Development of the Listening in Spatialized Noise-Sentences Test (LISN-S)

Sharon Cameron and Harvey Dillon

Objective: **The goals of this research were to develop and evaluate a new version of the Listening in Spatialized Noise Test (LISN®; Cameron, Dillon & Newall, 2006a) by incorporating a simplified and more objective response protocol to make the test suitable for assessing the ability of children as young as 5 yr to understand speech in background noise. The LISN-Sentences test (LISN-S; Cameron & Dillon, Reference Note 1) produces a three-dimensional auditory environment under headphones and is presented by using a personal computer. A simple repetition response protocol is used to determine speech reception thresholds (SRTs) for sentences presented in competing speech under various conditions. In four LISN-S conditions, the maskers are manipulated with respect to location (0° versus ±90° azimuth) and vocal quality of the speaker(s) of the stories (same as, or different than, the speaker of the target sentences). Performance is measured as two SRT measures and three "advantage" measures. These advantage measures represent the benefit in decibels gained when either talker, spatial, or both talker and spatial cues combined, are incorporated in the maskers. This use of difference scores minimizes the effects of between-listener variation in factors such as linguistic skills and general cognitive ability on LISN-S performance.**

Design: **An initial experiment was conducted to determine the relative intelligibility of the sentences used in the test. Up to 30 sentences were presented adaptively to 24 children ages 8 to 9 yr to estimate the SRT (eSRT). Fifty sentences each were then presented at each participant's eSRT, eSRT +2 dB, and eSRT −2 dB. Psychometric functions were fitted and the sentences were adjusted in amplitude for equal intelligibility. After adjustment, intelligibility increased across sentences by approximately 17% for each 1 dB increase in signal-to-noise ratio (SNR). A second experiment was conducted to gather normative data on the LISN-S from 82 children with normal hearing, ages 5 to 11 yr.**

Results: **For the 82 children in the normative data study, regression analysis showed that there was a strong trend of decreasing SRT and increasing advantage as age increased across all LISN-S performance measures. Analysis of variance revealed that significant differences in performance were most pronounced between the 5-yr-olds and the other age groups on the LISN-S measures that assess the ability to use spatial cues to understand speech in background noise, suggesting that binaural processing skills are still developing at age 5 yr. Inter-participant variation in performance on the various SRT and advantage measures was minimal for all groups, including the 5- and 6-yr-olds who exhibited standard deviations ranging from only 1.0 dB to 1.8 dB across measures. The intra-participant standard error ranged from 0.6 dB to 2.0 dB across age groups and conditions. Total time taken to administer all four LISN-S conditions was on average 12 minutes.**

Conclusions: **The LISN-S provides a quick, objective method of measuring a child's ability to understand speech in background noise. The small degree of inter- and intra-participant variation in the 5- and 6-yr-old children suggests that the test is capable of assessing auditory processing in this age group. However, because there appears to be a strong developmental curve in binaural processing skills in the 5-yr-olds, it is suggested that the LISN-S be used clinically with children from 6 yr of age. Cut-off scores, calculated as 2 standard deviations below the mean adjusted for age, were calculated for each performance measure for children ages 6 to 11 yr. These scores, which represent the level below which performance on the LISN-S is considered to be outside normal limits, will be used to in future studies with children with suspected central auditory processing disorder.**

(Ear & Hearing 2007;28;196–211)

Central auditory processing disorder (CAPD) is an auditory-specific perceptual deficit in the processing of auditory stimuli that occurs despite normal intellectual capacity and normal peripheral hearing thresholds. It is hypothesized that a major cause of CAPD is a deficit in the accurate representation of auditory space (Jerger, 1998), which may result from an inability to utilize cues such as the timing and loudness of signals arriving at the two ears to separate targeted sounds (such as a teacher's voice) from background noise—a skill that requires appropriate binaural interactions to occur.

This ability was previously investigated in a group of 10 children with suspected CAPD, using a prototype version of the Listening in Spatialized Noise–Sentences Test (LISN-S), the LISN-Continuous Discourse test (LISN-CD; Cameron, Dillon & Newall,

National Acoustic Laboratories, Chatswood, Australia (S.C., H.D.), Macquarie University, North Ryde, Australia (S.C.).

0196/0202/07/2802-0196/0 • Ear & Hearing • Copyright © 2007 by Lippincott Williams & Wilkins • Printed in the U.S.A.

2005, 2006c). Like the LISN-S, the LISN-CD produces a three-dimensional auditory environment under headphones, whereby the speech stimuli appear to be emanating from various positions in auditory space around the listener. This offers an alternative to free-field testing, which is limited by factors such as listener head movement (which can affect the sound at the eardrum by several dB), replication of loudspeaker and listener placement between clinics, and the effects of reverberation between clinics.

On the LISN-CD, the signal-to-noise ratio (SNR) required by the CAPD group to understand a target children's story in the presence of spatially separated distracter sentences was significantly higher ($p < 0.000001$) than for a group of 48 normal-hearing, age-matched control subjects, with 9 of the 10 children in the CAPD group outside normal limits on this task (on average by 5 standard deviations). The implication and significance of these results are that the inability to adequately combine information at the two ears to directionally suppress noise coming from nontarget directions is a major cause of CAPD and presumably of listening difficulties in the classroom.

In the administration of the LISN-CD, the child was required to listen to the target story in background discourse. In a three-alternative, forced-choice adaptive procedure, the participant's task was to indicate the amplitude at which the target story was either easy to understand, just understandable, or too hard to understand, in the presence of the distracters by pointing to the appropriate picture on a response card. The listener's threshold in any condition was calculated across the final four of seven reversals between "easy to understand" and "too hard to understand." Once a practice or listening trial was completed, the children were required to provide details of the stories that they had heard. This procedure was implemented to keep the children engaged in the task and to ensure that they did not point to the "just understandable" response card icon unless they could still basically understand the story. However, the administration procedure of the LISN-CD limited use of this test to children from age 7 yr.

The aim of the current project was to develop a version of the LISN that could be used clinically to assess various skills used to understand speech in background noise, including binaural interaction ability, in children as young as 5 yr of age. Identification of such deficits in young children would allow for the earliest possible implementation of environmental modifications, compensation strategies, and remediation.

The LISN-S uses simplified administration procedures and a straightforward sentence-repetition response protocol to measure speech reception thresholds (SRTs) adaptively. This response protocol was considered to be more objective than that used by the LISN-CD, whereby the child is required to provide details of the target story at the completion of the listening trial, because story details can vary from child to child. For the purposes of this investigation, SRT can be defined as the SNR that yields 50% intelligibility. In the administration of an adaptive test, the presentation of each successive item in the test is determined by the person's responses to previous items, and the investigator determines the rules that will dictate when to increase or decrease the level of the target or the masker. The least complicated strategy is a simple up-and-down procedure (Bode & Carhart, 1974), whereby the stimulus level and/or the SNR is decreased after a positive response or increased after a negative response.

Adaptive procedures are recommended in the administration of speech intelligibility tests because they concentrate presentation levels in the range that yields the smallest within-participant standard deviations (SD) in SRT and slope estimates (Brand & Kollmeier, 2002), and ceiling and floor effects can be avoided (Mackie & Dermody, 1986). According to Brand & Kollmeier (2002), the typical slope values of the performance-intensity (PI) function for speech intelligibility tests that use sentences as the target stimuli in listeners with normal hearing is 20% per decibel. At least 20 sentences are required to provide an SRT with an SD of 1 dB (Brand & Kollmeier, 2002; Vaillancourt et al., 2005).

In adaptive tests, the threshold is usually based on the average of the test levels (e.g., sound pressure levels [SPLs] or SNRs) at which the test was administered, excluding those items near the start of the test where the test level has not yet started traversing the region containing the true underlying threshold. Because the threshold is a simple mean of the individual test levels, the reliability of the threshold (i.e., the within-participant SD of test scores that would be expected were the entire test to be repeated many times under identical conditions) can be estimated by calculating the standard error of the mean (SE) of the test levels used to determine the threshold. As the goal of the tester is often to obtain a threshold with some specified accuracy, it is reasonable to stop the adaptive test whenever the SE decreases below some predetermined value.

The SE stopping rule has been used successfully in a number of applications using computer-adaptive testing, including tests of reading comprehension in individuals with English as a second language (Lilley, Barker & Britten, 2004; Young, Shermis, Brutten & Perkins, 1996) and subjective judgments to evaluate health outcomes (Cook, Roddey, O'Malley & Gartsman, 2005; Dijkers, 2003).

When an SE stopping rule is applied, all participants are tested to the same level of precision, regardless of the number of items required to reach the specified SE (Lilley et al., 2004; Lunz, Bergstrom & Gershon, 1994). Chae, Kang, Jeon & Linacre (2000) report that computer-adaptive testing using an SE stopping rule can reduce testing time while maintaining the same level of reliability. Shorter testing times reduce fatigue, a factor that can significantly affect a participant's test results. Such a benefit was considered particularly important when developing a test for children with suspected CAPD who may be subjected to relatively lengthy test sessions involving multiple assessment tools.

Another important consideration in the development of speech intelligibility tests is equivalency of test items. In a study on the Revised Speech in Noise (RSIN) test, Cox, Gray & Alexander (2001) reported significant differences in performance across test lists, and it was suggested that a listener's score for a particular condition be multiplied by a weighting factor for that condition to correct for the problem of equivalence.

In the development of the Hearing in Noise Test (HINT), Nilsson, Soli & Sullivan (1994) adjusted the mean-squared amplitude of each sentence to equate intelligibility when presented in spectrally matched noise to 78 listeners with normal hearing. Groups of six to eight listeners were tested at fixed SNRs. Average percent correct scores for each sentence were calculated from the number of words repeated correctly. The difference in an individual sentence's percent intelligibility score from the overall mean was used to make an adjustment to the mean-squared amplitude level of the sentence. After rescaling of the sentences, the testing was repeated with another set of participants and the rescaling was repeated, with 51% of the sentences falling within ±1 dB of the average.

An additional 18 adults participated in a study on inter-list reliability on the HINT. Sentences were presented using an adaptive procedure in 2 dB steps. Whole sentences were scored as either correct or incorrect, and the first sentence was presented below threshold. There were 12 sentences in each of 22 lists. Nilsson et al. (1994) reported that mean presentation levels stabilized after the fifth sentence, suggesting that five sentences are sufficient for practice.

Test efficiency can also be affected by scoring method. Brand & Kollmeier (2002) performed Monte Carlo simulations to compare efficiency for sentence tests that use individual word scoring versus tests that score whether or not the complete sentence is repeated correctly. It was found that the complete sentence scoring method is the poorer condition for efficient estimates of SD because the adaptive procedures cannot work properly and the fitting is based on too few samples. This conclusion is consistent with sampling theory wherein the random measurement error inherent in any score decreases as the number of items giving rise to that score increases (Hagerman, 1976). The authors concluded that as word scoring only takes a little more experimental effort than sentence scoring this effort seems to be justified by the gain in efficiency and accuracy.

Another consideration to be addressed in the development of speech intelligibility tests is the age and linguistic abilities of the target population to be assessed. Bess (1983) notes that whereas certain adult lists can be used with some success in the evaluation of children the selection of speech materials for children should be made on the basis of the level of a child's speech and language competence. The Bamford-Kowal-Bench sentences test (BKB, Bamford & Wilson, 1979) is an example of a speech intelligibility test developed specifically for children. The BKB sentences contain mainly Stage 3 and some Stage 2 clause structures as described in the Language Assessment, Remediation and Screening Procedure (LARSP; Crystal, 1989), and are suitable for children from 4.6 yr of age (Kowal, 1979). The LISN-S sentences were also developed along these guidelines.

Finally, the need to control for the influence of confounding factors such as language skills on test performance was a major influence on the LISN-S design. Jerger & Allen (1998) commented that abnormal performance on global measures of word recognition that are used to assess CAPD could reflect either an auditory disorder with good linguistic skills or a linguistic disorder with good auditory skills.

In the presentation of the LISN-S, the competing speech is manipulated in respect to location (0° versus ±90°) and the vocal quality of the speaker(s) (same as, or different than, the speaker of the target sentences). This results in four listening conditions, (1) the same voice-0° condition (or "low-cue SRT," in which no spatial or vocal cues are provided), (2) the different voices-0° condition, (3) the same voice-±90° condition, and (4) the different voices- ±90° condition (or "high-cue SRT," in which both spatial and different talker cues are provided). Performance on the test is measured as a difference score, or the advantage in dB, gained from the listener's use of spatial and/or talker cues to distinguish the target stimulus from the distracters (see Fig. 1). The use of difference scores to measure performance on the LISN-S minimizes the influence of linguistic skills on test performance. For example, as linguistic

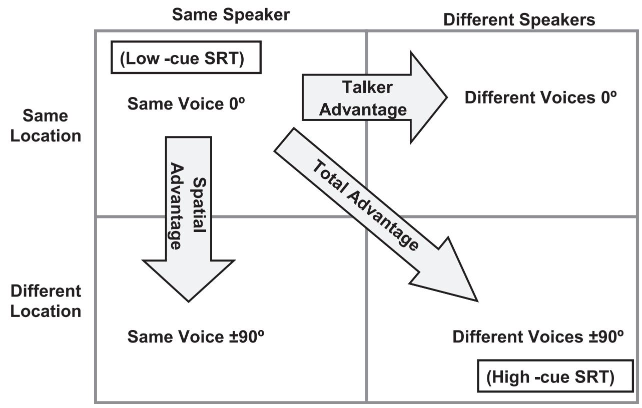

Fig. 1. LISN-S SRT and advantage measures.

skills affect both the SRT when the distracters are presented at 0°, and the SRT when they are spatially separated at ±90°, these skills will have minimal effect on the difference between the SRTs in these two conditions.

HYPOTHESES

Previous research studies such as Picard & Bradley (2001) and Jamieson et al. (2004) have shown that speech recognition in background noise improves with increasing age (for a review of the literature on this topic see Cameron, Dillon & Newall, 2006b). Cameron et al. (2006b) also found a linear trend of improved performance but no significant differences between age groups on the various performance measures of the LISN-CD test. It is therefore hypothesized that there will be a linear trend of improved performance on the LISN-S as a function of age. It is also hypothesized that the spatial, talker, and total advantage measured with the objective repetition task will be similar to that measured with the more subjective rating method used in the LISN-CD, whereby the children were required to indicate the intelligibility level of a target story (easy, just understandable, too hard to understand) by pointing to an icon on a response card.

DEVELOPMENT OF THE LISN-S

LISN-S Software Development

The LISN-S graphical user interface and signal processing application program were developed in the C# programming language and were based on the LISN-CD software described in Cameron, Dillon & Newall (2006a). An image of the playback screen used to administer the LISN-S is provided in Figure 2.

Speech Stimuli

The target sentences were developed by the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation (CRC HEAR) and are used under license from HearWorks Pty Limited. The sentences were written by a number of Australian registered speech pathologists specializing in rehabilitation of children with hearing loss. Each sentence was constructed in accordance with the criteria used in the development of the original BKB sentences (Kowal, 1979).

A total of 180 sentences were recorded for the LISN-S. A total of 120 sentences were used in the final version of the test. Examples of some sentences used in the LISN-S appear as Appendix A. Two stories are used for the LISN-S distracter discourse. The distracter stories, which were written for young school children by a published Australian children's novelist, are entitled "Loopy Lizard's Tail" and "The Great Big Tiny Traffic Jam." An extract from "Loopy Lizard's Tail" appears as Appendix B.

Recording

The Australian version of the LISN-S target sentences and distracter stories were recorded at the Speech, Hearing, and Language Research Centre (SHLRC) at Macquarie University, Sydney. The target sentences and both distracter stories were recorded by the first author (Female 1). The distracter stories were also recorded by two female linguists from SHLRC (Female 2 and Female 3). All stimuli were produced with a general Australian accent. General Australian English is the stereotypical variety of Australian English used by the majority of Australians and it dominates the accents found in contemporary Australian-made films and television programs. All speakers were of the same ethnicity, of approximately the same age, and came from the same region of Australia (Sydney, New South Wales). These qualifications were implemented to prevent listeners from detecting differences between the stories on the basis of confounding variables such as accent.

For both the target sentences and the distracter discourse, the speakers were instructed to speak with a normal clear voice, using a standard dialect, while maintaining a normal rhythm of speech and avoiding placing emphasis on key words. Specifically, clarity, pace, and effort were maintained across words.

Editing

The analog signal was recorded directly onto hard disk. The standard sampling frequency used in compact disk (CD) recordings of 44.1 kHz with a 16 bit

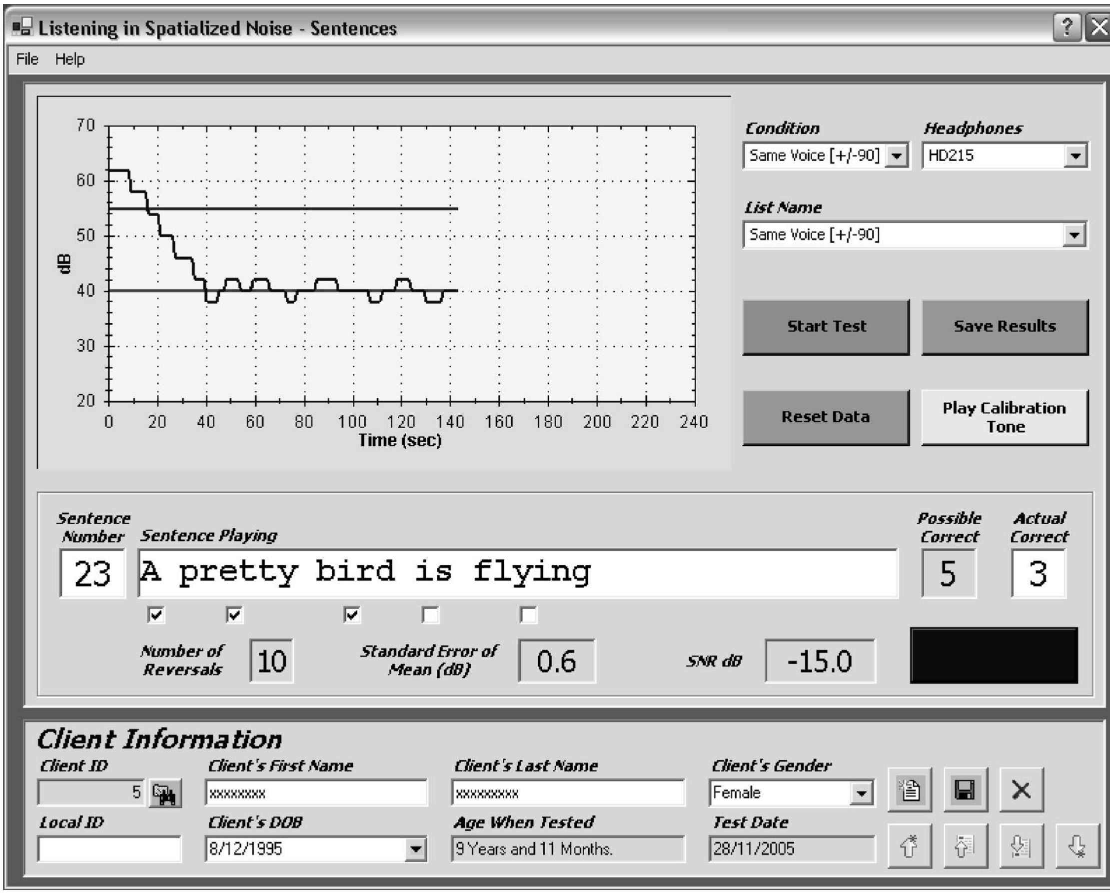

Fig. 2. LISN-S Playback Screen. Graph shows the history of the target level as the range of correct responses from greater than 50% correct to less than 50% correct is repeatedly traversed. Top horizontal line shows the level of the distracters; lower horizontal line shows the average level of the targets during the stable region.

digitization was used. The individual target sentences and distracter discourse were extracted from the recordings and edited using Adobe Audition Version 1.5. A silent period of 100 msec was inserted immediately preceding and after each distracter story. Extraneous pauses were removed during the editing process to ensure that the stories ran smoothly and at a constant intensity level. The stories were approximately 2 minutes and 30 sec in length.

Level Normalization

The root mean square (RMS) levels of each sentence and the distracter stories were ascertained by using Adobe Audition 1.5. These RMS levels were then averaged (in dB) across all stimuli. The amplitude of each sentence and distracter was then adjusted by a correction factor equal to the average total RMS, minus sentence/distracter RMS, minus 7 dB. The 7 dB correction was subtracted to ensure that no clipping occurred when the distracters were convolved with head-related transfer functions (HRTFs) at ±90° azimuth. The final corrected RMS level (before convolution) for all sentences was –27.1 dB re: digital full scale.

Convolution

Each sentence recorded by Female 1 was convolved with HRTFs recorded at 0° azimuth. "Loopy Lizard's Tail" (recorded by Female 1 and Female 2) was convolved with HRTFs recorded at 0° azimuth and –90° azimuth. "The Great Big Tiny Traffic Jam" (recorded by Female 1 and Female 3) was convolved with HRTFs recorded at 0° azimuth and +90° azimuth. The HRTFs were recorded in a chamber, anechoic above 50 Hz, using a Knowles Electronics Manikin for Acoustic Research (KEMAR), including a Zwislocki coupler and half-inch microphone, as described in Cameron et al. (2006a). Knowles Electronics small-sized pinnae were fitted to simulate the outer ear. The HRTFs were produced from swept sine waves ranging in frequency from 50 Hz to 20,000 Hz, which were produced by a signal analyzer and presented from a single loudspeaker positioned 1 meter from the center point of KEMAR's head.

The speech files were synthesized by using the LISN convolution program, developed using MATLAB software (MathWorks Inc, 2002), described in Cameron et al. (2006a). In summary, the various speech files were converted to the frequency domain by a fast Fourier transform (FFT) and then multiplied by the HRTFs, as well as an inverse headphone response that is described in the postequalization procedure section below. An inverse FFT (iFFT) was then applied to convert the signals back into the time domain for playback.

Postequalization Procedure

A postequalization procedure was also implemented to correct for the response of the headphones used during playback. Swept sine wave signals were played through Sennheiser HD215 circumaural audiometric high-frequency headphones on a KEMAR and recorded by a Stanford Research Systems two-channel network signal analyzer to measure the headphone-to-eardrum transfer function (HpTF). A filter with the inverse transfer function of the HpTF was developed, as described in Cameron et al. (2006a). The inverse HpTF was convolved with the spatialized LISN-S speech materials, effectively cancelling out the HpTF that occurs during playback. The convolved, postequalized stimuli were saved as .wav files.

Stimulus Generation

The convolved and hence spatialized speech files were stored in a LISN subdirectory for subsequent playback. The playback screen and related programs retrieved the spatialized target and distracter speech files and combined and scaled these stimuli in decibels to produce a binaural output signal. The right and left ear components of the binaural signal were assigned to the right and left channels of the computer sound card, respectively.

Calibration

The mean RMS level of the combined distracters (averaged across the recordings made by Female 1 and Females 2 and 3) at 0° was –22.7 dB and at ±90° was –21.6 dB. A 1 kHz reference tone was also established as a .wav file and was set to 10 dB greater than the average of the combined total RMS levels of the two distracter files convolved with HRTFs at 0°, that is, –12.7 dB.

The LISN-S is administered by using a PC, and the stimuli are presented through Sennheiser HD215 headphones that are connected directly to the headphone socket of the PC. To determine the exact output levels in millivolts required to present the LISN-S stimuli at a designated level in dB SPL, the various LISN-S stimuli were presented through the left and right ear headphones to a Brüel and Kjær type 4153 artificial ear, using a flat plate adaptor. Equivalent dBV RMS levels were measured directly from the headphone socket of the PC, using a Stanford Research Systems two-channel network signal analyzer.

When the 1 kHz reference tone was activated from the playback screen and the volume of the PC was adjusted until the electrical level of the calibration signal was 21 mV (as measured by a voltmeter), the dB SPL in each ear of the two combined distracters at 0° azimuth (recorded by Female 1 and Females 2 and 3) matched the corresponding level on the LISN-S competition slider bar. Similarly, the dB SPL of the target (recorded by Female 1) matched the corresponding level on the LISN-S target slider bar. Thus, for these headphones, daily calibration was achieved by adjusting the PC volume control until the electrical level of the calibration signal was 21 mV.

EXPERIMENT 1

Sentence Equivalence Study

The following study was conducted to determine the relative intelligibility of the LISN-S sentences and to adjust the level of the sentences for equal intelligibility. Approval to conduct the Sentence Equivalence Study was obtained from the relevant ethics review committees for human research.

Participants

Data were collected from 24 children with normal hearing, ages 8 yr, 0 mo, to 9 yr, 10 mo (mean, age 8 yr, 10 mo). There were 11 boys and 13 girls. The children were recruited from mainstream classrooms of a local primary school in Sydney, Australia. The participants were included in the study if they had Australian English as a first language, no history of hearing disorders (such as episodes of otitis media requiring ventilation tubes), and no reported learning or attention disorders. On the day of testing, all participants had pure-tone thresholds of ≤15 dB HL at 500 to 4000 Hz, and ≤20 dB HL at 250 and 8000 Hz, as well as normal Type A tympanograms and 1000 Hz ipsilateral acoustic reflexes present at 95 dB HL.

Materials

The LISN-S stimuli were administered with the use of a Toshiba Tecra M3 PC and Sennheiser HD215 headphones. Output levels were controlled with the calibration instrument described in the development section.

Design and Procedure

Testing was carried out during school hours in an Isuzu NPS300 mobile laboratory featuring an acoustically treated sound studio. Each participant heard target sentences initially presented at a level of 62 dB SPL (as measured in a Brüel and Kjær type 4153 artificial ear). Competing children's stories, looped during playback, were presented at a constant level of 55 dB SPL. The target and competing signals were presented simultaneously to both ears. The stimuli were all presented in the "same voice-0°" condition, whereby the target sentences and distracter stories are all spoken by the same female speaker and were processed with the HRTFs appropriate to a source at 0° azimuth (directly in front of the listener). The participant's task was to repeat the words heard in each sentence. A 1000 Hz tone burst, 200 msec in length, was presented before each sentence to alert the listener that a sentence would be presented. A silent gap of 500 msec separated the tone burst from the onset of the sentence. The tone burst was presented at a constant level of 55 dB SPL.

Up to 30 sentences were presented to estimate the SRT. The level of the target sentences was adjusted adaptively by the software to estimate the SRT (eSRT). If the listener responded with more than 50 percent of words in the sentence correctly identified the amplitude of the next sentence was reduced. If the listener responded with less than 50 percent of words correctly identified the amplitude of the next sentence was increased. The levels were adjusted in 4 dB steps until the first reversal in performance was recorded and in 2 dB steps thereafter. All sentences in the trial to form the initial estimate of SRT had an odd number of words so that SNR changed after each response. These sentences were not used in any subsequent study.

Independent midpoint target levels were calculated as the average of the level of a positive-going reversal and the level of the succeeding negative-going reversal. The trial to establish eSRT continued until the number of independent midpoints was $\geq$3 (with 2 reversals provided for practice) and the SE of the independent midpoints was less than 1 dB. The SE of the independent midpoints was calculated by using the SD of the independent midpoints using the formula SE = SD/sqrt n, where SD = sqrt $[\Sigma([x-]^2)/(n-1)]$, and where N = the number of independent midpoints.

An additional 150 sentences were presented at fixed SNRs, and the SRTs were used to determine the relative intelligibility of the sentences. For each participant, 50 sentences were presented at his or her eSRT, 50 sentences were presented at their eSRT +2 dB, and 50 sentences were presented at their eSRT –2 dB. The sentences assigned to each SNR were counterbalanced across participants. Logit curves were fitted for each sentence using least squares regression based on the equation: exp $(a - b * \text{SNR})/(1 + \exp(a - b * \text{SNR}))$. The dependent variable was the proportion of words correct at each SNR for a particular sentence averaged across participants. All analyses were made with Statistica 7.0.

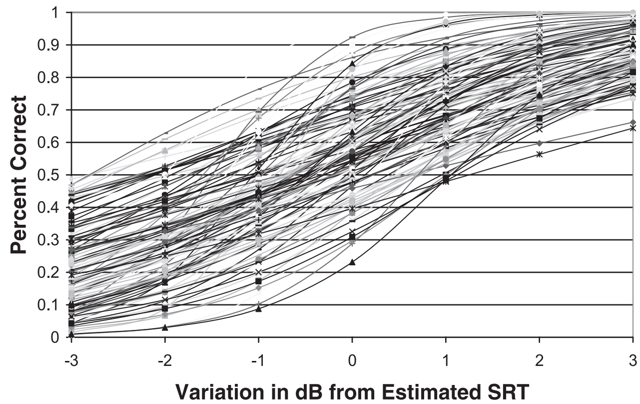

Fig. 3. Logit curves for the 120 LISN-S sentences before adjustment for equal intelligibility.

The resulting b values are related to the slope of the steepest portion of the curve for each sentence. The median b value across sentences was -0.594, or 15% per dB (calculated as $-b/4$). The ratio of a/b for any sentence (referred to as r) represents the SRT (that is, the SNR needed to achieve 50% correct identification of words in that sentence). The median value of r (r_{med}) was -0.7 dB. The sentences were adjusted in dB for equal intelligibility. The required adjustment for any sentence was calculated as r minus r_{med}. A sentence was discarded if (a) the required adjustment was too great, i.e., $r - r_{med} < -2.2$ dB or $> +2.2$ dB; (b) the slope was too shallow ($<$ 6% per dB), i.e., $-b/4 < 0.06$, or $b < -0.25$; or (c) the slope was too steep (50% per dB), i.e., $-b/4 > 0.5$, or $b > -2$.

Based on these criteria, 30 sentences were discarded. Logit curves for the remaining 120 unadjusted sentences are shown in Figure 3. The remaining sentences were adjusted in amplitude for equal intelligibility and used in the normative data study. The mean slope of the retained sentences was 17% per dB. Logit curves for the sentences after adjustment are shown in Figure 4.

EXPERIMENT 2: NORMATIVE DATA STUDY

Approval to conduct the normative data study was obtained from the relevant ethics review committees for human research.

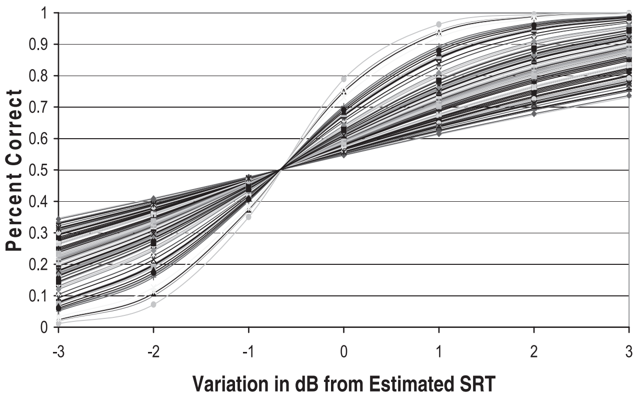

Fig. 4. Logit curves for the 120 LISN-S sentences after adjustment for equal intelligibility.

Participants

Data were collected from 82 children with normal hearing, ages 5 yr, 0 mo, to 11 yr, 11 mo, who were recruited from a local primary school. Participant details are provided in Table 1. Inclusion criteria were as per Experiment 1.

Exclusions

The results from one child (additional to the 82 children described above) were excluded from the study, based on teacher and parental reports obtained after testing had been completed that suggested that the child may have a CAPD. The performance of P28, age 9 yr and 2 mo, was notably poorer than for the other participants in his age group. No history of hearing or developmental disorders was reported on his consent form, although P28 reported during the test session that he had been having difficulty listening in the classroom. In a subsequent interview, his mother reported that P28 had only recently told her that he had been having listening difficulties. P28's peripheral hearing assessment was well within normal limits. The school principal and school counselor reported that P28 was a very bright child who had not progressed as expected at school.

P28 was well within normal limits on the LISN-S low-cue SRT and talker advantage measures. He was, however, 5.5 SD from the mean on the high-cue SRT, 6.5 SD from the mean on the spatial advantage measure, and 5 SD from the mean on the LISN-S total advantage measure. This pattern of results was typical of the children diagnosed with binaural interaction deficits on the LISN-CD CAPD study (Cameron et al. 2005, 2006c). It was decided that P28 would be excluded from the normative data results.

TABLE 1. Details of the 82 participants in the LISN-S normative data study

Age group	*N*	Male	Female	Min age (yr, mo)	Max age (yr, mo)	Mean age (yr, mo)
5	12	6	6	5, 1	5, 11	5, 5
6	12	5	7	6, 0	6, 11	6, 6
7	12	5	7	7, 1	7, 11	7, 5
8	12	3	9	8, 0	8, 11	8, 5
9	12	5	7	9, 0	9, 11	9, 5
10	12	4	8	10, 0	10, 11	10, 5
11	10	4	6	11, 1	11, 11	11, 6

Design and Procedure

The materials used in the Normative Data Study are as described for the Sentence Equivalence Study. Testing was again carried out during school hours in an Isuzu NPS300 mobile laboratory featuring an acoustically treated sound studio. The LISN-S target sentences were initially presented at a level of 62 dB SPL. The competing discourse was presented at a constant level of 55 dB SPL. The participant's task was to repeat as many words as possible heard in each sentence. The instructions provided to each participant are attached as Appendix C.

Up to 30 sentences were presented in each of four conditions of distracter location and voice: same voice at 0° (SV0°), same voice at ±90° (SV ±90°), different voices at 0° (DV0°), and different voices at ±90° (DV ±90°). The organization of the target sentences and distracter stories is provided in Table 2. The presentation order of the LISN-S conditions was counterbalanced among participants using a Latin-square protocol to enable analysis of the effect of practice on performance.

The SNR was adjusted adaptively in each condition to determine each participant's SRT. The SNR was not adjusted if a response of exactly 50% correct was recorded (for example, three of six words correctly identified). A minimum of five sentences were provided as practice, however, practice continued until one reversal in performance was recorded. Testing ceased in a particular condition when the listener had either (a) completed the entire 30 sentences in any one condition, or (b) completed the practice sentences plus a minimum of an additional 17 scored sentences, and their SE, calculated over the scored sentences, was less than 1 dB.

The formula for the SE stopping rule was modified from the usual formula on the basis of a Monte Carlo simulation of an adaptive test sequence. The simulation used an underlying assumed psychometric function comprising a logit curve with a midpoint slope of 15% per dB. Individual trials were carried out by comparing the value of the psychometric function at the current level to a number drawn randomly from a uniform distribution within

TABLE 2. Organization of target sentences and distracter stories for each LISN-S condition in the Normative Data Study

LISN-S condition	Distracter speaker	Distracter location	Distracter story
Same voice	Female 1	0°	Loopy Lizard's Tail
0°	Female 1	0°	The Great Big Tiny Traffic Jam
Same voice	Female 1	−90°	Loopy Lizard's Tail
+ and −90°	Female 1	+90°	The Great Big Tiny Traffic Jam
Different voices	Female 2	0°	Loopy Lizard's Tail
0°	Female 3	0°	The Great Big Tiny Traffic Jam
Different voices	Female 2	−90°	Loopy Lizard's Tail
+ and −90°	Female 3	+90°	The Great Big Tiny Traffic Jam

In each condition, the target sentences were spoken by Female 1 and presented at 0° azimuth (directly in front of the speaker).

the range 0 to 1. If the random number exceeded the value of the psychometric function, the test level was increased by the step size (2 dB). Otherwise the test level was decreased by the step size. The next trial was carried out with the new test level.

To simulate the possibility of the participant getting exactly 50% of items correct, an additional random number in the range 0 to 1 was generated and subtracted from the current value of the psychometric function. If the absolute value of this number was less than 0.1, the test level was left unchanged for the next trial. (This simulates no change of level in 20% of trials when the current level is near the true threshold and a decreasing proportion of "no-change" outcomes as the current level traverses below or above the true threshold.)

Thresholds were based on runs of 20 trials each, and the complete simulation comprised 100 such runs. For each run, the SE of all test levels was calculated as the SD of test levels within the run divided by the square root of 20. These SE values were averaged across the 100 runs to provide the SE estimate. For each run, the mean threshold was also determined and the SD of these means across all runs was calculated.

The simulation revealed that the reliability estimates obtained by the SE method were approximately half the reliability estimates obtained by directly calculating the SD of the mean threshold across runs. Repetition of the entire simulation 30 times resulted in the ratio of SE estimate (i.e., averaged across the 100 estimated SE values) to the true SD (i.e., calculated from the actual means of each of the 100 runs) ranging from 0.41 to 0.52 with a mean value of 0.46. The ratio was insensitive to the number of trials in the run.

The reason for this underestimate of variability is that the "samples" (i.e., the test levels for each trial) are not independent due to each test level being different from the previous one by at most one step size. Changing the parameters in the simulation confirmed this: When the step size was made very large and the probability of having no change in level on any trial was set to zero, the ratio of the SE estimate to the true SD approached 1.0. Conversely, the ratio approached zero as the step size was made smaller and smaller.

Consequently, use of a stopping rule based on the SE estimate obtained from a single run of 20 or so trials (which is all that is available in clinical practice) must include some allowance for the under-estimation revealed by the simulation. For the characteristics of the LISN-S test (a slope of approximately 15% per dB and a step size of 2 dB) the true SD of mean thresholds is approximately twice that estimated from the variability of test levels obtained within a single run. The SE stopping rule used in the LISN-S was therefore adjusted by a multiplier of two to estimate the true SE of the mean, as shown by the formula SE = 2 * SD / sqrt n, where n represents the total number of *scored* sentences.

RESULTS

LISN-S Conditions

Table 3 details the mean SRTs, inter-participant standard deviations and intra-participant SEs, in dB, as well as the number of sentences presented (excluding practice), for the four LISN-S distracter conditions: SV0°, SV ±90°, DV0°, and DV ±90°. Age groups were combined. An analysis of variance (ANOVA) of mean SRT was performed for the repeated-measures factors of distracter location (0°

TABLE 3. Average SRT (expressed as a SNR), inter-participant standard deviations, intra-participant standard errors in dB, and average number of sentences presented (excluding practice) for the 82 children, for each of the four LISN-S distracter location conditions

Condition	SRTdB	SDdB	SEdB	No. of sentences
Same voice				
0°	−0.5	1.3	0.86	19
±90	−11.9	2.8	0.92	19
Different voices				
0°	−4.1	2.1	0.96	19
±90	−13.4	2.7	0.97	20

versus ±90°) and distracter voice (same versus different). An α level of 0.05 was used for all comparisons. The Greenhouse-Geisser correction factor was applied to the degrees of freedom of the main effects and interaction to ensure that violations of sphericity did not influence the significance levels calculated for any of the analyses.

There was a significant main effect for location, $F(1,81) = 394.24$, $p < 0.001$, with the ±90° condition resulting in a lower SRT than the 0° condition, averaged across distracter voice. An analysis of simple contrasts revealed that the ±90° location produced a significantly lower mean SRT than the 0° location for both the same voice distracter, $F(1,81) = 2288.04$, $p < 0.001$, and the different voices distracter, $F(1,81) = 1762.80$, $p < 0.001$. There was also a significant main effect of voice, $F(1,81) = 3061.47$, $p < 0.001$, averaged across location. Simple contrasts revealed that the mean SRT of the same voice distracter was significantly higher than for the different voices distracter at both the 0° location, $F(1,81) = 341.26$, $p < 0.001$ and the ±90° location, $F(1,81) = 77.02$, $p < 0.001$.

The benefit from separation was influenced by the similarity of the voices of the speaker(s) of the distracters to the voice of the speaker of the target sentences. This is referred to as the spatial separation advantage, which is calculated by subtracting the SRT in the 0° location from that in the ±90° location for a particular condition of distracter voice. The interaction term in the ANOVA indicated that the spatial separation advantage in the same voice condition (11.3 dB) was significantly greater than the spatial separation advantage in the different voices conditions (9.3 dB) ($F(1,81) = 55.98$, $p < 0.001$). The spatial separation advantage for the same voice condition is referred to as the LISN-S spatial advantage measure.

Effect of Age on LISN Performance Measures

The mean SRT and advantage measures for the children ages 5 to 11 yr are illustrated in Figure 5, (a) to (e). There was a trend of decreasing SRT and increasing advantage, as age increased across measures. The inter-participant SDs of the measures ranged from 0.8 dB for the 11-yr-olds on the spatial advantage measure to 2.5 dB for the 10-yr-olds on the spatial advantage measure. The range of standard deviations for the 5-yr-olds was 1.0 dB on the low-cue SRT measure to 1.8 dB on the high-cue SRT measure.

Separate ANOVAs were performed to determine the effect of age on the performance measures. As the five measures were derived from the four basic LISN-S conditions (SV0°, SV ±90°, DV0°, and DV ±90°), the α level of 0.05 was multiplied by 4/5 to give an adjusted level of 0.04 to avoid inflating the Type I error rate. The most significant differences in performance were observed between the 5-yr-olds and the other age groups.

For the low-cue SRT there was a significant main effect of age ($F(6,75) = 11.46$, $p < 0.001$). Post hoc tests using Tukey's HSD revealed that the 5-yr-olds had a significantly higher (less favorable) SRT than the 7-yr-olds ($p = 0.015$), 8-yr-olds ($p = 0.007$), and 9, 10, and 11-yr-olds ($p < 0.001$). The 6-yr-olds had a significantly higher SRT than the 9-yr-olds ($p = 0.010$); 10-yr-olds ($p = 0.033$), and 11-yr-olds ($p < 0.001$). The 7-yr-olds had a significantly higher SRT than the 11-yr-olds ($p = 0.010$), as did the 8-yr-olds ($p = 0.021$). No differences in SRTs were significant between other combinations of age groups.

There was also a significant main effect of age for the high-cue SRT ($F(6,75) = 27.35$, $p < 0.001$). Post hoc tests revealed that the 5-yr-olds required a significantly higher SRT than all other age groups ($p < 0.001$). The 6-yr-olds also had a less favorable SRT than children ages 8 ($p = 0.013$), 9 ($p = 0.002$), and 10 and 11 ($p < 0.001$). The 7-yr-olds required a higher SRT than the 9, 10, and 11-yr-olds ($p = 0.030$; 0.002, and <0.001, respectively). No differences in thresholds were significant between other combinations of age groups.

Interestingly, there was no significant main effect of age for the talker advantage measure ($F(6,75) = 1.367$, $p = 0.239$). There was, however, a significant main effect of age for the spatial advantage measure ($F(6,75) = 8.77$, $p < 0.001$). The 5-yr-olds required a significantly higher SRT than all other age groups except the 6-yr-olds ($p = 0.006$ for 7-yr-olds, <0.001 for 8-yr-olds, 0.003 for 9-yr-olds, and <0.001 for 10- and 11-yr-olds). The 6-yr-olds also needed a higher SRT than the 8-yr-olds ($p = 0.011$) and 11-yr-olds ($p = 0.005$). No differences in thresholds were significant between other combinations of age groups.

Finally, there was a significant main effect of age for the total advantage measure, $F(6,75) = 11.70$, $p < 0.001$. Again, post hoc tests revealed that the 5-yr-olds required a significantly higher SRT than all other age groups ($p = 0.019$ for 6-yr-olds, 0.006 for 7-yr-olds and <0.001 for all other age groups). The 6-yr-olds also needed a higher SRT than the 10-yr-olds ($p = 0.024$) and 11-yr-olds ($p = 0.036$). No other differences in total advantage were significant among the children.

Gender Effects

An analysis was conducted to investigate gender effects in the children. Mean scores and SDs for the 50 girls and 32 boys on the various LISN-S SRT and

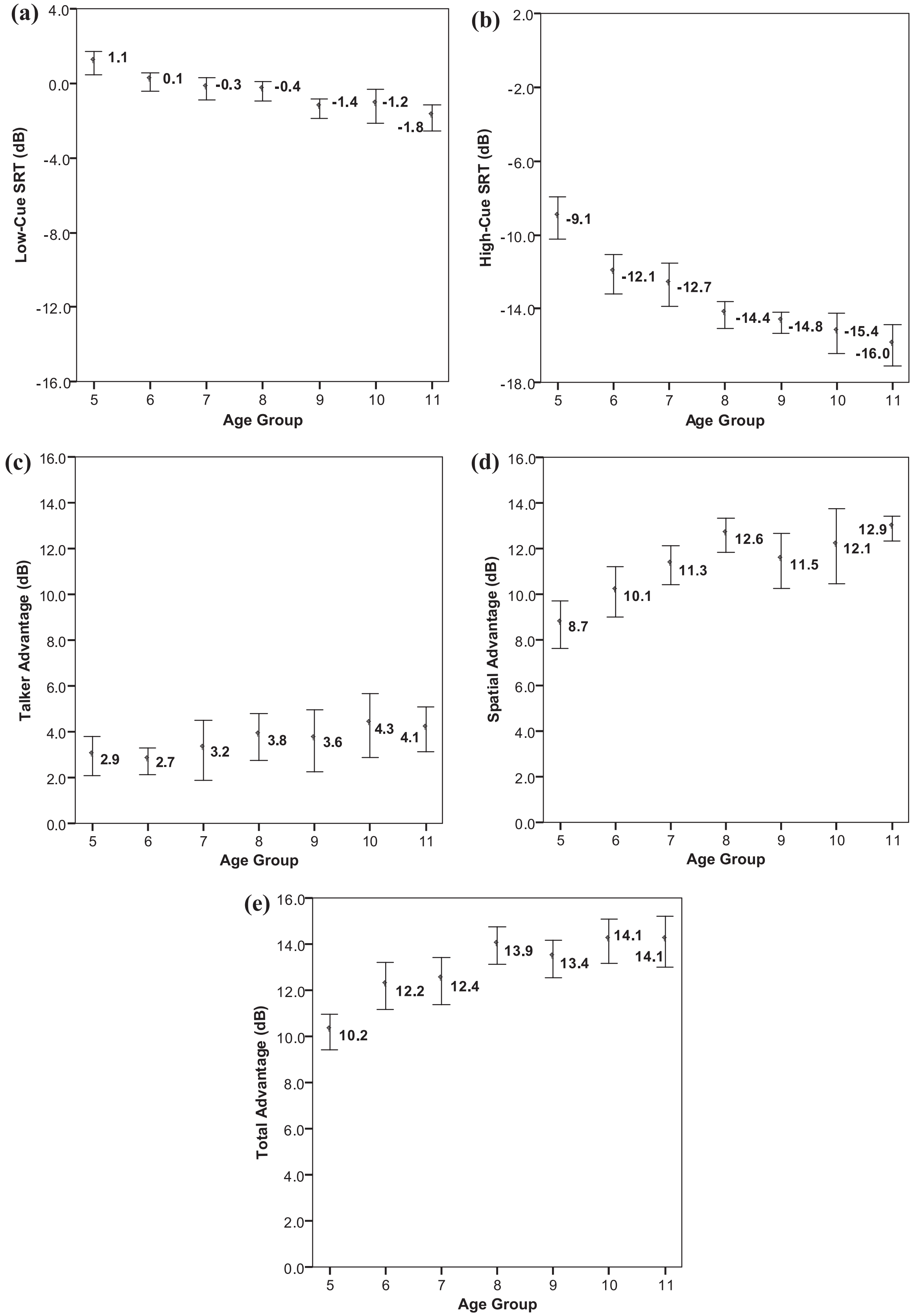

Fig. 5. Results on the various LISN-S SRT and advantage measures for children ages 5 yr, 0 mo, to 11 yr, 11 mo. (a), low-cue SRT; (b), high-cue SRT; (c), talker advantage; (d), spatial advantage; (e), total advantage. Error bars represent 95% confidence intervals from the mean.

TABLE 4. Mean SRT and advantage measure scores for 32 boys and 50 girls ages 5 to 11 yr, together with results of ANOVA investigating effects of gender on LISN-S performance, with age as a covariate

	Male		Female			
Measure	Mean dB	SD dB	Mean dB	SD dB	F(1,79)	*p*
Low-cue SRT	−0.6	1.5	−0.5	1.2	0.38	0.541
High-cue SRT	−13.2	2.9	−13.6	2.5	0.03	0.856
Talker advantage	3.4	1.7	3.5	1.8	0.01	0.911
Spatial advantage	11.0	2.4	11.4	1.9	0.38	0.541
Total advantage	12.7	2.1	13.0	1.8	0.36	0.553

advantage measures are provided in Table 4, along with the results of ANOVAs that were performed with each measure as the dependant variable, a fixed factor of gender, and age as a covariate. There was no significant effect of gender for any LISN measure.

Practice Effects

The effect of practice on performance on the LISN-S was examined for the 5- to 11-yr-olds. To determine whether practice improved performance the mean SRTs were compared for the four basic LISN-S conditions (SV0°, SV ±90°, DV0°, and DV ±90°) as a function of presentation order (first, second, third, or fourth). Participants in each age group completed the various conditions in exactly the same order. Age groups were combined to provide sufficient numbers in each condition and task combination to calculate meaningful inferential statistics. Table 5 shows the mean thresholds in dB for each LISN-S condition as a function of presentation order.

One-way ANOVAs revealed no significant differences in mean SRTs as a factor of presentation order for either the SV0° condition [$F(3,78) = 1.24$, $p = 0.302$]; the SV ±90° [$F(3,78) = 0.85$, $p = 0.469$]; or the DV ±90° condition [$F(3,78) = 0.04$, $p = 0.990$]. There was, however, a significant difference for the DV 0° condition [$F(3,78) = 5.80$, $p = 0.001$]. Post hoc tests using Tukey's HSD revealed that at –2.9 dB, the first presentation resulted in a significantly higher SRT than the third and fourth presentations at −4.6 dB ($p = 0.026$) and –5.2 dB ($p = 0.001$), respectively. However, there was no significant difference in SRT between the second presentation (at –3.7 dB) and any other subsequent presentation.

Standard Error and Time Analysis and Distribution of Data

As discussed in the Methods section, testing in any one LISN-S condition was terminated once a participant had completed 30 sentences, or once 17 sentences had been completed (plus at least five practice sentences including one reversal) and his or her SE was less than 1 dB. In the present study, the median SE ranged from 0.86 dB in the SV0° condition to 0.98 dB in the DV ±90° condition, with a range of 0.59 to 1.95 dB across all age groups. Normal probability-probability plots revealed that the data followed a normal distribution for all SRT and advantage measures. The median time taken to complete any LISN-S condition was 3 minutes and 4 sec (mean 3 minutes, 11 sec). Total time taken to complete the testing was on average approximately 12 minutes, plus 5 minutes for instructions and breaks. An average of 19 sentences was presented in each of the four LISN-S conditions, with an average of five sentences provided as practice.

Regression Analysis and LISN-S Cut-off Scores

As a strong trend of improved performance with increasing age was found for the various LISN-S SRT and advantage measures, it was determined that cut-off scores, calculated as 2 SD below the mean, would need to be adjusted for age for each performance measure. These cut-off scores represent the level below which performance on the LISN-S is considered to be outside normal limits. The data from the 5-yr-olds were therefore not included in the calculation of cut-off scores. As noted

TABLE 5. Mean SRTs and standard deviations of each LISN-S condition, as a function of presentation order for the 82 children ages 5 to 11 yr

	First presentation		Second presentation		Third presentation		Fourth presentation	
Condition	SRT dB	SD dB	SRT dB	SD dB	SRT dB	SD dB	SRT dB	SD dB
Same voice								
0°	−0.3	1.4	−0.2	1.3	−0.7	1.4	−0.9	1.2
±90	−11.5	2.2	−11.3	3.0	−12.4	3.2	−12.2	2.5
Different voices								
0°	−2.9	1.4	−3.7	1.7	−4.6	2.3	−5.2	2.0
±90	−13.4	2.9	−13.3	2.2	−13.6	2.3	−13.4	3.3

Age groups are combined.

TABLE 6. Data used in calculation of LISN cut-off scores for the 70 children ages 6 to 11 yr in the LISN-S normative data study

Measure	Mean dB	SD (Residuals) dB	Intercept	B-Value	r^2
Low-cue SRT	−0.8	0.98	2.50	−0.37	0.294
High-cue SRT	−14.2	1.50	−7.17	−0.78	0.442
Talker advantage	3.6	1.73	1.03	0.29	0.073
Spatial advantage	11.7	1.73	7.67	0.45	0.164
Total advantage	13.3	1.47	9.70	0.41	0.181

All r^2 values are significant at $p < 0.05$.

in the section on the effect of age on SRT and advantage measures, there were significant differences between the 5-yr-olds and most other age groups on nearly all measures except talker advantage, suggesting that there is a strong developmental curve in the ability to use binaural cues to understand speech in background noise.

A regression analysis was conducted with SRT for each measure as the independent variable and age (ranging from 6.00 to 11.92 yr) as the dependent variable. The cut-off scores were adjusted for age using the formula: cut-off score = intercept + (B-value * age) – (2 * SD of residuals from the age-corrected trend lines). All regression data are presented in Table 6. Figure 6 provides scatterplots of the regression analysis showing the individual data points.

DISCUSSION

The results of the normative data study showed that there was a strong trend of decreasing SRT, and increasing advantage, as age increased across all LISN-S performance measures. The 5-yr-olds exhibited significant differences in performance compared with nearly all other age groups on all performance measures except the "talker advantage" measure. This may suggest that the 5-yr-olds are able to use talker cues in the auditory environment to suppress noise and attend to a target. However, their ability to use binaural interaction skills to detect, perceive, and compare small differences in the arrival time and intensity of signals reaching both ears to locate the source of a sound presented in background noise may not be as developed as for other age groups. Inter- and intra-participant variation of the LISN-S was minimal, even in the 5-yr-old group, suggesting that it will be a suitable test for assessing children's ability to use talker and spatial cues to understand speech in background noise.

In respect to practice effects, as the first presentation of the LISN-S DV0° condition resulted in a SRT that was significantly higher than for the third and fourth presentations, it is suggested that the DV0° condition should not be presented first. It is recommended that presentation order be (1) DV ±90°; (2) SV ± 90°; (3) DV0°; and (4) SV0°. This configuration represents a gradient from "easy" to "difficult" and controls for the practice effects that were demonstrated for the DV0° condition.

It must be acknowledged that only a quarter of the participants in the current study received the LISN-S in the recommended order. To determine the effect of change of presentation order on performance, a regression analysis was conducted with presentation order (1, 2, 3, or 4) as the dependent variable and SRT in each LISN-S condition as the independent variable. The adjustment to the normative data needed to test in the recommended order was calculated as the number of steps away from the order midpoint (2.5), multiplied by the B-value (dB/step). As the calculated adjustments ranged from 0 dB in the high-cue SRT condition (DV ±90°) to only 0.24 dB in the talker advantage condition, it was considered that adjusting for a change in presentation order in the calculation of cut-off scores from the normative data was unnecessary.

As hypothesized, whereas there are some differences in the precise results, the similarities between the LISN-S and LISN-CD (Cameron et al., 2006b) are marked. Both tests show that the spatial advantage is greater than the talker advantage, and that total advantage is not much larger than the spatial advantage.

Whereas it was a goal of this project to develop an assessment tool that could be used with children as young as 5 yr, and whereas the inter- and intra-participant variation on the various LISN-S performance measures in this group are minimal, it is the belief of the authors that it may be prudent to use the LISN-S only with children ages from 6 yr, 0 mo. This age group has generally had more experience with assessment tasks due to their previous school experiences. Many of the children younger than 6 yr of age required a great deal of organization and frequent breaks to avoid fatigue during testing, whereas the other age groups exhibited no difficulty with the task.

Total time taken to administer the test stimuli was on average approximately 12 minutes, with about 5 minutes required for instructions and breaks. The use of the SE stopping rule contributes to the time-effective nature of test presentation without compromising the reliability of the data. These factors, together with the fact that the adaptive adjustment of SNR is automatically performed by the software and is effected simply by the clinician entering the number of words correct in each sentence, makes the LISN-S a practical tool for clinical assessment purposes. To ensure ease of

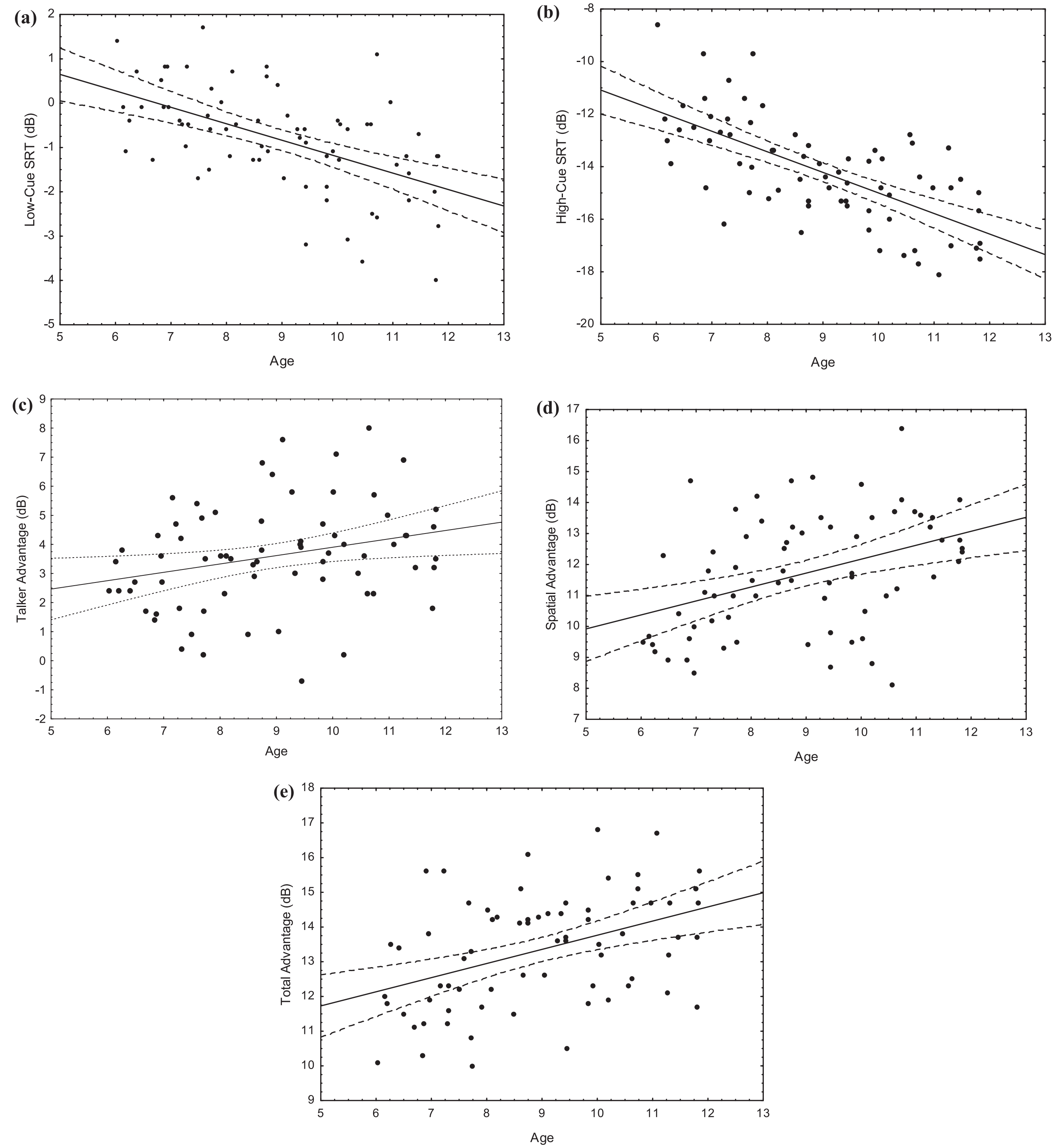

Fig. 6. Scatterplots of linear regression of LISN-S SRT and advantage measures on age for children ages 6 yr, 0 mo, to 11 yr, 11 mo. (a), low-cue SRT; (b), high-cue SRT; (c), talker advantage; (d), spatial advantage; (e), total advantage. Prediction lines represent 95% intervals from the mean.

calibration a simple calibration device has been developed specifically for use with the LISN-S.

Conclusion and Future Work

It is concluded that the LISN-S is a fast and efficient assessment tool that may be used clinically to evaluate the various processes children use to understand speech in background noise, including binaural interaction ability. Future studies are planned to collect data from children with suspected CAPD on the LISN-S, as well as CAPD test battery consisting of a dichotic task, a gap detection test, a masking level difference task, and a temporal sequencing test, to measure the proportion of children with suspected CAPD who have spatial and/or different talker detection deficits. A test-retest reliability study is also planned to determine the amount of both measurement error and systematic change in performance between two trials of the LISN-S. The results will also be utilized in studies to assess the effectiveness of the *LISN & Learn* training tool, which is currently being developed to remediate binaural interaction deficits.

APPENDIX A

Practice Sentences: LISN-S Same Voice 0° Condition

1. The boys are watching the game.
2. A dog is hiding the bone.
3. Two girls went to the shops.
4. A painting hangs on the wall.
5. Some people go to the gym.

APPENDIX B

Extract From LISN-S Distracter Continuous Discourse Presented at 0° and –90°

"Loopy Lizard's Tail" by Vashti Farrer

Loopy lizard was on his way home. His mother had told him not to dawdle, but Loopy wanted to play.

So he ran across a path and through the grass, where he pretended to hide.

Then he scampered up a wall and peered into a crack to see who lived there.

Then Loopy stopped very still in the sun, as if he were asleep, just to feel how warm it was.

Suddenly a big, fierce dog came down the path.

"He must be fierce," thought Loopy, "because he has big teeth."

And he started to run as fast as he could along the path, to get out of the dog's way. But the big dog chased him.

APPENDIX C

LISN-S Instructions

1. You are going to hear some sentences over these headphones.
2. The sentences are said by a lady called "Miss Smith."
3. Miss Smith will sound as if she is standing just in front of you.
4. There will be a "beep" before each sentence so you will know when it is about to start.
5. Your job is to repeat back the sentence that Miss Smith says.
6. I'll pretend to be Miss Smith, and I want you to repeat the sentence you hear.
7. "Beep." "The dog had a bone."
8. Child repeats "The dog had a bone."
9. Good, that's easy, isn't it? But there's a trick. At the same time that Miss Smith is telling you the sentence there are some very tricky people talking at the same time.
10. Sometimes the tricky people sound like they are standing right next to Miss Smith, sometimes they will sound like they are standing next to you.
11. No matter where the tricky people are, I don't want you to listen to them.
12. Just listen for the "beep" and the sentence.
13. Miss Smith always starts out louder than the tricky people, so you shouldn't have any trouble hearing her.
14. But sometime the tricky people get loud. If you only hear a bit of the sentence I want you to tell me all the words that you hear.
15. So if you just heard "dog" and "bone," what would you say?
16. Child repeats "dog" and "bone."
17. Great. If you don't hear Miss Smith at all, just shake your head and I'll go straight on to the next sentence.
18. Once you've heard the sentence tell me what you've heard straight away so you don't forget it.
19. In the first lot of sentences, the tricky people will be standing right next to you. Don't listen to them. Just concentrate on Miss Smith in front.
20. The tricky people start first and then Miss Smith starts a few seconds later. Ready?
21. Describe where the tricky people are before each listening condition.
    a. "Same Voice-±90°" Condition: "Now the tricky people will be next to you again, but their voices will be a bit different. Ignore them and just listen for Miss Smith."
    b. "Different Voices-0°" Condition: "Now the tricky people will be next to Miss Smith. Just listen for the beep and the sentence."
    c. "Same Voice-0°" Condition: "Now the tricky people will be next to Miss Smith, and their voices will be very similar to Miss Smith's voice. So you will have to listen very hard for the "beep" and Miss Smith."

ACKNOWLEDGMENTS

The authors would like to thank the Speech, Hearing, and Language Research Centre (SHLRC) of Macquarie University for their assistance in the recording of the LISN-S speech stimuli and for use of the mobile laboratory for data collection. The distracter children's stories used in this project were provided by Ms. Vashti Farrer. The contribution of Stephen Cameron in the production of the LISN-S software program is also sincerely appreciated. Finally, the authors would like to thank the Catholic Schools Office and the New South Wales Department of Education and Training for allowing us to conduct the LISN-S studies in their schools. This project was funded in part by the National Health and Medical Research Council (NHMRC) of Australia. Parts of this paper were presented at the "Audiology Now" Conference, Minneapolis, Minnesota, April 7, 2006.

Address for correspondence: Dr. Sharon Cameron, National Acoustic Laboratories, 126 Greville Street, Chatswood, NSW, 2067, Australia. E-mail: Sharon.Cameron@nal.gov.au.

Received March 22, 2006; accepted October 16, 2006.

References

Bamford, J., Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In: Bench, J., Bamford, J. eds. Speech-hearing tests and the spoken language of hearing-impaired children. London, UK: Academic Press, 146–187.

Bess, F. H. (1983). Speech Recognition. In: Konkle, D. F., and Rintelmann, W. F., eds. Principles of speech audiometry. Baltimore, MD: University Park Press, 127–201.

Bode, D. L., Carhart, R. (1974). Stability and accuracy of adaptive tests of speech discrimination. *Journal of the Acoustical Society of America*, *56*, 963–970.

Brand, T., Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *Journal of the Acoustical Society of America*, *111*, 2801–2810.

Cameron, S., Dillon, H., Newall, P. (2005). Three case studies of children with suspected auditory processing disorder. *Australian and New Zealand Journal of Audiology*, *27*, 97–112.

Cameron, S., Dillon, H., Newall, P. (2006a). Development and evaluation of the Listening In Spatialized Noise Test. *Ear and Hearing*, *27*, 30–42.

Cameron, S., Dillon, H., Newall, P. (2006b). Listening in Spatialized Noise Test: Normative data for children. *International Journal of Audiology*, *45*, 99–108.

Cameron, S., Dillon, H., Newall, P. (2006c). The Listening in Spatialized Noise Test: An auditory processing disorder study. *Journal of the American Academy of Audiology*, *17*, 304–318.

Chae, S., Kang, U., Jeon, E., Linacre, J. M. (2000). Development of Computerized Middle School Achievement Test. Seoul, South Korea: Komesa Press.

Cook, K. F., Roddey, T. S., O'Malley, K. J., Gartsman, G. M. (2005). Development of a flexilevel scale for use with computer-adaptive testing for assessing shoulder function. *Journal of Shoulder and Elbow Surgery 14*, 90–94.

Cox, R. M., Gray, G. A., Alexander, G. C. (2001). Evaluation of a Revised Speech in Noise (RSIN) test. *Journal of the American Academy of Audiology*, *12*, 423–432.

Crystal, D. (1989). Grammatical analysis of language disability. London, UK: Cole and Whurr Limited.

Dijkers, M. P. (2003). A computer adaptive testing simulation applied to the FIM instrument motor component. *Archives of Physical Medicine and Rehabilitation*, *84*, 384–393.

Hagerman, B. (1976). Reliability in the determination of speech discrimination. *Scandinavian Audiology*, *5*, 219–228.

Jamieson, D. G., Kranjc, G., Yu, K., Hodgetts W. E. (2004). Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology*, *15*, 508–517.

Jerger, J. (1998). Controversial issues in central auditory processing disorders. *Seminars in Hearing*, *19*, 393–398.

Jerger, S., Allen, J. (1998) How global behavioural tests of central auditory processing may complicate management. In: Bess F, ed. *Children with hearing impairment.* Nashville, TN: Bill Wilkerson Centre Press, 163–177.

Kowal, A. (1979). Sentence list construction and pilot test. In: Bench, J., and Bamford, J. eds. *Speech-hearing tests and the spoken language of hearing-impaired children.* London, UK: Academic Press, 110–145.

Lilley, M., Barker, T., Britton, C. (2004). The development and evaluation of a software prototype for computer-adaptive testing. *Computers & Education*, *43*, 109–123.

Lunz, M. E., Bergstrom, B. A., Gershon, R. C. (1994). Computer Adaptive Testing. *International Journal of Educational Research*, *21*, 623–634.

Mackie, K., Dermody, P. (1986). Use of a monosyllabic adaptive speech test (MAST) with young children. *Journal of Speech and Hearing Research*, *29*, 275–281.

MathWorks, Inc. (2002a). MATLAB (Version 6.5.6) [Computer software]. Natick, MA: The MathWorks, Inc.

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

Picard, M., Bradley, J. S. (2001). Revisiting speech interference in classrooms. *Audiology*, *40*, 221–244.

Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., Soli, S. D., Giguère, C. (2005). Adaptation of the HINT (hearing in noise test) for adult 25 Canadian Francophone populations. *International Journal of Audiology*, *44*, 358–369.

Young, R., Shermis, M. D., Brutten, S. R., Perkins, K. (1996). From conventional to computer-adaptive testing for ESL reading comprehension. *System*, *24*, 23–40.

Reference Note

1 Cameron, S., Dillon, H. (2006). Listening in Spatialized Noise Test (LISN) – Sentences (Version 1.0.0) [Computer software]. Sydney, NSW: National Acoustic Laboratories.