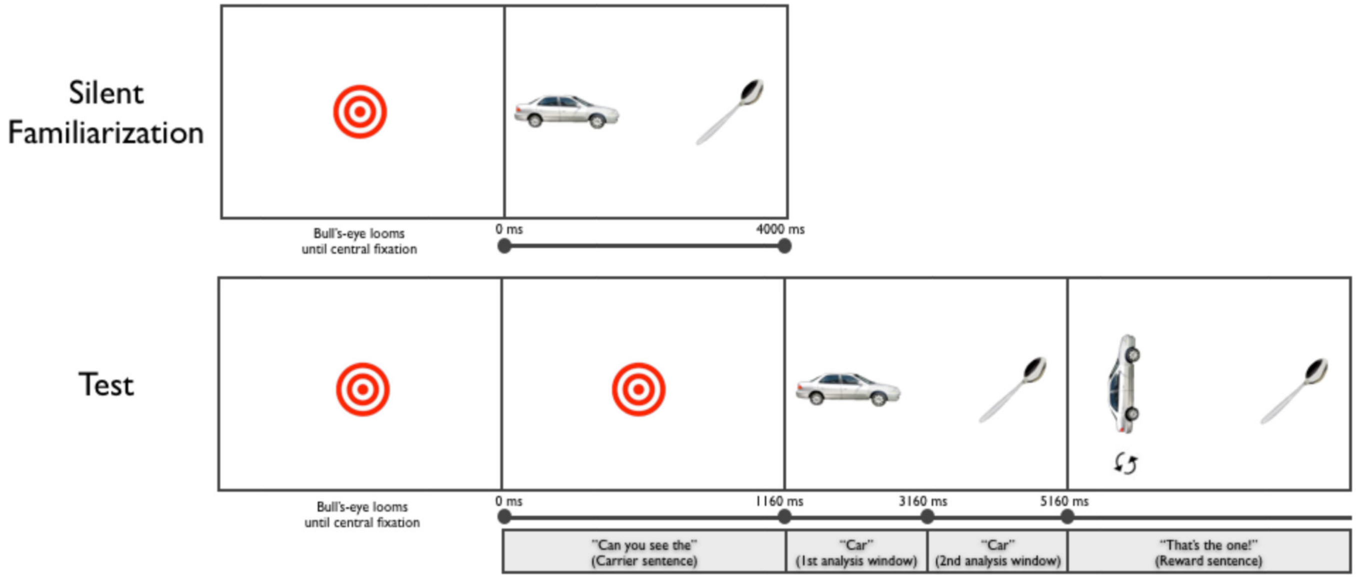

Figure 2.
Trial procedure used in this experiment. Each trial comprised a silent familiarization followed by a test phase, followed by a reward-sentence phase.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

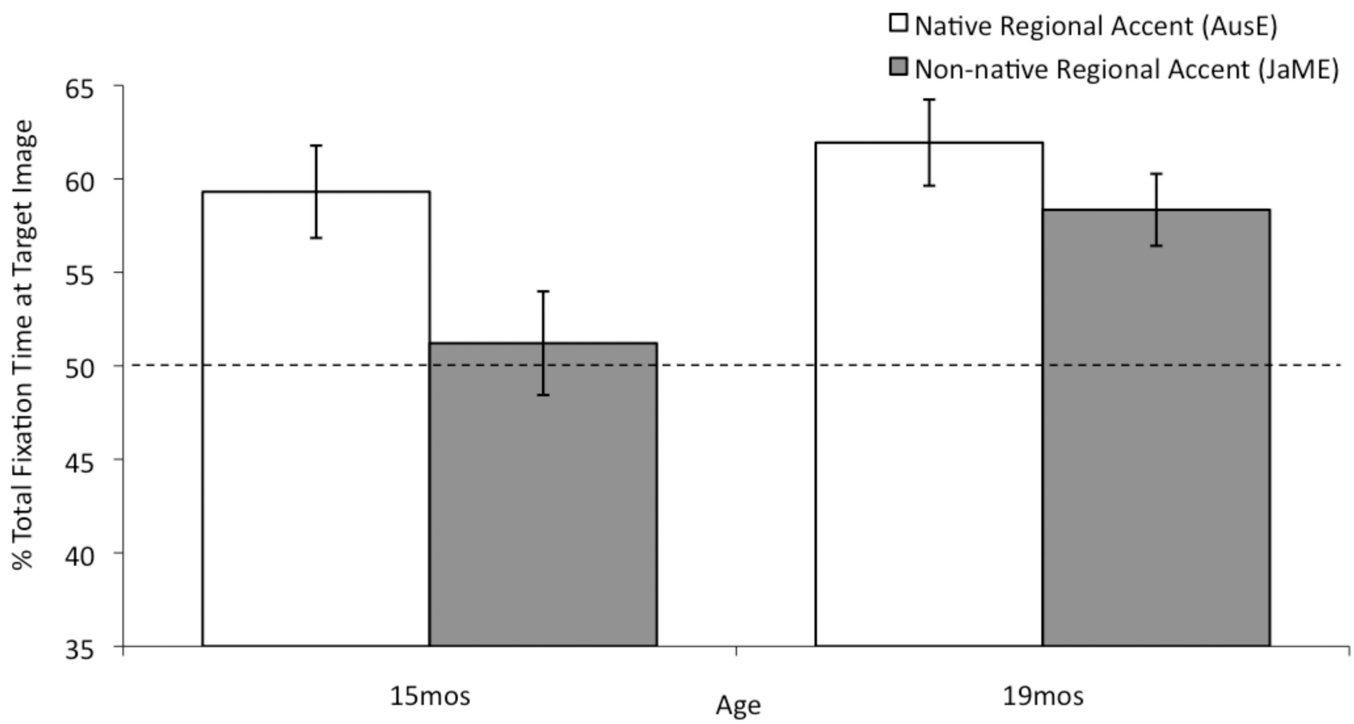

Figure 3.
Mean percentage of fixation time to the target image. This was calculated by dividing total fixation time on the target image by the sum of fixation time on the target and distractor images. Both 15- and 19-month-olds fixated to the target image more than chance (50%) when target words were spoken in the native regional accent (Australian English), but only 19-month-olds fixated to the target image more than chance also when target words were spoken in the non-native regional accent (Jamaican Mesolect English). Error bars represent standard error of the mean.

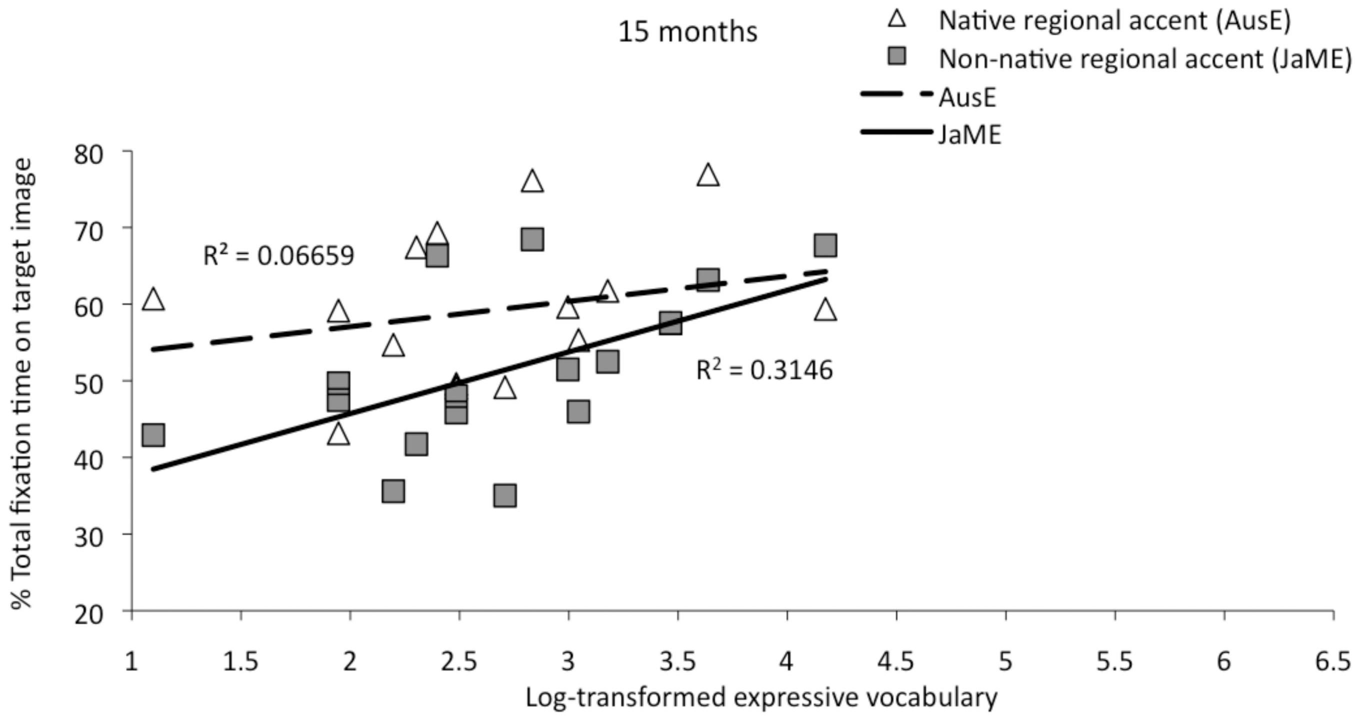

Figure 4.

Mean percentage of fixation time to the target image versus log-transformed expressive vocabulary size at 15 months. Although 15-month-olds did not fixate on the target image more than chance in the non-native regional accent (Jamaican Mesolect English), log-transformed expressive vocabulary size is positively correlated with total fixation time on the target image in the non-native regional accent. Log-transformed expressive vocabulary size did not predict their performance in the native regional accent (Australian English).

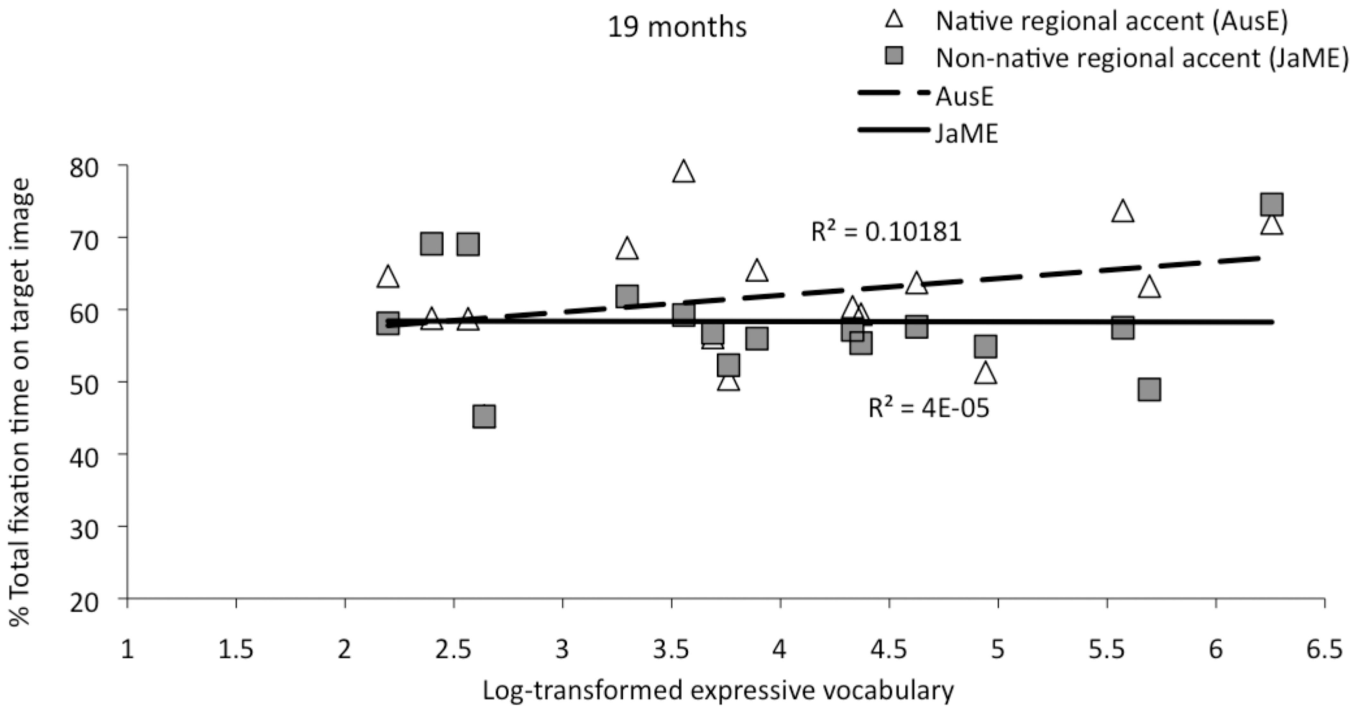

Figure 5.

Mean percentage of fixation time the target image versus log-transformed expressive vocabulary size at 19 months. Nineteen-month-olds fixated on the target image more than chance in both the native (Australian English) and non-native (Jamaican Mesolect English) regional accent, but expressive vocabulary size did not predict their performance in either regional accent.

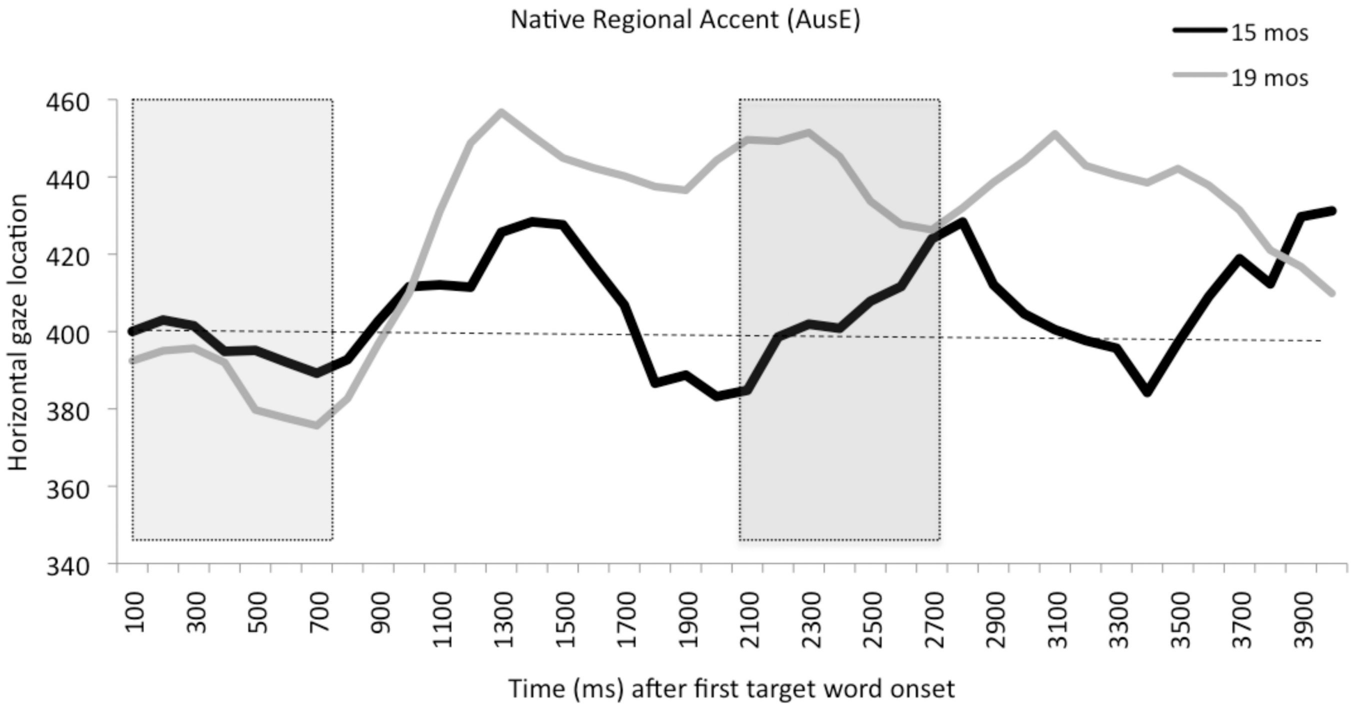

Figure 6.

Horizontal location of raw gaze for 15- and 19-month-olds in the native regional accent condition from onset of the first repetition of the target word. Onset of the second repetition of the target word occurred at 2000 ms, and shaded areas indicate the presentation of auditory stimuli. Data is normalized so that gaze above 400 reflects looking to the target image side of the screen, and gaze below 400 indicates looking to the distractor image side of the screen. While analysis did not reveal any differences in looking patterns across ages, analysis of total fixation time on the target image revealed both 15- and 19-month-olds identified target images in the native regional accent (Figure 3).

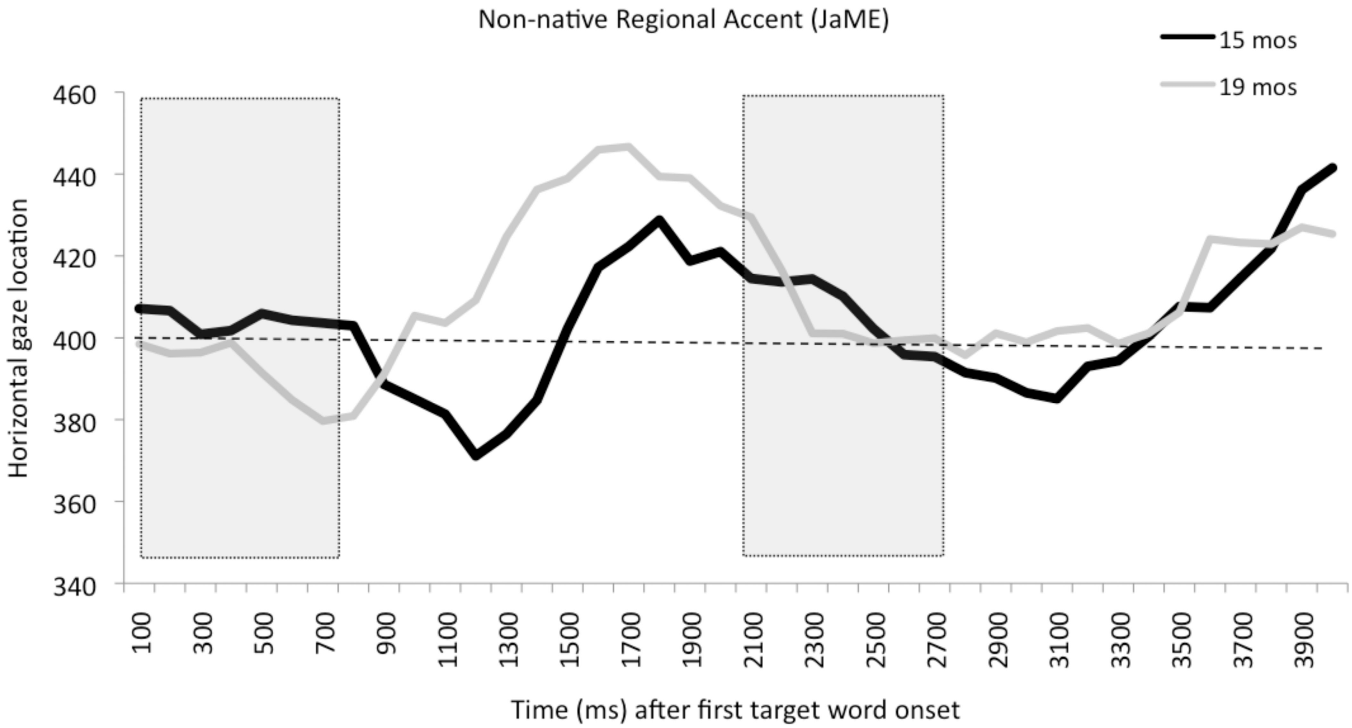

Figure 7.

Horizontal location of raw gaze for 15- and 19-month-olds in the non-native regional accent condition from onset of the first repetition of the target word. Onset of the second repetition of the target word occurred at 2000 ms, and shaded areas indicate the presentation of auditory stimuli. Data is normalized so that gaze above 400 reflects looking to the target image side of the screen, and gaze below 400 indicates looking to the distractor image side of the screen. While analysis did not reveal any differences in looking patterns across ages, analysis of total fixation time on the target image revealed that 19-, but not 15-month-olds, identified target images in the non-native regional accent (Figure 3).

Table 1

Target words, their phonetic realizations in the native (AusE) and non-native (JaME) regional accent, and their frequency in 15-month-olds' receptive vocabularies

	Phonetic realization				Phonetic realization		
Target Word	AusE	JaME	Freq (%)	Target word	AusE	JaME	Freq (%)
baby	[pæɪbiː]	[peᵊbiː]	89.1	cat	[k̟ʰæt]	[kʰjat]	65.6
ball	[pɔl]	[paːl]	95.3	doggy	[tɔg̟iː]	[taːgi]	87.5
bathtub	[pɐːθtɐb]	[pa˺ttʊb]	81.2	door	[toː]	[toɹ]	73.4
bear	[peː]	[piːɹ]	51.6	flower	[flæɔwə]	[flɔwa]	64.1
birdy	[pɜːdiː]	[pɚdi]	82.8	hair	[heː]	[iːɹ]	70.3
boat	[pəʉt]	[pɔːt]	*	mouth	[mæᵊθ]	[mɔᶷt]	67.2
bottle	[pɒtɫ]	[paːk̩l]	84.4	paper	[pʰæɪpə]	[pʰjepa]	37.5
button	[pɐtən]	[pɔʔən]	37.5	spoon	[spʉːn]	[spuːn]	68.7
car	[kʰɐː]	[kʰʲaːɹ]	82.8	toothbrush	[tʰʉːθbɹɐʃ]	[tʰuːtbɹaʃ]	65.6

Note. AusE = Australian English; JaME = Jamaican Mesolect English

*Receptive frequency for target word boat at 15 months is not available. However, it has a 41% occurrence in 16-month-olds' expressive vocabularies, which implies a notably higher frequency in receptive vocabulary. Moreover, it corresponds exactly to the average frequency for all our target words in 16-month-olds' expressive vocabularies: 41% (Dale & Fenson, 1996).