Spectral Characteristics of Speech at the Ear: Implications for Amplification in Children

This study examined the long- and short-term spectral characteristics of speech simultaneously recorded at the ear and at a reference microphone position (30 cm at 0° azimuth). Twenty adults and 26 children (2–4 years of age) with normal hearing were asked to produce 9 short sentences in a quiet environment. Long-term average speech spectra (LTASS) were calculated for the concatenated sentences, and short-term spectra were calculated for selected phonemes within the sentences (/m/, /n/, /s/, /ʃ/, /ɻ/, /u/, /i/). Relative to the reference microphone position, the LTASS at the ear showed higher amplitudes for frequencies below 1 kHz and lower amplitudes for frequencies above 2 kHz for both groups. At both microphone positions, the short-term spectra of the children's phonemes revealed reduced amplitudes for /s/ and /ʃ/ and for vowel energy above 2 kHz relative to the adults' phonemes. The results of this study suggest that, for listeners with hearing loss (a) the talker's own voice through a hearing instrument would contain lower overall energy at frequencies above 2 kHz relative to speech originating in front of the talker, (b) a child's own speech would contain even lower energy above 2 kHz because of adult–child differences in overall amplitude, and (c) frequency regions important to normal speech development (e.g., high-frequency energy in the phonemes /s/ and /ʃ/) may not be amplified sufficiently by many hearing instruments.

KEY WORDS: amplification, hearing loss, children, speech

For most hearing losses, the primary goal of amplification is to provide a sufficiently audible speech signal. For young children with hearing loss, however, the goal of amplification is more complex. Although communication is the desired result, young children with hearing loss must first use amplification to learn speech and language skills, including the accurate production of speech. One theory of speech production suggests that children with normal hearing learn to produce speech, in part, from the auditory feedback of their own voices (Borden, 1980). Harper and Huie (2000) illustrated the importance of auditory feedback in a group of 3- to 6-year-old children with normal hearing who were asked to recite the alphabet while listening to their own voices played back at them at different time delays. They found that the children slowed their speaking rate and experienced disruptions in the flow of speech when auditory feedback was delayed by only a few hundred milliseconds. The authors concluded that young children do monitor their own speech productions and are sensitive to at least one form of auditory anomaly.
Although little direct evidence is available regarding the impact of hearing loss on a child’s ability to use auditory feedback, indirect evidence suggests a link between the two. For example, Oller and Eilers (1988) argued that the poor quality and delayed onset of babbling in deaf infants may be due to a lack of auditory stimulation for the behavior. Although children with mild-to-moderate hearing losses receive somewhat more auditory stimulation, they have been found to demonstrate significant speech and language disorders over time (Deal & Haas, 1996; Elfenbein, Hardin-Jones, & Davis, 1994; Schonweiler, Ptok, & Radu, 1998). Several studies have demonstrated a more immediate effect of impaired auditory stimulation in that the speech production of prelingually deafened children changes when the input from their cochlear implants is discontinued for brief periods (Higgins, McCleary, & Schulte, 1999; Richardson, Busby, Blamey, Dowell, & Clark, 1993). These indirect examples suggest that a close examination of the amplified speech signal may be warranted for young hearing-aid users.

There are two issues that may interact to influence self-monitoring in children with hearing loss: (1) differences in the acoustic characteristics of children’s speech relative to those of adults and (2) differences in the acoustic characteristics of a talker’s own speech relative to speech received in typical face-to-face conversations. Because many hearing-aid fitting algorithms assume that the primary signal of interest is adult speech at 0° azimuth at a distance of 1 meter (Byrne & Dillon, 1986; Cox & Moore, 1988), those algorithms may not be sufficient to accommodate the acoustic differences in children’s speech relative to that of adults or to the variations between a talker’s own voice and that of another individual.

In terms of acoustic characteristics, children’s speech is known to differ from that of adults in a number of ways. Specifically, children’s speech is higher in formant and fundamental frequencies (Kent & Read, 1992). Phonologically, young children have less experience with speech and are less intelligible than older children or adults. Lee, Potamianos, and Narayanan (1999) suggested that these phonological differences are due in large part to (a) differences in the physical dimensions of the vocal tract, (b) the inexact use of articulators during motor control development, and (c) variability in articulation while children experiment with speech production.

For self-monitoring purposes, it is important to consider the acoustic characteristics of one’s own speech received at the ear. Cornelisse, Gagné, and Seewald (1991) recorded running speech in groups of men, women, and 8- to 12-year-old children at two microphone locations simultaneously. Results showed that the long-term average spectra of speech (LTASS) measured at the ear differed from the spectra measured at the reference microphone position (30 cm at 0° azimuth) in all three groups. Specifically, the spectra at the ear showed greater energy below 1 kHz and less energy above 2 kHz than the spectra measured at 0° azimuth. Although the greater low frequency energy at the ear can be explained by the proximity of the sound source to the microphone, the authors suggested that the 10-dB loss in energy at 4 kHz may have been due to the “directional radiation characteristics of the mouth.” Although these differences may have little influence on the development of speech production in children with normal hearing, the audibility of important high-frequency speech sounds may be markedly reduced in children with hearing loss.

The purpose of the present study was to determine the extent to which the spectral characteristics of young children’s speech measured at the ear differ from those at typical conversational distances. Specifically, the spectral characteristics recorded at the ear were compared to those recorded at a standard reference microphone position (30 cm at 0° azimuth) in both children and adults. The long-term spectra of speech were calculated and compared to determine adult-child differences as well as microphone location differences. The short-term spectra of selected phonemes also were calculated and compared across groups to determine the vulnerability of high-frequency phonemes to reduced audibility at the ear.

**Method**

**Participants**

Twenty adults and 26 children participated in this study. The adult group consisted of 10 women and 10 men between the ages of 21 and 45 years ($M = 33$ years, 8 months). The children’s group consisted of ten 4-year-olds, eight 3-year-olds, and eight 2-year-olds, with equal numbers of boys and girls ($M = 3$ years, 9 months). Normal hearing was confirmed in the 2-year-old children using otoacoustic emissions and acoustic immittance measures. Thresholds of all other participants were measured using pure-tone audiometry at 0.5, 1.0, 2.0, and 4.0 kHz. All but one 3-year-old child had hearing thresholds of 20 dB HL or less bilaterally. For that child, a mild conductive hearing loss was detected in one ear (~35–45 dB HL); because speech production appeared to be unaffected, however, she was not excluded. All participants were native speakers of American English with no noticeable regional dialects.

---

1 Attempts were made to recruit equal numbers of 2-, 3-, and 4-year-olds; however, many younger children were unable or unwilling to participate in the task and were excluded.
Materials

Nine short sentences appropriate for 2- to 4-year-old children were developed for this study and contained at least three repetitions of the phonemes /m/, /n/, /s/, /l/, /l/, /a/, /u/, and /i/. These phonemes were selected to allow for phonemic analyses in both the low- and high-frequency spectral regions (Boothroyd, Erickson, & Medwetsky, 1994) where the long-term spectra are known to differ at the ear (Cornelisse et al., 1991). Table 1 lists the sentences as well as the phonemes used for later analysis.

Apparatus

Simultaneous recordings were made using a head-worn microphone (Shure, WBH53B) positioned 1 inch from the right corner of the talker’s lips and a miniature condenser microphone (AKG, C417L) positioned just above the talker’s left ear. Both microphones were attached to a baseball cap using adjustable Velcro strips. The head-worn microphone was positioned out of the talker’s breath stream, and the bill of the baseball cap was positioned behind the talker. The frequency responses of both microphones were flat through 10 kHz. The signal from each microphone was amplified, digitized at a sampling rate of 20 kHz, and low-pass filtered at 10 kHz. Small level differences between the microphones were eliminated by filtering (TDT, PF1) the amplified output from 0.2 to 10 kHz as measured in a test box. To compare the results across conditions in dB SPL, the amplified output of each microphone was calibrated using a 1-kHz reference tone of known sound pressure level and voltage.

Procedure

Each talker was seated in a sound-treated room and instructed to read or repeat the list of sentences in a typical manner. When necessary, talkers were re-instructed to speak normally. The adults familiarized themselves with the list of sentences before recording. The children were prompted by an examiner to repeat each sentence as often as necessary to obtain two acceptable productions. Only those 3- and 4-year-old children who were able to repeat the sentences in full were included in the study. Two-year-old children were allowed to repeat the sentences in portions, if necessary.

Table 1. Sentences produced by children and adults. Bolded and underlined letters (consonants and vowels, respectively) indicate those phonemes selected for short-term analyses.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Give me your socks.</td>
<td>/g/ /m/ /s/ /s/</td>
</tr>
<tr>
<td>She got a shot.</td>
<td>/s/ /e/ /t/</td>
</tr>
<tr>
<td>Where’s your knee?</td>
<td>/w/ /e/ /r/ /s/ /u/ /i/</td>
</tr>
<tr>
<td>Soup is good food.</td>
<td>/s/ /u/ /u/ /i/ /d/ /f/</td>
</tr>
<tr>
<td>Mommy has big feet.</td>
<td>/m/ /o/ /m/ /a/ /h/ /b/ /i/ /f/</td>
</tr>
<tr>
<td>Knock, knock, who’s there?</td>
<td>/k/ /n/ /k/ /n/ /k/ /n/ /w/ /z/ /s/ /z/ /s/</td>
</tr>
<tr>
<td>I see the fox.</td>
<td>/i/ /s/ /e/ /t/ /f/ /o/ /k/</td>
</tr>
<tr>
<td>My shoes are new.</td>
<td>/m/ /i/ /s/ /h/ /u/ /e/ /s/ /n/ /e/</td>
</tr>
<tr>
<td>Goodnight moon.</td>
<td>/g/ /d/ /n/ /i/ /g/ /t/ /t/ /n/ /o/ /n/ /t/ /i/ /o/ /u/ /n/</td>
</tr>
</tbody>
</table>

The boundaries of the 16 sentences (8 sentences × 2 repetitions) were acoustically and visually identified (using CoolEdit, Syntrillium Software Corp.). The sentences were then extracted and concatenated for later analysis of the long-term spectra. The target phonemes within each sentence (see Table 1) were also identified and analyzed. Because the phonemes were embedded in running speech, some phonemes were preceded and followed by substantial vocalic transitions. In those cases, small segments of the preceding or following transitions were included in the analyses. The sentences were analyzed in seventeen 1/3-octave bands (over a frequency range of 178 to 8980 Hz) with a 100-ms Hanning window (50% overlap of adjacent windows). The phonemes were analyzed in the same way except that a 20-ms Hanning window was used.

To correct the speech spectra recorded from the head-worn microphone to a reference position of 30 cm at 0° azimuth, the continuous discourse of 20 of the 46 talkers (5 each of men, women, boys, and girls) was recorded. Each talker spoke for 2 minutes while wearing the head-worn microphone and facing a second microphone positioned at a distance of 30 cm at 0° azimuth. The children repeated sentences of a story read to them, whereas the adults simply read aloud. The speech samples then were edited to include only the talker’s voices and analyzed in seventeen 1/3-octave bands with a 100-ms Hanning window (50% overlap of adjacent windows). Differences between the spectrum levels measured at each microphone were used as correction factors that were applied to the spectra of the sentences obtained from the head-worn microphone.

Results and Discussion

Long-Term Spectra

The LTASS for the adults and children are shown in Figure 1 along with the noise floor of the sound-treated room and measurement system. The spectra measured at the reference microphone position (30 cm at 0° azimuth) are given in the top panel and the spectra at the ear are given in the bottom panel. For both microphone positions, the amplitude of the children’s speech was lower than that of the adults. Cornelisse et al. (1991) reported no difference between the vocal levels for children relative to adult males; however, the children in
their study were considerably older (8–12 years) than the children in the present study (2–4 years).

Figure 2 compares the LTASS for each microphone position by group. In general, the LTASS measured at the ear was higher in amplitude in the low frequencies and lower in amplitude in the high frequencies. Relatively little difference was observed between the LTASS measured at the ear and at the reference position in the midfrequencies for each group. Interestingly, the LTASS at the two positions overlapped between 1 and 2.5 kHz for the male and female adult talkers and between approximately 1.8 and 3 kHz for the children. The LTASS at the two microphone positions in the present study were consistent with those reported by Cornelisse et al. (1991).
**Short-Term Spectra**

The short-term spectra measured at the ear are displayed for the two nasal phonemes (/m/ and /n/) in Figure 3. The parameter in each panel is group. Relatively few differences were observed across groups in the production of the nasal phonemes. As shown in Figure 4, small differences were observed across groups in the production of the three vowels (/a/, /u/, and /i/). Although the vowel formants are obscured in the averaged spectrum, observation of the individual spectra confirmed that the formant frequencies were consistent with those reported by Peterson and Barney (1952) for men, women, and children talkers. The low-frequency notch observed in the children's /u/ and /i/ vowel spectra also was reported for the women and children in the Cornelisse et al. (1991) study.

**Figure 3.** Short-term spectra measured at the ear for the phonemes /m/ (top panel) and /n/ (bottom panel). The parameter in each panel is group.

**Figure 4.** Short-term spectra measured at the ear for the phonemes /a/ (top panel), /u/ (middle panel), and /i/ (bottom panel). The parameter in each panel is group.
The averaged formant amplitudes observed for the vowel /a/ (top panel) were similar to those reported by Huber, Stathopoulos, Curione, Ash, and Johnson (1999) for young children and adults. More substantial variations in formant amplitude were observed across groups for the vowels /u/ and /i/ (middle and lower panels, respectively). Specifically, amplitude in the region of the third formant (1–5 kHz) of both vowels was substantially lower in level for the children than for the male and female adult talkers. Because the vowel spectra measured at the reference microphone also showed similarly low amplitudes in this region, the differences were not considered to be related to microphone position but rather to adult–child differences. Unfortunately, few data are available regarding the formant amplitudes of young children for the vowels /u/ and /i/. One possible explanation for these adult–child differences is that the glottal waveforms of children are known to be more sinusoidal than those of adults, with shorter closed periods for more symmetric waveforms. As a result, children have steeper source spectra, reducing the high-frequency energy in vowels (Formby & Monsen, 1982).

The short-term spectra measured at the ear for the three voiceless fricatives (/s/, /ʃ/, and /f/) are displayed in Figure 5. No systematic differences were observed across groups in the production of /f/ (bottom panel). Although the spectral peaks of /s/ and /ʃ/ are somewhat broader than those of the individual spectra due to averaging, these spectra are consistent with group data reported for adult male and female talkers recorded and analyzed under similar conditions (Boothroyd et al., 1994). In general, the mean amplitude and peak frequency differed across groups for both fricatives. Similar peak amplitudes were observed for the /s/ and /ʃ/ fricatives produced by the male and female adults, whereas substantially lower peak amplitudes were observed for both fricatives produced by the children. Higher peak frequencies also were observed for the adult females for both /s/ and /ʃ/ (7.3 and 4.5 kHz, respectively) compared to the adult males (5.4 and 3 kHz, respectively). Although the peak frequencies of the children’s /ʃ/ were similar to those of the adult females, the peaks of /s/ for these children appeared to occur at a frequency beyond the limits of the 1/3-octave-band analysis (>8 kHz).

Discussion

The purpose of the present study was to determine the extent to which the spectral characteristics of young children’s speech measured at the ear differed from those of spectral characteristics reported for adult talkers at typical conversational distances (Byrne & Dillon, 1986;
Cox & Moore, 1988). Because young children are in the process of developing speech perception and production skills, knowledge of the acoustic characteristics of speech at the ear may be particularly important when considering the amplification needs of young children with hearing loss. Specifically, adequate amplification of a child’s own voice for self-monitoring purposes may be as important as the amplification of other talkers (children or adults).

The results of the present study confirm and extend those of Cornelisse et al. (1991) to younger children (ages 2–4) and to specific phonemes. Although the long-term speech spectra of the 8- to 12-year-old children in the Cornelisse study were similar to the adult spectra, adult–child differences were observed in the present study. Specifically, the children’s vocal levels were lower in amplitude at both the ear and at the reference microphone position, suggesting that these differences were due to age. Amplitude differences also were observed for the high-frequency formants of the vowels /u/ and /i/ and for the high-frequency energy of the fricatives /s/ and /ʃ/ relative to the adults. For both vowels and fricatives, the amplitude of children’s speech measured at the ear was substantially lower than that of adults for frequencies above approximately 2 kHz. Systematic adult–child differences were not observed for the vowel /a/, the fricative /ʃ/, or the nasals /m/ and /n/.

Overall, the relative loss of high-frequency energy at the ear is similar for both adults and children (see Figure 2). However, because children do not talk as loudly as adults, the absolute level of a child’s own voice at the ear is reduced even further relative to that of an adult (see Figure 1). The low levels of speech produced by a child may be complicated further by the algorithms used to select the amplification characteristics for a particular hearing loss. Specifically, fitting algorithms based on the long-term spectra of adult speech at a 0° azimuth may not be adequate to amplify the low-amplitude, high-frequency components of speech produced by young children. This point is illustrated in Figure 6. The solid line represents the long-term spectrum of the children’s speech measured at the ear in the present study. The comparison functions represent the presumed input used in fitting algorithms that include accommodations for amplification in children. The dashed line represents a combination of the ear-level LTASS for children and the average of the adult male, adult female, and child reference LTASS developed by Cornelisse et al. (1991). The dotted line is the long-term spectrum developed by Byrne et al. (1994) for which no adjustments for gender, age, or vocal effort were made. Although the shape of the Cornelisse spectrum is similar to that of the children’s spectrum in the current study, both spectra overestimate the levels of speech at the ear of young children by as much as 7 to 9 dB for frequencies above 2 kHz. Because the assumed high-frequency input is higher than that actually measured at the ear in the current study, the prescribed gain will be less than that needed for self-monitoring purposes, particularly when using the Byrne spectrum.

Although children with normal hearing sensitivity have little difficulty developing speech production skills under these conditions, low levels of high-frequency energy may seriously affect the speech perception and production skills of children with hearing loss. This problem is illustrated in Figure 7. The hearing thresholds (in dB SPL) of a hypothetical child (open circles) are plotted relative to the amplified spectra of /s/ produced by a 3-year-old boy from the present study. For reference, the lightly dotted line represents the unaided spectrum of /ʃ/ produced by a 3-year-old boy from the present study. The heavy dotted line shows the spectrum of /s/ after linear amplification, and the dashed lines represent nonlinear amplification through three digital devices that varied considerably in compression characteristics.2 Specifically, the devices differed in compression threshold (48 to 60 dB SPL), compression channels (2 to 4), and release times (50 to 5000 ms). In this

---

2 Each hearing aid was set according to DSL 4.1, which assumes the Cornelisse et al. (1991) spectrum (Seewald et al., 1997). The hearing thresholds and amplified spectrum of /ʃ/ include an estimate of the closed-ear resonance characteristics using an age-appropriate average real-ear-to-coupler difference.
example, only a narrow frequency region of the child’s own /s/ was above threshold after linear amplification (approximately 10 dB). The low compression thresholds in the digital instruments enhanced the audibility of this low-level phoneme by 2 to 5 dB and increased the overall bandwidth. Despite the substantial differences in the signal processing characteristics of the digital devices, only small spectral differences in the amplified output were observed. It is also worth noting that, in all cases, amplification resulted in a shift in the spectral peak of /s/ from >8 kHz downward to between 3 and 4 kHz.3 Although the bandwidth of current hearing instruments is wider than ever before, the high-frequency gain in most behind-the-ear instruments drops off precipitously above 5 kHz, well below the frequencies of peak energy for /s/ in both children and adult female talkers. Spectrally, the peak energy of /s/, when low-pass filtered in this way, is similar to the spectral peak of /ʃ/ (see Figure 5). This similarity may make it difficult to differentiate between the two phonemes, explaining, in part, the production problems unique to children with hearing loss.

For example, Elfenbein et al. (1994) examined the speech and language skills of children with mild, moderate, and severe sensorineural hearing losses and found that even children with the mildest hearing losses exhibited significant misarticulation of fricatives. Often, these fricatives were completely omitted by the children.

In summary, the results of this study suggest that the speech of young children differs spectrally from that of adults when measured at the ear relative to a standard conversational distance. Because the input to a hearing instrument is often assumed to have spectral characteristics similar to that of an adult male, a child’s own speech may be underamplified, particularly in the high frequencies. Coupled with the limited bandwidth of hearing instruments, a young child with hearing loss likely will perceive the low-level phonemes of speech inconsistently during the development of speech and language. There is a need for further investigation of how various hearing-aid processing schemes may be used to improve the self-monitoring capabilities of children with hearing loss.

Acknowledgments

This work was supported by Grant R01 DC04300 from the National Institutes of Health. We gratefully acknowledge the contributions of Maureen Higgins and Mary Pat Moeller on previous versions of this paper.

References


3 The probe microphone techniques used in the clinic and in research with young children may underestimate the high-frequency gain at the tympanic membrane. Although probe-tube placement is not critical at low and midfrequencies, deep insertion of the probe tube is necessary for high-frequency measurements. However, probe-tube placement at the tympanic membrane is not feasible with young children in clinical practice.

Figure 7. Example showing the audibility of a child’s /s/ after amplification through linear (heavy dotted line) and nonlinear (dashed lines) hearing devices set according to DSL 4.1 targets. The lightly dotted lines represent the unaided spectrum of /s/ produced by a 3-year-old child. A hypothetical moderate hearing loss is represented by the open circles.


Received June 17, 2002

Accepted January 22, 2003

DOI: 10.1044/1092-4388(2003/051)

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