Speech intelligibility and spatial release from masking in young children[a]

Ruth Y. Litovsky[b]
Waisman Center, University of Wisconsin—Madison, 1500 Highland Avenue, Madison, Wisconsin 53705

(Received 13 August 2003; accepted for publication 27 January 2005)

Children between the ages of 4 and 7 and adults were tested in free field on speech intelligibility using a four-alternative forced choice paradigm with spondees. Target speech was presented from front (0°); speech or modulated speech-shaped-noise competitors were either in front or on the right (90°). Speech reception thresholds were measured adaptively using a three-down/one-up algorithm. The primary difference between children and adults was seen in elevated thresholds in children in quiet and in all masked conditions. For both age groups, masking was greater with the speech-noise versus speech competitor and with two versus one competitor(s). Masking was also greater when the competitors were located in *front* compared with the *right*. The amount of masking did not differ across the two age groups. Spatial release from masking was similar in the two age groups, except for in the one-speech condition, when it was greater in children than adults. These findings suggest that, similar to adults, young children are able to utilize spatial and/or head shadow cues to segregate sounds in noisy environments. The potential utility of the measures used here for studying hearing-impaired children is also discussed. *© 2005 Acoustical Society of America.* [DOI: 10.1121/1.1873913]

PACS numbers: 43.66.Pn, 43.66.Qp, 43.71.Ft [AK] Pages: 3091–3099

I. INTRODUCTION

Children spend numerous hours every day in complex auditory environments, such as classrooms, where multiple sounds that vary in content and direction typically co-occur. In addition to voices of adults and children, instructional aids, environmental sounds, and reverberation are standard aspects of acoustic environments in classrooms. Some work indicates that children learn best in relatively quiet environments, and often have difficulty hearing speech in the presence of distracting sounds (Crandell, 1993; Yacullo and Hawkins, 1987; Papso and Blood, 1989). Psychophysical studies in which stimuli were presented over headphones have shown that, compared with adults, preschool listeners exhibit poorer attentional selectivity on auditory tasks (e.g., Stellmack *et al.*, 1997; Oh *et al.*, 2001) and reduced unmasking for tone detection under dichotic conditions (Wightman *et al.*, 2003; Hall *et al.*, 2004).

Also under headphones, it has been found that in the presence of two-talker maskers speech reception thresholds are higher in children than adults, and for both age groups thresholds are higher in the presence of two-talker maskers than with speech-shaped noise maskers (Hall *et al.*, 2002). Headphone stimulus presentation is limited, however, because spatial cues that are known to be important for sound segregation in realistic environments are missing. Studies with adults have shown that the ability to segregate target speech from competing speech and/or noise is determined by a complex set of auditory computations that involve both monaural and binaural processes (Hawley *et al.*, 1999, 2004; Bronkhorst, 2000; Culling *et al.*, 2004). Spatial cues in particular play a key role in facilitating source segregation. Speech intelligibility improves by up to 12 dB when the target speech and competing sounds are spatially separated, resulting in ‘‘spatial release from masking’’ (Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1992; Nilsson *et al.*, 1994; Koehnke and Besing, 1996; Peissig and Kollmeier, 1997; Hawley *et al.*, 1999, 2004; Shinn-Cunningham *et al.*, 2001; Litovsky *et al.*, 2002).

The extent to which children demonstrate spatial release from masking for speech is poorly understood. Of particular interest in the present study is the effect of number of maskers, as well as their content, on the extent to which young children experience spatial release from masking. In adult listeners spatial release from masking is especially large for multiple (two or more) maskers that carry linguistic content or context (i.e., speech or reversed speech), and relatively small for a single, nonspeech masker such as speech-shaped noise [Hawley *et al.*, (2004); see also Bronkhorst (2000) for review]. The authors of those works have concluded that release from masking as provided by spatial cues is particularly effective when the auditory environment is complex. The concept of ‘‘informational masking’’ has been invoked to explain this phenomenon, whereby, in the presence of maskers that are harder to ignore, spatial cues become important for sound source segregation. In this case, maskers that are multiple in number and/or that carry information resembling that contained in the target result in greater spatial release from masking (e.g., Brungart 2001; Freyman *et al.*, 2001; Arbogast *et al.*, 2002; Durlach *et al.*, 2003).

Several studies have reported that speech masking in children depends on the masker type (Papso and Blood,

---

[a] Select portions of these data were presented at the 143rd Meeting of the Acoustical Society of America, Pittsburgh, PA, and at the 24th Meeting of the Association for Research in Otolaryngology, Tampa, FL.

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

TABLE I. List of conditions tested for children (nine subjects per condition).

Group	No. of competitors	Age range (years. months$\pm$SD)	Competitor type	Conditions
1	1	5.4$\pm$1.1	Speech	Quiet, 1 front, 1 right
2	1	5.6$\pm$1.2	Speech-noise	Quiet, 1 front, 1 right
3	2	5.8$\pm$1	Speech	Quiet, 2 front, 2 right
4	2	5.6$\pm$1	Speech-noise	Quiet, 2 front, 2 right

1989; Hall *et al.*, 2002, 2004). However, the effect of number and spatial cues, and the possible contribution of these stimulus parameters to spatial release from masking, remain poorly understood. Binaural abilities in children are adultlike on measures of binaural masking level differences (Nozza *et al.*, 1988; Moore *et al.*, 1991) and minimum audible angle (Litovsky, 1997). Since spatial cues are known to play a key role in speech understanding for adults, it is important to understand how young children comprehend speech in realistic, multi-source acoustic environments, and the conditions that enable them to benefit from spatial cues. The research paradigm used here may ultimately also be useful in evaluating performance of hearing-impaired children. Noisy environments are particularly problematic for children with a history of otitis media (e.g., Hall *et al.*, 2003; Moore *et al.*, 2003; Roberts *et al.*, 2004) and for hearing aid and cochlear implant users (e.g., Dawson *et al.*, 2004; Eisenberg *et al.*, 2004; Litovsky *et al.*, 2004). Because the important task of hearing speech in noise can be a daily struggle for many of these children, ultimately their performance on these measures can assist with diagnosis and fitting strategies.

In the present study the task involved a four-alternative forced-choice (4AFC) word discrimination paradigm. Subjects selected a picture that matched the speech target from an array of four pictures that appeared on a computer monitor. Other tests such as the HINT-C (Nilsson *et al.*, 1994) may be usable for measuring speech intelligibility in noise in children as young as 6 years, but are difficult to implement with younger children. The test protocol described here was specifically designed to enable the study of speech intelligibility in noise in children as young as 4 years old, an age at which many children begin to spend a significant number of hours in noisy environments such as preschool classrooms.

II. METHODS

A. Subjects

A total of 36 volunteer children were recruited from local public schools and the general community (14 males and 22 females), and all subjects completed testing on the three required conditions. Subjects ranged in age from 4.5 to 7.5 years (average and standard deviation=5.5$\pm$1 years; see also Table I).[1] All were native speakers of English with no known auditory dysfunction or other cognitive disorders. According to the parents' report, none of the children were on medication or had known illness or ear infections on the day of testing, and none of the children had a known history of hearing loss. Total testing time for each listener was approximately 45 min.

Nine paid adult volunteers, with normal hearing as verified by standard audiometric testing for frequencies between 250 and 8000 Hz, and English as their first language, were also tested. Since testing was much less time consuming with adults than with children, a within-subject design was used whereby each subject participated in all conditions that pertained to the four groups of children.

B. Testing chamber, materials apparatus

Testing was conducted in a single-walled sound booth (3.6×4 m) with carpeting. This room had a reverberation time (T_{60}) = 250 ms and ambient noise levels averaging 35 dB SPL. During testing, subjects were always seated in the center of the room, with loudspeakers (Radio Shack Minimus 7) placed at 15.24 cm above ear level for children (ear level for adults) and at a distance of 1.67 m from the center of the subject's head. All stimuli were prerecorded, digitized, and stored on a laptop computer (Winbook). In the one-competitor conditions, the target and competing sound were fed to separate channels of a two-channel soundcard (Digigram VX Pocket), amplified (Crown D-75), and presented to separate loudspeakers. When both target and competitor were presented from the front position, the speakers were placed next to one another, with their centers at $\pm 2°$, with their medial walls nearly touching. Each loudspeaker subtended 4° in the horizontal dimension, hence strictly speaking, speakers were separated by 4°. In the two-competitor condition, when both occurred from the front, they were presented from the same loudspeaker. Target stimulus selection, level controls, and output as well as response acquisition were achieved using Matlab. A picture book containing four target pictures per page was placed on a small table in front of the subject.

C. Stimuli

Stimuli consisted of *target words* and *competing sentences*. Targets comprised a closed set of 25 spondaic words from CID W-1 obtained from Auditech and spoken by a male talker. Although a larger set of words is available, the subset chosen for the present study consisted of words that were easily represented with a visual illustration and readily recognized as such during pilot testing of 20, 4 to 5 year-old children (a list of the target words used is shown in the Appendix). The root-mean-square levels were equalized for all target words using Matlab software. The competitors were either speech or modulated speech-shaped noise. Competing sentences were taken from the Harvard IEEE list (Rothauser *et al.*, 1969) and 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.’’ Ten such sentences were used, and these were presented in a random order during testing. Speech-noise was made based on the ten competitor sentences and also played in a random order during testing. These interferers were filtered to match the long-term spectrum of the speech competitors, calculated for each talker separately. The noise samples were scaled to the same root-mean-square value and cut to the same length as the matching speech competitor. The envelope was then extracted from the speech competitor and was used to modulate the noise tokens, giving the same coarse temporal structure as the speech. The envelope of running speech was extracted using a method similar to that described by Festen and Plomp (1990) in which a rectified version of the waveform is low-pass filtered. A first-order Butterworth low-pass filter was used with a 3-dB cutoff at 40 Hz.

D. Design

The target words were always presented from the front (0°). Competitors were presented from either front or side (90°). Four groups of children with nine subjects per group were tested (see Table I). The side condition was always with competitor(s) on the right. Each child subject was randomly assigned to a group that was tested on one combination of type (speech or speech-noise) and number (1 or 2) of competitor(s). The subject was then tested on three conditions: (1) *quiet*: no competitor(s), (2) *front*: target and competitor(s) in front, and (3) *right*: target in front and competitor(s) at 90° on the right; the order of conditions was randomized using a Latin-square design. For the adult group, testing was conducted in a single 2-h session, with the order of the nine conditions randomized for each listener.

For each condition one adaptive track was measured. When two competitors were presented they were of the same type, but different samples were used for the two sources; in the two-speech conditions the same female voice was presented, speaking two different sentences, and in the two-speech-noise conditions two different segments of the noise were presented.

E. Familiarization

The present study was not aimed at testing children’s vocabulary, but rather their speech intelligibility for known words. The 25 words were selected from the spondee list after pilot testing indicated that 20, 4 to 5 year-old children were either familiar with the words or could easily ascertain their meaning after one presentation. For each of the 25 words a commissioned artist-drawn picture was used to visually represent the meaning of the word. Prior to testing, subjects underwent a familiarization session (approximately 5 min in duration) in which they were presented with the picture-word combinations and tested to insure that they associated each of the pictures with their intended auditory target.

F. Speech reception threshold estimation

The test involved a single interval 4AFC discrimination procedure. On each trial, the child viewed a set of four pictures from the set of 25 picture-word matches. A word matching one of the pictures was randomly selected and presented from the front speaker. A leading phrase such as ‘‘Point to the picture of the...’’ or ‘‘Where is the...’’ preceded each target word. The child was asked to select the picture matching the heard word, and to guess if not sure or if the word was not audible. The randomization process ensured that for every subject, on average, all 25 words were selected an equal number of times. The experimenter entered the child’s response into the computer. Following correct responses, feedback was provided in the form of 3-s musical clips from popular children’s music. Approximately 20 clips were digitized and stored on the computer, and randomly selected on correct-feedback trials. Following incorrect responses, feedback was provided in the form of a brief phrase such as ‘‘Let’s try another one’’ or ‘‘That must have been difficult.’’ Five such phrases were digitized and stored on the computer, and randomly selected on incorrect-feedback trials.

An adaptive tracking method was used to vary the level of the target signal, such that correct responses result in level decrement and incorrect responses result in level increment. The algorithm includes the following rules: (1) Level is initially reduced in steps of 8 dB, until the first incorrect response. (2) Following the first incorrect response a three-down/one-up rule is used, whereby level is decremented following three consecutive correct responses and level is incremented following a single incorrect response. (3) Following each reversal the step size is halved. (4) The minimum step size is 2 dB. (5) A step size that has been used twice in a row in the same direction is doubled. For instance, if the level was decreased from 40 to 36 (step=4) and then again from 36 to 32 (step=4), continued decrease in level would result in the next level being 24 (step=8). (6) After three consecutive incorrect responses a ‘‘probe’’ trial is presented at the original level of 60 dB. If the probe results in a correct response the algorithm resumes at the last trial before the probe was presented. If more than three consecutive probes are required, testing is terminated and the subject’s data are not included in the final sample. (7) Testing is terminated following five reversals.

For each subject, speech-reception-thresholds (SRTs) were measured for each condition. At the start of each SRT measurement, the level of the target was initially 60 dB SPL. When competitors were present (non-*quiet* conditions), the level of each competitor was fixed at 60 dB SPL, such that the overall level of the competitors was increased by approximately 3 dB when two competitors were presented compared with the one-competitor conditions. Thus, the adaptive track began with a signal-to-noise ratio of 0 dB in the one-competitor cases and −3 dB in the two-competitor cases.

Results were analyzed using a constrained maximum-likelihood method of parameter estimation outlined by Wichmann and Hill (2001a, b). All the data from each experimental run for each participant were fit to a logistic function.

Thresholds were calculated by taking the inverse of the function at a specific probability level. In our 4AFC task, using an adaptive three-down/one-up procedure, the lower bound of the psychometric function was fixed at the level of chance performance, 0.25, and the threshold level corresponded to the point on the psychometric function where performance was approximately 79.4% correct. Biased estimates of threshold can occur. Bias can be introduced by the sampling scheme used and lapses in listener attention. Wichmann and Hill (2001a, b) demonstrated that bias associated with lapses was easily overcome by introducing a highly constrained parameter to control the upper bound of the psychometric function. This approach was used to assess our data. The upper bound of the psychometric function was constrained within a narrow range (0.06) as suggested by Wichmann and Hill (2001b). As the authors suggest, under some circumstances, bias introduced by the sampling scheme may be more problematic to avoid even when a hundred trials are obtained per level visited. The possibility of biased threshold estimates due to our sampling scheme was assessed by comparing the thresholds obtained using the constrained maximum-likelihood method with traditional threshold estimates based on the last three reversals in each experimental run. A repeated measured *t*-test on quiet thresholds for the 36 children tested revealed no statistically significant difference between the estimated threshold values obtained using the ML approach versus the traditional approach $[t(35)=1.37, p>0.05$, two tailed].

III. RESULTS

SRTs were statistically analyzed for the children groups using a mixed-design analysis of variance (ANOVA) with two between-subjects variables (number of competitors, competitor type) and one within-subjects variable (condition). Significant main effects of number $[F(1,32)=4.05; p<0.05]$ and condition $[F(2,32)=119.57, p<0.0001]$ were found, but there was no effect of type. Significant interactions were found for condition with number $[F(2,64)=66.50; p<0.03]$ and condition with type $[F(2,64)=162.01; p<0.001]$. Scheffe's *posthoc* contrasts (significance value $p<0.05$) showed that SRTs in *quiet* were significantly lower than SRTs in either *front* or *right*. Children tested with two competitors had significantly higher SRTs than those tested with one competitor for the *front* and *right* conditions (further comparisons between *front* and *right* are described below with regard to spatial release from masking). Finally, for reasons that are not clear, SRTs on the quiet conditions were lower in the two speech-noise groups than in the groups tested with the speech competitors. Adult data were analyzed with a one-way ANOVA for the nine conditions, which revealed a significant main effect $[F(8,8)=3.77; p<0.05]$. Scheffe's *posthoc* contrasts $[F(8,8); Fp<0.01]$ revealed that *quiet* SRTs were lower than SRTs on all other conditions. Child and adult SRTs were compared with independent *t*-tests for each of the nine conditions; since the *quiet* condition was tested for each of the child groups, a total of 12 comparisons were conducted. The Bonferroni adjustment for multiple comparisons as described by Uitenbroek (1997) was applied (df=16, criterion of $t>3.34$ and $p<0.004$). Significant differences were found for all 12 comparisons, suggesting that adults' SRTs were lower than those of children for all conditions tested.

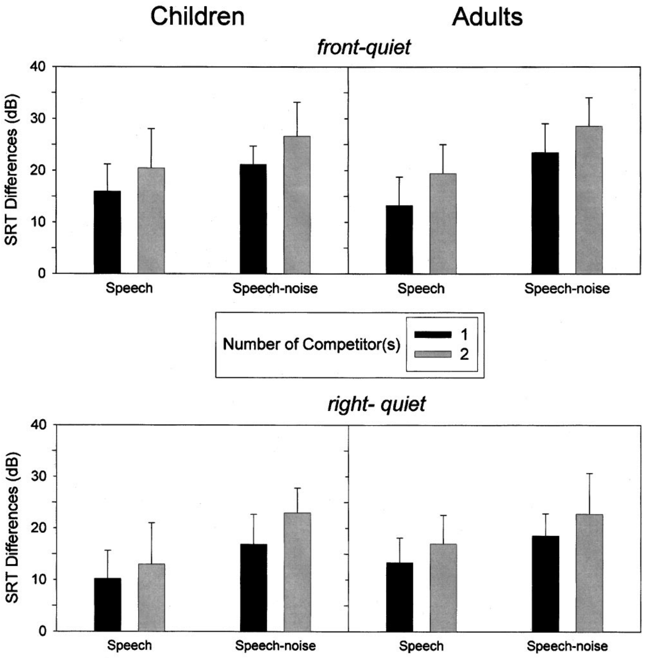

FIG. 1. Average (±SD, dB SPL) differences between speech reception thresholds (SRTs) in the masked and quiet conditions. Data are plotted for *front* (top panels) and *right* (bottom panels) conditions, for children (left panels) and adults (right panels). Each panel compares difference values for the speech and speech-noise competitors when the number of competitor(s) was either one (black bars) or two (gray bars).

Figure 1 shows group means (±SD) for masking (differences between masked and quiet SRTs). For each subject masking amounts for front and right were obtained by subtracting *quiet* SRTs from *front* and *right* SRTs, respectively. To place the masking values into context, average (±SD) SRTs for all groups and conditions are listed in Table II. Statistical analyses on the amount of masking for the child groups were conducted with a three-way mixed-design ANOVA treating condition (*front minus quiet, right minus quiet*) as the within-subjects variable and competitor type and number as the between-subjects variables. A significant effect of condition $[F(1,32)=29.13; p<0.0001]$ suggests

TABLE II. Mean (±SD) speech reception thresholds (in dB SPL)[a]

Group	Quiet	Front	Right
Children			
1 speech	26.02(3.81)	41.81 (6.31)	36.64 (6.48)
2 speech	27.32(5.25)	47.75 (6.30)	40.33 (6.29)
1 speech-noise	23.25(5.56)	44.37 (6.50)	40.13 (3.89)
2 speech-noise	21.45(3.3)	48.01 (2.07)	44.41 (7.18)
Adults			
1 speech	3.84(3.18)	16.71 (5.66)	16.86 (3.84)
2 speech		23.35 (4.41)	20.43 (4.01)
1 speech-noise		27.39 (5.28)	22.25 (4.82)
2 speech-noise		32.82 (4.40)	27.60 (8.65)

[a] It is important to recall that each child was tested on three conditions (quiet, front, right) for one masker type, and that each adult was tested on all nine conditions, hence only one entry in Table II for adult quiet thresholds.

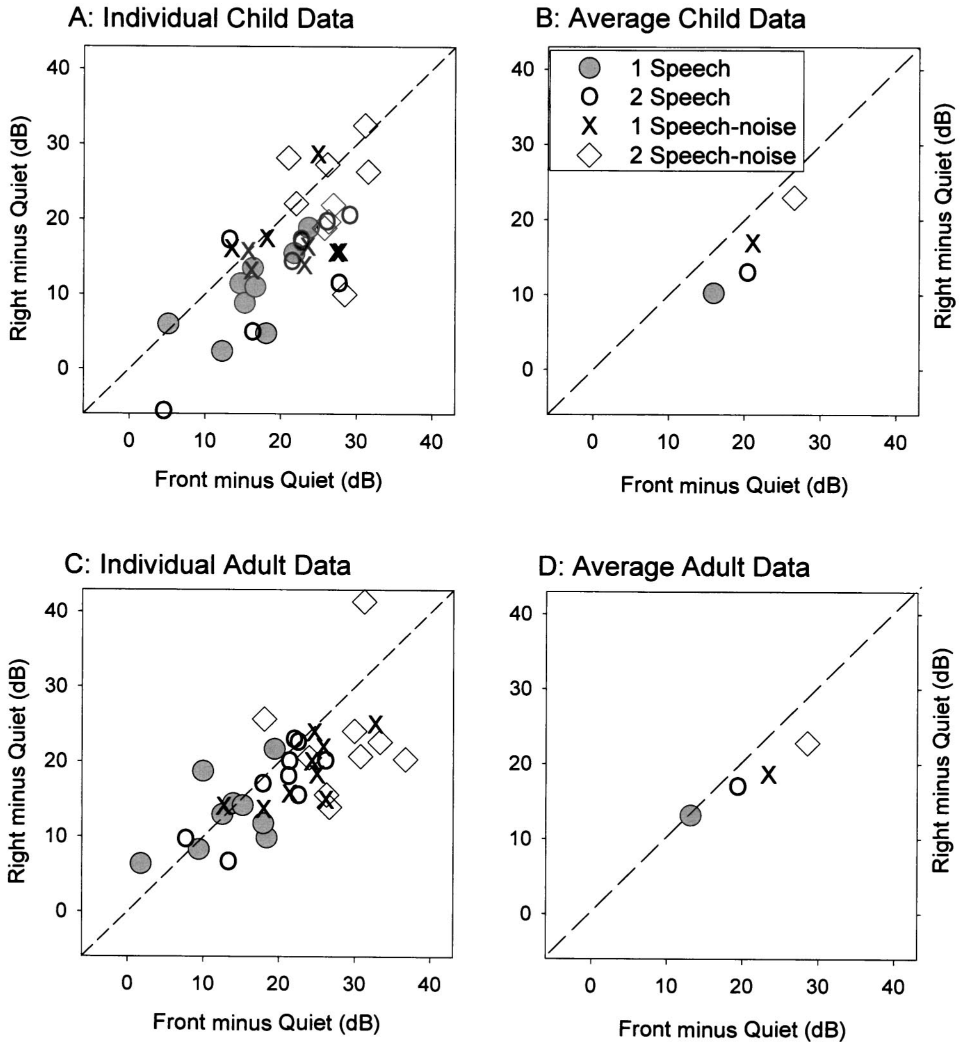

FIG. 2. Masking amounts (differences between masked and quiet thresholds) for the *Right minus Quiet* conditions are plotted vs. *Front minus Quiet* conditions. Panels (A) and (C) show data for children and adults, respectively; each symbol denotes data from an individual subject, and the four different symbols refer to the type/number combination of competitor(s). The diagonal lines denote equality between the two variables. Panels (B) and (D) show average group data from (A) and (C), respectively, for the four conditions tested.

that masking in the *front minus quiet* condition was higher than in *right minus quiet*. Significant effects of type [$F(1,32)=15.51$; $p<0.0001$] and number [$F(1,32)=6.95$; $p<0.013$] further suggest that masking was greater for two competitors than one, and greater for the speech-noise competitor compared with speech. There were no significant interactions. For the adult subjects, a three-way repeated measures ANOVA (condition×type×number) suggested, similar to the children, that masking was greater in the *front* versus *right* conditions [$F(1,8)=27.72$; $p<0.001$], greater with speech-noise than speech [$F(1,8)=30.72$; $p<0.001$] and greater for two compared with one competitor [$F(1,8)=16.71$; $p<0.004$]. Masking data for child and adult groups were compared with independent *t*-tests for each competitor location/type/number combination, and the Bonferroni correction for eight comparisons was applied (Uitenbroek, 1997). None of the comparisons yielded a significant difference in masking between the child and adult groups, and none of the interactions were significant.

Spatial release from masking was defined as the difference between front masking (*front minus quiet*) and right masking (*right minus quiet*). Figure 2 shows individual points for *right minus quiet* plotted versus *front minus quiet* for all subjects and conditions tested. If no spatial release from masking occurred, the points would be expected to fall along the diagonal. Points falling below the diagonal would be indicative of spatial release from masking. Alternatively, points falling above the diagonal would represent cases in which thresholds were higher when the competitors were on the right rather than in front. The majority of individual data points in Fig. 2 are below the diagonal, and average points for all but one group are also indicative of spatial release from masking.

Figure 3 summarizes the findings for spatial release from masking. For children, group average values are between 3.6 and 7.5 dB; the overall average for all 36 children is 5.25 dB. For adults, group averages range from 0 to 5.2 dB with an overall average of 3.34 dB. Children's data were analyzed with a two-way between-subjects ANOVA (type ×number), revealing no significant main effects or interactions. This lack of an effect may not surprising given the large intersubject variability, which is notable in Fig. 3(A); while some children had spatial release from masking values greater than 10 dB, other children had values near 0, and a small number had negative values. Adult data were analyzed with a two-way repeated measures ANOVA (type×number), also revealing no significant effects or interactions. Finally,

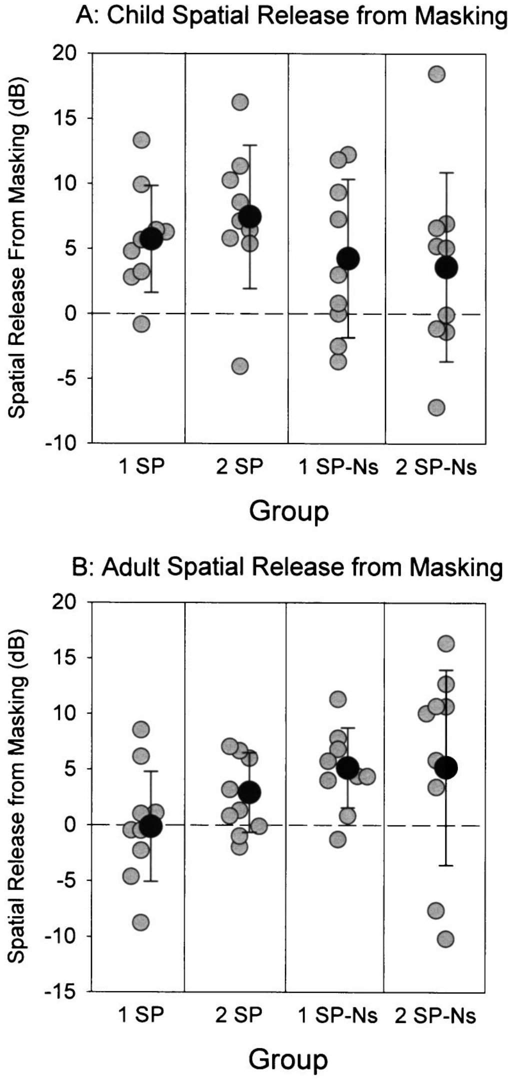

FIG. 3. Spatial release from masking values are shown for children and adults in panels (A) and (B), respectively. Each panel shows values grouped by competitor type/number condition (on the *x*-axis labels SP and Sp-Ns refer to the speech and speech-noise conditions, respectively). Individual values appear in gray circles, and group averages (±SD) are shown in black circles. When necessary to avoid overlap of data points, in some cases there was a slight shifting along the *x* axis.

to compare spatial release from masking for children and adults independent *t*-tests were conducted for each type/number combination, with the Bonferroni correction for four contrasts applied (Uitenbroek, 1997). The only significant difference between groups was for the one-speech competitor condition, in which the average spatial release from masking in adults is 0, compared with an average value of 5.7 for the child group.

IV. DISCUSSION

Speech intelligibility in quiet and in the presence of competing sounds and the ability to benefit from spatial separation of the speech and competitor(s) were investigated in children and adults. Although extensively studied in adults, to date this area of research has been minimal in children. This study may therefore be helpful towards improving our understanding of children's ability to hear and learn in noisy and reverberant environments, especially given that such abilities are known to be compromised compared with abilities measured under quiet condition (e.g., ANSI, 2002; Yacullo and Hawkins, 1987; Knecht *et al.*, 2002). The results can be summarized as follows: (1) Adults' SRTs were lower than those of the children for all conditions. (2) For both age groups masking was significantly greater with speech-noise than with speech and with two competitors compared with one. (3) The amount of masking did not differ across the two age groups. (4) The amount of spatial release from masking was similar for children and adults on all but one condition. (5) The number or type of competitor did not affect the size of spatial release from masking for either age group.

A. SRTs and masking amount

The primary age difference was that of higher SRTs in children than adults, in quiet and in all masked conditions. This age effect is consistent with existing developmental psychoacoustic literature, which has shown that children ages 4 to 7 typically have higher tone detection thresholds compared with adults (e.g., Buss *et al.*, 1999; Oh *et al.*, 2001). Similarly, recognition of spondee words such as those used here in temporally modulated noise has been shown to produce higher thresholds in 5 to 10 year-old children than in adults (Hall *et al.*, 2002).

The age effect found here can be attributed to a combination of peripheral and central mechanisms. Peripherally, frequency resolution is highly similar to that of adults by 5 years of age (Allen *et al.*, 1989; Hall and Grose, 1991; Veloso *et al.*, 1990). However, young children appear to integrate auditory information over a greater number of auditory channels than adults, suggesting that their ability to extract auditory cues, and in the present study to identify target words at low signal levels, is likely to be still developing (e.g., Hall *et al.*, 1997; Buss *et al.*, 1999; Hartley *et al.*, 2000; Oh *et al.*, 2001). Immaturity of central auditory processes and the adoption of listening strategies that are nonoptimal or less efficient than adults (Allen and Wightman, 1994; Lutfi *et al.*, 2003) may have also affected SRTs. Finally, differences in thresholds may represent age-related differences in the ability to take advantage of hearing partial word segments and to ''fill in'' the remainder of the target word. Anecdotal reports from adults suggest that they relied heavily on this strategy at low signal levels. The ability to adopt this strategy can most likely be attributed to adults' having more experience and better-developed language skills, including the ability to parse phonetic, semantic, and lexical aspects of speech (Fletcher and MacWhinney, 1995).

Of interest is the lack of an age effect for the amount of masking. Previous studies have typically shown that adults experience reduced masking compared with children (e.g., Buss *et al.*, 1999; Oh *et al.*, 2001; Papso and Blood, 1989; Hall *et al.*, 2002). Although this explanation may not be entirely satisfying, the lack of an age-related masking effect may be attributed to the task itself. In the current study, using the 4AFC task, *quiet* thresholds were extremely low in adults. In contrast, adults tested on the same measure using identical stimuli, but with a 25AFC did not reveal such low

SRTs in *quiet*, but continued to show lower masked SRTs. The amount of masking in the 25AFC task was therefore lower in adults than children (Johnstone and Litovsky, 2005). When increasing task difficulty for adults, a more realistic story with regard to age-related masking differences may emerge, suggesting the importance of equating for difficulty of the task when comparing perceptual abilities across age groups.

B. Competitor type

SRTs did not differ for the two types of competitors for children, but were higher with speech-noise than speech for the adults, which may be in part due to greater statistical power in the adult within-subjects comparisons. For both age groups, masking was greater with speech-noise than speech. These findings are consistent with other findings in adults in a one-masker paradigm, whereby greater amounts of masking were reported in the presence of speech-noise compared with speech (e.g., Hawley *et al.*, 2004). This has been attributed to greater amounts of overlap in the energies of the speech-noise masker and the target, resulting in the reduction of F0 discrimination. However, in previous work, as the number of maskers increased, speech became a more potent masker, an explanation involving informational masking and linguistic interference from multiple speech maskers was invoked to account for the increased interference from speech (e.g., Bronkhorst, 2000; Hawley *et al.*, 2004). Here, there was no interaction of type and number of competitors, which may be explained by stimulus differences across studies. Studies such as those of Hawley *et al.* (2004) typically use male voices for both the target and competitors, whereas here the target was a male voice and the competitor was spoken by a female. The differences in voice pitch, quality, and ongoing F0 differences provided a robust cue for source segregation in the presence of speech competitors, regardless of the number of competitors. The speech-noise competitor, having momentary dips in amplitude but no ongoing changes in frequency, served as a more potent masker whose effect was greater than that of speech. With same-gender competitors it is highly likely that speech would have produced masking at least as great, if not larger than the speech-noise competitor (e.g., Brungart *et al.*, 2001). Finally, the differences in masking amounts for the child groups may be accounted for by the fact that, for reasons that are not entirely clear, but probably due to random variation within the population, SRTs on the quiet conditions were lower in the two speech-noisé groups than in the groups tested with the speech competitors.

C. Number of competitors

For both children and adults, masking was significantly greater for two compared with one competitor(s), and the interactions of number with location (front versus right) were not significant. Averaged over all competitor types and numbers, the addition of a second competing sound resulted in increased masking of 4.7 dB for children and 4.8 for adults. Two interpretations can be considered here. First, in the presence of competitors with envelope modulations such as those used here, listeners may be better able to take advantage of the modulations and ‘‘listen in the gaps’’ in the presence of a single competitor. As a second competitor is added the signal contains fewer gaps, thereby decreasing opportunities of ‘‘gap listening’’ (e.g., Festen and Plomp, 1990; Hawley *et al.*, 2004). Second, consider the possible role of ‘‘informational’’ masking. In recent years this term has been used extensively in the auditory literature to explain masking phenomena that cannot be attributed solely to peripheral auditory mechanisms (e.g., Neff and Green, 1987; Lutfi, 1990; Kidd *et al.*, 2003). In the speech intelligibility literature, one of the conditions under which informational masking has been thought to occur is when the addition of a second masker elevates thresholds by more than the 3 dB expected simply from the added energy in the presence of a second masker (e.g., Brungart *et al.*, 2001; Hawley *et al.*, 2004; Durlach *et al.*, 2003). This threshold elevation may result from the increased complexity of the listening environment, possibly due to uncertainty on the part of the listener as to what aspects of the stimulus to ignore and what aspects to pay attention to. Although difficult to evaluate numerically, this component of masking may have been present here to some extent, and more direct tests of the effect in children would be important to pursue in future studies.

D. Spatial release from masking

Measures of spatial release from masking did not statistically differ across age groups, nor were there effects of competitor type and number. The only effect was the lack of spatial release from masking in the one-speech condition in adults, compared with 5.7 dB in children. The adult data differ from other free field studies in adults, in which spatial release from masking for speech was reported to be at least 3 dB for a single competing talker and as high as 12 dB for multiple talkers (Bronkhorst, 2000; Hawley *et al.*, 2004). The lack of release from masking found here with the one-speech competitor is likely due to the nature of the task and stimuli; the use of a fairly easy 4AFC task in combination with different-gender talkers for the target and competitor most likely created a relatively simple listening situation for adults.

Spatial cues are thought to be especially useful in challenging conditions when nonspatial cues are difficult to access (Peissig and Kollmeier, 1997; Bronkhorst, 2000; Durlach *et al.*, 2003; Freyman *et al.*, 2004). In the adult group tested here, spatial cues were beneficial in the conditions that created greater amounts of *front* masking (two-speech, one-speech-noise and two-speech-noise). The lack of a location effect in the one-speech condition is likely due to the general ease of listening to spondees when the competitor consists of a single, different-gender talker. In that condition, spatial cues did not help to reduce masking in the *right* condition, since masking was already relatively small in that condition. In contrast with adults, in children the one-speech *front* condition did present a challenging situation, probably because children are less able to take advantage of the different-gender competitor to hear the target speech. Thus, spatial cues were indeed relevant to the children so as to produce a robust improvement in the *right* condition compared with the *front*. These findings suggest that, while tasks that are more complex, using sentence material and/or same-gender stimuli may be more appropriate for measuring spatial release from masking in adults, the task used here is a good tool for measuring the ability of young children to negotiate complex auditory environments.

The finding that, overall, spatial release from masking in children is similar to that in adults is consistent with work showing that preschool-age children perform similar to adults on measures of binaural masking level differences (Nozza *et al.*, 1988; Moore *et al.*, 1991) and minimum audible angle [Litovsky (1997); for review see Litovsky and Ashmead (1997)]. This finding implies that for a simple closed-set task young children are able to utilize spatial and/or head shadow cues to the same extent as adults in order to segregate sounds in noisy environments. That is not to say that children would be expected to perform similar to adults on all measures of speech intelligibility in noise. Given recent findings that children exhibit poorer attentional selectivity on auditory tasks (e.g., Oh *et al.*, 2001), and reduced unmasking for tone detection under dichotic conditions (Wightman *et al.*, 2003; Hall *et al.*, 2004), the possibility remains that age differences would be seen under more demanding conditions, such as an open-set test or with same-gender target and competitors. Those differences, however, would not be attributable to age-dependent binaural abilities, but rather to other central processes such as auditory attention.

E. Conclusions

Young children require higher signal levels than adults to identify spondees in a simple 4AFC task, and these age-related differences may be mediated by both peripheral and central auditory processes. The fact that young children can benefit from spatial separation of the target speech and competing sources suggests that in a complex acoustic environment, such as a noisy classroom, they might find it easier to attain information if the source of interest is spatially segregated from noise sources. Although, the extent to which this is true with real-world sounds may depend on duration, complexity and type of sounds, and the demand on attentional resources that various sounds may require. Finally, the test used here (developed by Litovsky, 2003) is designed to also be used in pediatric clinical settings where young children are often fitted with hearing aids or cochlear implants, with little knowledge about the efficacy of the fittings in noisy environments. This test may offer a way to evaluate the abilities in children with hearing aids and cochlear implants to function in noisy environments, and may, for example, be useful in assessing the extent to which children obtain a benefit from bilateral fitting strategies (Litovsky *et al.*, 2004).

ACKNOWLEDGMENTS

The author is grateful to Aarti Dalal and Gerald Ng for assistance with programming and data collection, and to Patti Johnstone and Shelly Godar for helping with data analysis. The author is also grateful to Dr. Joseph Hall for initially suggesting the use of spondees in a forced choice paradigm, and to Dr. Adelbert Bronkhorst and an anonymous reviewer for helpful suggestions during the review process. This work was supported by NIDCD (Grant Nos. DC00100 and DC0055469), National Organization for Hearing Research, and the Deafness Research Foundation. Portions of the data were collected while R. Litovsky was at Boston University, Hearing Research Center.

APPENDIX: LIST OF SPONDEE WORDS USED IN THE PRESENT EXPERIMENT

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

---

[1] The lower limit of 4.5 years is slightly conservative, and was based on pilot testing which suggested that by that age all children were familiar with the majority of the target words. The upper limit of 7.5 is somewhat smaller than the 10-year limit used in a number of other works (e.g., Oh *et al.*, 2001; Hall *et al.*, 2002), but similar to that used in studies on auditory attention in young children, in which there do not appear to be developmental effects within the age range (e.g., Stellmack *et al.*, 1997; Oh *et al.*, 2001).

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., Wightman, F., Kistler, D., and Dolan, T. (**1989**). ''Frequency resolution in children,'' J. Speech Hear. Res. **32**, 317–322.

American National Standards Institute (**2002**). ''Standard for acoustical characteristics of classrooms in the United States,'' ANSI—S12.60.

Arbogast, T. L., Mason, C. R., and Kidd, G. (**2002**). ''The effect of spatial separation on informational and energetic masking of speech,'' J. Acoust. Soc. Am. **112**, 2086–2098.

Bronkhorst, A. (**2000**). ''The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' Acta. Acust. Acust. **86**, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1992**). ''Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing,'' J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). ''Informational and energetic masking effects in the perception of multiple talkers,'' J. Acoust. Soc. Am. **110**, 2527–2538.

Buss, E., Hall, III, J. W., 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.

Crandell, C. C. (**1993**). ''Speech recognition in noise by children with minimal degrees of sensorineural hearing loss,'' Ear Hear. **14**, 210–216.

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.

Dawson, P. W., Decker, J. A., and Psarros, C. E. (**2004**). ''Optimizing dynamic range in children using the nucleus cochlear implant,'' Ear Hear. **25**, 230–241.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, Jr., G. (**2003**). ''Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity,'' J. Acoust. Soc. Am. **114**, 368–379.

Eisenberg, L. S., Kirk, K. I., Martinez, A. S., Ying, E. A., and Miyamoto, R. T. (**2004**). ''Communication abilities of children with aided residual hearing: comparison with cochlear implant users,'' Arch. Otolaryngol. Head Neck Surg. **130**, 563–569.

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.

Fletcher, P., and MacWhinney, B. (**1995**). *Handbook of Child Language* (Blackwell, Oxford, UK).

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). ''Spatial release from informational masking in speech recognition,'' J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). ''Effect of number of masking talkers and auditory priming on informational masking in speech recognition,'' J. Acoust. Soc. Am. **115**, 2246–2256.

Hall, J. W., Buss, E., Grose, J. H., and Dev, M. B. (**2004**). ''Developmental effects in the masking-level difference,'' J. Speech Lang. Hear. Res. **47**, 13–20.

Hall, III, J. W., 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**, 651–60.

Hall, III, J. W., Grose, J. H., and Dev, M. B. (**1997**). ''Auditory development in complex tasks of comodulation masking release,'' J. Speech Lang. Hear. Res. **40**, 946–954.

Hall, III, J. W., 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, III, J. W., Grose, J. H., Buss, E., Dev, M. B., Drake, A. F., and Pillsbury, H. C. (**2003**). ''The effect of otitis media with effusion on perceptual masking,'' Arch. Otolaryngol. Head Neck Surg. **129**, 1056–1062.

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 Colburn, H. S. (**1999**). ''Speech intelligibility and localization in complex environments,'' J. Acoust. Soc. Am. **105**, 3436–3448.

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

Johnstone, P., and Litovsky, R. Y. (**2005**). ''Speech intelligibility and spatial release from masking in children and adults for various types of interfering sounds,'' J. Acoust. Soc. Am. (in press).

Kidd, Jr., G., Mason, C. R., and Richards, V. M. (**2003**). ''Multiple bursts, multiple looks, and stream coherence in the release from informational masking,'' J. Acoust. Soc. Am. **114**, 2835–2845.

Koehnke, J., and Besing, J. M. (**1996**). ''A procedure note for testing speech intelligibility in a virtual listening environment,'' Ear Hear. **17**, 211–217.

Knecht, H. A., Nelson, P. B., Whitelaw, G. M., and Feth, L. L. (**2002**). ''Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements,'' Am. J. Audiol. **11**, 65–71.

Litovsky, R. (**1997**). ''Developmental changes in the precedence effect: Estimates of Minimal Audible Angle,'' J. Acoust. Soc. Am. **102**, 1739–1745.

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

Litovsky, R., and Ashmead, D. (**1997**). ''Developmental aspects of binaural and spatial hearing,'' in *Binaural and Spatial Hearing*, edited by R. H. Gilkey and T. R. Anderson (Earlbaum, Hillsdale, NJ), pp. 571–592.

Litovsky, R. Y., Fligor, B., and Tramo, M. (**2002**). ''Functional role of the human inferior colliculus in binaural hearing,'' Hear. Res. **165**, 177–188.

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.

Lutfi, R. A. (**1990**). ''How much masking is informational masking?'' J. Acoust. Soc. Am. **88**, 2607–2610.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). ''One factor underlies individual differences in auditory informational masking within and across age groups,'' Percept. Psychophys. **65**, 396–406.

Moore, D. R., Hutchings, M., and Meyer, S. (**1991**). ''Binaural masking level differences in children with a history of otitis media,'' Audiology **30**, 91–101.

Moore, D. R., Hartley, D. E., and Hogan, S. C. (**2003**). ''Effects of otitis media with effusion (OME) on central auditory function,'' Int. J. Pediatr. Otorhinolaryngol. **67**, S63–S67.

Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 409–415.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). ''Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise,'' J. Acoust. Soc. Am. **95**, 1085–1099.

Nozza, R. J., Wagner, E. F., and Crandell, M. A. (**1988**). ''Binaural release from masking for a speech sound in infants, preschoolers, and adults,'' J. Speech Hear. Res. **31**, 212–218.

Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). ''Children's detection of pure-tone signals with random multiple 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**, 337–338.

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., and Mimpen, A. M. (**1981**). ''Effect of the orientation of the speaker's head and the azimuth on a noise source on the speech reception threshold for sentences,'' Acustica **48**, 325–328.

Roberts, J., Hunter, L., Gravel, J., Rosenfeld, R., Berman, S., Haggard, M., Hall, III, J., Lannon, C., Moore, D., Vernon-Feagans, L., and Wallace, I. (**2004**). ''Otitis media, hearing loss, and language learning: controversies and current research,'' J. Dev. Behav. Pediatr. **25**, 110–122.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). ''IEEE Recommended Practice for Speech Quality Measurements,'' IEEE Trans. Audio Electroacoust. **17**, 227–246.

Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. (**2001**). ''Spatial unmasking of nearby speech sources in a simulated anechoic environment,'' J. Acoust. Soc. Am. **110**, 1118–1129.

Stellmack, M. A., Willihnganz, M. S., Wightman, F. L., and Lutfi, R. A. (**1997**). ''Spectral weights in level discrimination by preschool children: analytic listening conditions,'' J. Acoust. Soc. Am. **101**, 2811–2821.

Uitenbroek, D. G. (**1997**). ''SISA Binomial,'' Southampton: D. G. Uitenbroek. Retrieved 1 January, 2004, from the World Wide Web: http://home.clara.net/sisa/binomial.htm.

Veloso, K., Hall, III, J. W., and Grose, J. H. (**1990**). ''Frequency selectivity and comodulation masking release in adults and in 6-year-old children,'' J. Speech Hear. Res. **33**, 96–102.

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

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

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.

Yacullo, W. S., and Hawkins, D. B. (**1987**). ''Speech recognition in noise and reverberation by school-age children,'' Audiology **26**, 235–246.