Journal of Speech and Hearing Research, Volume 31, 102–107, March 1988

COMPOSITE SPEECH SPECTRUM FOR HEARING AID GAIN PRESCRIPTIONS

ROBYN M. COX JEFFREY N. MOORE
Memphis State University

Average long-term RMS 1/3-octave band speech spectra were generated for 30 male and 30 female talkers. The two spectra were significantly different in both low and high frequency bands but were similar in the mid-frequency region. It was concluded that a single spectrum could validly be used to represent both male and female speech in the frequency region important for hearing aid gain prescriptions: 250 Hz through 6300 Hz. In addition, the male and female spectra were compared with analogous spectra reported by Byrne (1977) and Pearsons, Bennett, and Fidell (1977). For each sex, significant differences were found among the three spectra in a few frequency bands. The best estimate of the average speech spectrum for each sex was obtained from a weighted average of the three sets of data, excluding the significantly different data points. The long-term RMS 1/3-octave band speech spectrum for male and female talkers combined was derived for use in hearing aid gain prescriptions.

Hearing aid gain prescriptions often incorporate an adjustment to compensate for the fact that normal speech contains more low frequency energy than high frequency energy (e.g., Berger, 1976; Byrne & Tonisson, 1976; Byrne & Dillon, 1986; Cox, 1983; Pascoe, 1978). These prescriptions generally provide for less low frequency gain and greater high frequency gain.

The level and shape of the long term RMS speech spectrum that is assumed when a prescriptive procedure is formulated affects the absolute amount of gain prescribed at each frequency when that procedure is used. Unfortunately, different long term RMS speech spectra have been assumed by different investigators. Typically, each investigator has performed an independent study of the long term RMS speech spectrum and used these data in developing the gain prescription tables or formulae. Therefore, different procedures result in different gain prescriptions, not because the procedures are fundamentally different but because they make different assumptions about the long term RMS speech spectrum. This factor tends to obscure the real differences (if any exist) among various prescriptive procedures and makes comparisons among them difficult to perform.

These considerations suggest that it would be desirable for all hearing aid gain prescription procedures to use the same long term RMS speech spectrum. This paper describes the derivation of a long term RMS speech spectrum that has been developed for use in hearing aid gain prescription procedures.

Two types of multi-talker speech spectra have been employed in hearing aid gain prescription procedures: They will be referred to as simultaneous spectra and sequential spectra. A simultaneous spectrum is obtained by measuring the long term RMS spectrum produced by several talkers recorded at one time, talking together. This type of spectrum has been employed in hearing aid gain prescriptions by Cox (1983) and Pascoe (1978). In contrast, a sequential spectrum is defined here as obtained by measuring the long term RMS speech spectrum for each of several individual talkers and arithmetically averaging the obtained levels across talkers. This type of spectrum is found in the hearing aid gain prescription procedure described by Byrne and Tonisson (1976).

For the purposes of hearing aid gain prescription, a sequential type of spectrum seems more appropriate than a simultaneous spectrum because the sequential spectrum accurately represents the average levels in the speech of individual talkers. Furthermore, it is convenient to express the long term RMS speech spectrum in 1/3-octave band levels.

A literature review revealed only two reports of 1/3-octave band sequential speech spectra for relatively large groups of male and female English-speaking talkers. Byrne (1977) reported sequential spectra for 15 male and 15 female talkers producing continuous speech with "normal" vocal effort. Pearsons, Bennett, and Fidell (1977) reported sequential spectra for male talkers and female talkers at each of several levels of vocal effort. Although these authors did not report the number of talkers contributing to the spectra, the numbers contributing to the spectra for "normal" vocal effort appear to be 42 men and 27 women.

  0022-4685/88/3101-0102$01.00/0

The spectra reported for male talkers in these two investigations are considerably different. The Byrne (1977) spectrum is several decibels higher than the Pearsons et al. (1977) spectrum below 400 Hz and several decibels lower in the 400–4000 Hz range. At 1000 Hz, the difference between the two spectra is 9 dB (see Figure 2). Equally large differences are found between the two investigations in the data for the female speech spectra: these two spectra are rather similar below 400 Hz but diverge above this frequency with a maximum difference of 10.5 dB at 1250 Hz (see Figure 3).

Because differences in assumed long term speech spectra typically translate into equal differences in prescribed gain, the differences between the spectra reported in these two investigations are large enough to have a significant impact on the outcome of a hearing aid gain prescription. These differences might be attributable simply to the different samples of talkers contributing to the data. In this case, the best estimate of the long term RMS speech spectrum would be obtained by combining the two sets of data. However, it also seemed possible that the differences might be partly attributable to the fact that the talkers in the Byrne study were producing Australian English while those in the Pearsons et al. study were producing American English. In addition, close examination of the two reports suggests that the mean overall levels of speech were quite different in the two studies. In the Byrne study, the mean overall level for male talkers appears to be about 74 dB at 1 m, whereas the corresponding level for the male talkers in the Pearsons et al. study was 61 dB. For the female talkers, the mean overall level at 1 m in the Byrne study seems to be about 72 dB, whereas the corresponding level in the Pearsons et al. study was 57.5 dB. Because the speech spectrum is known to change as intensity increases (Licklider, Hawley, & Walkling, 1955; Pearsons et al., 1977), it seemed possible that the higher levels apparently produced by the talkers in the Byrne study might have affected the measured spectra. However, it should be remembered that talkers in both studies were instructed to produce speech with "normal" vocal effort. The apparent differences in overall speech levels may be the result of different measurement procedures.

Finally, the two investigations employed very different speech samples. In the Byrne study, the talkers read a 2–2.5 min passage from a popular magazine. In the Pearsons et al. study, the talkers repeated the passage "Joe took father's shoe bench out; she was waiting at my lawn" for 10 seconds. As Tarnoczy (1956) has pointed out, the "Joe . . . lawn" passage is not representative of the phoneme distribution in English and, therefore, may result in a long-term spectrum that is different from one obtained using a more representative passage [because the phoneme distributions in spoken and written speech are very similar (Fletcher, 1953, p. 96) it is reasonable to assume that the 2-min passage of popular reading material used by Byrne would have a phoneme distribution that is rather representative of everyday speech]. Benson and Hirsh (1953) compared long-term octave-band speech spectra for talkers reading a newspaper article, and repeating the "Joe . . . lawn" passage. They concluded that the spectra were equivalent. However, small differences between the two spectra may be discerned in their data. The more detailed 1/3-octave band analysis might conceivably reveal more noteworthy differences between spectra derived from these two types of speech material.

To evaluate the differences between the Byrne (1977) spectra and those derived by Pearsons et al. (1977), and to provide a basis for deriving best estimates of the average long term RMS 1/3-octave band speech spectra for male and female talkers, an investigation was undertaken to generate a third pair of sequential speech spectra, one for male talkers and one for female talkers. Additional research questions were: (a) Can the same spectrum be used to represent both male and female talkers, and (b) what is the intertalker variability in long term RMS speech spectra among talkers whose speech has been normalized to the same overall level?

METHOD

Subjects

Thirty male and 30 female talkers provided speech samples for analysis. Ages ranged from 20 to 57 with mean ages of 34 years (men) and 27 years (women). All were native talkers of American English and were able to read fluently from a written text.

Recording the Speech Samples

Environment. Speech samples were recorded in a double-walled sound treated room, 1.8 m x 1.8 m, with a mean reverberation time (250–8000 Hz) of 52 ms. Ambient noise levels in the room were 53 dB(C)/19 dB(A). Post hoc analysis indicated that ambient noise was well below the speech level in each 1/3-octave band, with the exception of the female talker levels in the 80 Hz band where the ambient noise/average spectrum difference was 8 dB.

Procedure. Each talker provided a 2-min sample of continuous speech. The talker was instructed to read a passage of text derived from a children's educational reading source. The subject was asked to talk with normal vocal effort, to disregard any reading errors, and to continue to read at a natural rate until instructed to stop.

Instrumentation. The talker was seated in the middle of the room. A 2.54 cm, field-calibrated microphone (Bruel & Kjaer, type 4145) was located 30 cm from the talker's mouth at a 0-degree azimuth. The microphone output was amplified and high-pass filtered (cutoff frequency=50 Hz) by a precision sound level meter with associated preamplifier (Larson Davis, model 800B). The sound pressure level (integrated RMS) of the 2-min speech sample was determined using the sound level meter. The AC output of the sound level meter was recorded (Panasonic VCR, model 6810) on magnetic tape

(Scotch Color Plus). The frequency response of the recording system was flat, +/− 1 dB, from 50 Hz to 16 kHz.

Analysis. To derive the talker's speech spectrum, the recorded speech sample was replayed, bandpass filtered from 50 Hz to 15 kHz (Wavetek, model 753A), and analyzed using a Signal Analyzer (Hewlett Packard, model 3561A) that digitally synthesized ANSI Class III 1/3-octave band filters. A Hanning window was used to acquire the data. The overall RMS level and the RMS level in each 1/3-octave band from 80 Hz to 12.5 kHz were obtained (dB re 1 volt) for the 2-min speech sample. The sample provided 70 to 75 separate 1/3-octave band analyses, each derived using 1.6 s of speech. These analyses were combined on an RMS basis to produce the final spectrum.

RESULTS

Comparison of Male and Female Speech Spectra

The mean integrated sound pressure level at 1 m for the male talkers was 61 dB (SD=3.6 dB) and the corresponding level for the female talkers was 59 dB (SD=3.2 dB).

Each subject's 1/3-octave band spectrum was expressed in dB relative to that subject's overall speech level. Thus, all talkers were normalized to the same overall speech level. This type of presentation focuses on the shape of the speech spectrum and negates differences among talkers in overall intensity. Figure 1A gives the average 1/3-octave band spectra derived for male and female talkers. Data for the 80-Hz band have been omitted from the female spectrum because, as noted earlier, the ambient noise was only 8 dB less than the average level for these data. Figure 1B illustrates the unbiased standard deviations associated with these data.

The two average spectra are very similar in the range from 400 to 5000 Hz. However, the small mean differences seen for the 800 and 1000 Hz 1/3-octave bands were found to be statistically significant [$t(58)>2.6$, $p<.01$]. Male/female mean differences were also statistically significant [$t(58)>3.5$, $p<.001$] for the 315 Hz 1/3-octave band and for all 1/3-octave bands below 250 Hz and above 6.3 kHz.

Comparison With Previous Investigations

To provide a basis for generating the best estimates of the average 1/3-octave band speech spectra for male and female talkers, it was important to evaluate the differences between the average spectra found in this investigation and those reported in the investigations of Pearsons et al. (1977) and Byrne (1977). To test the significance of the differences among the three spectra for each sex, it was necessary to have an estimate of intersubject variability for each set of data. Because neither of the previous reports included this information for their samples, the intersubject variability observed in this study (Figure 1B) was used as an estimate of the variability in the previous studies. Comparison with variability data from other studies suggested that this expedient would be reasonably accurate (see later discussion). Using this value for intersubject variability and the number of talkers contributing to the average spectra in each study, it was possible to estimate, for each 1/3-octave band, the standard deviation of the sampling distribution of differences between pairs of means (Ferguson, 1959, p. 146). The t distribution was used to estimate the maximum nonsignificant difference between two means ($p < .01$).

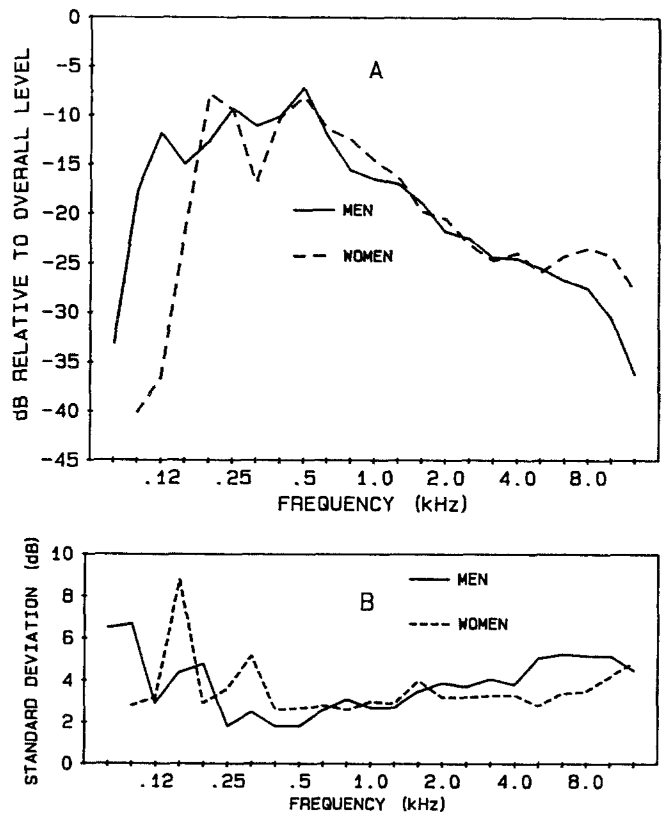

FIGURE 1.(A). Average long-term RMS 1/3-octave band speech spectra for 30 male talkers and 30 female talkers. The level in each 1/3-octave band is plotted relative to the overall speech level. (B). Unbiased intersubject standard deviations associated with long-term RMS 1/3-octave band speech levels for male talkers and female talkers.

In the bands from 125 Hz to 10 kHz, the three spectra for male talkers were compared, in pairs, to determine if the differences between them in each 1/3-octave band were large enough to justify a conclusion that they were drawn from different populations. Because a large number of comparisons were made, there was some potential for spurious significant differences. To control this, only significant differences that were observed in two or more contiguous 1/3-octave bands were accepted as indicative of true differences between the spectra. In sum, if the mean 1/3-octave level for one spectrum was significantly different ($p < .01$) from both of the other spectra in two or more contiguous 1/3-octave bands, then it was considered to represent a different population for that frequency region. It was hoped that this procedure would identify any specific frequency regions in which the various spectra were not typical of talkers of American English, perhaps because of the talker's accent or the particular speech materials used.

The three male talker average spectra are illustrated in Figure 2. The method of comparison described above revealed that the Byrne spectrum was significantly lower than the other two spectra in the 800 Hz, 1000 Hz, and 1250 Hz bands. No other significantly different frequency regions were revealed. A best estimate of the average long term RMS spectrum for male talkers was determined using the following procedure: (a) For all 1/3-octave bands where the differences among the three spectra were not significant, it was assumed that the differences were the result of sampling error and the best estimate of the population value was obtained from a weighted mean of all three sets of data; (b) in the 1/3-octave bands where one spectrum was found to differ significantly from the other two, the best estimate of the population value was obtained from a weighted mean of the two nondifferent spectra; (c) for 1/3-octave bands having only one or two data values (<125 Hz and >10 kHz), the best estimate was obtained from a weighted mean of all data.

Figure 3 is an illustration of the three average spectra for the female talkers. The method described above was used to evaluate the differences among these three spectra. The pairwise comparisons revealed that the Byrne spectrum was significantly lower than the other two spectra in the 1 kHz through 2 kHz bands, and the Pearsons et al. spectrum was significantly higher than the other two spectra in the 3 kHz through 5 kHz frequency region. On the basis of these results, the best estimate of the average long term RMS spectrum for female talkers was determined using the rules followed to construct the male talker spectrum.

Figure 4A illustrates the best estimate of the average long term RMS 1/3-octave band speech spectrum for male talkers and Figure 4B gives the corresponding spectrum for female talkers. Each of these figures also displays the 95% confidence limits determined for the spectrum (see Ferguson, 1959, p. 72, for method of calculating variance of combined samples).

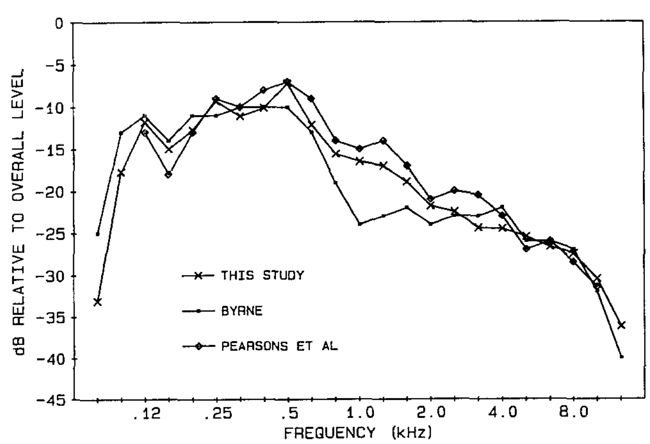

FIGURE 2. Average long-term RMS 1/3-octave band speech spectra for male talkers from three investigations.

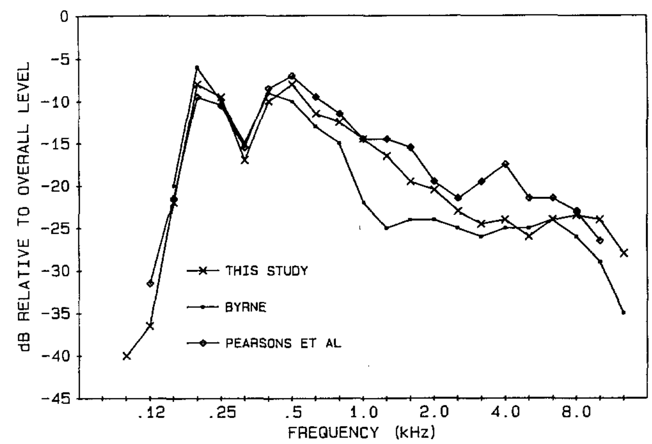

FIGURE 3. Average long-term RMS 1/3-octave band speech spectra for female talkers from three investigations.

DISCUSSION

The mean levels observed in this investigation for normal conversational speech produced by male and female talkers were almost identical to those reported by

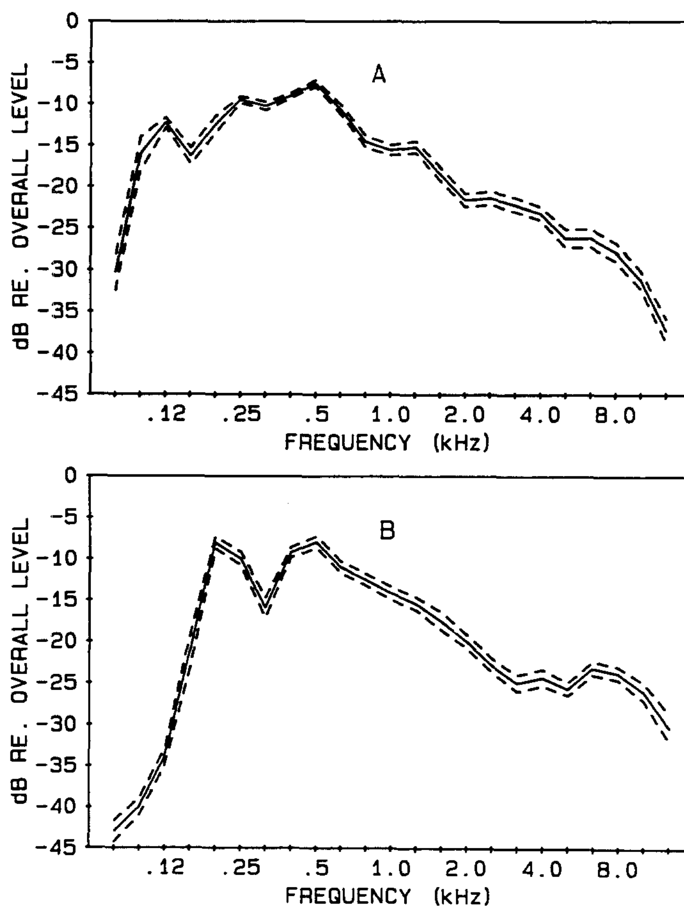

FIGURE 4. (A). Best estimate of the average long-term RMS 1/3-octave band speech spectrum for male talkers. The dashed lines show the limits of the 95% confidence interval around the average value. (B). Best estimate of the average long-term RMS 1/3-octave band speech spectrum for female talkers. The dashed lines show the limits of the 95% confidence interval around the average value.

Pearsons et al. (1977). In addition, these data are consistent with the finding reported by Benson and Hirsh (1953), Pearsons et al., and Byrne (1977), that the mean difference in overall speech level between male and female talkers producing speech with normal vocal effort is about 2 dB. The intersubject standard deviation of roughly 3.5 dB was quite similar to the 4 dB reported by Pearsons et al. for A-weighted Leq overall speech levels.

The pattern of male/female mean differences shown in Figure 1 is remarkably similar to the pattern observable in the spectra reported by both Byrne (1977) and Pearsons et al. (1977), although neither of these reports includes statistical tests of the mean male/female differences. These results suggest that the same average speech spectrum could be used to represent both male and female talkers in the frequency range important for hearing aid gain prescription: 250 Hz to 6300 Hz. Although male talkers produce significantly higher levels than female talkers in the 315 Hz 1/3-octave band, this frequency is not typically included in hearing aid gain prescriptions. Hence, for application to hearing aid gain prescriptions, this difference may be ignored. In addition, although there were significant differences between male and female talkers in the 800 Hz and 1000 Hz 1/3-octave bands, these differences were only 2–3 dB and, thus, an average spectrum would not be seriously in error.

Comparisons among the three male spectra illustrated in Figure 2 and the three female spectra illustrated in Figure 3 revealed that the differences among them were largely insignificant and, thus, probably due to sampling error. In general, the outcome of this investigation supports the notion that different dialects of English have similar long-term speech spectra. One exception to this outcome is the difference between the Byrne spectrum and the other two spectra, for both sexes, in the 1 kHz to 2 kHz frequency region. It is tempting to postulate that this difference may have been due to the different vowels encountered in Australian and American English (for example, this frequency region encompasses the second and third formants of the American English vowel /ɝ/, which does not occur in Australian English).

The source of the difference between the Pearsons et al. female spectrum and the other two female spectra in the 3 kHz to 5 kHz is more obscure. The phonemes /ʃ/ and /ʒ/ produce high output in this frequency region and analysis of the "Joe . . . lawn" passage reveals that these sounds occur about six times more frequently in it than in written speech. If one concludes that this outcome resulted from the different speech texts used in the Pearsons et al. study compared to the other two studies, then it is also necessary to suggest that the effect is limited to female talkers because no corresponding significant difference was observed among the male talker spectra. In fact, Fletcher (1953, p. 78) has suggested that female talkers emphasize sibilant sounds more than male talkers do. Additional support for this position may be found in the data of Benson and Hirsh (1953) who reported long-term octave-band speech spectra. In their data, there is a 3 to 4 dB mean difference in the 1400 Hz through 6 kHz region for female talkers between the speech spectrum derived from the "Joe . . . lawn" passage and that derived from the newspaper article. This difference is less discernible between the corresponding male talker spectra. In the present study, reporting 1/3-octave band spectra, the significant mean difference between the female spectra for the "Joe . . . lawn" passage and the other passages is 4 to 7 dB.

The confidence limits for the two best estimate spectra reported in Figure 4 indicate that the spectra derived by combining data from three studies are rather accurate: The 95% confidence intervals are seldom wider than 2 dB. It is important to keep in mind that these confidence limits (as well as the evaluation of the differences among the spectra from the three studies) were generated with the assumption that the unbiased estimate of population intersubject variability derived from the sample of talkers reported in the present study could validly be applied to the Byrne (1977) and Pearsons et al. (1977) samples. Comparison of the variability data from this investigation with the limited available reports of similar data suggests that this assumption is acceptable. Dunn and White (1940) reported extreme differences of 18 dB in some bands (octave or 1/2-octave) between the speech spectra of two talkers of the same sex. This suggests a maximum standard deviation on the order of 3 to 4 dB (assuming that the two observations lay at opposite extremes of a normal distribution). Benson and Hirsh (1953) reported standard deviations for octave bands of speech. Their values ranged from 1 dB to 4 dB with female talker variability increasing to 5 to 6 dB in the lowest and highest bands. The standard deviations measured in the present study, encompassing 1.8 dB to 8.8 dB, with a mean value of 3.7 dB, were slightly larger than those reported in these earlier studies. This would be expected because the earlier studies employed wider analysis bandwidths than the 1/3-octave bands used in this study. In summary, the variability observed in the present investigation was consistent with expectations derived from earlier comparable reports.

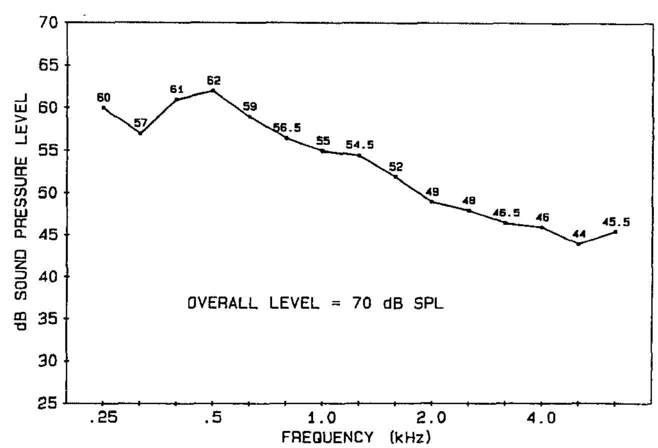

FIGURE 5. Long-term RMS 1/3-octave band speech spectrum, male and female talkers combined, for an overall speech level of 70 dB SPL.

Application to Hearing Aid Gain Prescription

For the prescription of hearing aid gain, assumption of an overall speech level of about 70 dB results in gain values that are rather similar to those used in everyday listening (Farrell, Jansen, & Sweetman, 1979; Reid, Smiarowski, & McPherson, 1977; Walden, Schuchman, & Sedge, 1977). An overall speech level of 70 dB SPL corresponds more closely to a raised voice than to a normal voice level (Pearsons et al., 1977). However, the data of Pearsons et al. indicate that the shape of the average 1/3-octave band speech spectrum is the same for both normal and raised voice levels. Figure 5 gives the mean (unweighted) long-term RMS 1/3-octave speech spectrum for men and women derived from Figures 4A and 4B and adjusted to represent an overall level of 70 dB SPL. This spectrum could be applied in most hearing aid gain prescription schemes that compensate for the long term average speech spectrum.

ACKNOWLEDGMENTS

This work was supported, in part, by Veterans Administration Rehabilitation, Research, and Development Project #344.

REFERENCES

BENSON, R., & HIRSH, I. (1953). Some variables in audio spectrometry. *Journal of the Acoustical Society of America, 25,* 499–505.

BERGER, K. W. (1976). Prescription of hearing aids: A rationale. *Journal of the American Auditory Society, 2,* 71–78.

BYRNE, D. (1977). The speech spectrum—Some aspects of its significance for hearing aid selection and evaluation. *British Journal of Audiology, 11,* 40–46.

BYRNE, D. & DILLON, H. (1986). The national acoustics laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing, 7,* 257–265.

BYRNE, D., & TONISSON, W. (1976). Selecting the gain of hearing aids for persons with sensorineural hearing impairments. *Scandinavian Audiology, 5,* 51–59.

COX, R. M. (1983). Using ULCL measures to find frequency/gain and SSPL 90. *Hearing Instruments, 34*(7), 17–21,39.

DUNN, H. K., & WHITE, S. D. (1940). Statistical measurements on conversational speech. *Journal of the Acoustical Society of America, 11,* 278–288.

FARRELL, D. E., JANSEN, J. W., & SWEETMAN, R. H. (1979, November). *Clinical vs. real-life gain setting for experienced hearing aid wearers.* Paper presented at the National Convention of the American Speech-Language-Hearing Association, Atlanta, GA.

FERGUSON, G. A. (1959). *Statistical analysis in psychology and education.* New York, NY: McGraw-Hill.

FLETCHER, H. (1953). *Speech and hearing in communication.* Huntington, NY: Robert E. Krieger Publishing.

LICKLIDER, J. C. R., HAWLEY, M., & WALKLING, R. A. (1955). Influences of variations in speech intensity and other factors upon the speech spectrum. *Journal of the Acoustical Society of America, 27,* 207.

PASCOE, D. P. (1978, June). An approach to hearing aid selection. *Hearing Instruments,* pp. 12–16, 36.

PEARSONS, K. S., BENNETT, R. L., & FIDELL, S. (1977). *Speech levels in various noise environments* (EPA Report No. 600/1–77–025). Washington, DC: Environmental Protection Agency. (NTIS No. PB 270 05).

REID, L., SMIAROWSKI, R. A., & MCPHERSON, L. T. (1977). Hearing aid gain-setting performance of experienced users. *Archives of Otolaryngology, 103,* 203–205.

TARNOCZY, T. (1956). Determination of the speech spectrum through measurements of superimposed samples. *Journal of the Acoustical Society of America, 28,* 1270–1275.

WALDEN, B. E., SCHUCHMAN, G. I. & SEDGE, R. K. (1977). The reliability and validity of the comfort level method of setting hearing aid gain. *Journal of Speech and Hearing Disorders, 42,* 455–461.

Received January 14, 1987
Accepted May 26, 1987

Requests for reprints should be sent to Robyn M. Cox, Memphis Speech and Hearing Center, 807 Jefferson Avenue, Memphis, TN 38105.