Speech intelligibility in free field: Spatial unmasking in preschool children

Soha N. Garadat and Ruth Y. Litovsky[a]
Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin 53705

(Received 16 January 2006; revised 22 November 2006; accepted 22 November 2006)

This study introduces a new test (CRISP-Jr.) for measuring speech intelligibility and spatial release from masking (SRM) in young children ages 2.5–4 years. Study 1 examined whether thresholds, masking, and SRM obtained with a test designed for older children (CRISP) and CRISP-Jr. are comparable in 4 to 5-year-old children. Thresholds were measured for target speech in front, in quiet, and with a different-sex masker either in front or on the right. CRISP-Jr. yielded higher speech reception thresholds (SRTs) than CRISP, but the amount of masking and SRM did not differ across the tests. In study 2, CRISP-Jr. was extended to a group of 3-year-old children. Results showed that while SRTs were higher in the younger group, there were no age differences in masking and SRM. These findings indicate that children as young as 3 years old are able to use spatial cues in sound source segregation, which suggests that some of the auditory mechanisms that mediate this ability develop early in life. In addition, the findings suggest that measures of SRM in young children are not limited to a particular set of stimuli. These tests have potentially useful applications in clinical settings, where bilateral fittings of amplification devices are evaluated. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2409863]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.71.Ft, 43.71.Gv [AJO] Pages: 1047–1055

I. INTRODUCTION

Auditory environments in which children spend a majority of their time are typically acoustically complex; multiple acoustic signals are likely to coincide in time, location, and spectrum. In general, to achieve a high level of performance, children must segregate multitudes of signals into their basic components to make sense of the “what” and “where” aspects of the auditory environment. The phenomenon by which the auditory system extracts a distinct message in the presence of other competing sounds is known in general terms as the “cocktail party effect” (Cherry, 1953; Pollack and Pickett, 1958). During the many years since the problem was originally described there has been increasing interest in identifying the mechanisms by which the auditory system teases apart co-occurring sounds and facilitates speech understanding in noisy environments.

Given the complex nature of the “listening-in-noise” problem, it is important to better understand the mechanisms by which children are able to navigate in a complex auditory environment, a good example of which is a classroom. The vast majority of what is known about this ability is derived from studies on adult listeners. A number of studies that simulated aspects of complex auditory environments have shown that speech recognition in noise improves when the target source is spatially separated from the competing sounds (e.g., Hirsh, 1950; Dirks and Wilson, 1969; Plomp, 1976; Bronkhorst and Plomp, 1988; Festen and Plomp, 1990; Yost *et al.*, 1996; Peissig and Kollmeier, 1997; Noble *et al.*, 1997; Freyman *et al.*, 1999; Drullman and Bronkhorst, 2000). This improvement has been referred to as spatial release from masking (SRM), which has been shown to incorporate both binaural and monaural components of auditory processing (Kidd *et al.*, 1998; Culling *et al.*, 2004; Hawley *et al.*, 1999, 2004; Lin and Feng, 2003).

In contrast, little is known about how it is that children succeed in resolving auditory information in complex environments, and the developmental time course of this ability. This ability most likely includes mechanisms that mediate simple aspects of spatial hearing known to reach full maturation in early childhood. One example is the mechanism that contributes to discrimination of sound location for single-source sounds, which is adult-like by 5 years of age (Litovsky, 1997). Other examples exist for abilities that are not inherently spatial, such as frequency resolution, which is adult-like by age 6 (Allen *et al.*, 1989; Hall and Grose, 1991). On the other hand, mechanisms that are still developing are likely to limit children’s performance, such that they would appear to be poorer at extracting auditory information compared with adults. In particular, measures that focus on temporal resolution abilities suggest a more protracted developmental progression that extends into middle-to-late childhood. These include gap detection (Wightman *et al.*, 1989), backward masking (Hartley *et al.*, 2000), amplitude modulation detection at various rates (Hall and Grose, 1994), and masking level difference (Hall and Grose, 1990).

Studies that focused on auditory masking suggest that children have poorer performance in the presence of competing noise compared to that in *quiet* conditions. In general, it has been reported that children require a higher signal-to-noise ratio (SNR) than adults for comparable performance, in other words, they exhibit a greater degree of masking (Elliott *et al.*, 1979; Papso and Blood, 1989). Children also appear to have greater difficulty than adults in extracting a target signal embedded in background noise when listening to multiple

---

[a]Electronic mail: litovsky@waisman.wisc.edu

sound sources (Hall *et al.*, 2002) or under conditions of perceptual masking and signal uncertainty (Allen and Wightman, 1995; Oh *et al.*, 2001; Wightman *et al.*, 2003).

Despite the rich and growing literature on auditory development, an area of research that remains relatively unexplored concerns how children exploit spatial cues in segregating complex signals. In a recent study, Litovsky (2005) showed that children 4.5–7 years of age are able to take advantage of spatial cues to segregate target speech in the presence of competing speech or modulated speech-shaped-noise. Although the children's average speech reception thresholds (SRTs) were higher than those of adults, the amount of masking (differences in SRT between quiet thresholds and conditions in which competitors are present) and SRM (differences in SRTs between masking when the competitors were in front versus on the side) were similar in children and adults, suggesting that by the time children reach school-age some aspects of source segregation are well developed. In a follow-up study, Johnstone and Litovsky (2006) found that, when the maskers consist of time-reversed speech, the novelty of the stimulus rendered the task significantly more difficult for the children, but not for adults, suggesting the importance of nonsensory factors such as informational masking in this task. Further work is needed to determine the robustness of these findings in younger children, and the developmental ontogeny of spatial unmasking. This area of research can not only give some insight into how children perform in environments such as classrooms, but might also be able to provide appropriate tools for evaluating children's performance using listening prostheses such as hearing aids and cochlear implants (Litovsky *et al.*, 2004, 2006).

The present study evaluated SRTs, masking, and SRM in two groups of children, ages 3 and 4–5 years. Although age-appropriate tests for measuring language and vocabulary development are available (e.g., the Reynell Developmental Language Scales III; Ball, 1999), none are geared toward measuring speech intelligibility in a challenging masked situation. In addition, the test used by Litovsky (2005), known as CRISP, uses spondees as targets (e.g., "cowboy," "bluejay," and "bird-nest"). That test was geared toward children with a minimum of 4 years and contains vocabulary that is not age-appropriate for the average younger child. In order to evaluate speech unmasking and SRM in younger children, the design of a novel stimulus set was required. A new test named CRISP-Jr. was developed for this purpose, with a corpus of target words that are estimated to be within the vocabulary of typically developing, normal-hearing 2.5 year olds. Ultimately, this test can also be extended to populations of children who may be older but whose language skills are delayed relative to their age-matched peers.

This paper presents results from two studies. The first study was designed to compare values of SRTs, masking, and SRM using the new set of stimuli (CRISP-Jr.) and the previously used corpus (CRISP). In a second study, the CRISP-Jr. test was extended to a group of 3-year-old children, in order to better understand the developmental trajectory of sound source segregation.

II. STUDY 1

This study utilized two sets of stimuli to measure SRTs, masking, and SRM in a group of 4- to 5-year-old children. Stimuli were the spondees from the CRISP test (Litovsky, 2005) and a new corpus of words designed for measuring speech intelligibility in children as young as 2.5 years of age, otherwise known as CRISP-Jr. The main purpose of this study was to determine whether perceptual measures are comparable and can be generalized across the two different tests and sets of stimuli. The differences were in the auditory stimuli with the matching pictures, as well as the animations and feedback that are used to entertain and engage the children during testing. Otherwise, algorithms for varying sound level, threshold estimation, and stimulus delivery are identical in the two tests. SRTs were measured in a group of 4- to 5-year-old children. For each test, three conditions were used that enabled estimation of masking and SRM.

A. Method

1. Participants

Ten subjects, 4 females and 6 males, all native speakers of English, participated. The age range was 4 years, 2 months to 5 years, 6 months (mean age of 4 years, 11 months). All subjects had normal hearing sensitivity as indicated by pure-tone, air conduction thresholds of 20 dB or better for frequencies ranging from 500 to 8000 Hz, and normal tympanograms. None of the children had ear infection or known illness, nor were any of the children taking medication on the day of testing (as reported by the parent/guardian). All subjects were also reported to be healthy and free of neurological disorders, and were right-handed. Every one of the recruited participants was able to complete the required measurements, with no exceptions, drop-outs, or replacements for other reasons.

2. Stimuli

The target stimuli in the CRISP test consist of a closed set of 25 words from the Children's Spondee list (CID W-1 test) recorded with a male voice. Targets in the CRISP-Jr. test consist of a closed set of 16 words, recorded with a different male voice. The target words in CRISP-Jr. were chosen to be within the receptive language and vocabulary of average 2.5- to 3.0-year-old children, and therefore consist of names of objects and/or body parts. The word list was inspired by the well-known Mr. Potato Head® game, in which the player can attach or remove body parts onto a toy shaped like a potato. This game has been used in some clinical and research settings with live voice to evaluate language acquisition of very young children (e.g., Robbins, 1994; Svirsky *et al.*, 2004). Here we used a computerized "listening game," based on a similar platform to that described by Litovsky (2003, 2005), with the presumption that young children who have normal hearing and cognitive abilities can easily identify body parts and related items. The test, however, was developed for the purpose of measuring SRTs, and the words being tested are *known* to each child prior to testing (see Sec. II A 4). Appendices A and B include lists of the words used in the two tests, respectively. It is important to note that the

CRISP-Jr. list included 12 monosyllabic and 4 bi-syllabic words. Although this may not be ideal, we felt that it was important to increase the size of the target corpus. During individual trials, the options presented to the child on the computer screen in the 4-AFC procedure (see the following) were set up such that all the words had the same number of syllables. Although systematic testing was not conducted to determine whether some of the words were more difficult than others to identify, the procedures ensured that subjects were well familiarized with all targets prior to testing.

The competitor stimuli in both tests were sentences from the Harvard IEEE corpus (Rothauser *et al.*, 1969), recorded with a female voice. Examples of sentences are: "Glue the sheet to the dark blue background," "Two blue fish swam in the tank" and "The meal was cooked before the bell rang." Thirty such sentences were strung together, and segments were randomly chosen and played for a duration of 6 s during each trial. The timing was such that the target words occurred approximately 1.5 s after the onset of the competing sentence. All stimuli for this study were pre-recorded and digitized using a sampling rate of 44.1 kHz and stored on a laptop computer. The levels for all stimuli were rms equalized using MATLAB software.

3. Test setup and procedure

Testing was conducted in a carpeted double-walled sound booth (2.75 m × 3.25 m) with reverberation time (T_{60}) = 250 ms. Subjects were seated in the middle of the room with speakers mounted on a stand at a distance of 1.2 m from the center of the subject's head. The target and the competitor stimuli were fed to separate audio channels of a laptop computer, amplified (Crown D-75) and played through separate loudspeakers (Cambridge Soundworks, Center/Surround IV). The target stimuli were always presented from the front (0° azimuth), while the competitor was presented from either the front (0° azimuth) or the right (90° azimuth). Prior to each testing session, stimuli were calibrated (Larsen-Davis System 824). A computer monitor was placed directly below the speaker at 0° azimuth (to avoid obstruction with wave propagation in the room). The child sat on a chair in front of a small table (covered with foam), on which the computer mouse and keyboard were placed. During testing the child was engaged in a computerized "listening game" and all visual stimuli (e.g., picture representations of the targets, puzzles, and animations) were presented from the monitor, which was particularly helpful in maintaining the head centered and minimizing head movements.

For both tests, the task involved 4-AFC. On each trial, following the target presentation, four randomly selected pictures, including the target, appeared simultaneously on the screen. The pictures were arranged in a 2 × 2 grid of equal-size squares, and the square containing the target picture was randomly chosen from trial to trial. A defined set of rules was applied in order to eliminate the possibility of words with similar initial sounds occurring in the same interval (e.g., "hands" and "hat"). On each trial a leading phrase such as "point to the…" or "show me…" preceded the target presentation. Subjects were instructed to listen to the target word (male voice) on each trial, and to select the one picture on the computer monitor that matched the heard word. A verbal response was also required to ensure subjects' correct identification of the targets, and the answer was entered into the computer by an examiner. In the unlikely event that a child pointed to one picture but verbally reported another, the trial was repeated (with a new stimulus). Feedback was provided for both correct and incorrect responses. Following each correct response, a brief musical clip was presented or a missing piece from a puzzle was added to the computer display. Following incorrect responses a phrase such as "that must have been difficult" or "let's try a different one" was presented from the front speaker.

Using the aforementioned setup, measurements were obtained from each child on the two tests. For each test the following three conditions were included: (1) *quiet* with target at 0° azimuth and no competitor, (2) *front* with target and competitor both at 0° azimuth, and (3) *right* with target at 0° azimuth and competitor at 90° azimuth. For each subject, the order of the test (CRISP or CRIPS-Jr.) was randomized using a Latin-square design, and the order of the three conditions (quiet, front, and right) within each test was then also randomized. The study was completed in one session that lasted approximately 2 h per subject, including frequent breaks.

4. Familiarization

Prior to each test, children underwent a familiarization procedure that lasted 5–10 min. First, the target words were each presented along with their associated pictures, and then the child was tested on his/her ability to correctly identify the targets. Words that were not easily identified were eliminated from that subject's target corpus. In a typical case, a child was already familiar with all targets, or was able to quickly associate the auditory stimulus with the matching picture. In three cases one to two targets were discarded from the CRISP-Jr. corpus and in eight cases one to four targets were discarded from the CRISP corpus due to incorrect identification after the familiarization step.

5. Stimulus levels and threshold estimation

The level of the competitor was fixed at 60 dB SPL. Both tests incorporated an adaptive tracking procedure to vary the level of the target signal. The target level was set to 60 dB SPL (0 dB SNR) at the beginning of each adaptive track. The algorithm for varying the target level included a set of rules that are outlined in greater detail elsewhere (Litovsky, 2005). Briefly, during the initial portion of the adaptive track, the target level was decreased by 8 dB following each correct response. After the first incorrect response, a modified adaptive three-down/one-up algorithm was used, with the following rules: Following each reversal, the step size is halved, with the minimum step size set to 2 dB. If the same step size is used twice in a row in the same direction, the next step size is doubled in value. Testing is terminated following four reversals.

Thresholds were estimated using a constrained maximum-likelihood method of parameter estimation (MLE) which has been described by Wichmann and Hill (2001a, b). On average, the MLE approach has been shown to yield

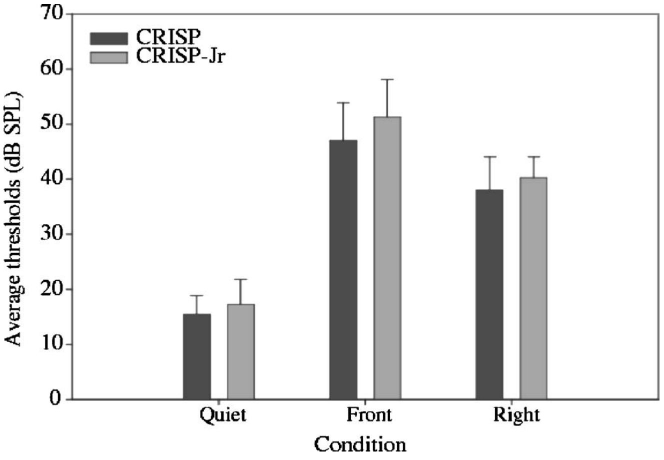

FIG. 1. Speech reception thresholds (SRTs in dB SPL) +s.d. are plotted for the quiet, front, and right conditions for ten subjects aged 4 to 5 years. Dark bars represent data collected with CRISP and light bars show data collected with CRISP-Jr.

comparable results to the traditional approach which estimates thresholds based on the average of certain number of reversal points; however, the MLE method has the advantage of producing smaller group variances (Litovsky, 2005). This approach is particularly suitable in pediatric research given that a smaller number of reversals can be obtained. In addition, because lapses of attention can lead to incorrect responses when the child would otherwise provide a correct response, values obtained on single trials can have disproportionate weighting in reversal-based threshold estimation. By using MLE, these issues can be minimized given that thresholds are estimated via a method that places greater weight on levels with the largest number of trials visited at a given adaptive track.

In this study, all the data from each experimental run for each participant were fit to a logistic function. Thresholds were calculated by taking the level of the speech signal at a specific probability level. Given that an adaptive, three-down/one-up procedure was employed, thresholds were estimated targeting performance level of 79.4% on the psychometric function (as estimated by Levitt, 1971).

B. Results

Group average SRTs are compared for the three conditions and two tests in Fig. 1. The results were subjected to a two-way repeated measures analysis of variance (ANOVA) with test (CRISP and CRISP-Jr.) and condition (quiet, front, and right) as within-subject variables. A main effect was found for test $[F(1,9)=8.872;\ p<0.05]$, suggesting that CRISP-Jr. yielded higher SRTs than CRISP. In addition, there was a main effect for condition $[F(2,9)=109.463,\ p<0.0001]$. Scheffe's post hoc comparisons showed that SRTs were lower for *quiet* than *right* $[F(2,9)=100.23,\ p<0.0001]$ and *front* $[F(2,9)=208.54,\ p<0.0001]$. This indicates that masking occurred for both competitor locations. SRTs were also significantly higher for *front* than *right* $[F(2,9)=19.62,\ p<0.001]$, which suggests that SRM occurred and that these children benefited from the spatial separation of the target and competitor. There was no significant interaction between test and condition, suggesting that the effects described thus far occurred similarly for CRISP and CRISP-Jr.

The amount of masking for the two conditions is defined as the difference in SRTs between *front* or *right*, and *quiet*, respectively (e.g., $\mathrm{SRT}_{\mathrm{front}}-\mathrm{SRT}_{\mathrm{quiet}}$ and $\mathrm{SRT}_{\mathrm{right}}-\mathrm{SRT}_{\mathrm{quiet}}$). Figure 2 shows both individual and group data for the amount of masking on each test in the *front* and *right* conditions. A two-way repeated-measure ANOVA (test × competitor location) revealed no statistically significant difference between the two tests. However, masking was significantly greater in *front* than *right* $[F(1,9)=51.735;\ p<0.0005]$. There was no significant interaction effect, which suggests that the amount of SRM (front masking – right masking) was similar for the CRISP and CRISP-Jr. tests, averaging 9.02 and 11.09 dB, respectively. An interesting finding, discussed later, is that subjects are not always consistent in their performance across the two tests; while some subjects exhibited higher thresholds with CRISP, others showed the opposite. This is highlighted for select cases by connecting the masked threshold lines in the two tests in Fig. 2.

In summary, results thus far suggest that the CRISP-Jr.

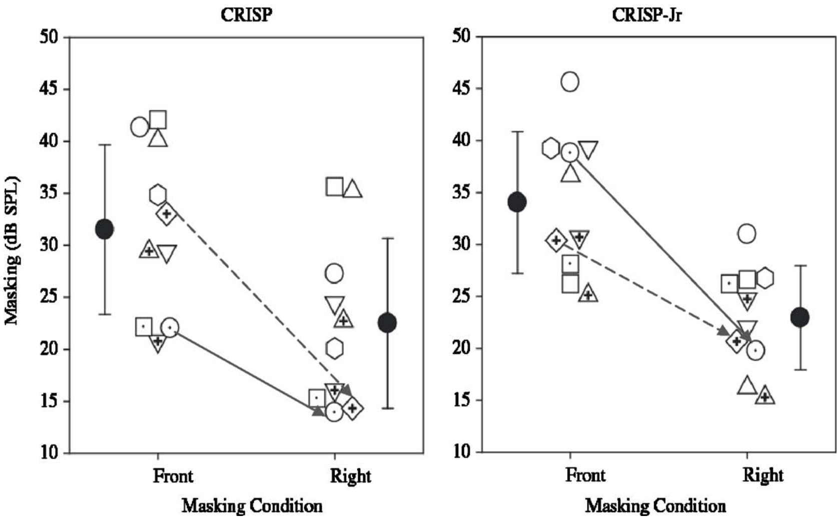

FIG. 2. Amounts of masking for front and right conditions are displayed for both CRISP (left panel) and CRISP-Jr. (right panel). Each subject's data are represented by a different symbol; closed circles represent group data (±s.d.) for each condition. The connecting lines are used to demonstrate examples of individual subjects' masking in the front vs right conditions. Comparisons across the two panels shows these subjects' inconsistent performance across the two tests; while some children exhibited higher masked thresholds with CRISP, for others CRISP-Jr. elicited higher thresholds.

test produces similar results to those obtained with CRISP, in particular for measures of masking and SRM. This testing paradigm may, therefore, provide a useful tool for investigating these same effects in younger children.

III. STUDY 2

This study was aimed at extending the CRISP-Jr. testing tool to a group of children under the age of four. Results from this young age group were compared to those obtained from the older group on the same test (CRISP-Jr.).

A. Method

1. Participants

Ten children (3 females and 7 males) participated in this study. All children were native speakers of English and were in the age range of 3 years, 4 months to 3 years, 10 months (mean age of 3 years, 7 months). Thresholds measurements were obtained separately in each ear using a standard audiometric clinical approach, whereby the child was asked to lift a finger or say "I hear it" every time that a tone was heard. A one-down one-up rule was used to adaptively vary the tone level. None of the children had ear infection or known illness, or were taking medication on the day of testing as reported by parents. Nine children were reported to be right handed and one was left handed. In order to obtain 10 participants 11 children were recruited (one child was unable to complete the necessary measurements due to lack of cooperation or lack of interest in the task). Otherwise, there was no need for averaging, exclusion, or filling in of missing data points.

B. Stimuli and procedure

Stimuli and testing apparatus were identical to those described in detail in experiment 1. Each child was tested on three conditions: *quiet*, *front*, and *right*, with the order of conditions randomized. The study was completed in one session that lasted approximately 1 h, including breaks.

1. Results

CRISP-Jr. SRTs were calculated using the same MLE procedure described in study 1. Average SRTs for the two age groups are shown in Fig. 3. Results were subjected to a two-way ANOVA, with age as the between-subjects variable and condition as the within-subjects variable. A significant main effect for age was found $[F(1,18)=24.582;\ p<0.0005]$ such that across conditions, SRTs were an average of 10 dB higher in the 3-year-old age group compared with the 4 to 5 year olds. This finding suggests that children's ability to identify speech in a forced-choice paradigm improves with a small increase in age during the preschool years. A significant main effect for condition was also found $[F(2,18)=114.43,\ p<0.0005]$, with Scheffe's post hoc contrasts showing that SRTs in the *quiet* condition were lower than *front* $[F(2,18)=216.55,\ p<0.0001]$ as well as *right* $[F(2,18)=108.09,\ p<0.0001]$. As was found in study 1, this result suggests again the occurrence of masking regardless of the location of the competing sound. Finally, SRTs were significantly higher in *front* compared with *right* $[F(2,18)=18.65,\ p<0.0001]$. No significant interactions were found.

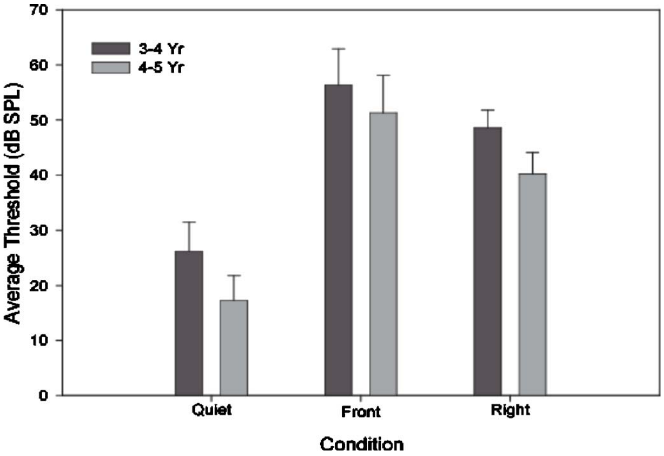

FIG. 3. Speech reception thresholds (SRTs in dB SPL) +s.d. elicited with CRISP-Jr. are plotted for the quiet, front, and right conditions for children grouped into two age groups, 3 and 4 to 5 year olds. Dark bars represent data collected from the younger group and light bars show data collected from the older group.

Amount of masking was computed by subtracting SRTs in the *front* and *right* from *quiet*. A two-way ANOVA (age × competitor location) showed no statistically significant difference between the two age groups in the amount of masking. A main effect of competitor location was found $[F(1,18)=35.035;\ p<0.0005]$, with masking from the front being higher than the right. This finding, together with the result that SRTs were higher in the front than right, indicates the occurrence of SRM. The lack of interaction suggests that SRM was similar in both groups, or that the variability was too large to reveal age-related effects. Average SRM values of the 3 and 4 to 5 year olds were 7.7 (±7.2 s.d.) dB, and 11 (±7.1 s.d.) dB, respectively. These are plotted in Fig. 4 along with the individual subjects' results, to demonstrate the large intersubject variability within each group.

A regression analysis was conducted between the amount of SRM and SRTs obtained when the competitor was in one of the two masking conditions (*front* or *right*). Results plotted in Fig. 5 (left panel) support the statistical finding of

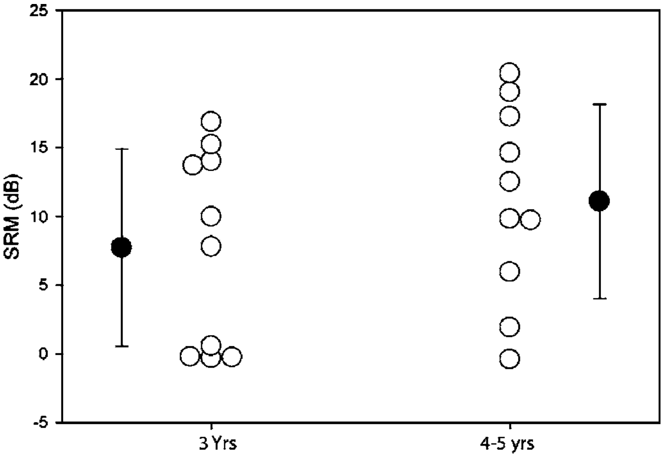

FIG. 4. Individual SRM values (open circles) and mean data (closed circles) are plotted for 20 children distributed into two age groups 3 years and 4 to 5 years old tested on CRISP-Jr.

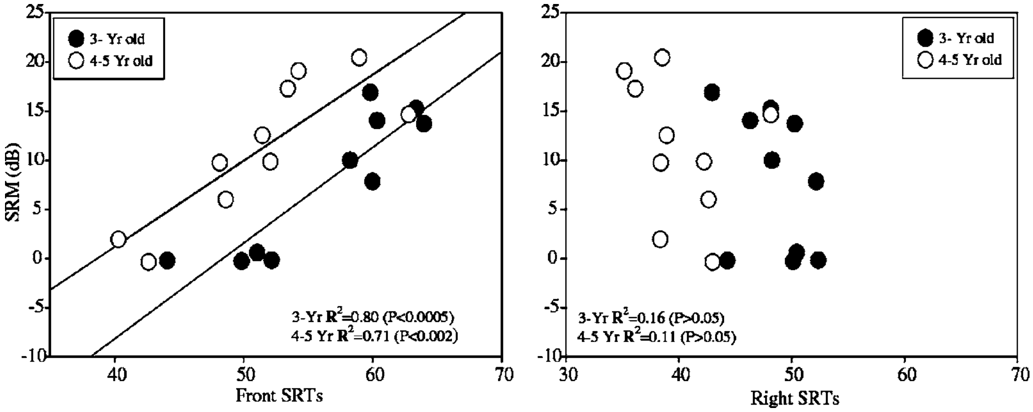

FIG. 5. Left panel displays SRM (in dB SPL) plotted against SRTs in the front condition. The right panel displays SRM against SRTs in the right condition. Data are shown for 20 subjects ranging in age between 3 and 5 years. The results of the regression analysis are shown at the bottom of each panel.

a significant regression with the *front* for both the 3-year-old ($p < 0.0005$) and the 4- to 5-year-old ($p < 0.002$) groups. Furthermore, two distinct regression lines can be noted due to the higher SRTs exhibited by the younger group of children. A similar analysis (Fig. 5, right panel) when the competitor was on the *right* was nonsignificant ($p > 0.05$) for both age groups. This suggests that spatial separation between the target and masking source is particularly beneficial for children whose SRTs are high when both target and competitor are in front (i.e., in absence of spatial cues). In contrast, children whose *front* SRTs are initially low do not gain much extra benefit when spatial cues are introduced in the *right* condition.

IV. DISCUSSION

When listeners are faced with a complex array of sounds in a room, one of the ways in which they can separate important speech signals from noise is if the two sources arrive from different locations. In adults, spatial separation of speech from noise provides 3–12 dB improvement in SRT, depending on the type of stimulus, task, and whether stimuli are presented under binaural or monaural listening modes (e.g., Bronkhorst, 2000; Culling *et al.*, 2004; Hawley *et al.*, 2004). To the extent that spatial hearing mechanisms are well-developed in young children, similar benefits might also be observed. This is an important issue to determine, given the vast number of hours that children spend in noisy multisource situations. In addition, there may be insights into potential benefits of amplification fitting strategies in hearing impaired children who receive bilateral hearing aids and/or cochlear implants (e.g., Litovsky *et al.*, 2004, 2006). For these reasons, the present study was aimed at extending the testing age to a younger group to evaluate the developmental trajectory of sound source segregation.

A. Speech reception thresholds and masking

SRTs measured here suggest that 3 year olds perform worse than 4 to 5 year olds. Our other work (Litovsky, 2005; Johnstone and Litovsky, 2006) suggest that SRTs continue to improve beyond age 7 compared with adults. However, on these tasks the amount of masking appears to be adult-like at the youngest ages. The overall increase in thresholds is consistent with other findings on age-related changes in thresholds (e.g., Lenihan *et al.*, 1971; Fior, 1972; Roche *et al.*, 1978; Schneider *et al.*, 1986; Werner, 1996). Generally, it is believed that children's poor performance on certain auditory tasks is reflective of immaturity at central levels in the auditory pathway, as neurodevelopmental research suggests that central maturation continues throughout childhood (Ponton *et al.*, 2000; Moore and Guan, 2001; Rakic and Yakovlev, 1968). The neuroanatomical and physiological measures do not however identify which specific mechanisms are immature, and leave open the question as to what might account for the age differences in SRTs seen here.

Performance is most likely influenced by a complex interaction between various processes, including those borne in the auditory system, as well as attention, working memory, and other nonsensory factors. To the extent that differences in thresholds across age groups may be attributable to maturational changes in auditory processes, it is noteworthy that, despite the increased mean for the younger children, variance was generally similar in the two age groups (Fig. 3).

It is unlikely, however, that the results shown here are affected by age-related differences in language acquisition. Although language measures were not obtained, this factor and its effects were minimized if not eliminated by employing the familiarization procedure which ensured that all target words were known by each participant prior to testing.

The two groups showed similar amount of elevation in thresholds in the presence of competing speech. Although this study did not evaluate adults' performance on CRISP-Jr., differences in performance between adults and children is assumed because two previous studies that applied the CRISP testing paradigm have shown adults' performance to be significantly better than the children on the masked conditions, even when task difficulty is equated (Litovsky *et al.*, 2005; Johnstone and Litovsky, 2006). Other paradigms have also demonstrated that children are more susceptible to noise than adults, and that they require more favorable SNR than adults for comparable performance (e.g. Elliott *et al.*, 1979; and Nittrouer and Boothroyd, 1990). Children appear to adopt nonoptimal listening strategies when distracting noise is present (Allen and Wightman, 1994; Lutfi *et al.*, 2003) and they are very likely to have poor processing efficiency caused by higher internal noise than adults (Hill *et al.*, 2004).

The mechanism for suppressing this central noise has been suggested to improve with age as demonstrated in improved ability to segregate sounds with similar acoustical characteristics (Allen and Nelles, 1996).

B. Spatial release from masking

The results from the present study showed that by age three children can use spatial cues to segregate different auditory streams, and the amount of SRM for that age group was not statistically different from that of the older group. Previous studies (Litovsky, 2005; Johnstone and Litovsky, 2006), in which SRM was measured with CRISP, showed that children ages 4.5–7 did not differ significantly from adults. By extension, studies on this subject imply that by 3 years of age some of the auditory mechanisms that mediate SRM are well developed.

These current findings appear to be at variance with previous reports that showed a protracted developmental period when the task involves segregation of multiple auditory signals (e.g. Oh *et al.*, 2001; Hall *et al.*, 2005; Wightman *et al.*, 2003; Wightman and Kistler, 2005). In an attempt to explain the difference between the present findings and previous reports several possibilities should be considered. First, the differences may be due to the testing paradigm. Previous studies investigated speech unmasking in children under dichotic conditions in which *spatial cues* are limited to a single cue, inter-ear separation. By comparison, SRM as measured here was most likely influenced by a multitude of spatial cues that include both inter-aural binaural cues, as well as monaural cues such as level and spectra. The results from these two different paradigms are therefore not directly comparable, nor are they inconsistent with one another. A fuller exploration of the cues that mediate source segregation would thus help resolve some of these issues. Second, discrepancies in the results across studies may also be attributable to differences in the testing tools utilized across studies. The current study was aimed at using a very engaging testing tool that was especially developed for preschoolers. The tool was very successful in attracting the preschoolers and maintaining their attention throughout the testing. Other published work has also been successful at demonstrating that with the use of age-appropriate tests children can selectively attend to specific auditory information and suppress competing information, an ability that seems to emerge in early life and it is somewhat adult-like as early as 3 years of age (Sanders *et al.*, 2006).

Findings from this study also showed that individual subjects vary in amount of SRM. For instance, while some children showed SRM values that are greater than 15 dB, others had values near 0 dB, and a small group of children showed negative SRM (better SRTs in the front condition). This variability is consistent with previous reports on children's performance on auditory tasks (Litovsky, 1997; Buss *et al.*, 1999; Oh *et al.*, 2001; Litovsky, 2005). However, given that each SRM data point obtained in the current study represents a single adaptive track, the extent of the contribution of intrasubject variability to these results cannot be addressed or ruled out. Future studies should address this important issue.

For both groups of children there appeared to be a robust relationship between individuals' performance in *front* (when the target and the competitor are spatially near) and the amount of SRM. That is, SRM was generally greater when performance in front was worse. It is possible that young children adopt different listening strategies under nonoptimal listening situations. While some children are likely to use both spatial and perceptual cues such as differences in speakers' fundamental frequencies (F0), others are unable to exploit such nonspatial cues and for those children spatial cues are more potent. It is important to note that for those children who are good users of nonspatial cues, it is less likely to see robust SRM since the effect of spatial separation is minimized by the low (better) *front* SRTs. However, the extent to which children could differ in taking advantage of these perceptual cues or when this ability emerges needs to be investigated by future studies.

C. Effect of test

In the 4- to 5-year-old group, SRTs were somewhat higher when using CRISP-Jr. compared with CRISP. This result could be attributed to differences in the speech material; CRISP-Jr. contains primarily single-syllabic words while the stimuli in CRISP are bi-syllabic (spondees). Although the monosyllabic words are simpler, and therefore suitable for young children, they may require higher signal level to be fully identifiable. This is consistent with reports that monosyllabic words are the least understandable on the continuum of meaningful words (e.g., Hirsh *et al.*, 1954), and that speech intelligibility (i.e., SRTs measured in the present study) is expected to improve as the number of syllables per word increases. This does not negate the utility of the CRISP-Jr. test as an appropriate tool for testing very young children, because of the limited vocabulary of most young children. This issue was successfully overcome here. In fact, using single-syllabic words has the advantage of minimizing the redundancy of the speech signal, which eventually makes the test more sensitive for certain measures such as binaural integration (Smith and Resnick, 1972).

The SRT issue is also less fundamental when one considers the fact that the amounts of masking and SRM were similar for the CRISP-Jr. and the CRISP tests. This finding suggests that masking and SRM in young children are both robust phenomena that do not depend on the exact stimulus corpus or stimulus presentation platform. Similarly, the individual variability in amount of SRM was not restricted to one test. Most interesting perhaps is the finding that, while some children were better able to use spatial cues in the CRISP-Jr. test, for other children the CRISP test was more effective in eliciting SRM (see Fig. 2). It is possible that this type of inconsistent performance across the two tests emerged out of perceptual elements triggered by the difficulty of particular stimuli for some children but not for others. This may have elements akin to informational masking, whereby SRM increases when the challenge to the auditory system is greater. One aspect of informational masking is the stimulus uncertainty that occurs when the background masker is unknown or unpredictable (e.g., Durlach *et al.*, 2003; Brungart and Simpson, 2004; Hawley *et al.*, 2004). Each of the tasks used here placed the children in a somewhat different complex auditory environment, and it is possible that each child's approach to the problem varied in ways that created more uncertainty for some children in one task, and for other children on the alternate task. These differences may have led to increased informational masking in situations that caused greater uncertainty. Resolving this potential issue is beyond the scope of the present study, but is an issue that can be better addressed with more extensive testing in future work.

V. SUMMARY AND CONCLUSIONS

This study was aimed at evaluating the developmental ontogeny of sound source segregation in young children, ages 3 through 5 years, using child-friendly testing tools (CRISP and CRISP-Jr.). Results showed that there are age-related differences in SRTs, but not in amount of masking or SRM. In addition, the two tests yielded similar amount of SRM, which indicates that SRM is a robust phenomenon that does not depend on the exact stimulus corpus or stimulus presentation platform. These findings further suggest that children as young as 3 years old are able to use spatial cues to segregate different auditory streams, implying that some of the mechanisms that mediate spatial source segregation develop by early childhood. However, in complex auditory environments, individual children seem to adopt different listening strategies; while some children use both spatial and nonspatial cues, others seem to rely more heavily on spatial cues, which might be primarily monaural, although that remains to be better understood. These results underscore the importance of spatial separation between the target and the competing noise in educational settings such as classrooms. Furthermore, because CRISP and CRISP-Jr., can be used to evaluate benefits of spatial cues in source segregation, these tests may prove to be useful for evaluating hearing aids and cochlear implant fitting for usage in noisy environments. Finally, the CRISP-Jr. test could further be used for evaluating speech intelligibility in noise in children whose receptive language is delayed compared with age-matched peers.

ACKNOWLEDGMENTS

The authors are grateful to Diana Brown and Shelly Godar for helping in subjects' recruitment, and to Patti Johnstone and Gary Jones for helping with data analysis. We would also like to thank all of the parents and children who participated in the listening games, without whom this work would not be possible. This work was supported by NIH-NIDCD (Grant Nos. R01-DC003083 and R21-DC05469 to R.Y.L.).

APPENDIX A: LIST OF 25 SPONDEE WORDS USED IN CRISP

Hotdog
Ice cream
Birdnest
Cowboy
Dollhouse
Barnyard
Scarecrow
Railroad
Sidewalk
Rainbow
Cupcake
Birthday
Airplane
Eyebrow
Shoelace
Toothbrush
Hairbrush
Highchair
Necktie
Playground
Football
Baseball
Bluejay
Bathtub
Bedroom

APPENDIX B: LIST OF TARGET WORDS USED IN CRISP-JUNIOR

Mouth
Nose
Eyes
Cheeks
Hat
Feet
Ears
Hands
Socks
Shoes
Hair
Ball
Eyebrows
Toy car
Glasses
Balloons

Allen, P., and Nelles, J. (**1996**). "Development of auditory information integration abilities," J. Acoust. Soc. Am. **100**, 1043–1051.
Allen, P., and Wightman, F. (**1994**). "Psychometric functions for children's detection of tones in noise," J. Speech Hear. Res. **37**, 205–215.
Allen, P., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.
Allen, P., Wightman, F. L., Kistler, D. J., and Dolan, T. R. (**1989**). "Frequency resolution in children," J. Speech Hear. Res., **32**, 317–322.
Ball, M. J. (**1999**). "Reynell developmental language scales III: A quick and easy LARSP," Int. J. Lang Commun. Disord. **34**, 171–174.
Bronkhorst, A. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. 86, 117–128.
Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.
Brungart, D. S., and Simpson, B. D. (**2004**). "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty," J. Acoust. Soc. Am. **115**, 301–310.
Buss, E., Hall, J. W., III, Grose, J. H., and Dev, M. B. (**1999**). "Development of adult-like performance in backward, simultaneous, and forward masking," J. Speech Lang. Hear. Res. **42**, 844–849.
Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with

one and two ears," J. Acoust. Soc. Am. **25**, 975–979.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Dirks, D. D., and Wilson, R. H. (**1969**). "The effect of spatially separated sound sources on speech intelligibility," J. Speech Hear. Res. **12**, 5–38.

Drullman, R., and Bronkhorst, A. (**2000**). "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," J. Acoust. Soc. Am. **107**, 2224–2235.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G. Jr. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Elliott, L. L., Conners, S., Kille, E., Levin, S., Ball, K., and Katz, D. (**1979**). "Children's understanding of monosyllabic nouns in quiet and in noise," J. Acoust. Soc. Am. **66**, 12–21.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Fior, R. (**1972**). "Physiological maturation of auditory function between 3 and 13 years of age," Audiology **11**, 317–321.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Hall, J. W., III, Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-talker masker and a speech shaped noise masker in adults and children," Ear Hear. **23**, 159–165.

Hall, J. W., III, and Grose, J. H. (**1990**). "The masking level difference in children," J. Am. Acad. Audiol **1**, 81–88.

Hall, J. W., III, and Grose, J. H. (**1991**). "Notched noise measures of frequency selectivity in adults and children using fixed masker level and fixed signal level presentation," J. Speech Hear. Res. **34**, 651–660.

Hall, J. W., III, and Grose, J. H. (**1994**). "Development of temporal resolution in children as measured by the temporal resolution transfer function," J. Acoust. Soc. Am. **96**, 150–154.

Hall, J. W., III, Buss, E., and Grose, J. H. (**2005**). "Informational masking release in children and adults," J. Acoust. Soc. Am. **118**, 1605–1613.

Hartley, D. E., Wright, B. A., Hogan, S. C., and Moore, D. R. (**2000**). "Age-related improvements in auditory backward and simultaneous masking in 6- to 10-year-old children," J. Speech Lang. Hear. Res. **43**, 1402–1415.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefits of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Hill, P. R., Hartley, D. E., Glasberg, B. R., Moore, B. C., and Moore, D. R. (**2004**). "Auditory processing efficiency and temporal resolution in children and adults," J. Speech Lang. Hear. Res. **47**, 1022–1029.

Hirsh, I. J. (**1950**). "The relation between localization and intelligibility," J. Acoust. Soc. Am. **22**, 196–200.

Hirsh, I. J., Reynolds, E. G., and Joseph, J. (**1954**). "Intelligibility of different speech material," J. Acoust. Soc. Am. **26**, 530–538.

Johnstone, P., and Litovsky, R. (**2006**). "Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults," J. Acoust. Soc. Am. **120**, 2177–2189.

Kidd, G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.

Lenihan, J. M. A., Christie, J. F., Russell, T. S., Orr, N. M., Hamilton, M. D., and Konx, E. C. (**1971**). "The threshold of hearing in school children," J. Laryngol. Otol. **85**, 375–385.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lin, W. Y., and Feng, A. S. (**2003**). "GABA is involved in spatial unmasking in the frog auditory midbrain," J. Neurosci. **23**, 8143–8151.

Litovsky, R. Y. (**1997**). "Developmental changes in the precedence effect: Estimate of minimum audible angel," J. Acoust. Soc. Am. **102**, 1739–1745.

Litovsky, R. Y. (**2003**). "Method and system for rapid and reliable testing of speech intelligibility in children," U.S. Patent No. 6,584,440.

Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., and Yu, G. (**2004**). "Bilateral cochlear implants in adults and children," Arch. Otolaryngol. Head Neck Surg. **130**, 648–655.

Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.

Litovsky, R. Y., Johnstone, P., and Godar, S. (**2006**). "Benefits of bilateral cochlear implants and/or hearing aids in children," Int. J. Audiol. **45**, 78–91.

Lutfi, R. A., Kistler, D. J., Oh, E. L., and Callahan, M. R. (**2003**). "One factor underlies individual differences in auditory informational masking within and across age groups," Prog. Aerosp. Sci. **65**, 396–406.

Moore, J. K., and Guan, Y. L. (**2001**). "Cytoarchitectural and axonal maturation in human auditory cortex," J. Assoc. Res. Otolaryngol. **2**, 297–311.

Nittrouer, S., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705–2715.

Noble, W., Byrne, D., and Ter-Horst, K. (**1997**). "Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing-impaired listeners," J. Acoust. Soc. Am. **102**, 2343–2352.

Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.

Papso, C. F., and Blood, I. M. (**1989**). "Word recognition skills of children and adults in background noise," Ear Hear. **10**, 235–236.

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.

Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acustica **34**, 200–211.

Pollack, I., and Pickett, J. M. (**1958**). "Stereophonic listening and speech intelligibility against voice babble," J. Acoust. Soc. Am. **30**, 131–133.

Ponton, C. W., Eggermont, J. J., Kwong, B., and Don, M. (**2000**). "Maturation of human central auditory system activity: Evidence from multichannel evoked potentials," Clin. Neurophysiol. **111**, 220–236.

Rakic, P., and Yakovlev, P. I. (**1968**). "Development of the corpus callosum and cavum septi in man," J. Comp. Neurol. **132**, 45–72.

Robbins, A. M. (**1994**). "The Mr. Potato head task," Indiana University School of Medicine, Indianapolis.

Roche, A. F., Siervogel, R. M., Himes, J. H., and Johnson, D. L. (**1978**). "Longitudinal study of hearing in children: Baseline data concerning auditory thresholds, noise exposure, and biological factors," J. Acoust. Soc. Am. **64**, 1593–1601.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbigert, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Sanders, L. D., Stevens, C., Coch, D., and Neville, H. J. (**2006**). "Selective auditory attention in 3- to 5-year-old children: An event-related potential study," Neuropsychologia **44**, 2126–2138.

Schneider, B. A., Trehub, S. E., Morrongiello, B. A., and Thorpe, L. A. (**1986**). "Auditory sensitivity in preschool children," J. Acoust. Soc. Am. **79**, 447–452.

Smith, B. B., and Resnick, D. M. (**1972**). "An auditory test for assessing brainstem integrity: Preliminary report," Laryngoscope **82**, 414–424.

Svirsky, M., Teoh, S., and Neuburger, H. (**2004**). "Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation," Audiol. Neuro-Otol. **9**, 224–233.

Werner, L. A. (**1996**). "The development of auditory behavior (or what the anatomists and physiologists have to explain)," Ear Hear. **17**, 438–446.

Wichmann, F. A., and Hill, J. (**2001a**). "The psychometric function. I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.

Wichmann, F. A., and Hill, J. (**2001b**). "The psychometric function. II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

Wightman, F., Allen, P., Dolan, T., Kistler, D., and Jamieson, D. (**1989**). "Temporal resolution in children," Child Dev. **60**, 611–624.

Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**, 3164–3176.

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.

Yost, W. A., Dye, R. H., and Sheft, S. (**1996**). "A simulated "cocktail party" with up to three sound sources," Percept. Psychophys. **58**, 1026–1036.