Assessing Toddlers' Speech-Sound Discrimination

Rachael Frush Holt and **Kaylah Lalonde**
Department of Speech and Hearing Sciences, Indiana University

Abstract

Objective—Valid and reliable methods for assessing speech perception in toddlers are lacking in the field, leading to conspicuous gaps in understanding how speech perception develops and limited clinical tools for assessing sensory aid benefit in toddlers. The objective of this investigation was to evaluate speech-sound discrimination in toddlers using modifications to the Change/No-Change procedure[1].

Methods—Normal-hearing 2- and 3-year-olds' discrimination of acoustically dissimilar ("easy") and similar ("hard") speech-sound contrasts were evaluated in a combined repeated measures and factorial design. Performance was measured in d'. Effects of contrast difficulty and age were examined, as was test-retest reliability, using repeated measures ANOVAs, planned post-hoc tests, and correlation analyses.

Results—The easy contrast (M=2.53) was discriminated better than the hard contrast (M=1.72) across all ages ($p < .0001$). The oldest group of children (M=3.13) discriminated the contrasts better than youngest (M=1.04; $p < .0001$) and the mid-age children (M=2.20; $p = .037$), who in turn discriminated the contrasts better than the youngest children ($p = .010$). Test-retest reliability was excellent ($r = .886$, $p < .0001$). Almost 90% of the children met the teaching criterion. The vast majority demonstrated the ability to be tested with the modified procedure and discriminated the contrasts. The few who did not were 2.5 years of age and younger.

Conclusions—The modifications implemented resulted, at least preliminarily, in a procedure that is reliable and sensitive to contrast difficulty and age in this young group of children, suggesting that these modifications are appropriate for this age group. With further development, the procedure holds promise for use in clinical populations who are believed to have core deficits in rapid phonological encoding, such as children with hearing loss or specific language impairment, children who are struggling to read, and second-language learners.

Keywords

Audiology; Toddler; Speech Perception

Corresponding Author: Rachael Frush Holt, Ph.D., Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, IN 47405, 812.855.6659 (phone), 812.855.5531 (fax), raholt@indiana.edu.
Author Addresses: Rachael Frush Holt, Ph.D., Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, IN 47405, 812.855.6659 (phone), 812.855.5531 (fax), raholt@indiana.edu
Kaylah Lalonde, B.A., Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, IN 47405, 812.856.6965 (phone), 812.855.5531 (fax), klalaonde@umail.iu.edu

Introduction

Accurate perception of speech requires some mastery of different levels of perception, often times simultaneously. A listener must detect the utterance, discriminate its component sounds from others, recognize the word or phrase, and ultimately map the perception to meaning in order to comprehend the message. Disruptions to the formation of phonological representations, such as those caused by hearing loss, auditory processing deficits or other disabilities, can negatively influence any task that requires the ability to form accurate representations of auditory signals. Disturbances at any level of perceptual processing can lead to delays in language, social, and educational development[2–4]. Despite the potentially devastating consequences of poor processing of speech, little is known about how speech perception develops during toddlerhood and childhood, even in typically developing populations[5]. One reason for this gap in understanding is that, with the exception of some preliminary research by Dawson et al.[6], valid and reliable methods for assessing speech discrimination in this age range are lacking[7–9]. Consequently, many investigators routinely omit 2-year-olds from research, instead including infants up to about 12-months-old and then jumping to 3- and/or 5-year-olds[e.g., 10–11], despite this critical age of rapid language development[5,12–13]. The purpose of this investigation is to evaluate speech-sound discrimination of 2- and 3-year-old children, who often are neglected in the literature, using novel, developmentally appropriate modifications to a procedure that has been used to evaluate speech-sound discrimination in older children – the Change/No-Change procedure[1].

Toddlers tend to be difficult to test for several reasons: their attention spans are shorter than their older peers; their language is not as fully developed as their older peers; they are not fully intelligible until age 4 years[14–15]; and they are more mobile than infants. These developmental factors limit what stimuli, procedures, and response tasks can be used. Eisenberg, Boothroyd and Martinez have begun to develop new methods for testing this population. Their test battery is based on early work by Boothroyd on the Speech Pattern Contrast (SPAC) Test[16], with modifications to accommodate children at different stages of development. At present, the two published tests developed by this group of researchers that have been used with toddlers are the Visual Reinforcement Assessment of the Perception of Speech Pattern Contrasts (VRASPAC)[17] and the Imitative Test of Speech Pattern Contrast (IMSPAC)[18–19].

The first prototype of the VRASPAC combines three procedures that have been used to test young children: the visually reinforced, repeating auditory background discrimination paradigm used with infants[20], the Change/No-Change procedure[1] which itself is based on Eilers et al.'s approach, and the SPAC. The VRASPAC has been used with normal-hearing infants and toddlers between the ages of 7 and 34 months and those with hearing loss between the ages of 9 and 38 months[9, 19]. Listeners are conditioned to turn their heads to a lighted, animated toy when they detect a change in a nonsense syllable stimulus array. The syllable contrast increases in difficulty, beginning with vowel height and moving to vowel place, consonant voicing, consonant manner/continuance, and, finally to consonant place. Martinez et al.[9] reported that regardless of hearing status, all of the infants (9 with normal hearing and 11 with hearing loss) that were tested on the VRASPAC were able to reliably discriminate the vowel height contrast, but those with hearing losses greater than 60 dB HL had difficulty discriminating vowel place, which relies on perceiving second formant transitions. In contrast, performance for the consonant contrasts was extremely variable, even for the infants and toddlers with normal hearing. Further, performance was best for the infants under 12 months of age, but declined for children older than 12 months. Finally, the VRASPAC was sensitive to changes in discrimination performance after cochlear implantation in a single 16-month-old. These mixed results suggest that in its current form

the VRASPAC is most useful in infants younger than 12 months of age, a conclusion that is consistent in some respects with the early work on the traditional clinical threshold-seeking visual reinforcement audiometry (VRA) procedure used with infants. Primus and Thompson[21] reported that 11- to 13-month-olds and 22- to 26-month-olds could be conditioned at the same rate, but that the reinforcer maintained the operant in the 11- to 13-month-old infants for significantly more test trials than it did for the 22- to 26-month-olds. In other words, the habituation rate increased with increasing age. Although using novel visual reinforcers slowed the rate of habituation, the 22- to 26-month-old infants still had lower rates of responding than the younger infants.

The IMSPAC assesses young children's ability to convey discrimination of phonologically contrastive information through imitations of syllables. In the original version of the test, the imitations were recorded and played back to adult subjects. In an effort to make test administration more efficient, a modified version, the On-line IMSPAC (OlimSPAC), was developed[22]. Eisenberg et al. presented nonsense syllables in the auditory-only and audiovisual modalities to 30 children (10 with normal hearing, 10 with hearing aids, and 10 with cochlear implants) ranging in age from 2.75 to 7.9 years. During testing, each syllable was presented individually eight times and the child was instructed to imitate what she/he perceived. Responses were scored for feature correct (vowel height, vowel place, consonant voicing, consonant manner/continuance, and consonant place). The results suggest that the ability to imitate speech sounds did not vary significantly over the age range tested in the investigation. Further, normal-hearing children achieved approximately 90% correct, indicating that they were unable to complete the test with perfect accuracy. Finally, the test was sensitive to different degrees of hearing loss and to the inherent difficulty of the vowel or consonant feature contrast.

A potential limitation of the IMSPAC/OlimSPAC is that one has to make the assumption that incorrect imitation reflects an error in perception. Only 50% of the speech of average normal-hearing 2-year-olds is intelligible[14–15] and children with hearing loss have even poorer articulation. Even normal-hearing children are not fully intelligible until 4 years of age[14–15]. In a recent report, Boothroyd, Eisenberg, and Martinez[23] evaluated developmental effects on OlimSPAC performance in 34 children with normal hearing. Four of the youngest children (ages 1;10 to 2;10 [month;year]) were unable or unwilling to do the task, leaving 30 children (ages 2;7 to 6;7) who completed the protocol. Six different individual vowel-consonant-vowel syllable contrasts were presented eight times. Children imitated each syllable and the tester, who was blind to the presented stimulus, selected which of eight alternatives the child uttered. Note that responses on the OlimSPAC are scored for contrast feature correct and not phoneme correct. Therefore, if a child imitated the feature correctly (e.g., consonant voicing), but the phoneme itself was not imitated correctly, the response was scored as correct. Across all contrasts, percent correct (corrected for chance) increased up to 4 years of age, at which scores approached ceiling. Further, younger children's performance was influence more by the type of feature contrast than were older children, suggesting that phonological development contributes to task performance. Finally, all 26 children who were 3 years of age or older performed above chance on at least five of the six contrasts. The results suggest that normal-hearing children who are 3 years of age or older are likely to be able to complete the OlimSPAC, but that phonological development plays a significant role in performance of young children on the test. Further work is needed to evaluate developmental effects in children with hearing loss, who typically struggle with intelligibility more than their normal-hearing peers[23].

An approach that has been used often with infants, but is a bit more novel in toddlers and young children, is to use eye tracking to evaluate what children look at in a speech perception task. Newton et al.[24] recently investigated the utility of measuring direction and

duration of eye gaze in a minimal pairs "discrimination" paradigm, termed Speech by Eye (SP-EYE). Children with normal hearing ages 2 through 7 years were presented with 30 sets of pictures of monosyllabic items that differed in consonant voicing, consonant place of articulation, consonant manner, vowel, or by consonant voicing, place and manner all at once. During each trial the child heard the name of the target picture presented auditory-only and then eye movement was recorded for a 2-second interval. The primary measure of accuracy, how often the child's first gaze shift was to the target picture, resulted in no significant effects: children did not look to the target first significantly more often than the distractor. In fact, overall, children looked to the distractor first more often than to the target (although the difference was not significant). The authors argue that this was because the children were not given the instruction to look at the item that was named. They did, however, report that children looked significantly longer at the target than the distractor picture. Furthermore, no age effects or contrast difficulty effects were found. Finally, although the authors describe the task as one of discrimination, it really is a recognition and comprehension task because it requires the child to accurately perceive the presented item, map it to an item in their lexicon and identify it as one of the pictured items on the screen by looking to that item.

The results from the IMSPAC and the VRASPAC suggest that whereas they are effective measures of speech perception for some children, other children, particularly those who are 2 and 3 years of age, might benefit from further test modifications or a slightly different approach. The results from the SP-EYE suggest that it does not assess speech discrimination in its current form, and that more work is needed before it could be a sensitive measure of speech recognition. The overall objective of the current study is to evaluate speech-sound discrimination in 2- and 3-year-olds, the population of children for whom speech-sound discrimination measures are lacking, using a modified version of the Change/No-Change procedure[1].

Recently Holt and Carney[25] successfully used the Change/No-Change procedure to evaluate speech discrimination in children as young as 4 years of age. The procedure—originally implemented by Sussman and Carney[1]—was adapted from the visually reinforced, repeating auditory background discrimination paradigm used with infants[20]. During a trial, listeners are presented with an auditory string of nonsense syllables that either change (e.g., "ra ra la la") or remain identical (e.g., "ra ra ra ra"). The first stimuli presented are standards ("ra ra") and the second are comparisons (either "la la" or "ra ra"). The listener indicates whether the standard and comparison are the same or different by using a developmentally appropriate motor response task (e.g., pressing a button or registering a response on a touch screen monitor). The methodology is particularly useful for young children with the most limited language and auditory skills because it does not require a linguistic response. This is a particularly important problem as the average age at cochlear implantation and hearing aid fitting continues to decrease, because these children will have limited language skills, due to their chronological age and/or their limited auditory experience. Another advantage of the procedure is that it can be used to assess not only discrimination ability, but also the strength of the perceptual representation[25–26], which is important for certain models of speech perception (e.g., Trace[27], PARSYN[28], and Shortlist[29]). The procedure requires the listener to act upon a neural representation of the encoding of speech input, and thus provides a window into children's early sensory encoding of speech. Measurement of early sensory encoding is important for children with hearing loss or other populations with underspecified phonological representations, such as children with specific language impairment[30–31], children experiencing reading deficits[32–33] and second-language learners[34–35]. If children cannot do this task, they will be unable to perform higher-level tasks, such classification and categorization. Furthermore, the procedure is not influenced by speech-motor ability, it uses a developmentally appropriate response, it provides multiple

types of reinforcement to keep the child motivated and attentive, it is flexible – clinicians or researchers can employ any desired contrast, and it uses a fun, engaging interface. The procedure's validity and reliability have been demonstrated in normal-hearing and hearing-impaired adults and children[1, 6, 25–26, 36–40], although its success was limited in children younger than 4 years of age[6]. Therefore, the form of the Change/No-Change procedure that has been used with young children is limited in its applicability to toddlers.

In this investigation we implemented modifications to the Change/No-Change procedure, such as employing a developmentally appropriate gross motor response and using various forms of reinforcement, to allow assessment of toddlers' speech-sound discrimination. We hypothesize that toddlers' relative sensitivity to perceptually "easy" (acoustically distinct) and perceptually "hard" (acoustically similar) speech-sound contrasts will emerge by implementing these novel modifications to the procedure. Further, we expect that older toddlers will demonstrate better sensitivity to the speech-sound contrasts than the younger toddlers. Finally, we hypothesize that toddlers' performance will be reliable in that scores at test will be related to those obtained at retest.

Method

Participants

Out of a total of 43 English-speaking children, 30 met the criteria to participate in the study. The 30 children who met the study criteria were between the ages of 2;2 and 3;11 and were stratified into three groups based on chronological age. More 2-year-olds than 3-year-olds were recruited because it was anticipated that developmental effects would be more influential in the younger toddlers relative to the older toddlers. The youngest group of toddlers ("Young," n = 10) were comprised of younger 2-year-olds with an average age of 2;4. The second group of toddlers ("Mid-Age," n = 10) consisted of older 2-year-olds with an average age of 2;9. Finally, the oldest group ("Old," n = 10) was comprised of primarily 3-year-olds with an average age of 3;4. Participant demographics are displayed in Table 1.

Participants passed a binaural hearing screening using conditioned play audiometry at 20 dB HL for audiometric frequencies between and including 500 and 8000 Hz, and 25 dB HL at 250 Hz (re: American National Standards Institute, 2004) or a 4-frequency distortion-product otoacoustic emission (DPOAE) screening. The DPOAE screening was used with the 2-year-olds, because conditioned play audiometry is most appropriate for children 3 years and older (e.g., Madell, 2008), and also with some of the young 3-year-olds, because it reduced the total testing time. In addition, all participants passed a speech-language screening. The Early Language Milestone Scale – 2[41] was used to screen speech and language development in the 2-year-olds using the pass/fail scoring method and the Preschool Language Scales – 4 Screening[42] was used with the 3-year-olds. The 13 children who failed to meet the study criteria did so for the following reasons: failed the hearing screening (n=4), failed the speech-language screening (n=2), were unable to learn the procedure (n=4), or the parents opted not to return for further testing due to family problems or because the child cried or was uncooperative (n=3). Note that the investigators never used lack of cooperation as a reason for not inviting families back for return visits; these decisions were made independently by the families.

Stimuli

Three sets of natural, digitally recorded nonsense syllable contrasts were used in the investigation. Nonsense syllables were selected to reduce the effects of word knowledge on the task, because we anticipate using this procedure with toddlers with limited speech and language abilities in the future, such as children with hearing loss. A training contrast

consisting of a long /u/ vs. a short /ga/, was used to evaluate whether the children understood the task and to ensure continued task understanding and attention after each test and retest condition. Two test contrasts, a perceptually easy one (/ba/ vs. /bu/) and a perceptually hard one (/sa/ vs. /ʃa/), also were used. The vowels in /ba/ and /bu/ differ in height, and thus the first formant, as well as in lip rounding, whereas the fricatives in /sa/ and /ʃa/ differ in place of articulation. The easy contrast was selected to be maximally contrastive and thus, should be easier to discriminate than the hard contrast. Vowel height tends to be a relatively easy contrast for most listeners, even young children with hearing loss, to discriminate[9] and consonant place errors are common in consonant-vowel discrimination tasks, even in listeners with normal hearing[43]. The fricative pair has been used in previous research on the Change/No-Change procedure[25–26] and has been used extensively by Nittrouer and colleagues[44–48] in their investigations on the developmental weighting shift theory of speech perception.

Stimulus Recording—Twenty highly intelligible, natural productions of each syllable were used as stimuli in the discrimination procedure. The use of multiple tokens almost eliminates the chance that listeners will use unanticipated subphonemic cues from any given token. The utterances of three native-English females with standard General American dialects were recorded in a double-walled sound booth with a Senheiser head-mounted condenser microphone and a Marantz digital recorder using 16-bit, 44.1 kHz sampling rate. To ensure that the tokens were highly intelligible, each speaker produced at least 70 tokens of each syllable. To avoid vocal fatigue on a particular token, the speakers produced approximately 10 tokens of each syllable and then repeated the sequence six times. The speakers were told to produce all of the tokens in the same manner and approximate duration using a pleasant-sounding voice. In addition, they were given a photo of a 2-year-old child reading a book and were told to utter each syllable as though they were identifying an item in the book to the child. This instruction was meant to help ensure that the stimuli were uttered in a child-directed fashion. All of the tokens were edited into individual .wav files and the total rms was normalized to have equal intensity in Audition v. 2.0[49]. For each speaker, the mean duration of the tokens was calculated. For each syllable, only tokens that were within 50 ms of the mean duration and that were void of any extraneous audible sounds (e.g., lip closure) were included in the stimulus selection experiments. Table A1 in Appendix A displays the descriptive data for the subset of tokens that were within 50 ms of the mean of each speaker's syllables.

Two stimulus selection experiments were then carried out on this subset of tokens. The purpose of the first stimulus selection experiment was to identify the speaker whose utterances were most intelligible. The purpose of the second stimulus selection experiment was to select 20 tokens of each syllable that best exemplified the intended target from the speaker selected in the first stimulus selection experiment. A description of the two preliminary experiments and their respective results are included in Appendix A.

Discrimination Experiment Stimuli—The stimuli presented to the children consisted of two standard, followed by two comparison syllables, using an inter-stimulus interval of 100 ms. All four syllables in each trial were randomly drawn from the twenty most highly intelligible tokens from Speaker 1 (as identified during the stimulus selection experiments). During no-change trials, the standard and comparison were the same syllable but not the same tokens. In other words, no-change trials consisted of four randomly chosen tokens of the same syllable. During change trials, the standard and comparison stimuli were different. Therefore, change trials consisted of 2 randomly chosen tokens of the standard syllable followed by 2 randomly chosen tokens of the comparison. Eighteen different change and 18 no-change trials were created for each syllable contrast. The standard used for each contrast was randomized across participants.

Equipment

E-Prime2 software[50] on an Intel desktop computer was used to present stimuli and record responses. The auditory signal was routed through a GSI-61 audiometer to two GSI speakers located at +/− 45 degrees azimuth relative to the listener in a double-walled sound booth. Speech was presented at an overall level of 65 dB A at the location of the listener's head. Similar sound field presentation methods have been used with other age groups in previous implementations of the Change/No-Change procedure[25–26]. Sound field presentation was necessary, because children were unlikely to remain attentive while wearing earphones for the entire duration of a one-hour session and headphones with cords would impede the gross motor response described in the Procedures section. Stimuli were presented at signal levels well above threshold; therefore, small variations in sound pressure level due to variations in head placement are not likely to influence results. Calibration was checked each day of testing.

We presented the user interface, the centering/listening cue, complex, visual feedback, as well as a visual method for keeping track of the number of trials on a 19" Elo Touchsystems touch-screen monitor (see Figure 1 for a diagram of the test set-up). The experimenter recorded the child's responses using the touch-screen monitor. The child stood or sat on a 4' × 5' SoftTile interlocking foam mat on the floor of the sound booth. The mat was wood grain patterned, except for two red pieces with circle-shaped cut-outs and a purple piece with a star-shaped cut out. The star, referred to as the "listening star," was set in the middle of the back half of the mat. The two red circles, one to the front/left and the other to the front/right of the star, displayed pictures corresponding to the no-change and change responses respectively. The pictures were clip-art images from Microsoft Office ironed on to white fabric using Printworks white t-shirt transfers. During the teaching and training phases, the no-change picture consisted of four identical images in a row (e.g. four cows) and the change picture consisted of two sets of two different images (e.g. two cows and two frogs). For test and retest conditions, children chose response pictures from two or three alternatives (e.g., bananas and flowers, dogs and cupcakes, or apples and crayons). The toddlers were taught the procedure using live-voice presentations of "moooo" and "ribbit." Therefore, the pictures were semantically related to the stimuli ("moooo" – cow picture; "ribbit" – frog picture) during the teaching phase. However, the pictures were not semantically related to the stimuli during the training, test and retest conditions to strip the task of as much required language knowledge as possible.

Procedures

Although the Change/No-Change procedure has been used successfully with 4- and 5-year-old children[25–26] and with some limited success in 3-year-olds[6], many procedural adjustments were required for it to have a chance of successfully evaluating speech-sound discrimination in toddlers. The following modifications were implemented in the current investigation to the Change/No-Change procedure as used by Holt and Carney[25], based on toddlers' developmental constraints (see Table 2 for a summary). First, the number of trials and conditions were reduced to accommodate toddlers' shorter attention spans[51–55]: 36 trials in each of 4 conditions (versus 50 trials in each of 12 conditions in the previous version of the procedure). Second, two additional forms of reinforcement for correct answers that have been used successfully in testing auditory detection thresholds in toddlers were employed– animated, visual reinforcers[21] and tangible reinforcers[56] (e.g., mini M&M's and Cheerios). Third, a large gross motor response of jumping to, stepping to, or placing a beanbag on one of two response mats on the floor was used to capitalize on toddlers' desire to exploit their newfound mobility[57]. Fourth, the stimuli were presented in quiet to accommodate toddlers' less mature auditory systems and listening strategies[53, 58–62]. Fifth, a teaching session was employed to learn the response task. Sixth, post-test training-contrast trials were

administered to preliminarily evaluate task understanding at the end of each condition. And finally, digitized, natural speech tokens were used because they provide a richer signal than synthetic stimuli and richer speech signals are less cognitively demanding[63].

Participants were run in a combined repeated measures and factorial design. All participants completed live-voice teaching trials, followed by recorded training trials. Then each completed the recorded easy and hard contrasts (order of contrast difficulty was randomized). Participants either were tested on the easy or hard contrasts randomly at retest to reduce testing time (Table 3 displays the order of the conditions, along with their contents). Each condition consisted of 36 trials (half were change trials), which was selected based on a power analysis to determine the number of trials required to be sensitive to differences in discrimination abilities. Most children required two or three 1-hour visits to complete the protocol. Four children (one Young and three Mid-Age toddlers) required four 1-hour visits to complete the protocol.

Teaching Phase—Before testing began, the experimenter taught the child to respond to live-voice, familiar animal sounds that are maximally contrastive auditorally ("mooooo" v. "ribbit"), and pictures that semantically mapped to the animal sounds (cows and frogs, respectively). Marzolf and DeLoache[64] have shown that initial testing using iconic symbols leads to better performance on symbolically harder tasks that follow. Performance in all testing conditions should therefore benefit from the teaching session. The child stood on the listening star on the response mat, facing the response spaces and the touch screen monitor, while the tester sat off to the side (see Figure 1). The tester instructed the child to jump to, step to, or place a beanbag on the set of pictures of all cows if she/he heard, "moooo, moooo, moooo, moooo" and to direct her/his action to the set of pictures of two cows and two frogs if she/he heard, "moooo, moooo, ribbit, ribbit." Successive approximations of the motor response were reinforced until the child completed five trials in a row correctly (consistent with Trehub et al.'s criterion[11]). All but four of the toddlers successfully met the teaching performance criterion. These four participants were some of the youngest recruited – 2;0, 2;0, 2;2, and 2;6 – and did not complete any further testing (and thus were not included in the 30 who completed the protocol).

Training Phase—Following the teaching session, participants completed the training contrast – recorded long /u/ vs. short /ga/. The cow and frog pictures continued to be used during the training phase. During training (as well as the test and retest conditions), a picture of a woman with her hand cupped around her ear appeared on the monitor, cueing the child to listen. The experimenter began the trial when the child was in a ready state (quietly standing or sitting on the listening star). The child responded with a gross motor task and the experimenter recorded the child's responses on the touchscreen monitor. If a child chose both pictures, the experimenter reminded the child to choose the one picture that corresponded to the stimulus. Correct responses were reinforced with a 3-second animated video (complex, visual reinforcer), which varied from trial to trial. Both correct and incorrect responses were followed by verbal encouragement/praise and a puzzle piece filling in on the screen. Each puzzle consisted of 12 pieces, such that three puzzles were completed in the training phase (as well as the test and retest conditions). Typically after each puzzle was completed, the child received a small snack (tangible reinforcer), although verbal encouragement and breaks were adapted to each individual child.

Test and Retest Conditions—Following the training phase, participants completed the test contrasts – a perceptually easy contrast (/ba/ vs. /bu/) and a perceptually hard contrast (/sa/ vs. /ʃa/). The order of easy or hard test contrast was randomized. Participants were randomly retested on either the easy or the hard contrast to reduce testing time. After each test and retest condition, the child was presented with four of the training contrasts (two

change and two no-change trials). The four post-test training-contrast trials were included to attempt to evaluate whether poor performance in the test/retest conditions was primarily due to not being able to complete the procedure or to not discriminating the test/retest contrasts well. The procedure for presenting the stimuli, responding on each trial, as well as use of reinforcements that was described for the training phase was the same as that used the test and retest conditions with two exceptions. First, the response pictures were selected by each participant from several sets of options. Second, an additional 4-piece puzzle followed test and retest conditions, corresponding to the post-test training-contrast trials.

Results

Performance was measured in d'. P(C) will also be reported for the reader to get a better sense for proportion correct in each condition. Both individual and group data were evaluated across the training and test/retest contrasts. In addition to examining performance across the conditions, we also evaluated performance on the first half of the trials within each condition relative to the second half of the trials. This analysis was carried out to broadly examine effects of learning and fatigue. Performance for the first half was measured by calculating d' of the first 9 change trials and the first 9 no-change trials and performance during the second half was measured by calculating d' of all remaining trials. Using this method, first- and second-half trials sometimes overlapped temporally due to the random presentation of stimuli. Post-test trials were not included in this calculation. A three-way ANOVA (main effects: first/second half of trials, age group [Young/Mid-Age/Old], and contrast [training/easy test/hard test]) were used to evaluate performance between the first and second halves of testing across age groups and speech-sound contrasts. Overall, performance did not differ between the two halves ($p = 0.780$) and no interactions were significant. This suggests that there were no group differences in relative performance during the first and second half of testing for a particular contrast in any the age groups or for any contrast. Therefore, the split-half data only will be used for looking at trends in performance of those children who had low discrimination scores on individual contrasts.

Perceptual Difficulty and Age Effects

Overall Results—Figure 2 displays the average d' scores on the easy and hard test contrasts as a function of age group. A 2-way repeated measures ANOVA (within subject factor: test contrast [easy/hard]; between subject factor: age [Young/Mid-Age/Old]) was carried out on the test data. The results of the ANOVA are displayed in Table 4. The toddlers showed better discrimination of the easy contrast than the hard contrast, $F(1,27) = 20.192$, $p < .0001$. In order to visualize individual performance differences on the perceptually easy and hard test contrasts, Figure 3 displays a scatterplot of individual participants' performance. Five of the children displayed the unexpected performance trend of better discrimination sensitivity of the perceptually hard contrast than the easy contrast and three children had the same discrimination sensitivity for both the perceptually easy and perceptually hard contrasts (see the three data points directly on the diagonal). Therefore, 22 of the children had better discrimination sensitivity for the perceptually easy than the perceptually hard contrast and for many of them the differences were quite substantial. Additionally, the older toddlers discriminated the contrasts better than the younger toddlers, $F(2,27) = 12.254$, $p < .0001$. Planned post-hoc LSD t tests revealed that all three age groups performed differently from each other: the Old group had better discrimination than the Mid-Age group ($p = .037$) and the Young group ($p < .0001$), and the Mid-Age group was better at discriminating the contrasts than the Young group ($p = .01$). The interaction between contrast difficulty and participant age was not significant ($p = .257$).

Training Contrast—Individual performance, as well as group means for the Young, Mid-Age and Old groups, are displayed in Figures 4 through 6 for the training contrast (long /u/ vs. /ga/), the easy contrast (/ba/ vs. /bu/), and the hard contrast (/sa/ vs. /ʃa/), respectively. Individual participants are ordered by increasing age at first test in Figures 4 through 6.

On the training contrast (Figure 4), all of the Old toddlers (who were primarily age 3 years and older) had high d' scores (the lowest was 0.94) and thus, as a group showed excellent discrimination of the training contrast (mean d' = 3.07, mean P(C) = 89.44). The mean d' score for the Mid-Age group (comprised of older 2-year-olds) also was high (2.69, mean P(C) = 85.56) and lowest individual score in that group was 1.05, demonstrating that the Mid-Age toddlers could discriminate the training contrast. In contrast to the older age groups, only six individual children in the Young group (who consisted of young 2-year-olds) showed clear evidence of training contrast discrimination and understanding of the procedure. Of the four toddlers with negative d' scores, two were highly negative. This suggests that these two toddlers might have been "flipping" their responses in a relatively consistent way: responding "change" on no-change trials and "no-change" on change trials. Because of these highly negative scores, as a group the youngest children – the young 2-years-olds – had a mean d' of 0.68 (mean P(C) of 63.61). The split-half analysis of the individuals in the Young group revealed that the toddler with the largest negative d' score on the training contrast (age 2;3;21) performed at chance on the first half of the training trials (d' = 0), but on the second half of the training trials had a d' of −1.56. Although this is based on a limited number of trials (n=18), these data suggest some benefit from practice, although confusion on use of the response mat cannot be ruled out. There were no split-half trends for the other toddlers with low d' scores on the training contrast.

Perceptually Easy Contrast—Similar to the training contrast, all of the toddlers in the Old group easily discriminated the easy speech-sound contrast (see Figure 5), with a group mean d' of 3.68 (mean P(C) = 94.44). All of the Mid-Age toddlers showed evidence that they could discriminate the easy contrast, and thus, as a group had a high mean d' score of 2.40 (mean P(C) = 83.89). The lowest-performing Mid-Age toddler (age 2;9;25; d' = 0.57) showed a slight learning trend with scores increasing from 0.30 to 0.90 on the first and second half of the easy contrast trials, respectively. This child could discriminate the training contrast well (see Figure 4) and in fact, responded correctly to three of the four post-test training-contrast trials.

As a group the Young toddlers displayed evidence of discrimination of the easy contrast (mean d' = 1.50; mean P(C) = 70.83). One Young toddler (age 2;6;0) continued to be one of the lower-performing toddlers in this young group, but showed evidence of discrimination of the easy contrast (d' = 0.71). In addition, this toddler attained a perfect score on all four of the post-test training contrast trials, suggesting that she/he understood the procedure, but required more practice to learn it. Only one of the Young toddlers clearly could not discriminate the easy contrast (age 2;3;21; d' = −0.16). This was the same Young toddler who had a highly negative d' score on the training contrast. Further, she/he was at chance on the post-test training contrast trials. The split-half data did not show any clear trends for this toddler. Together, these results suggest that this Young toddler (age 2;3;21) was struggling more than the others to complete the procedure.

Perceptually Hard Contrast—Individual and group mean data for the perceptually hard contrast appear in Figure 6. The two oldest age groups showed clear evidence of discrimination of the hard contrast (Old group mean d' = 2.57; Old group mean P(C) = 85.56; Mid-Age group mean d' = 2.00; Mid-Age group P(C) = 79.17). Unlike the easy contrast, there were two 3-year-olds who struggled somewhat with discriminating the hard contrast. These children, ages 3;2;17 and 3;3;27, had d' scores on the hard contrast of 0.45

and 0.48, respectively. Further, both showed a slight learning trend in their split-half data: the younger of the two children improved from .29 to .62 and the older child improved from chance to 0.86 from the first to the second half of the hard contrast trials. Upon post-test training-contrast testing, these two 3-year-olds each achieved three or four of the four trials correct. Together, these results suggest that they could do the task, but were struggling to discriminate the difficult place-of-articulation contrast.

One toddler (age 2;9;21) in the Mid-Age group was struggling to some extent with discriminating the hard contrast, as well (d' = 0.48). The split-half data for this Mid-Age toddler revealed that d' decreased from 1.56 to −1.28 from the first to the second half of the trials. Interpreting this second toddler's split-half data is difficult – either she/he began confusing the response pads in the second half of the trials, fatigue set in, or there was some combination of factors (or the low number of trials that occurs when the condition was split into two sections could reduce sensitivity of the measure). This toddler only responded correctly to 1 of the 4 post-test training-contrast trials, even though she/he had no difficulty discriminating either the training or easy contrasts. Together, these performance trends suggest that this child was likely experiencing fatigue by the end of the hard contrast condition.

Three of the toddlers in the Young group clearly were unable to show discrimination of the hard contrast – ages 2;2;23, 2;4;9, and 2;4;26 (d' = 0.0, −.15, and −.14, respectively). None of these toddlers had any noteworthy trends in their split-half data. Two additional toddlers in the Young group struggled more than the others to discriminate the hard contrast (age 2;3;19 and 2;7;8, d' = 0.42). The younger of the these two (age 2;3;19) showed improved performance on the first half to the second half of the hard contrast trials (−.29 to 1.19), suggesting a possible learning effect with this harder contrast. Further, this toddler never had any difficulty discriminating the training or easy contrasts and in fact, responded correctly to 3 of the 4 post-test training-contrast trials. Together, this suggests that this child was able to perform the procedure, but was having a harder time discriminating the perceptually hard contrast than the easy or the training contrasts. The older of these two toddlers (age 2;7;8) had consistent performance during the first and second halves of the hard contrast trials and responded correctly to 3 of the 4 post-test training-contrast trials. Further, this toddler never demonstrated difficulty discriminating the training or the easy contrast. Together, this suggests that she/he too was able to do the task, but was struggling somewhat to detect the spectral difference between the hard contrast stimuli. Finally, one toddler (age 2;6;0) in the Young group had a d' of −0.73 on the hard contrast and only responded correctly to 1 of the 4 post-test-training contrast trials, and struggled discriminating the training contrast. Overall, at age 2;6;0, this was the oldest toddler to show difficulty mastering the task. As a group, the Young toddlers struggled with the hard contrast achieving a mean d' of 0.57 (mean P(C) = 58.89).

Test-Retest Reliability

Scores at test and retest are displayed in a scatterplot in Figure 7. The unfilled triangles display data from three toddlers who scored at ceiling at test and/or retest and the unfilled circle represents a single participant who was unable to return for retest for 4 months and subsequently moved to the next chronological age group at retest. These four participants' data were not included in the correlation analysis for test-retest reliability. Pearson's correlation between test and retest revealed highly significant reliability across the two testing sessions, $r = .886$, $p < .0001$.

Discussion

The purpose of the current investigation was to evaluate speech-sound discrimination in normal-hearing 2- and 3-year-olds using a selected set of modifications to the Change/No-Change procedure. The results suggest that the modified procedure is sensitive to different perceptual abilities across speech-sound contrasts, as well as to developmental differences across the children. Because there is no gold standard test to which to compare these results, revealing that the modified procedure is sensitive to acoustic similarity of the speech-sound stimuli and to developmental effects is a critical first step in demonstrating that the method is a valid approach to assessing speech perception in toddlers. Further, individual data reveal that of the 90% of toddlers who met the teaching criterion, the vast majority were able to complete the procedure; only two of the youngest participants struggled with the task. Moreover, the results demonstrate that the modified procedure is highly reliable across testing sessions, despite this population's notoriously variable performance. Although the investigation did not evaluate which modifications (or if all of them) were responsible for making it a viable procedure to use with toddlers, the developmentally appropriate response task and the multiple forms of reinforcement, including tangible rewards, are likely contenders, because these are primarily what makes this modified procedure different from others that have been used with limited success in this population.

To address task difficulty versus perceptual contrast difficulty, we examined performance across conditions and between the post-test training-contrast trials. To balance number of post-test trials with length of testing, only four post-test training-contrast trials were used. Although there might not have been enough trials to be adequately sensitive to performance differences, other procedures, such as the VRASPAC[9], the OlimSPAC[22]), and the SP-EYE[24] use a similar number of test trials as we used in our post-test training-contrast trials. The VRASPAC used up to 5 trials per speech-sound contrast[9], as did the SP-EYE[24] and the On-line IMSPAC uses 8 trials per speech-sound contrast[22]. Therefore, we contend that as a first attempt to tease these issues apart, use of performance on post-test training-contrast trials is appropriate. Because the vast majority of the children, even some of the youngest ones, displayed evidence of discrimination of the contrasts, task difficulty versus perceptual difficulty need only be addressed for a few children. Those children with low scores (but that were above chance) on the hard test contrast, but not on the easy and/or training contrasts, were not likely having difficulty carrying out the procedure, because they were able to complete it in easier conditions. Further, they typically responded correctly to three or four of the post-test training-contrast trials after the hard condition. Therefore, these children could do the procedure, but were struggling to consistently discriminate the difficult place-of-articulation contrast. One toddler (age 2;3;21) never clearly demonstrated the ability to complete the procedure. At no time did this child show clear evidence of either discrimination or task understanding. The 2-year-olds who had low discrimination scores in some conditions seemed to show similar overall performance trends: their training and hard test contrast scores were low, whereas there was some evidence of discrimination on the easy contrast. Perhaps the performance on the easy contrast was an anomaly or perhaps these children could not discriminate the hard contrast well, but were showing some learning effects. Certainly, there is evidence for some learning effects overall and for individual children: there is a trend for the training contrast performance to be lower for all of the age groups than the easy contrast, even though the training contrast is a much easier contrast to perceive. We are unable to tease apart these alternate explanations but future work will address this issue of possible learning effects, by determining when performance plateaus.

A reliable procedure that can evaluate speech perception in toddlers could help begin filling in the glaring gaps in the developmental speech perception literature. In particular, it could begin addressing some inconsistent or conflicting conclusions about how speech perception

develops in infancy through childhood. For example, many investigators have reported that newborn infants are capable of discriminating phonemic contrasts categorically[1, 65–69]. In contrast, Nittrouer[70] has shown that large percentages of both infants and preschoolers are unable to discriminate native contrasts. In an attempt to reconcile these contradictory findings, Nittrouer et al.[42] proposed the Developmental Weighting Shift based on evidence that children weight acoustic properties upon which phonetic decisions are made differently from adults[44–45, 71–74]. The model assumes that infants have some general auditory boundaries that begin to shift in early childhood with language experience due to weights that children ascribe to different acoustic aspects of speech and that these weights are based on the linguistic decisions made by the child. Although the Developmental Weighting Shift, as well as other models of developmental speech perception, attempt to account for the pediatric speech perception data, gaps in understanding remain due to not being able to evaluate the toddler population effectively. The preliminary results reported here suggest that, with further development, the modified Change/No-Change procedure might be able to be used as one approach for investigating toddler speech perception.

Although the results are promising, there are some important limitations of the current form of the modified procedure. First, four of the youngest children (approximately 10% of the sample) were unable to meet the learning criterion of getting five correct responses in a row to the live-voice teaching contrast. Although studies on infant perception often report nearly 50% attrition rates[70, 75–76], if 10% of the normal-hearing toddlers are unable to learn the procedure, further modifications are needed before the procedure can be evaluated in children with hearing loss or other populations who have deficient or fragile phonological representations (e.g., children with specific language impairment[30–31], children experiencing reading deficits[32–33] and second-language learners[35–36]). Second, some children had more difficulty discriminating the training contrast than the easy contrast even though the training contrast is a much easier contrast to discriminate perceptually. This suggests that some of the toddlers were showing learning effects. Therefore, a better understanding of the learning trajectory on this task will help establish when performance has stabilized. Third, the toddlers who had difficulty with the training contrast (primarily those 2;6 and younger) had negative d' scores, suggesting that perhaps these children had significant difficulty with the task when the auditory stimuli no longer shared a direct relationship with the visual response pictures (as was the case with the teaching stimuli), and/or when the stimuli were no longer presented audiovisually via live-voice but instead were recorded auditory-only stimuli, and/ or when the contrast increased in difficulty. The current study design does not allow us to evaluate these possible explanations. Therefore, future research will evaluate further modifications, such as using response pictures that semantically map to the stimuli in test conditions (e.g., a picture of a sheep for /ba/ or a picture of a sock for /sa/), so as to eliminate the need for toddlers to make the cognitive leap from live-voice teaching contrasts that are semantically related to the response pictures to recorded training and testing contrasts that have no semantic relationship to the response pictures. Finally, the procedure in its current form requires more testing time that most clinics have for hearing aid fitting and cochlear implant programming verification. Analyses of the first 20 of the total 36 trials in each condition preliminary suggest that fewer trials in each condition would not likely compromise the validity and reliability of the procedure for assessing speech-sound discrimination in toddlers. Scores for the first twenty trials were highly correlated ($p = .0001$) with those from all 36 Training contrast ($r = .959$), Easy contrast ($r = .914$) and Hard contrast ($r = .930$) trials. A repeated measures ANOVA on only the first 20 trials in each condition suggests that listeners are still better at discriminating the easy than the hard contrast, $F(1, 27) = 11.713$, $p = .002$ and that older toddlers were more sensitive to the contrasts than younger toddlers, $F(2, 27) = 8.628$, $p = .001$. Finally, test-retest reliability for the first 20 trials in each condition was highly significant ($r = .724$, $p < .0001$). More work is necessary to carefully evaluate the effects of reducing the number of trials further.

Despite the limitations noted, the search for valid, reliable methods to test speech discrimination in toddlers remains essential. The lack of speech perception testing procedures for toddlers[7, 9] has led not only to gaps in knowledge of how children's speech perception develops, particularly during rapid periods of language acquisition[5, 12–13], but also has limited how clinical audiologists evaluate benefit from sensory aids in toddlers and in older children with limited language skills. Currently, best practices in pediatric hearing aid fitting involve using a prescriptive gain method developed specifically to address the listening needs of children, such as the Desired Sensation Level (DSL)[77–79] and DSL [i/o][80]. The goals of the DSL approaches are to make speech consistently audible, comfortable, and distortion free, and to provide the widest frequency spectrum possible[79]. Even though pediatric prescriptive gain methods call for the added gain that children require, they still use an audibility-based model of prescribing gain. This is probably appropriate for adults who have a solid, robust language base from which to draw when adjusting to amplification. But for toddlers, who by their very nature are developing a language base, the assumption that audibility alone is sufficient for learning language might not be accurate. In fact, pediatric hearing aid users with mild to severe hearing losses display delays in language, vocabulary, and psychosocial and adaptive skills[81]. Even children with minimal hearing losses are more than three times as likely to experience difficulties in academics, communication, and attention as their normal-hearing peers and the chances of repeating a grade is significantly higher for children with hearing loss than it is for normal-hearing children[82]. More effective and targeted early intervention could reduce the chances of academic failure for children with hearing aids.

Toddlers with cochlear implants also stand to benefit from additional assessment methods. Certainly spoken language testing of children with cochlear implants is far more common clinical practice than that of children with hearing aids, however, as the age at cochlear implantation continues to decline, more and more young children will require spoken language perception testing, which currently is limited in the toddler population.

Developing a method for assessing children's perceptual development at a younger age than is currently possible is critical to determining if adjustments are needed to children's hearing aids and cochlear implants, and would lead directly to intervention strategies for young children and their families. Children who demonstrate strong auditory skills on this discrimination task would most likely benefit from auditory/oral modes of communication, whereas those who show reduced auditory discrimination might benefit most from multisensory communication modes. The results also could reduce costly decision-making delays, as well as families' and early intervention workers' effort and frustration during the earliest stages of intervention, which are known to be critical in developing normal speech and language.

Conclusion

The modifications made to the Change/No-Change procedure implemented here resulted in one of the first procedures shown, at least preliminarily, to be sensitive to differences in toddlers' speech-sound discrimination. Further, the majority of toddlers tested could complete the procedure. Moreover, the results are highly reliable – a critical attribute should the procedure be used to track perceptual development or to evaluate pre- vs. post-sensory aid fitting performance. Future research will address the shortcomings of the current investigation by evaluating alternate teaching strategies and/or performance criteria, as well as implementing more developmentally appropriate modifications intended to promote transfer of learning to the testing phase and improving the percentage of young toddlers that can be tested. Finally, evaluating modification that transfer easily to clinical practice, such as reducing testing time, will be important in future investigations.

Acknowledgments

References

1. Sussman JE, Carney AE. Effects of transition length on the perception of stop consonants by children and adults. J Speech Hear Res. 1989; 36:380–395. [PubMed: 8487530]

2. Carney AE, Moeller MP. Treatment efficacy: hearing loss in children. J Speech Lang Hear Res. 1998; 41:S61–S84. [PubMed: 9493747]

3. Moeller MP, Hoover B, Putman C, Arbataitis K, Bohnenkamp G, Peterson B, et al. Vocalizations of infants with hearing loss compared with infants with normal hearing: Part I – phonetic development. Ear Hear. 2007; 28:605–627. [PubMed: 17804976]

4. Moeller MP, Hoover B, Putman C, Arbataitis K, Bohnenkamp G, Peterson B, et al. Vocalizations of infants with hearing loss compared with infants with normal hearing: part II – transition to words. Ear Hear. 2007; 28:628–642. [PubMed: 17804977]

5. Walley, AC. Speech perception in childhood. In: Pisoni, DB.; Remez, RE., editors. The Handbook of Speech Perception. Malden: Blackwell Publishing; 2005. p. 449-468.

6. Dawson PW, Nott PE, Clark GM, Cowan RSC. A modification of play audiometry to assess speech discrimination ability in severe-profoundly deaf 2–4 year old children. Ear Hear. 1998; 19:371–384. [PubMed: 9796646]

7. Eisenberg LS, Martinez AS, Boothroyd A. Assessing auditory capabilities in young children. Int J Pediatr Otorhinolaryngol. 2007; 71:1339–1350. [PubMed: 17604127]

8. Hnath-Chisolm TE, Laipply E, Boothroyd A. Age-related changes on a children's test of sensory-level speech perception capacity. J Speech Lang Hear Res. 1998; 41:94–106. [PubMed: 9493737]

9. Martinez A, Eisenberg L, Boothroyd A, Visser-Dumont L. Assessing speech pattern contrast perception in infants: early results on VRASPAC. Otol Neurotol. 2008; 29:183–188. [PubMed: 18223444]

10. Schneider BA, Trehub SE, Thorpe L. Developmental perspectives on the localization and detection of auditory signals. Percept Psychophys. 1991; 49:10–20. [PubMed: 2011447]

11. Trehub SE, Schneider BA, Henderson JL. Gap detection in infants, children, and adults. J Acoust Soc Am. 1995; 98:2532–2541. [PubMed: 7593935]

12. Bloom, L. Language Development from Two to Three. New York: Cambridge University Press; 1993.

13. Smith, ME. Studies in Child Welfare. Vol. vol. 3. Iowa City: University of Iowa; 1926. An investigation of the development of the sentence and the extent of vocabulary in young children.

14. Coplan J, Gleason JR. Unclear speech: recognition and significance of unintelligible speech in preschool children. Pediatrics. 1988; 82:447–452. [PubMed: 3405680]

15. Weiss, CE.; Lillywhite, HS. Communicative disorders. St Louis: Mosby; 1976.

16. Boothroyd A. Auditory perception of speech contrasts by subjects with sensorineural hearing loss. J Speech Hear Res. 1984; 27:134–144. [PubMed: 6716999]

17. Eisenberg, LS.; Martinez, AS.; Boothroyd, A. Perception of phonetic contrasts in infants: development of the VRASPAC. In: Miyamoto, RT., editor. International Congress Series 1273: Proceedings of the VIII International Cochlear Implant Conference; Elsevier; Amsterdam. 2004. p. 364-367.

18. Boothroyd A. Evaluation of speech production of the hearing impaired: some benefits of forced-choice testing. J Speech Hear Res. 1985; 28:185–196. [PubMed: 4010248]

19. Boothroyd, A.; Hanin, L.; Eran, O. Speech perception and production in children with hearing impairment. In: Bess, FH.; Gravel, JS.; Tharpe, AM., editors. Amplification for Children with Auditory Deficits. Nashville: Bill Wilkerson Center Press; 1996. p. 55-74.

20. Eilers RE, Wilson WR, Moore JM. Developmental changes in speech discrimination in infants. J Speech Hear Res. 1977; 20:766–780. [PubMed: 604689]

21. Primus MA, Thompson G. Response strength of young children in operant audiometry. J Speech Hear Res. 1985; 28:539–547. [PubMed: 4087889]

22. Eisenberg LS, Martinez AS, Boothroyd A. Auditory-visual and auditory-only perception of phonetic contrasts in children. Volta Rev. 2003; 103:327–346.

23. Boothroyd A, Eisenberg LS, Martinez AS. An on-line imitative test of speech-pattern perception (OlimSPAC): developmental effects in normally hearing children. J Speech Lang Hear Res. 2010; 41:S61–S84.

24. Newton C, Chiat S, Hald L. Evaluation of a novel technique for assessing speech discrimination in children. Clin Linguist Phon. 2008; 22:325–333. [PubMed: 18415732]

25. Holt RF, Carney AE. Developmental effects of multiple looks in speech sound discrimination. J Speech Lang Hear Res. 2007; 50:1404–1424. [PubMed: 18055765]

26. Holt RF, Carney AE. Multiple looks in speech sound discrimination in adults. J Speech Lang Hear Res. 2005; 48:922–943. [PubMed: 16378483]

27. McClelland JL, Elman JL. The TRACE model of speech perception. Cognit Psychol. 1986; 18:1–86. [PubMed: 3753912]

28. Luce PA, Goldinger SD, Auer ET, Vitevich MS. Phonetic priming, neighborhood activation, and PARSYN. Percept Psychophys. 2000; 62:615–625. [PubMed: 10909252]

29. Norris D. Shortlist: a connectionist model of continuous speech recognition. Cognition. 1994; 52:189–234.

30. Coady JA, Evans JL. Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). Lang Commun Disord. 2008; 43:1–40.

31. Gathercole SE, Baddeley AD. Phonological memory deficits in language disordered children: is there a causal connection? J Mem Lang. 1990; 29:336–360.

32. Snowling, MJ. Dyslexia. 2nd ed.. New York: Malden Blackwell Publishing; 2000.

33. Snowling M, Bishop DVM, Stothard SE. Is preschool language impairment a risk factor for dyslexia in adolescence? J Child Psychol Psychiatry. 2000; 41:587–600. [PubMed: 10946751]

34. Hu CF. Rate of acquiring and processing L2 color words in relation to L1 phonological awareness. Mod Lang J. 2008; 92:39–52.

35. Papagno C, Valentine T, Baddeley A. Phonological short-term memory and foreign language vocabulary learning. J Mem Lang. 1991; 30:331–347.

36. Carney, AE.; Osberger, MJ.; Miyamoto, RT.; Karasek, A.; Dettman, DL.; Johnson, DL. Speech perception along the sensory aid continuum: from hearing aids to cochlear implants. In: Feigin, J.; Stelmachowicz, P., editors. Proceedings of the 1991 National Conference on Pediatric Amplification; Boys Town National Research Hospital; Omaha. 1991. p. 93-113.

37. Carney AE, Osberger MJ, Carney E, Robbins AM, Renshaw J, Miyamoto RT. A comparison of speech discrimination with cochlear implants and tactile aids. J Acous Soc Am. 1993; 94:2036–2049.

38. Osberger MJ, Miyamoto RT, Zimmerman-Philips S, Kemink J, Stroer BS, Firszt J, et al. Independent evaluation of the speech perception abilities of children with the Nucleus 22-Channel cochlear implant system. Ear Hear. 1991; 12(Suppl.):66S–80S. [PubMed: 1955092]

39. Sussman JE. Stimulus ratio effects on speech discrimination by children and adults. J Speech Hear Res. 1991; 34:671–678. [PubMed: 2072692]

40. Sussman JE. Auditory processing in children's speech perception: Results of selective adaptation and discrimination tasks. J Speech Hear Res. 1993; 36:380–395. [PubMed: 8487530]

41. Coplan, J. Early Language Milestone Scale Scale-2. Austin, TX: Pro-Ed; 1993.

42. Zimmerman, IL.; Steiner, VG.; Pond, RE. Preschool Language Scale. 4th Ed.. San Antonio, TX: The Psychological Corporation; 2002.

43. Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. J Acoust Soc Am. 1955; 27:338–352.

44. Nittrouer S. Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. J Phon. 1992; 20:1–32.

45. Nittrouer S. Discriminability and perceptual weighting of some perceptual cues to speech perception by 3-year-olds. J Speech Hear Res. 1996; 39:278–287. [PubMed: 8729917]

46. Nittrouer S. The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM. J Speech Hear Res. 1996; 39:1059–1070. [PubMed: 8898258]

47. Nittrouer S, Manning C, Meyer G. The perceptual weighting of acoustic cues changes with linguistic experience. J Acoust Soc Am. 1993; 94:S1865.

48. Nittrouer S, Miller ME. Predicting developmental shifts in perceptual weighting schemes. J Acoust Soc Am. 1997; 101:2253–2266. [PubMed: 9104027]

49. Adobe Adobe Audition 2.0.[computer program]. San Jose, CA: Adobe Systems, Inc.; 2005.

50. Psychology Software Tools. E-Prime 2.0. [computer program]. Pittsburgh, PA: Psychology Software Tools; 2007.

51. Levy F. The development of sustained attention (vigilance) and inhibition in children: some normative data. J Child Psychol Psychiatry. 1980; 21:77–84. [PubMed: 7358807]

52. Enns, JT. The Development of Attention. Amsterdam: Elsevier; 1990.

53. Gomes H, Molholm S, Christodoulou C, Ritter W, Cowan N. The development of auditory attention in children. Front Biosci. 2000; 5:d108–d120. [PubMed: 10702373]

54. Ruff HA, Lawson KR. Development of sustained, focused attention in young children during free play. Dev Psychol. 1990; 26:85–93.

55. Ruff, HA.; Rothbart, MK. Attention in Early Development. New York: Oxford University Press; 1996.

56. Fulton RT, Gorzycki PA, Hull WL. Hearing assessment with young children. J Speech Hear Disord. 1975; 40:397–404. [PubMed: 1234953]

57. Gabbard, CP. Lifelong Motor Development. 4th ed.. San Francisco: Benjamin Cummings; 2004.

58. Boothroyd A. Auditory capacity of hearing-impaired children using hearing aids and cochlear implants: issues of efficacy and assessment. Scand Audiol. 1997; 26(Suppl. 46):17–25.

59. Ceponiene R, Cheour M, Naatanen R. Interstimulus interval and auditory event-related potentials in children: evidence from multiple generators. Electroencephalogr Clin Neurophysiol. 1998; 108:345–354. [PubMed: 9714376]

60. Gilley PM, Sharma A, Dorman M, Martin K. Developmental changes in refractoriness of the cortical auditory evoked potential. Clin Neurophysiol. 2005; 116:648–657. [PubMed: 15721079]

61. Shamra A, Kraus N, McGee TJ, Nicol TG. Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. Electroencephalogr Clin Neurophysiol. 1997; 104:540–545. [PubMed: 9402896]

62. Werner, LA.; Rubel, EW. Developmental Psychoacoustics. Washington, D.C.: American Psychological Association; 1992.

63. Luce PA, Feustel TC, Pisoni DB. Capacity demands in short-term memory for synthetic and natural speech. Hum Factors. 1983; 16:17–32. [PubMed: 6840769]

64. Marzolf DP, DeLoache JS. Transfer in young children's understanding of spatial representation. Child Dev. 1994; 65:1–15. [PubMed: 8131641]

65. Bertoncini J, Bijeljac-Babic R, Blumstein SE, Mehler J. Discrimination in neonates of very short CV's. J Acoust Soc Am. 1987; 82:31–37. [PubMed: 3624638]

66. Eimas PD. Auditory and linguistic processing of cues for place of articulation by infants. Percept Psychophys. 1974; 16:513–521.

67. Eimas PD. Auditory and phonetic coding of the cues for speech: Discrimination of the (r-l) distinction by young infants. Percept Psychophys. 1975; 18:341–347.

68. Eimas PD, Miller JL. Discrimination of the information for manner of articulation. Infant Behav Dev. 1980; 3:367–375.

69. Kuhl PK, Miller JD. Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. Percept Psychophys. 1982; 31:279–292. [PubMed: 7088673]

70. Nittrouer S. Challenging the notion of innate phonetic boundaries. J Acoust Soc Am. 2001; 110:1598–1605. [PubMed: 11572369]

71. Morrongiello BA, Robson RC, Best CT, Clifton RK. Trading relations in perception of speech by five-year-old children. J Exp Child Psychol. 1984; 37:231–250. [PubMed: 6726113]

72. Nittrouer S, Crowther C, Miller ME. The relative weighting of acoustic properties in the perception of s] + stop clusters by children and adults. Percept Psychophys. 1998; 60:51–64. [PubMed: 9503911]

73. Nittrouer S, Miller ME, Crowther CS, Manhart MJ. The effect of segmental order on fricative labeling by children and adults. Percept Psychophys. 2000; 62:266–284. [PubMed: 10723207]

74. Parnel MM, Amerman JD. Maturational influences on perception of coarticulatory effects. J Speech Hear Res. 1978; 21:682–701. [PubMed: 745369]

75. Bertoncini J, Bijeljac-Babic R, Jusczyk PW, Kennedy LJ, Mehler J. An investigation of young infants' perceptual representations of speech sounds. J Exp Psychol Gen. 1988; 117:21–33. [PubMed: 2966228]

76. Jusczyk PW, Pisoni DB, Mullennix J. Some consequences of stimulus variability on speech processing by two-month-old infants. Cognition. 1992; 43:253–291. [PubMed: 1643815]

77. Moodie K, Seewald R, Sinclair S. Procedure for predicting real-ear hearing aid performance in young children. Am J Audiol. 1994; 3:23–31.

78. Seewald R, Ross M, Spiro M. Selecting amplification characteristics for young hearing-impaired children. Ear Hear. 1985; 6:28–53.

79. Seewald R, Ross M, Stelmachowicz P. Selecting and verifying hearing aid performance characteristics for young children. J Acad Rehabil Audiol. 1987; 20:25–38.

80. Seewald, R.; Cornelisse, L.; Ramji, K.; Sinclair, S.; Moodie, K.; Jamieson, D. DSL v. 4 for Windows: a software implementation of the Desired Sensation Level (DSL[i/o]) method for fitting linear gain and wide-dynamic-range compression hearing instruments. London Ontario, Canada: Hearing Health Care Research Unit, The University of Western Ontario; 1997.

81. Wake M, Hughes EK, Poulakis Z, Collins C, Rickards FW. Outcomes of children with mild-profound congenital hearing loss at 7 to 8 years: a population study. Ear Hear. 2004; 25:1–8. [PubMed: 14770013]

82. Bess FH, Dodd-Murphy J, Parker RA. Children with minimal sensorineural hearing loss: prevalence, educational performance, and functional status. Ear Hear. 1998; 19:339–354. [PubMed: 9796643]

Appendix A

Preliminary Stimulus Selection Experiments: Stimulus Identification and Rating

In the first stimulus selection experiment, 10 normal-hearing adults who spoke English as their first or only language (mean age: 24.3 years; age range: 21–32 years) were presented with all of the tokens from each speaker in random order in the sound field at 65 dB A. The participants were seated in a double-walled sound booth in front of a touch screen monitor. On the monitor were seven squares: six contained the orthographic representation of each of the syllables ("ooo," "gah," "boo," "bah," "sa," and "sha") and the seventh square contained an "o" for "other." Participants were instructed to touch the box that represented the syllable presented on each trial or to touch the "o" box if the syllable did not match any of the six syllables options. The "other" option was used to discourage subjects from guessing if the spoken syllable was a poor representation of the target. Each response was coded as either correct or incorrect. Both the presentation of stimuli and recording of responses were controlled using E-Prime2 software[50]. Table A2 in this Appendix displays the total percent correct identification of each syllable for each of the three speakers. Overall, the first two speakers were highly intelligible across all six syllables. However, Speaker 1 was the most intelligible and there was less variability in her intelligibility across the six tokens, with no average rating falling below 96%. Therefore, Speaker 1's tokens were used in the second stimulus selection experiment.

The purpose of the second stimulus selection experiment was to select twenty tokens of each syllable from Speaker 1 based on how well they exemplified the intended target. Only those tokens that were identified with 100% accuracy in the first stimulus selection experiment were used in this, the second stimulus selection experiment. To make the training contrast maximally distinguishable, the /u/ tokens identified with perfect accuracy in the first stimulus selection experiment were elongated using Audition v. 2.0[49]. This artificial process is preferable to directing the speakers to extend the duration of /u/ tokens naturally, as doing so can affect the fundamental frequency contour of the long /u/ tokens, the relative vocal effort across the duration of the syllable, and the duration of other syllables. Each /u/ token was subjected to a 67% decrease in tempo using the time stretch effect, a stretch option in Audition v. 2.0 that does not alter pitch. This adjustment increased the duration of /u/ stimuli from an average of 531 ms to an average of 787 ms. Note that the value of 531 ms is not the same as the average value for the /u/ stimuli in Table A1, because in this phase of stimulus selection only those tokens identified with perfect accuracy were included in the duration calculation.

Seated in front of a touch screen monitor, 10 additional normal-hearing adults who spoke English as their first language (mean age: 24.4 years; age range: 22–37 years) were presented with all of the tokens from Speaker 1 that were identified with perfect accuracy in the first stimulus selection experiment, as well as the elongated /u/ tokens, in the sound field at 65 dB A. On each trial, an orthographic representation of the syllable appeared on the monitor. Underneath the syllable were seven squares labeled with the integers 1 through 7. Participants were instructed to rate how good of an example each syllable presentation was (1 = poor example; 7 = excellent example). Average ratings and standard deviations for the 20 highest-rated tokens of each syllable appear in Table A3 of this Appendix. Average ratings ranged from 5.53 to 6.40 (out of 7.00), suggesting the selected tokens were not only identified with perfect accuracy, but were also rated as very good exemplars of the intended targets. Therefore, these 20 highest-rated tokens for /ga/, long /u/, /ba/, /bu/, /sa/, and /ʃa/ from Speaker 1 were used as the final stimuli that were presented to the children in the discrimination procedure.

Table A4 in this Appendix displays the duration data for the final set of tokens used in the discrimination procedure. Paired-samples t-tests revealed that the durations between the sets of tokens within each contrast were not significantly different from each other for the hard contrast (/sa/ vs. /ʃa/, $p = .242$), approached significance for the easy contrast (/ba/ vs. /bu/, $p = .053$), and was significant for the training contrast (long /u/ vs. /ga/), $t(19) = 21.158$, $p < .0001$. The syllables used in the training contrast were intended to be different in length by design. That the easy contrast contained a potentially more noticeable duration cue ought to, if anything, make the easy contrast even easier to discriminate than the hard contrast.

Table A1

Descriptive data for the natural, recorded, digitized stimuli that were within 50 ms of the mean duration of each of the three speaker's syllable utterances

	Speaker 1		Speaker 2		Speaker 3	
Syllable	# tokens	Mean length (SD)	# tokens	Mean length (SD)	# tokens	Mean length (SD)
/u/	40	521 ms (29 ms)	40	711 ms (29 ms)	40	542 ms (28 ms)
/ga/	56	555 ms (27 ms)	46	682 ms (45 ms)	36	500 ms (24 ms)
/ba/	46	531 ms (31 ms)	38	640 ms (26 ms)	48	479 ms (23 ms)
/bu/	37	543 ms (26 ms)	46	624 ms (30 ms)	50	468 ms (28 ms)

	Speaker 1			Speaker 2			Speaker 3	
Syllable	# tokens	Mean length (SD)	# tokens	Mean length (SD)	# tokens	Mean length (SD)		
/sa/	44	680 ms (28 ms)	45	784 ms (25 ms)	29	631 ms (35 ms)		
/ʃa/	33	686 ms (29 ms)	40	758 ms (24 ms)	47	635 ms (28 ms)		

\# = Number; SD = Standard Deviation

Table A2

Mean percent correct identification for each syllable produced by the three speakers in the identification portion of the stimulus selection experiments

Speaker	/u/	/ga/	/ba/	/bu/	/sa/	/ʃa/	mean across syllables
Speaker 1	97.60	99.11	97.83	96.49	99.77	99.39	98.36
Speaker 2	99.50	98.89	97.37	94.13	100.00	98.54	98.07
Speaker 3	98.75	98.89	88.33	86.60	99.66	99.79	95.34

Table A3

"Goodness" ratings for the 20 highest-rated syllables spoken by Speaker 1 in the rating portion of the stimulus selection experiments.

	/u/	long /u/	/ga/	/ba/	/bu/	/sa/	/ʃa/
Mean rating (out of 7)	5.94	5.93	6.30	6.16	5.53	6.40	6.19
Standard deviation	0.25	0.26	0.18	0.20	0.49	0.17	0.17

Table A4

Mean durations and standard deviations for the final tokens used in the discrimination procedure

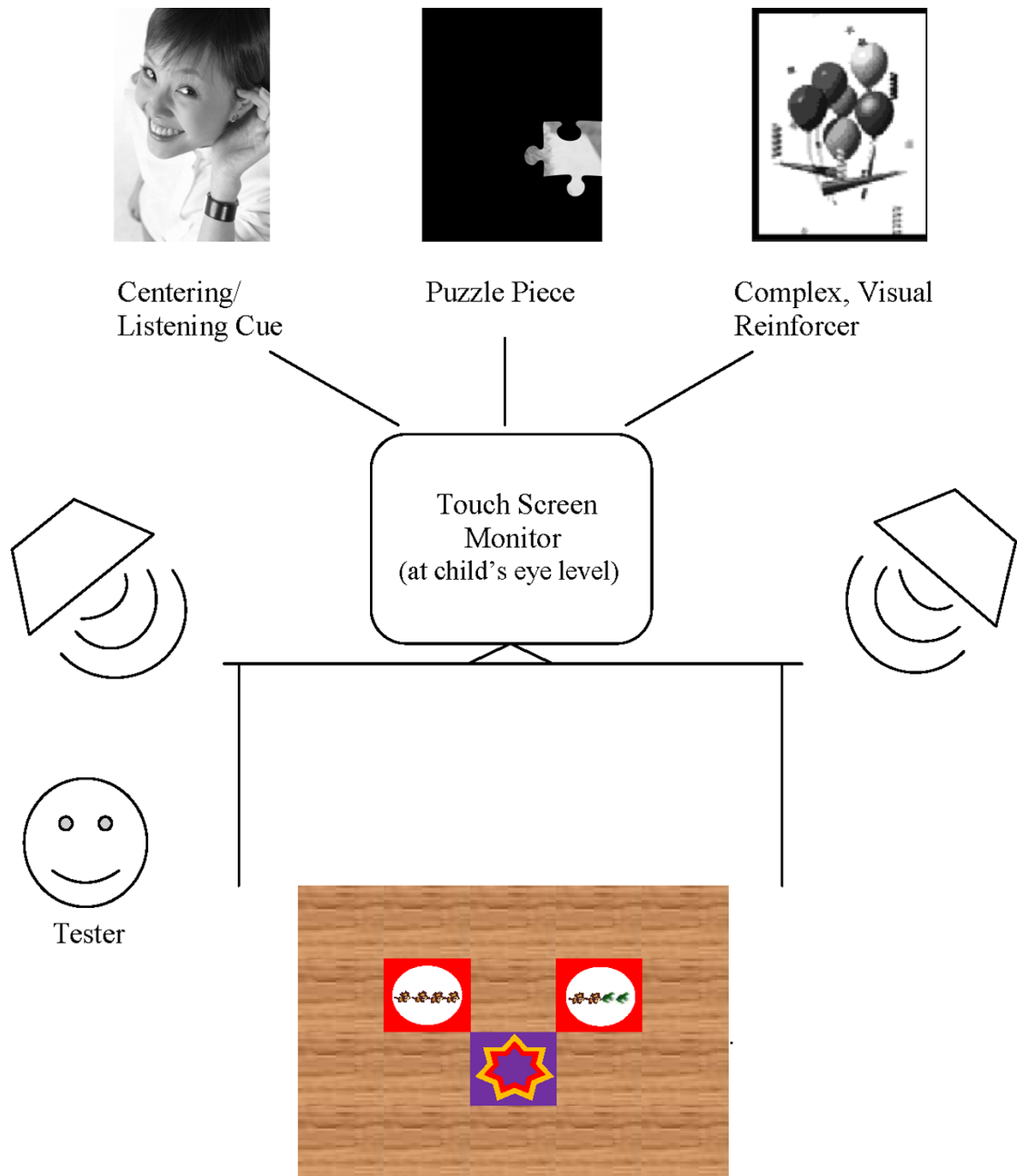|	long /u/	/ga/	/ba/	/bu/	/sa/	/ʃa/
Mean length (ms)	787	551	533	551	678	688
Standard deviation (ms)	49	23	34	17	29	27

Figure 1.
Schematic of the test set-up, including the 4' × 5' SoftTile interlocking foam mat that served as the response mat. This was placed on the floor of the sound booth. The mat had a wood grain pattern, except for two red pieces with circle-shaped cut-outs and a purple piece with a star-shaped cut out. The child stood on the "listening star" and a trial would begin when the child was in a ready state. The two red circles, one to the front/left and the other to the front/right of the star, displayed pictures corresponding to the no-change and change responses, respectively. In this figure, the pictures displayed (cows and frogs) were used in the teaching and training phases. A touch screen monitor was placed on a table at the child's eye level. The touch screen monitor displayed the centering/listening cue, the complex, visual

reinforcers, and the puzzle pieces. The monitor also served as a user interface for the Tester to register the child's responses. Speakers were placed at +/− 45 degrees azimuth. Note that figure is not to scale.
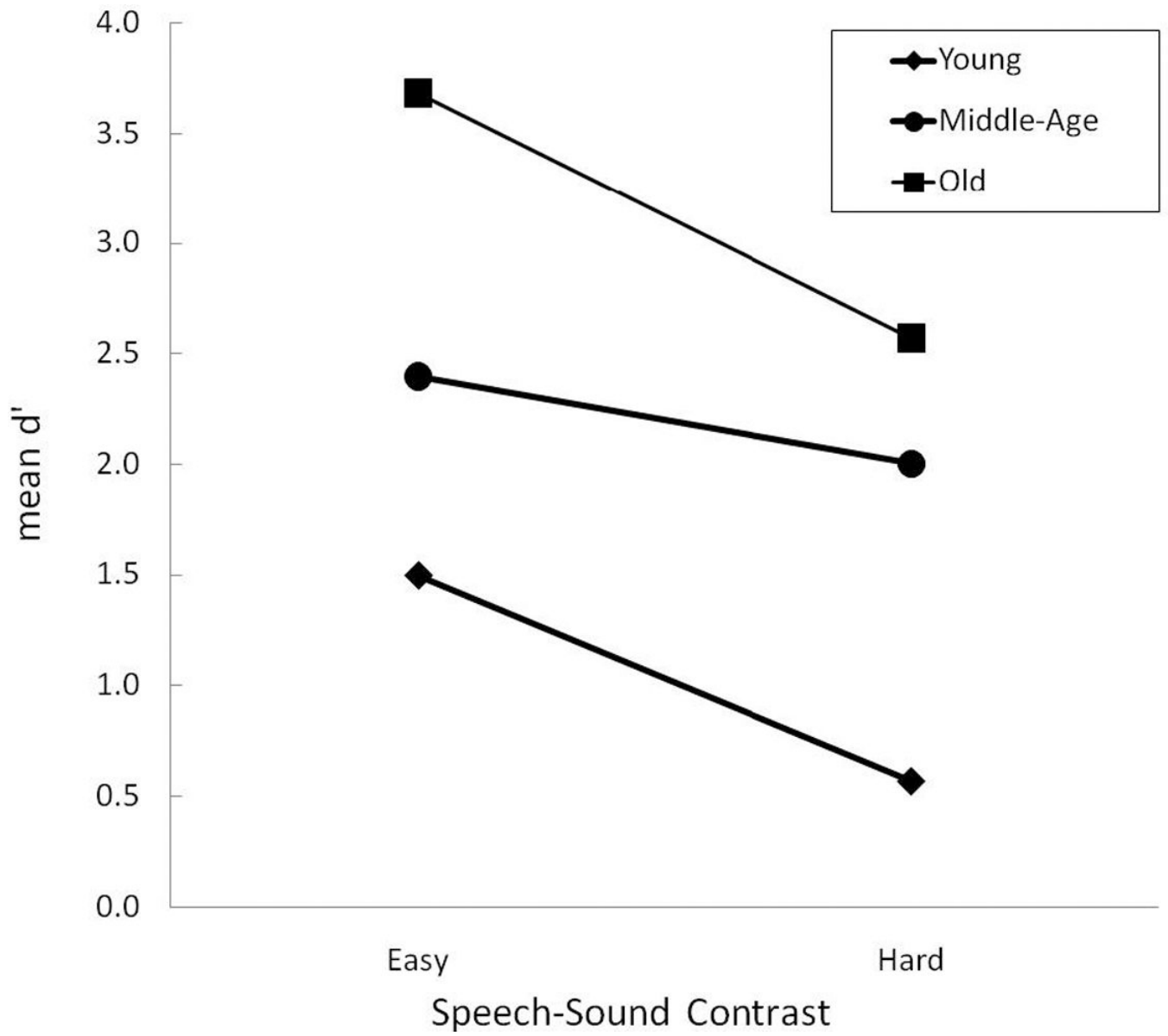
Figure 2.
Mean performance on the perceptually easy and hard test contrasts as a function of participant age group: Young (diamonds), Mid-Age (circles), and Old (squares).
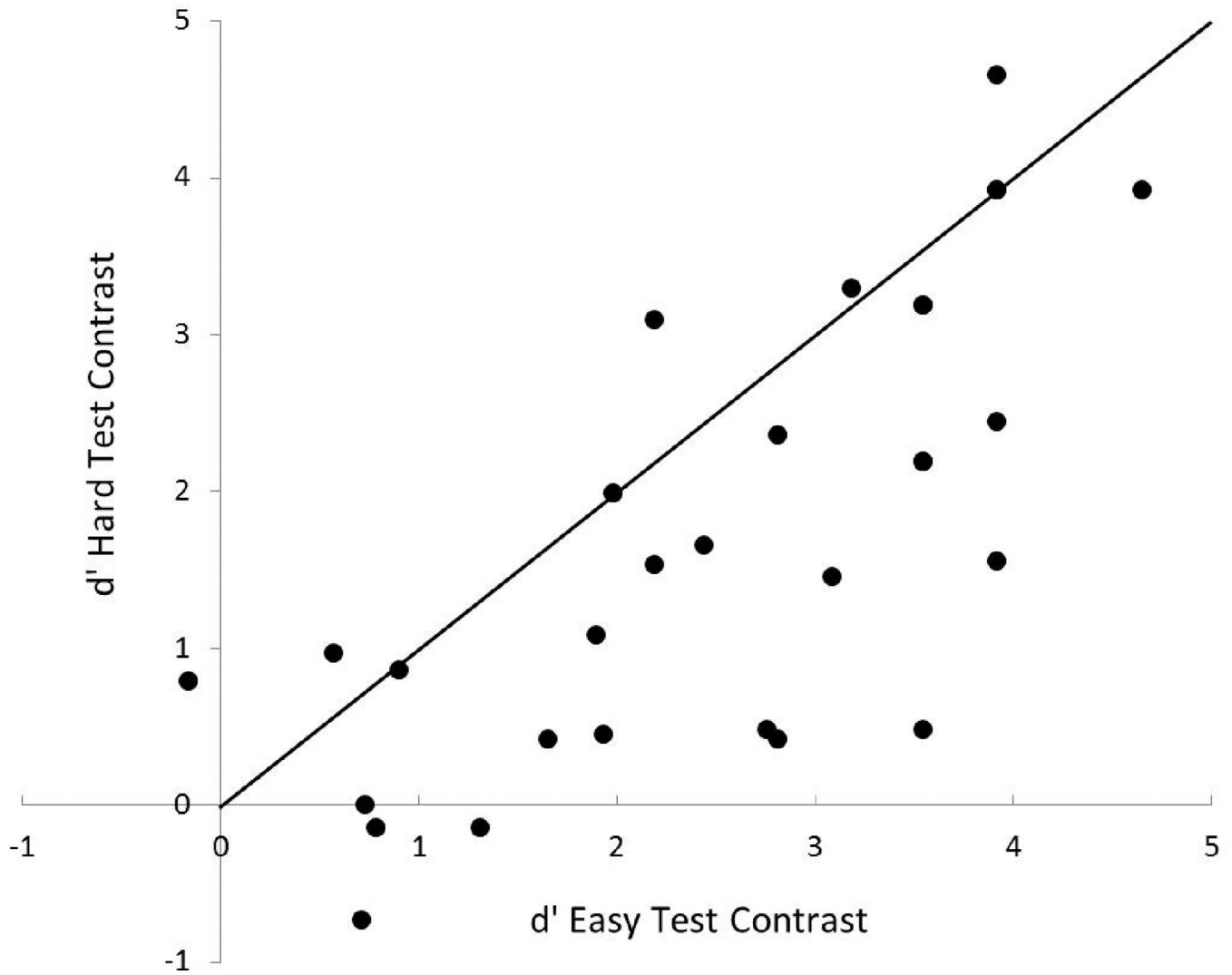
Figure 3.
Scatterplot of individual performance on the perceptually easy (/bu/ vs. /ba/) and hard (/sa/ vs. /ʃa/) test contrasts. The diagonal line displays what would be expected if performance were equivalent on the perceptually easy and hard contrasts. Note that there were 3 pairs of participants who had identical scores on the perceptually easy and hard contrasts, explaining why only 27 data points are plotted in the figure.
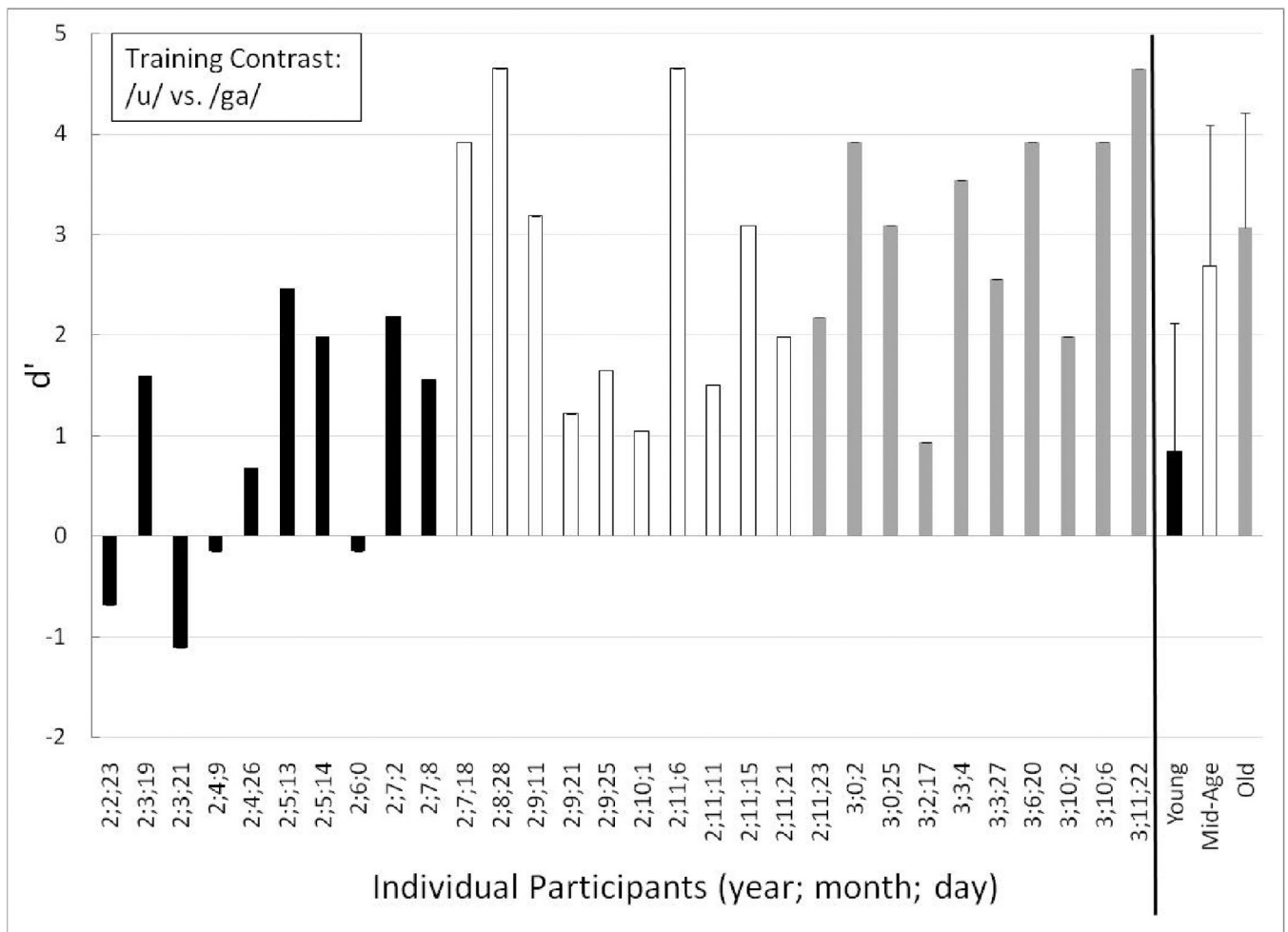
Figure 4.
Individual performance for the Young (black-filled bars), Mid-Age (unfilled bars) and Old toddlers (gray-filled bars) on the training contrast (long /u/ vs. /ga/). The children are ordered from youngest to oldest age at first testing session. At the far right are the mean (and +1 standard deviation) group data for each age group: Young (black-filled bars), Mid-Age (unfilled bars) and Old (gray-filled bars).
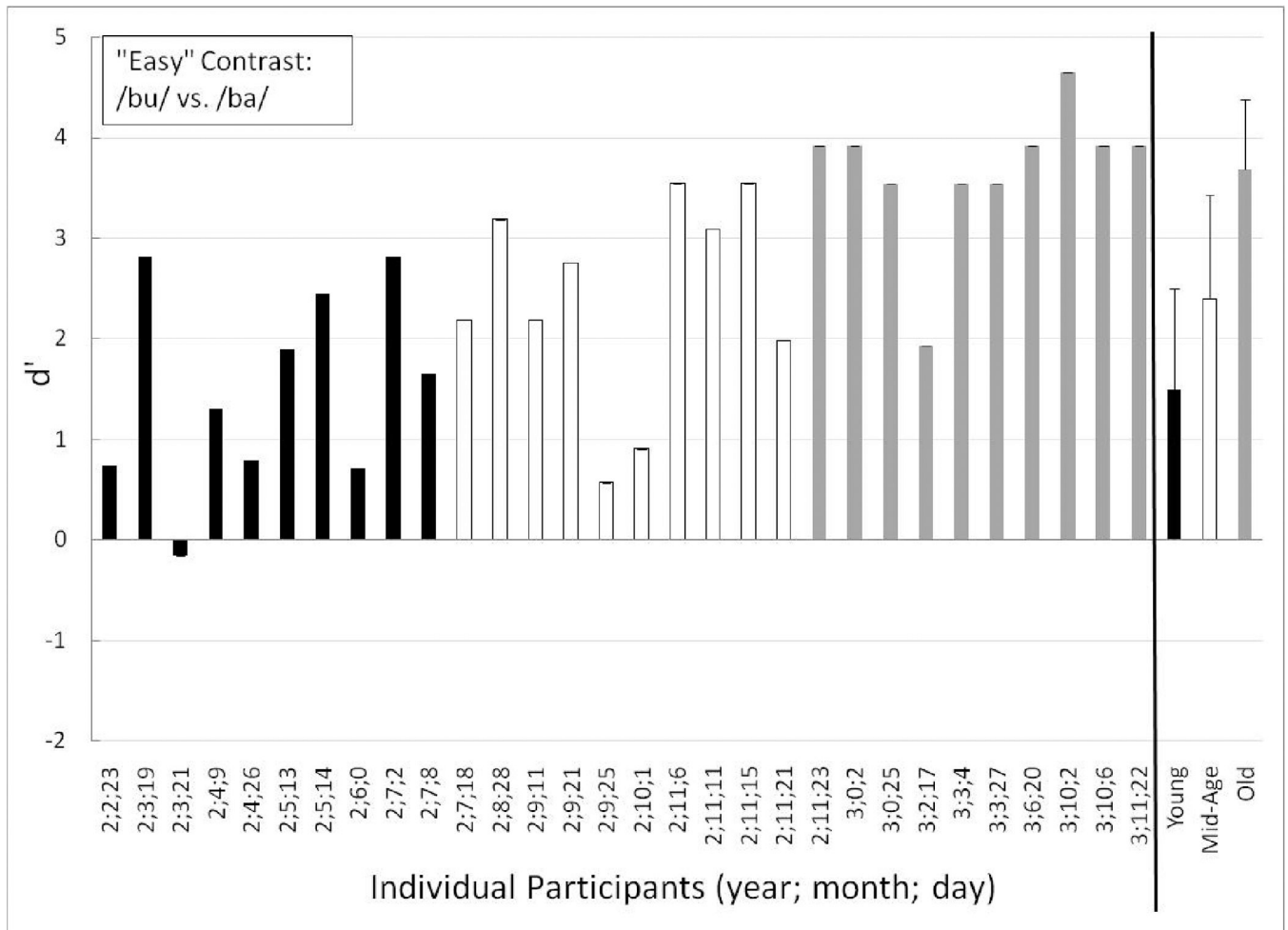
Figure 5.
Individual performance for the Young (black-filled bars), Mid-Age (unfilled bars) and Old toddlers (gray-filled bars) on the perceptually easy test contrast (/bu/ vs. /ba/). The children are ordered from youngest to oldest age at first testing session. At the far right are the mean (and +1 standard deviation) group data for each age group: Young (black-filled bars), Mid-Age (unfilled bars) and Old (gray-filled bars).

Figure 6.
Individual performance for the Young (black-filled bars), Mid-Age (unfilled bars) and Old toddlers (gray-filled bars) on the perceptually hard test contrast (/sa/ vs. /ʃa/). The children are ordered from youngest to oldest age at first testing session. At the far right are the mean (and +1 standard deviation) group data for each age group: Young (black-filled bars), Mid-Age (unfilled bars) and Old (gray-filled bars).
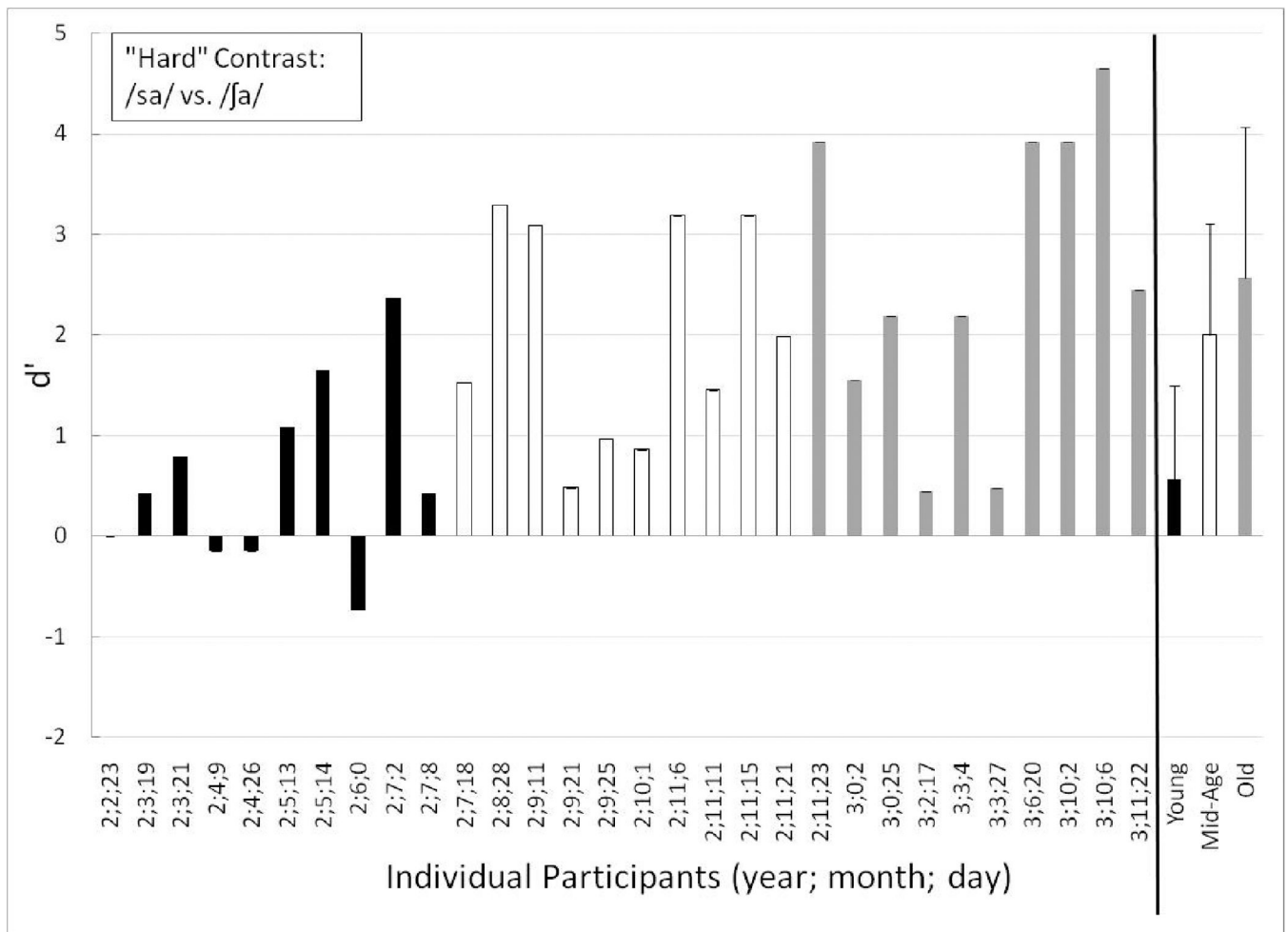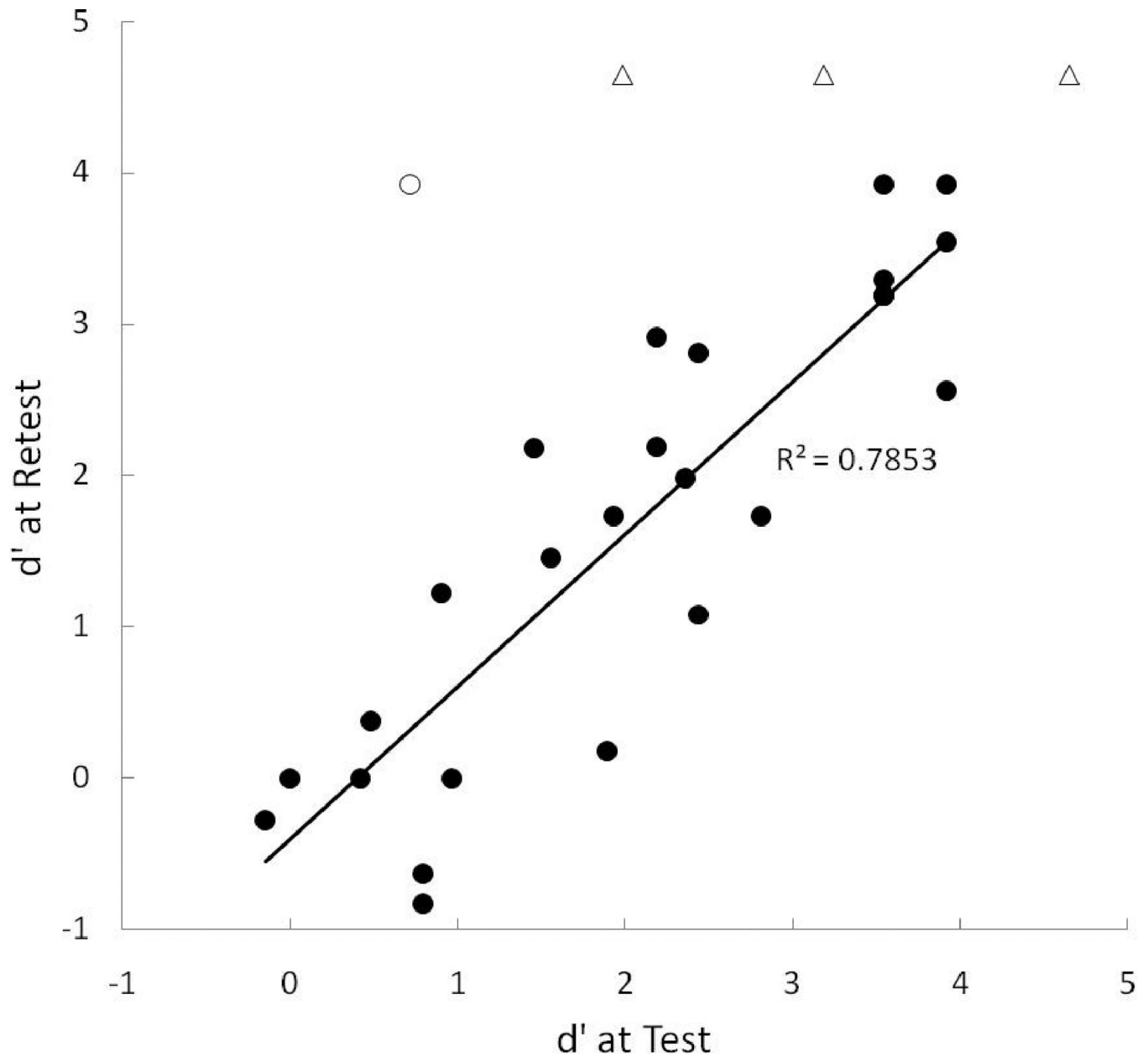
Figure 7.
Scatterplot of individual test and retest scores. The unfilled triangles display data from three toddlers who scored at ceiling at test and/or retest and the unfilled circle represents a single participant who had a 4-month delay between test and retest. The regression line was calculated with data from the toddlers who scored at ceiling and the child with the long test-retest delay excluded from the analysis.

Table 1

Toddler participant demographics

Group	n	Mean Age (year;month)	Age Range (year;month;day)	Number Female
Young	10	2;4	2;2;23 – 2;7;8	4
Mid-Age	10	2;9	2;7;18 – 2;11;21	5
Old	10	3;4	2;11;23 – 3;11;22	5

Table 2

Modifications implemented in the current investigation to the Change/No-Change procedure used by Holt and Carney (2007)

Modification	Previous Procedure	Current Procedure
Testing time reduced: Number of trials per condition	50 trials	50 trials
Testing time reduced: Number of conditions	12 conditions	4 conditions
Response method	Touch pictures on a touch screen monitor while sitting in a chair	Large gross motor response of jumping to, stepping to, or placing a beanbag on one of two response mats on the floor
Easier listening conditions	Stimuli presented in noise	Stimuli presented in quiet
Teach response task	No formal teaching session	Formal teaching session, requiring 5 consecutive correct responses
Evaluate task understanding at the end of each condition	No post-test trials	4 post-test training trials at the end of each condition
Additional reinforcements	Puzzle piece added on screen and verbal praise at the end of each trial	Puzzle piece added on screen and verbal praise at the end of each trial, as well as animated visual reinforcement after correct trials, and periodic tangible reinforcement
Stimulus	Synthetic tokens	Natural, digitized tokens

Table 3

Condition Order and Content of each Condition

Phase	Stimuli	Response Pictures
1. Teaching	Live-voice "moo" v. "ribbit"	Cows and frogs
2. Training	Recorded long /u/ v. /ga/	Cows and frogs
3. Testing	Recorded /ba/ v. /bu/ (perceptually easy) Recorded /sa/ v. /ʃa/ (perceptually hard) (Order randomized)	Chosen by child
4. Retest	Recorded /ba/ v. /bu/ OR /sa/ v. /ʃa/(Randomized)	Chosen by child

Table 4

2-way repeated measures ANOVA and planned post-hoc t-test results

Main Effect or Post-Hoc Comparison	p-value
Contrast	<.0001
Age	<.0001
Contrast X Age	.257
Young vs. Mid-Age	.010
Young vs. Old	<.0001
Mid-Age vs. Old	.037