Perception of Voiceless Fricatives by Normal-Hearing and Hearing-Impaired Children and Adults

Andrea L. Pittman
Patricia G. Stelmachowicz
Boys Town National Research Hospital
Omaha, NE

This study examined the perceptual-weighting strategies and performance-audibility functions of 11 moderately hearing-impaired (HI) children, 11 age-matched normal-hearing (NH) children, 11 moderately HI adults, and 11 NH adults. The purpose was to (a) determine the perceptual-weighting strategies of HI children relative to the other groups and (b) determine the audibility required by each group to achieve a criterion level of performance. Stimuli were 4 nonsense syllables (/us/, /uʃ/, /uf/, and /uθ/). The vowel, transition, and fricative segments of each nonsense syllable were identified along the temporal domain, and each segment was amplified randomly within each syllable during presentation. Point-biserial correlation coefficients were calculated using the amplitude variation of each segment and the correct and incorrect responses for the corresponding syllable. Results showed that for /us/ and /uʃ/, all four groups heavily weighted the fricative segments during perception, whereas the vowel and transition segments received little or no weight. For /uf/, relatively low weights were given to each segment by all four groups. For /uθ/, the NH children and adults weighted the transition segment more so than the vowel and fricative segments, whereas the HI children and adults weighted all three segments equally low. Performance-audibility functions of the fricative segments of /us/ and /uʃ/ were constructed for each group. In general, maximum performance for each group was reached at lower audibility levels for /s/ than for /ʃ/ and steeper functions were observed for the HI groups relative to the NH groups. A decision theory approach was used to confirm the audibility required by each group to achieve a ≥90% level of performance. Results showed both hearing sensitivity and age effects. The HI listeners required lower levels of audibility than the NH listeners to achieve similar levels of performance. Likewise, the adult listeners required lower levels of audibility than the children, although this difference was more substantial for the NH listeners than for the HI listeners.

KEY WORDS: speech perception, adults, children hearing loss, normal hearing

A number of studies have shown that the perception of speech develops throughout childhood and adolescence for individuals with normal hearing (NH). The age at which children reach adult-like performance appears to be a function of the complexity of the stimulus. For example, temporal and spectral discrimination reaches adult-like performance between 4 and 5 years of age (Allen & Wightman, 1992; Allen, Wightman, Kistler, & Dolan, 1989; Wightman, Allen, Dolan, Kistler, & Jamieson, 1989), perception of nonsense syllables around 7 years of age (Hnath-Chisolm, Laipply, & Boothroyd, 1998), identification of high-frequency fricatives as a function of bandwidth around 10 years of age (Kortekaas and Stelmachowicz, 2000), and perception of high- and low-predictability sentences between 15 and 17 years of age (Elliott, 1979).

It has been suggested that the strategies used during speech perception also may change as a function of age. Four- to 5-year-old NH children attended to dynamically changing vocalic-transition segments when identifying the consonant-vowel nonsense syllables /sV/ and /ʃV/, whereas older NH children (7 years of age) and adults attended to static fricative segments (Nittrouer & Crowther, 1998; Nittrouer, Crowther, & Miller, 1998; Nittrouer & Miller, 1997a, 1997b). The authors concluded that young children and adults use different acoustic cues to achieve the same perceptual goals and that children generally learn to weight (or attend to) those segments in an adult-like manner as they gain experience with speech.

Carney and Moeller (1998) have suggested that sensorineural hearing loss, among other factors (e.g., recurrent otitis media, poor speech, and language role models in the home), may reduce both the quality and quantity of speech and language experiences. As a result, children with moderate to severe sensorineural hearing loss may use listening strategies that differ from those of NH children. The recent results of Stelmachowicz, Hoover, Lewis, Kortekaas, and Pittman (2000), however, suggest otherwise. They examined the effects of stimulus context and audibility on sentence recognition in NH adults and children (ages 5, 6, 8, and 10 years of age) and in children with sensorineural hearing loss (ages 5 to 12 years). Listeners were asked to repeat four-word sentences from two lists that differed in semantic context. The results from the majority of the HI children fell within the range of the NH children, regardless of age when hearing loss was compensated for by providing equivalent audibility. Thus, the NH and HI children seemed to function similarly with regard to the benefits of semantic context. At lower levels of speech processing, however, it remains possible that hearing loss in early childhood may alter the normal perceptual-weighting strategies described by Nittrouer et al. (1997a, 1997b).

In the present study, the voiceless fricatives /s/, /ʃ/, /f/, and /θ/ were selected for identification in view of the difficulty HI listeners experience when identifying these phonemes (Dubno & Levitt, 1981), as well as the relative importance of each phoneme to speech and language comprehension. The fricatives were produced in a vowel-consonant context with the vowel /u/. The productions of three talkers were included because children are known to experience greater difficulty than adults when perceiving speech from multiple talkers (Ryalls & Pisoni, 1997) and because multiple talkers better represent the variety of individuals children must perceive in many listening situations (e.g., teachers, parents, siblings, and peers). The vowel, transition, and fricative segments of each syllable were identified so that the amplitude of each could be varied randomly within each syllable during the identification task. Perceptual weights were determined using point-biserial correlation coefficients calculated from the correct and incorrect responses of each listener and the amplitude variations of the three segments of each syllable. This method has been used previously with speech stimuli segmented into bands along the frequency domain to determine the spectral weighting strategies of both NH and HI adults (Doherty & Turner, 1996; Turner, Kwon, Tanaka, Knapp, Hubbartt, & Doherty, 1998). In the present study, the stimuli were segmented along the temporal domain to determine the relative weights associated with three specific segments of each syllable. One advantage to using a correlational analysis in this manner is that the contribution of each segment to perception can be determined without having to present a segment in isolation. Zeng & Turner (1990) found that by removing the fricative segment of a nonsense syllable and presenting the vowel and transition segments only, the fricatives /s/, /f/, /θ/, and /ʃ/ were often perceived as the stop consonants /d/, /b/, or /g/.

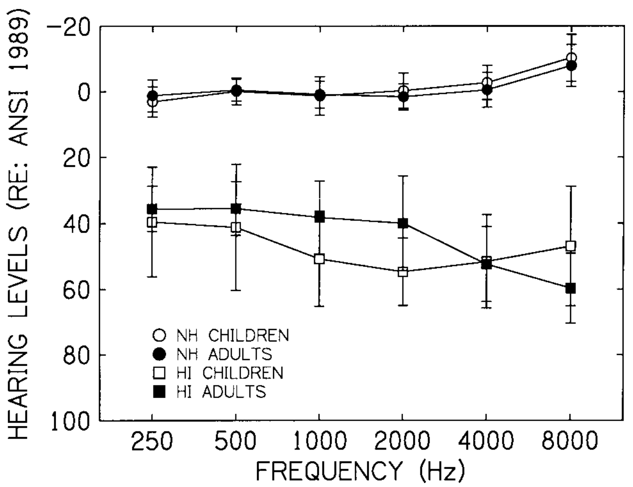
The purpose of the present study was twofold. First, the perceptual-weighting strategies of HI children were examined in relation to those of age-matched NH children, as well as NH and HI adults. Second, the audibility required to achieve a criterion level of performance was determined for the segment of each syllable receiving the highest weights by each group. If the development of speech perception is influenced negatively by hearing loss, then the perceptual-weighting strategies and performance-audibility functions of HI children could differ from those of (a) age-matched NH children whose perceptual skills are likely more developed due to increased exposure to speech and language, (b) HI adults who likely have adjusted their perceptual strategies to accommodate a hearing loss, and (c) NH adults who likely possess the most efficient perceptual skills.

Method

Listeners

A total of 44 listeners participated in this study. The two NH groups comprised 11 adults (22–42 years, *M* = 32 years, 5 months) and 11 children (8–11 years, *M* = 10 years, 1 month). The two HI groups consisted of 11 adults (54–69 years, *M* = 61 years, 11 months) and 11 children (8–11 years, *M* = 10 years, 3 months). NH listeners had pure-tone hearing levels ≤15 dB at octave frequencies from 0.25 through 8.0 kHz. Hearing-impaired listeners displayed relatively flat moderate sensorineural hearing losses. Figure 1 shows mean pure-tone thresholds (±1 standard deviation) for each group. Based upon audiological data and subjective history,

Figure 1. Mean hearing thresholds (in dB HL) as a function of frequency for the 11 NH children, 11 NH adults, 11 HI children, and 11 HI adults. Error bars represent ±1 *SD* from the mean.



all of the HI children had congenital losses, and all of the adults had acquired hearing loss later in life.

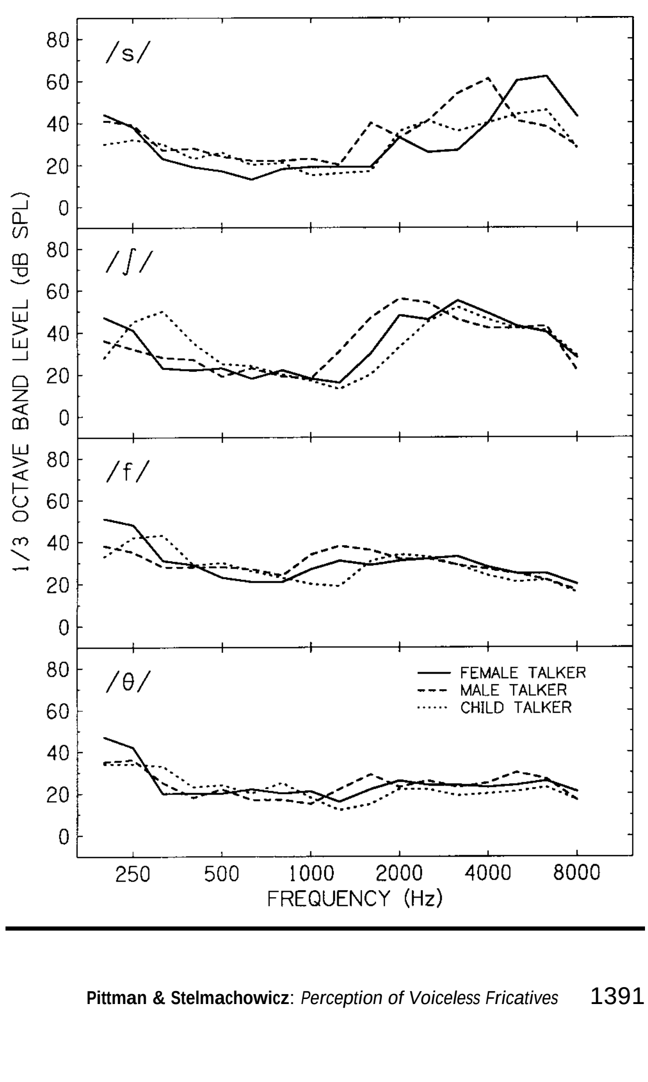
Stimuli

Several repetitions of the nonsense syllables /us/, /uʃ/, /uf/, and /uθ/ were recorded by three talkers: an adult male, an adult female, and an 11-year-old child. All recordings were made in a sound-treated room at normal vocal levels using a microphone with a flat frequency response to 10 kHz (AKG Acoustics, C 535 EB). The speech samples were amplified (Shure, M267), filtered at 9.5 kHz (Ithaco, 4302), and digitized at a sampling rate of 20 kHz. A total of twelve stimuli were chosen from these recordings by subjectively selecting the clearest sample of each nonsense syllable produced by each talker. Vowel, transition, and fricative segments were identified using a digital editing program (Cool Edit 96). The transition segment was determined by measuring the second formant peak frequency from the beginning of the fricative backward to the point where no further change in formant frequency was observed. All 12 nonsense syllables had measurable transitions with frequency changes of 43 to 769 Hz. The vowel and fricative were measured from the temporal boundaries of the transition to the beginning and end of each utterance, respectively. Average duration of each segment across talkers was 176.0 ms for the vowel (range: 143 to 250 ms), 79.5 ms for the transition (range: 44 to 139 ms), and 208.8 ms for the fricative (range: 163 to 267).[1]

Figure 2 shows the acoustic spectra of the fricative segments of each nonsense syllable by talker. The fricative-segment amplitude spectra were analyzed in 1/3-octave bands with a 20-ms Hanning window (50% overlap) after the syllables were transduced by an insert earphone (Etymotic, ER-2) and recorded in a 2-cm^3 coupler at 60 dB sound pressure level (SPL). The bandwidth for the fricative spectrum of /ʃ/ was nearly twice that of /s/. Fricative energy for /s/ peaked at 4 kHz for the male talker and at 6 kHz for the female and child talkers. Peak amplitude was lower for the child talker than for the male and female talkers. Lower peak frequencies for the fricative /ʃ/ were produced by the male talker (2 kHz) than for the female and child talkers (~3 kHz) with similar amplitudes for each talker. The fricatives /f/ and /θ/ were low amplitude stimuli with broad bandwidths and no apparent peak.

For each stimulus presentation, a computer program (MATLAB, Mathworks) was used to vary randomly the relative amplitude of each segment. Specifically, one nonsense syllable was presented in entirety per trial,

Figure 2. One-third octave band spectra (in dB SPL) for the fricative segments of /us/, /uʃ/, /uf/, and /uθ/ produced by the female (solid line), male (dashed line), and child talkers (dotted line).



[1] To prevent the short duration of three transition segments from being a cue to perception, the vowel-transition boundary was moved into the steady-state portion of the vowel segment to make the durations similar to others produced by the same talker.

but the relative amplitudes of the vowel, transition, and fricative segments within the syllable were varied randomly. These random amplitude variations provided a method for determining perceptual-weighting strategies, as well as the relation between performance and audibility. To reduce spectral splatter at the segment boundaries due to potential discontinuities in amplitude, a moving-average filter was applied to the temporal amplification profile to obtain smooth changes in level between segments. That is, the stimulus envelope was smoothed by a low-pass filter with a cut-off frequency of 100 Hz. The sampling rate of each stimulus also was varied randomly by ±10 Hz (in 1-Hz increments) during each presentation to minimize cues from small pitch variations within and across the three talkers.

Procedure

Listeners were seated in a sound-treated room and asked to identify each nonsense syllable presented monaurally through an insert earphone with a flat frequency response to 10 kHz (Etymotic, ER-2). For each trial, one of the 12 nonsense syllables was selected randomly. Each listener was instructed to ignore the gender of the talker and indicate which nonsense syllable was presented by selecting the appropriate number on a computer keyboard. Responses were entered by the tester for three children who were unwilling to use the keyboard. After each response, listeners were given visual feedback on the computer screen by displaying "Correct" or "Incorrect" for the adults, and "Right!" or "Try Again" for the children. The difficulty of the task and the number of possible stimuli (3 talkers × 4 nonsense syllables × 3 segments × 6 amplification levels = 216) limited the possibility of learning through feedback. The random amplification procedure also prevented the listeners from dichotomizing their responses as either /uf/ or /uθ/, if the fricative level was low, and /us/ and /uʃ/, if the fricative level was high relative to the vowel. Listeners were tested in six blocks of 200 stimuli each for a total of 1,200 presentations (100 stimuli per nonsense syllable and speech segment). Each listener completed the six blocks in about 2 hours. To reduce fatigue, children were tested in pairs; while one child completed a block of stimuli (about 15 minutes), the other was occupied with computer games, reading, coloring, etc.

The test parameters for each listener (presentation level, amplification range, and step size) were determined prior to testing. An automated program was used to calculate the short-term audibility (STA) values (discussed in the next section) for each speech segment and listener. Because the fricative segments of /us/ and /uθ/ represented the highest and lowest amplitude levels, respectively, for each talker and speech segment (including the vowel and transition segments), the test parameters were adjusted until a range of STA values were found that would likely result in a distribution of data points in the changing portion of each listener's performance-STA function. This helped to reduce the number of data points falling at floor or ceiling values, although the results of a few listeners fell at very high or very low performance levels regardless of these adjustments. For the NH listeners, 10-dB steps in relative segment level were used over a range of 60 and 50 dB for the adults and children, respectively. Due to the reduced dynamic ranges for the HI listeners, 5-dB steps in relative segment level were used over a range of 25 dB.

Short-Term Audibility (STA)

To express audibility with as much precision as possible, hearing thresholds were converted to dB SPL, and interpolation was used to estimate thresholds at 16 frequencies between 0.25 and 8.0 kHz using the formula

$$\theta(f) = \frac{\theta_{f_U} - \theta_{f_L}}{\log(f_U - f_L)} + \log\frac{f}{f_L} + \theta_{f_L}$$

where θ indicates hearing threshold in dB SPL, and f_U and f_L are the upper and lower frequency bands adjacent to the threshold being interpolated in Hz.

To calculate the STA of each of the three speech segments (vowel, transition, and fricative), each was first recorded in a 2-cm^3 coupler at 100 dB SPL. The amplitude of the vowel segments were reduced to prevent clipping and the fricative segments were amplified to bring them out of the noise floor of the recording microphone. The 1% peak amplitudes of each recorded segment was measured in 1/3-octave bands using a 20-ms Hanning window (50% overlap) and corrected subsequently for the amplitude changes made for recording purposes. Values were scaled to represent the presentation level of the stimuli during testing (60 dB SPL for the NH listeners and 90 to 100 dB SPL for the HI listeners).

Although the audibility of speech typically is calculated using the long-term speech spectrum, the goal in this study was to estimate audibility of the vowel, the transition, and the fricative segments of each syllable with as much precision as possible. For each of the 36 segments (3 talkers × 4 nonsense syllables × 3 segments) and relative amplitude values (7 for the NH adults and 6 for all other listeners), STA was calculated using a modification of the Speech Intelligibility Index (SII; American National Standards Institute, 1997). The modified formula was

$$STA = \frac{1}{30}\sum_{i=1}^{16}[TOB_i - \theta_i]W$$

where i is the center frequency of each 1/3-octave band, TOB is the 1% peak level for each 1/3-octave-band in the stimulus segment (re: 2-cm³ coupler), and θ is the hearing threshold in dB SPL referenced to the same coupler. This equation essentially calculated the signal-to-noise ratio for the stimulus (TOB_i) relative to the hearing threshold (θ_i) for each 1/3-octave band. These values are multiplied by the weight in each band (W) and summed. This signal-to-noise ratio was restricted to a range of 0 to 30 dB by the 1/30 multiplier at the beginning of the equation. Previously derived importance functions for nonsense syllables or sentences (American National Standards Institute, 1997; Duggirala, Studebaker, Pavlovic, & Sherbecoe, 1988; Studebaker & Sherbecoe, 1993) were considered inappropriate for STA measures due to the lack of weights for frequencies ≥5 kHz. Because the female and child talkers produced substantial energy at frequencies ≥6 kHz when producing /us/, the use of traditional importance functions would have resulted in little or no audibility for this fricative at all presentation levels despite rising performance levels. For those reasons, equal weights ($W = 0.0625$) were assumed for the sixteen 1/3-octave bands.

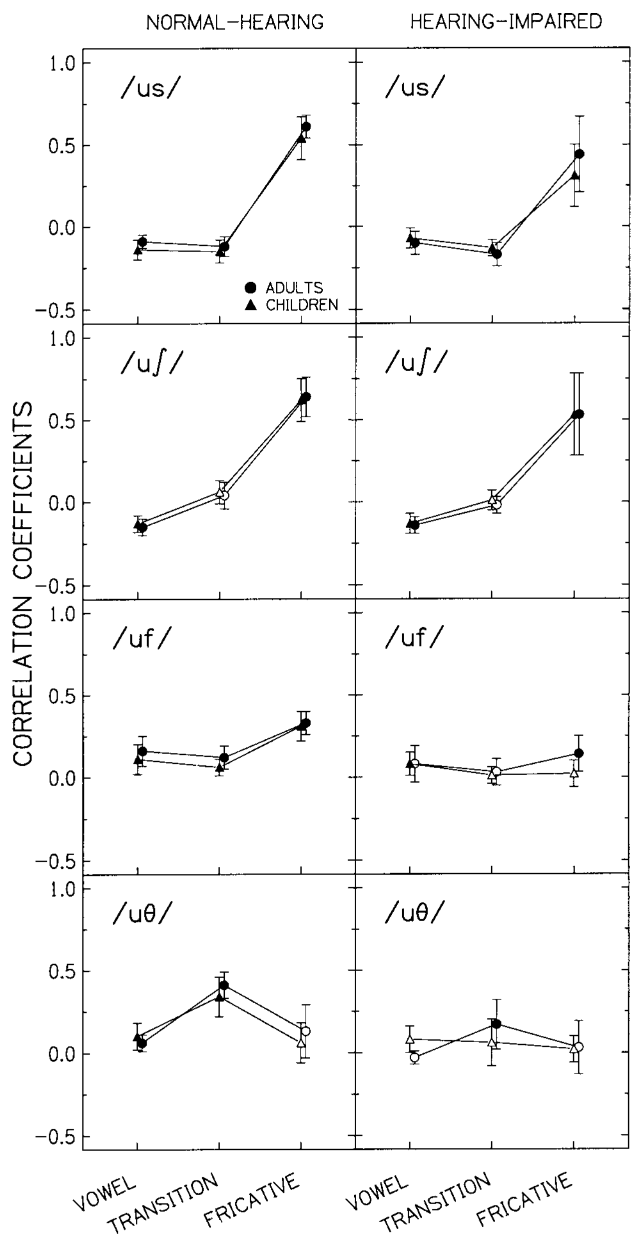
Results and Discussion

Perceptual-Weighting Strategies

For each listener, a point-biserial correlation coefficient was calculated from the responses to each stimulus (incorrect vs. correct) and the STA values of each stimulus segment. This correlation can be thought of in terms of the slope of a line drawn between the average STA values for the incorrect responses and the average STA values for the correct responses. A significant positive correlation indicates a higher average STA value for the correct responses than for the incorrect responses. In terms of the present study, a positive correlation indicates that performance *increased* as audibility increased. A negative correlation indicates that performance *decreased* as audibility increased. Correlation coefficients not significantly different from zero indicate that performance was not influenced by changes in audibility. Because audibility was varied randomly for the vowel, transition, and fricative segments of each nonsense syllable, correlation coefficients were calculated for each segment and interpreted in terms of the perceptual weight each segment contributes to perception. It is important to remember that correlation coefficients describe the relation between performance and audibility and that low correlation coefficients do not necessarily indicate poor recognition of the syllable.

Individual correlation coefficients were transformed to Z scores and pooled to obtain mean and standard deviations for each of the four groups. These values were transformed back to correlation coefficients and are shown in Figure 3 as a function of stimulus segment for each nonsense syllable. Coefficients for the NH and HI listeners are in the left and right panels, respectively. Data for adults are displayed as circles, and data for children are displayed as triangles. Filled symbols represent coefficients that are significantly different from zero ($p < .05$). For /us/ and /uʃ/, the performance of all four groups correlated most highly with the fricative segments of both syllables. Small negative correlation coefficients for the

Figure 3. Perceptual weights (displayed as correlation coefficients) for each nonsense syllable as a function of stimulus segment. NH groups are displayed to the left and HI groups to the right. Circles represent adults, and triangles represent children. Filled symbols are those correlation coefficients that were significantly different from zero.



vowel and transition segments of /us/ and the vowel segment of /uʃ/ also were found, indicating that performance decreased slightly as the level of those segments increased. In other words, as the level of the vowel segment in the nonsense syllable /uʃ/ increased, performance decreased somewhat. The same can be said for both the vowel and transition segments of /us/. For /uf/, relatively low weights were given to each segment by all four groups, although the NH children and adults appeared to weight the fricative segment of /uf/ more so than the vowel and transition. For /uθ/, the NH children and adults weighted the transition segment more so than the vowel and fricative segments and the HI children and adults weighted all three segments equally low. Despite some statistically significant differences for the segments of /uf/ and /uθ/ across the four groups, the effects were small compared to the weights for /uʃ/ and /us/. Therefore, further analyses were limited to the fricative segments of /uʃ/ and /us/ which showed the largest effects.

Performance-Audibility Functions

Recall that the amplitude of the fricative segments (as well as the vowel and transition segments) was varied randomly across trials, providing a wide range of audibility values. Figure 4 shows the data relating performance to STA for /ʃ/ and /s/ for the NH listeners (upper four panels) and the HI listeners (lower four panels). In general, performance increased as a function of STA for all four groups. The data were fitted originally with a standard SII transfer function. However, the reduced audibility range for the HI listeners prevented a good fit at low SII values. In addition, the maximum performance levels of two HI children for /s/ were substantially below those of the remaining HI children and served to decrease the goodness of fit for any function fitted to these data. A fit to the changing portion of each listener's function provided a more valid representation of the data.[2] The filled and open circles in each panel of Figure 4 represent the changing and plateau portions of each listener's functions, respectively. The data points identified as being in the changing portion of each function were fit with a quadratic function (solid lines) after the data were subjected to an exponential transformation. The data and quadratic fit were then transformed back to their original values. A goodness of fit measure (R^2) for each quadratic function is shown in each panel. Ninety-five percent confidence intervals were calculated for both /ʃ/ and /s/ for the NH adults only, and these are shown in each panel to facilitate comparison of each group to the NH adults.

[2] To identify the changing and plateau portion of each function, the following procedure was used: An automated iterative process provided linear fits to each of the two portions of the function until the total error variance of the two fits was minimized.

In general, maximum performance for each group was reached at lower audibility levels for /s/ than for /ʃ/, and steeper functions were observed for the HI groups relative to the NH groups. Surprisingly, the two HI groups required lower audibility levels than their NH counterparts to reach equivalent performance for both /ʃ/ and /s/. For example, to achieve 90% performance for /ʃ/, the HI children required an average STA of 0.10 whereas the NH children required an average STA of 0.23. Adult-child differences also were apparent. For both /ʃ/ and /s/, the NH children required higher STA levels than their adult counterparts to achieve similar levels of performance. These differences were not as apparent for the HI children and adults.

Recall that one purpose of this study was to determine the minimum audibility required by each group to achieve a criterion level of performance. Although the quadratic fits of each group fall within the 95% confidence intervals of the NH adult data, many of the data points for the NH children and the HI children and adults fall outside these confidence intervals (at both higher and lower STA levels). Furthermore, the use of a quadratic fit assumes that the data are normally distributed and have equal variances, which may not be true for all STA levels tested. Finally, these fits only include data on the changing portion of the functions, and it could be argued that all points should be considered when calculating the minimum STA for a certain level of performance. For these reasons, a decision theory approach (which makes no assumptions regarding the underlying distribution of the data) was used to determine the minimum STA criteria of each group to achieve a specific level of performance.

The data in each panel of Figure 4 were dichotomized into categories of good performance (those observations where performance was ≥90%) and poor performance (those observations where performance was <90%). This dichotomy of good and poor performance served as the gold standard by which the effectiveness of STA was measured. Recall that this data set represents multiple observations for each subject across three different talkers and a range of fricative amplitudes. For each STA level, hit, miss, correct rejection, and false alarm rates were calculated. Consistent with other clinical applications of decision theory, accuracy was calculated as the hit rate (those observations where poor performance was expected and occurred) and the correct rejection rate (those observations where good performance was expected and achieved). For the same STA values, error was calculated as the miss rate (those observations where good performance was expected but was *not* achieved) and the false alarm rate (those observations where poor performance was expected but *did not* occur).

Figure 4. Individual data points relating performance to STA for /s/ (left panels) and /ʃ/ (right panels). Filled and open symbols represent those data points in the changing and plateau portions of each individual's performance-STA function, respectively. Quadratic functions fit to the data corresponding to the changing portion of the performance-STA functions for each group are represented by the solid line. Goodness of fit measures (R^2) also are displayed for each function. The separate 90% confident intervals calculated for /s/ and /ʃ/ from the NH adult data are displayed in each panel for comparison.

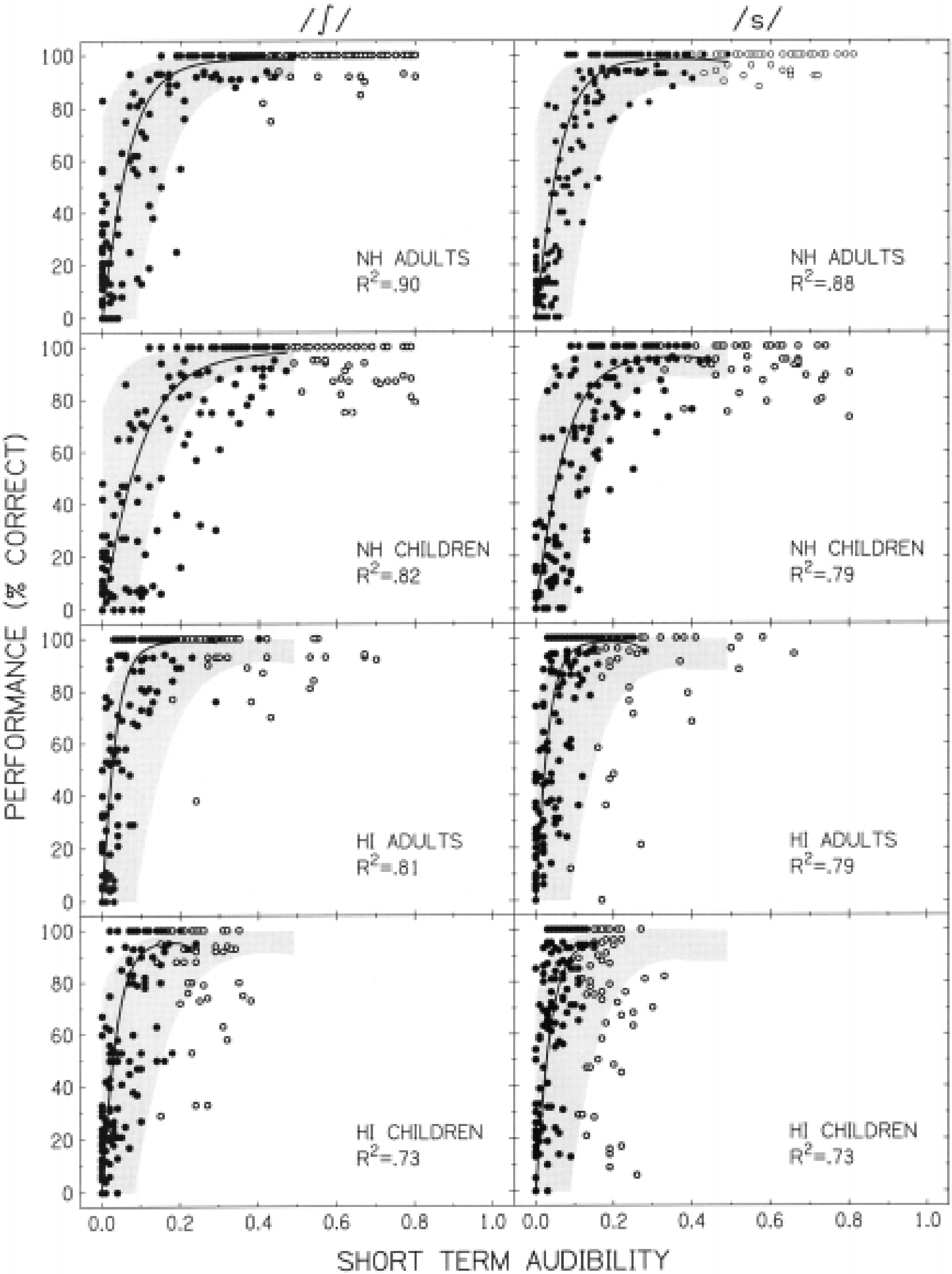


Figure 5 shows the frequency distributions of good and poor performance (dashed and dotted lines, respectively) as a function of STA for each group for the phoneme /ʃ/. Figure 6 shows the same information for the phoneme /s/. The corresponding hit, miss, correct rejection (CR), and false alarm (FA) rates are shown in each panel. Also shown is the calculation of the area under the relative operating characteristic curve (ROC area) relating hits to false alarms. This number varies between 0.5 and 1.0, with areas closer to 1.0 representing a large separation between the two distributions and areas closer to 0.5 indicating considerable overlap (chance). The STA criterion resulting in equally low miss and false alarm rates is shown by the vertical line in each panel. This "equal error" condition is used in clinical decision theory to determine the point of maximum sensitivity and specificity apart from any factors that may influence the final selection of a criterion value (e.g., the high cost of a miss or the false alarm).

As can be seen in the top left panel of Figure 5, there is some overlap between the two distributions for the NH adults, but the hit and correct rejection rates are reasonably high (90%) for an STA criterion of 0.21. Stated differently, as long as fricative audibility exceeded 0.21 (regardless of talker), 90% of the observations (correct rejections) resulted in good performance (≥90%). For both phonemes, however, the pattern for the remaining three groups differs from that of the NH adults. There is more overlap of the two distributions resulting in ROC areas that are lower, but still well above chance in most cases. Despite the greater variability in these three groups, some interesting patterns emerge. The two HI groups required lower minimum STA values than their NH counterparts to achieve good performance. Also, the minimum STA for the NH children was higher than for the NH adults. This same age effect was not observed for the two HI groups.

In this paradigm, false alarms were those observations where performance remained high despite low audibility levels and can be thought of as instances of overachievement. These observations are interesting but irrelevant since performance remained high. Misses, on

Figure 5. Distribution of good (dashed line) and poor (solid line) performance as a function of STA for the phoneme /ʃ/. Hit, miss, correct rejection (CR), false alarm (FA), relative operating characteristic curve area (ROC area) and number of observations (N) are given in each panel. Right going hatch marks represent misses and left going hatch marks represent false alarms. The vertical line represents the STA criteria required to minimize the miss and false alarm rates for each group.

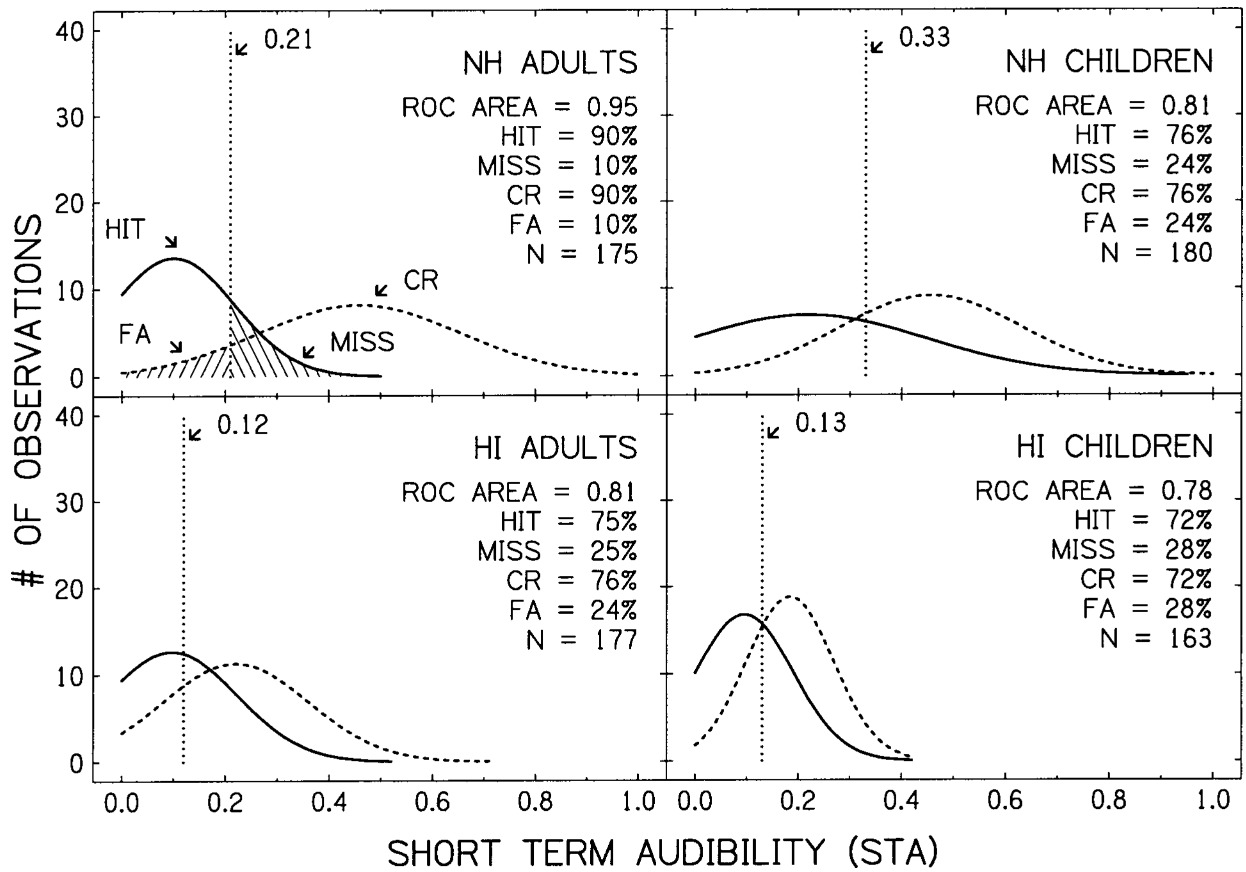
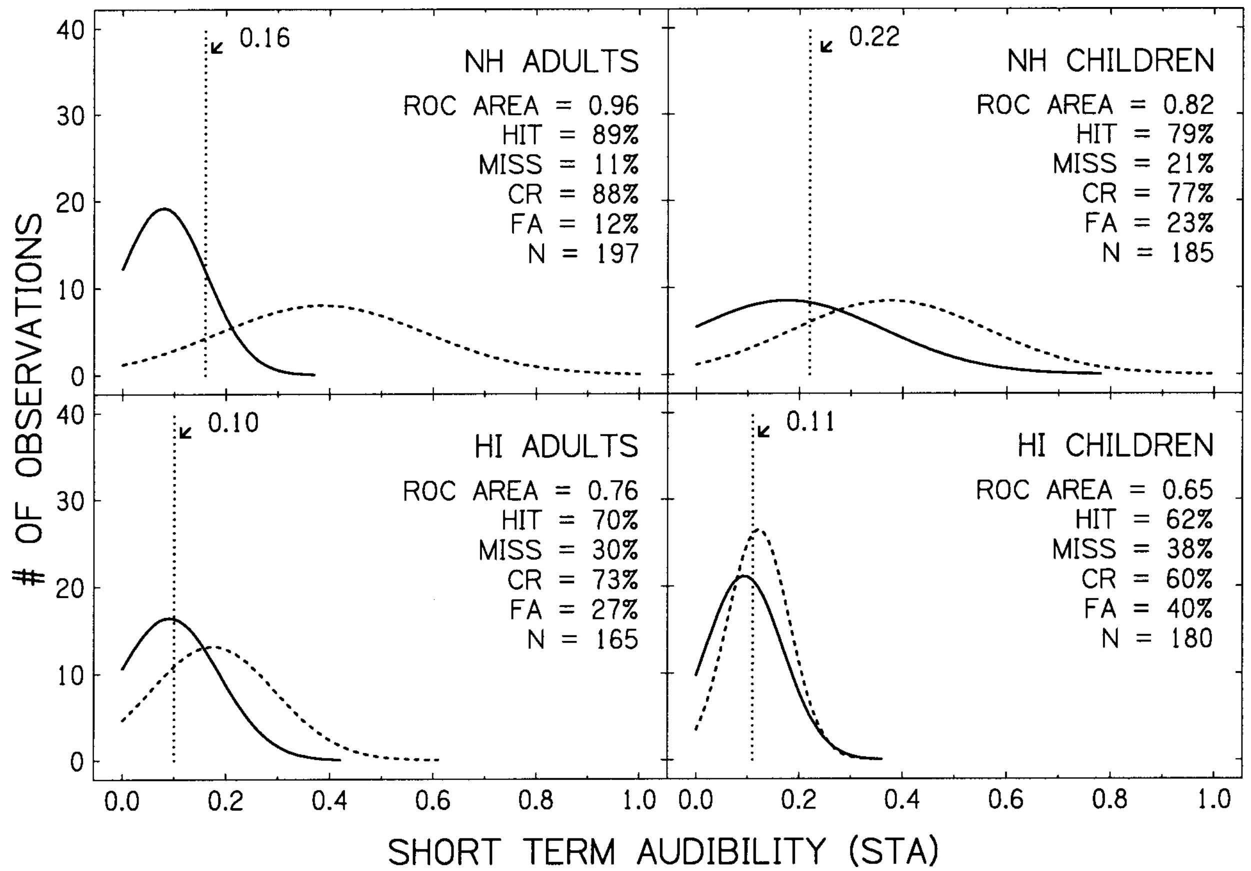


Figure 6. Same as Figure 5, but for the phoneme /s/.



the other hand, are those observations where poor performance occurred even though STA levels should have been sufficient to produce good performance. These observations represent instances of underperformance and are of particular interest in terms of amplification. In theory, the miss rate could be minimized for the HI subjects by increasing hearing aid gain and thus signal audibility. However, it is important to remember that substantial peak spectral energy was required to produce these low STA values. For the fricative /s/ produced by the male talker, peak spectral energy was approximately 25 dB above threshold at an STA of 0.10. For the female and child talkers, the peak spectral energy was about 22 and 10 dB above threshold, respectively, for the same STA.[3] When the degree and configuration of hearing loss provides sufficient dynamic range, audibility could be improved by increasing hearing aid gain. However, when there is insufficient dynamic range, potential loudness discomfort or reduced performance may result due to high compression ratios or peak clipping (Dreschler, 1988; Hawkins, 1980).

[3] The STA values appear low because they were calculated across 16 equally weighted frequency bands for stimuli with spectral energy concentrated in a relatively narrow high-frequency region.

General Discussion

The purpose of the present study was to identify the segments of speech used for perception and to characterize the relation between performance and audibility in NH and HI children and adults. The performance of children and adults was compared within their respective hearing categories to determine the influence of age on the perception of these syllables. Likewise, the performance of the NH groups and HI groups was compared within their respective age categories to determine the influence of hearing status. Although caution is warranted when applying these results to the perception of speech in general, several interesting results emerged.

In terms of the perceptual-weighting strategies, all four groups placed the greatest weight on the fricative segments of /us/ and /uʃ/ for perception. For /uf/, relatively low weights were given to each segment by all the HI listeners, whereas the NH listeners weighted the fricative segment more so than the vowel and transition segments. For /uθ/, the NH listeners weighted the transition segment more so than the vowel and fricative segments, whereas the HI listeners weighted all three segments equally low. These results are consistent with those reported in previous studies where the fricative segments of /us/ and /uʃ/ and the transition segments of /uθ/ were thought to help identify these phonemes in NH listeners (Harris, 1958; Hedrick, 1997; LaRiviere, Winitz, & Herriman, 1975; Zeng & Turner, 1990). Interestingly, the NH listeners appeared to rely on the audibility of the fricative segment of /uf/ rather than the transition segment, which was not predicted from the results of previous studies. The weighting strategies of the HI listeners also were consistent with previous studies that found similar performance for both NH and HI adults when using the fricative to identify /us/ and /uʃ/ and poorer performance for HI adults compared to NH adults when using the transition segment to differentiate /uf/ and /uθ/ (Hedrick, 1997; Zeng & Turner, 1990). Zeng and Turner (1990) suggested that, even though the transition segments may be audible to HI listeners, those listeners may not be able to use the transition segments as efficiently as the NH listeners due to a loss of discriminability.

The weighting strategies of young NH children in previous studies were found to differ from those of NH adults in that the children (aged 4 to 5 years) placed greater weight on the dynamically changing transition segments of /us/ and /uʃ/ instead of the steady-state fricative (Nittrouer & Crowther, 1998; Nittrouer, Crowther, & Miller, 1998; Nittrouer & Miller, 1997a,1997b). These investigators suggested that the children's limited speech perception experience may have contributed to their results. A priori, one hypothesis for the present study was that the HI children might also have a limited experience with speech perception and therefore possess similar weighting strategies to those of younger NH children. Instead, both the NH and HI children in this study assigned the most weight to the fricative segments of /us/ and /uʃ/, suggesting that differences in the perceptual development of these children due to the presence of hearing loss, if any, likely occurred earlier than 8 years of age.

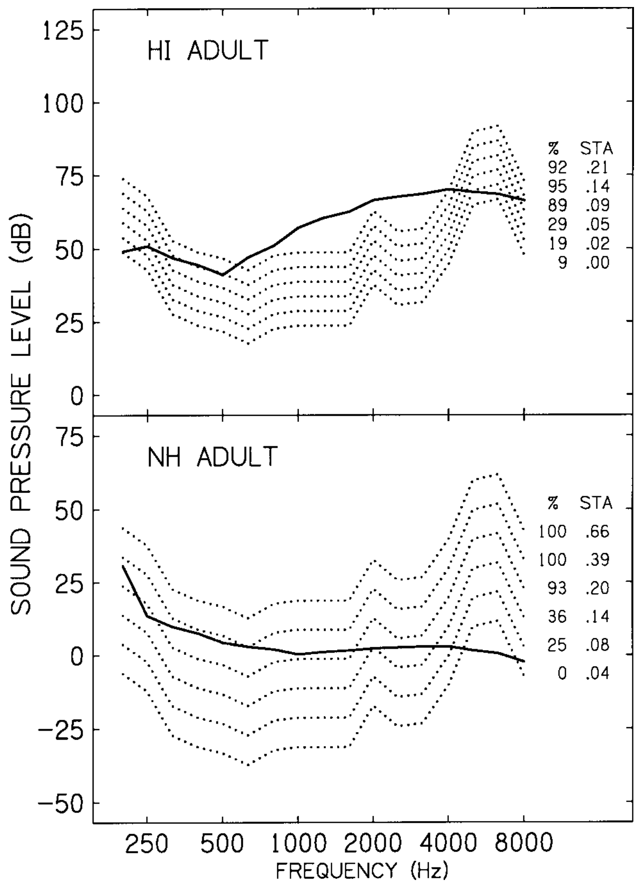
Two factors may be responsible for the low correlation between performance and audibility for the phonemes /f/ and /θ/ in the HI listeners. First, these phonemes were lower in amplitude than the phonemes /s/ and /ʃ/ resulting in a restricted range of audibility with which performance was correlated. Second, the phonemes /f/ and /θ/ occur at relatively low rates in conversational speech. Studies of conversational English have shown that of the 24 English consonants, /s/, /f/, /θ/, and /ʃ/ rank 3rd, 15th, 20th, and 21st in frequency of occurrence, respectively (Denes, 1963). Given the relatively low occurrence of /f/ and /θ/ in conversation, the HI children and adults may not have received sufficient exposure to, or encountered substantial difficulty with those consonants to form a perceptual strategy. The higher amplitude and wider frequency spectra of /ʃ/ probably distinguished it from the other fricatives during perception and would explain the high weights given to the fricative segment of this infrequently occurring consonant.

In terms of the audibility required for a given level of performance, previous studies have found that NH children require higher levels of audibility than NH adults (Stelmachowicz et al., 2000). For this reason, it was expected that the NH children would require higher audibility levels than the NH adults and, in general, this was the case. Previous studies also have found that HI adults require higher levels of audibility than NH adults to achieve equivalent levels of performance (Ching, Dillon, & Byrne, 1998; Dubno, Dirks, & Ellison, 1989; Hogan & Turner, 1998; Kamm, Dirks, & Bell, 1985; Turner & Robb, 1987). It was, therefore, expected that the two HI groups in the present study would require higher audibility levels than the two NH groups. Surprisingly, the HI listeners required less audibility than the NH listeners to achieve equal performance. The apparent contradiction between these results and those of previous studies may be due, in part, to the manner in which audibility was calculated in the current study. In previous studies, the long-term spectrum of speech was used to calculate an audibility index (American National Standards Institute, 1997). These spectra are weighted heavily by vowel energy and are relatively insensitive to the shorter duration components of speech. This method cannot provide an accurate estimate of the audibility of specific speech sounds, particularly for short duration consonants. In the present study, separate spectra were calculated for each speech segment and talker to determine the audibility of each during perception. Using this method, the HI listeners were found to require lower than expected audibility levels to achieve the criterion level of performance. The observation that HI listeners require less audibility than the NH listeners is consistent with the abnormal growth of response that is seen typically in ears with cochlear hearing loss. Specifically, activity along the basilar membrane may be influenced by a loss of compression in the cochlea that accompanies outer hair cell loss and produces steeper than normal growth of response functions (Nelson & Schroder, 1997; Oxenham & Plack, 1997). As for the surprisingly similar results for the HI children and adults, it is tempting to hypothesize that the rehabilitative strategies used with young HI children may have accelerated the normal course of speech perception development, yielding more adult-like patterns of performance at an earlier age.

Recall that decision theory was used to determine the minimum STA required by each group to achieve good performance. One advantage of this approach is that the variability in performance at any STA level can be more easily and accurately evaluated with decision theory than from functions fit to the performance-STA data. More importantly, data need not meet the assumptions of normal distributions and equal variances at all levels of STA. Using decision theory, it appears that STA is a good predictor of performance in NH adults. Good and poor performance for the HI listeners and NH children, however, were not as well separated using STA as a predictor variable. Specifically, if a group of NH adults were provided the minimum STA level for /ʃ/, 90% of them would be expected to achieve ≥90% recognition of that phoneme. In contrast, only about 72% of the HI children would be expected to perform as well. Although the separation of good and poor performance for the HI listeners and NH children is not as good as that of the NH adults, separation is well above chance in most cases and revealed interesting age and hearing loss effects that were not as easily observed in the functions fit to the performance-STA data.

Although this study was not designed to address the use of high frequency information in the perception of nonsense syllables, the data are relevant to this issue. In several previous studies it was found that adults with sensorineural hearing losses ≥55 dB HL at 4 kHz did not benefit from amplification at frequencies ≥4 kHz (Ching, Dillon, & Byrne, 1998; Hogan & Turner, 1998; Turner & Cummings, 1999). Specifically, recognition of nonsense syllables and sentences did not improve when the bandwidth of the speech signal included frequencies at or above 4 kHz. In some cases, the inclusion of signal energy above 4 kHz actually reduced recognition. Based on this information, the adults with hearing losses ≥55 dB HL at 4 kHz in the present study might be expected to show poorer performance than that of the NH adults for fricatives comprised primarily of high frequency energy. However, this was not the case. Figure 7 shows the hearing thresholds (solid lines) for one HI adult (upper panel) and one NH adult (lower panel). In each panel, the 1/3-octave-band levels of the fricative segment of /us/ produced by the female speaker are shown at six presentation levels (dotted lines). For this speaker, the acoustic energy of /s/ occurred at frequencies above 4 kHz. Recognition scores (% correct) corresponding to each presentation level and the associated STA values are given in each panel. This HI adult was able to achieve high levels of performance primarily using high-frequency information. In addition, the performance of both the NH and HI adults continued to improve from very low to very high recognition as STA increased. However, the HI adult required less than half the STA to achieve similarly high levels of performance. For example, the maximum score of 100% was achieved by the NH adult at a STA of 0.39, whereas the HI adult achieved a maximum score of 95% at a STA value of only 0.14. Of the eleven HI children and adults with hearing thresholds ≥55 dB HL at 4 kHz in the present study (including the adult

Figure 7. Hearing thresholds (in dB SPL) as a function of frequency for one HI adult and one NH adult as well as the 1/3-octave band spectra for /s/ produced by the female speaker. The recognition scores (in percent) and short-term audibility (STA) values corresponding to the six presentation levels are shown to the right of each spectrum.



described in Figure 7), 5 adults and 3 children were able to use the high-frequency information to achieve scores of 95 to 100% at STA values between 0.05 to 0.18. The performance of one HI adult reached 82% at a STA value of 0.11 before dropping to 48% at 0.20, suggesting a decrease in performance as audibility increased. The performance of one HI child continued to increase to a maximum of 80% at the highest audibility level presented (0.19) and another performed poorly (~30%) across all audibility levels (0.04 to 0.22).

When making a determination regarding the importance of high-frequency amplification, it is critical to recognize methodological differences between the current study and previous studies. For example, in the Hogan and Turner (1998) and Turner and Cummings (1999) studies, a much larger set of phonemes (23 and 24, respectively) were used. For the majority of subjects in those studies, performance remained constant as high-frequency audibility was increased. Since most of the phonemes in their stimulus set contained low-frequency acoustic information (e.g., /m/, /n/, /ŋ/, /ʤ/, /g/, /l/, /r/, /w/, /j/, etc.), it is possible that the lack of improvement may have been due to an insufficient number of consonants containing primarily high-frequency energy. In the Ching et al. study (1998), high-context sentence materials were used to evaluate speech perception in NH and HI adults. The authors acknowledged that language competence may have allowed their listeners to maintain relatively good performance despite the loss of high-frequency acoustic information through filtering. Infants and young children in the developing stages of language, however, may not perform as well. In the current study, a small number of low-context stimuli consisting primarily of high-frequency acoustic information were used. The rationale for these stimuli was to avoid the confounds of language acquisition and to focus on stimuli with acoustic characteristics in the frequency region of interest. Although these restrictions may limit the ability to draw definitive clinical recommendations, the relative importance of /s/ to both speech and language development should be mentioned. Specifically, /s/ plays an important role in English because of its high frequency of occurrence (ranked 3rd) and because of its role in the development of both verb and noun morphology when it occurs in word-final position (Rudmin, 1983). The fact that 3 of the 5 children with hearing losses ≥55 dB HL at 4 kHz showed improvement for the perception of /s/ with increases in audibility suggests that clinicians may wish to proceed with caution when considering the restriction of amplification to frequencies below 4 kHz.

Several aspects of the present study warrant further investigation. First, the HI children and adults in the present study had relatively flat hearing losses suggesting that they would be less likely to develop odd perceptual strategies than individuals with severely sloping losses or losses with an unusual configuration. For that reason, it would be of interest to compare the perceptual-weighting strategies of NH children and adults to those of HI children and adults with various configurations of hearing loss. Second, it would also be of interest to compare the perceptual-weighting strategies of younger HI children (<8 years) to those of age-matched NH children. Although, the HI children in the present study exhibited adult-like weighting strategies, younger HI children may exhibit less efficient weighting strategies than their NH peers (Nittrouer et al., 1997a, 1997b). Third, the data in this study were collapsed across a male, a female, and a child speaker. Although this was done to provide a more valid representation of everyday listening situations for children, the spectral characteristics of these 3 speakers differed substantially. Specifically, each speaker tended to produce energy in different frequency regions for the same phoneme. Although this study was not designed to evaluate speaker effects, inspection of the raw data suggests that those effects occurred. Because young children are more likely to listen to female and child speakers, further studies specifically designed to address speaker effects would be informative. Also, a study using multiple speakers of the same age or gender with similar spectral characteristics may improve the predictive value of the decision theory analysis by reducing variability in performance. Finally, the ability to predict performance may be improved by including other relevant measures of hearing with STA in a multivariate clinical decision analysis approach.

Acknowledgments

The authors would like to thank Reinier Kortekaas for his help with the design and implementation of this study, Douglas Keefe and Eric Odgard for their assistance with the analyses and interpretation of these data, and Dawna Lewis and Donna Neff for their many helpful comments on preliminary manuscripts. This project was supported by a grant from NIH.

References

Allen, P., & Wightman, F. (1992). Spectral pattern discrimination by children. *Journal of Speech and Hearing Research, 35*, 222–233.

Allen, P., Wightman, F., Kistler, D., & Dolan, T. (1989). Frequency resolution in children. *Journal of Speech and Hearing Research, 32*, 317–322.

American National Standards Institute. (1997). *Methods for calculation of the speech intelligibility index* (ANSI S3.5–1997). New York: Author.

Carney, A. E., & Moeller, M. P. (1998). Treatment efficacy: Hearing loss in children. *Journal of Speech, Language, and Hearing Research, 41*, S61–S84.

Ching, T. Y., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *Journal of the Acoustical Society of America, 103*, 1128–1140.

Denes, P. B. (1963). On the statistics of spoken English. *Journal of the Acoustical Society of America, 35*, 892–904.

Doherty, K. A., & Turner, C. W. (1996, November). *Weighting strategies listeners use to perceive speech.* Paper presented at American Speech-Language-Hearing Association Convention, Seattle, WA.

Dreschler, W. (1988). Dynamic-range reduction by peak clipping or compression and its effects on phoneme perception in hearing-impaired subjects. *Scandinavian Audiology, 17*, 45–51.

Dubno, J. R., Dirks, D. D., & Ellison, D. E. (1989). Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss: I. The contribution of selected frequency regions. *Journal of the Acoustical Society of America, 85*(1), 347–354.

Dubno, J. R., & Levitt, H. (1981). Predicting consonant confusions from acoustic analysis. *Journal of the Acoustical Society of America, 69*(1), 249–261.

Duggirala, V., Studebaker, G. A., Pavlovic, C. V., & Sherbecoe, R. L. (1988). Frequency importance functions for a feature recognition test material. *Journal of the Acoustical Society of America, 83*, 2372–2382.

Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America, 66*, 651–653.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech, 1*, 1–7.

Hawkins, D. B. (1980). Loudness discomfort levels: A clinical procedure for hearing aid evaluations. *Journal of Speech and Hearing Disorders, 45*, 3–15.

Hedrick, M. (1997). Effect of acoustic cues on labeling fricatives and affricates. *Journal of Speech, Language, and Hearing Research, 40*, 925–938.

Hnath-Chisolm, T. E., Laipply, E., & Boothroyd, A. (1998). Age-related changes on a children's test of sensory-level speech perception capacity. *Journal of Speech, Language, and Hearing Research, 41*, 94–106.

Hogan, C. A., & Turner, C. W. (1998). High-frequency audibility: Benefits for hearing-impaired listeners. *Journal of the Acoustical Society of America, 104*, 432–441.

Kamm, C. A., Dirks, D. D., & Bell, T. S. (1985). Speech recognition and the articulation index for normal and hearing-impaired listeners. *Journal of the Acoustical Society of America, 77*, 281–288.

Kortekaas, R. W. L., & Stelmachowicz, P. G. (2000). Bandwidth effects on children's perception of the inflectional morpheme /s/: Acoustical measurements, auditory detection, and clarity rating. *Journal of Speech, Language, and Hearing Research, 43*, 645–660.

LaRiviere, C., Winitz, H., & Herriman, E. (1975). The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research, 18*, 613–622.

Nelson, D. A., & Schroder, A. C. (1997). Linearized response growth inferred from growth-of-masking sloped in ears with cochlear hearing loss. *Journal of the Acoustical Society of America, 101*, 2186–2201.

Nittrouer, S., & Crowther, C. S. (1998). Examining the role of auditory sensitivity in the developmental weighting shift. *Journal of Speech, Language, and Hearing Research, 41*, 809–818.

Nittrouer, S., Crowther, C. S., & Miller, M. E. (1998). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America, 101*, 2253–2266.

Nittrouer, S., & Miller, M. E. (1997a). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America, 102*, 572–580.

Nittrouer, S., & Miller, M. E. (1997b). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America, 101*, 2253–2266.

Oxenham, A. J., & Plack, C. J. (1997). A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. *Journal of the Acoustical Society of America, 101*, 3666–3675.

Rudmin, F. (1983). The why and how of hearing /s/. *Volta Review, 85*, 263–269.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33*(3), 441–452.

Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas, R. W. L., & Pittman, A. L. (2000). The relation between stimulus context, speech audibility, and perception for normal and HI children. *Journal of Speech, Language, and Hearing Research, 43*, 902–914.

Studebaker, G. A., & Sherbecoe, R. L. (1993). Frequency-importance functions for speech recognition. In G. Studebaker & I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance* (pp. 185–204). Boston: Allyn and Bacon.

Turner, C. W., & Cummings, K. J. (1999). Speech audibility for listeners with high-frequency hearing loss. *American Journal of Audiology, 8*, 47–56.

Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., & Doherty, K. A. (1998). Frequency-weighting functions for broadband speech as estimated by a correlational method. *Journal of the Acoustical Society of America, 104*(Pt. 1), 1580–1585.

Turner, C. W., & Robb, M. P. (1987). Audibility and recognition of stop consonants in normal and hearing-impaired subjects. *Journal of the Acoustical Society of America, 81*, 1566–1573.

Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D. (1989). Temporal resolution in children. *Child Development, 60*(3), 611–624.

Zeng, F. G., & Turner, C. W. (1990). Recognition of voiceless fricatives by normal and hearing-impaired subjects. *Journal of Speech and Hearing Research, 33*, 440–449.

Received October 26, 1999

Accepted August 10, 2000

Contact author: Andrea Pittman, PhD, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE, 68131. Email: pittmana@boystown.org