Clinical Forum

Speech-Spectrum Analysis of Mandarin: Implications for Hearing-Aid Fittings in a Multi-Ethnic Society

June Antablin McCullough*
Clara Tu*
Henry L. Lew†

Abstract

Using similar recording and analysis techniques, the long-term average speech spectra of English and Mandarin were compared in order to identify differences that might influence hearing-aid fitting strategies. Despite the well-documented pitch contour differences between English and Mandarin, no significant difference was found in the long-term average spectral analysis. Implications for hearing-aid prescription are discussed.

Key Words: English, hearing aids, Mandarin, multi-ethnic, speech spectrum

The increasingly multi-ethnic makeup of our society challenges audiologists to re-examine many of the assumptions upon which current testing protocols are based. Hearing-aid prescriptive formulas, for example, often employ estimates of the long-term average speech spectrum of English for calculating a desired amount of amplification (Skinner et al, 1982; Cox, 1983; Byrne and Dillon, 1986). It is assumed that placing the average spectral energy of English at a specified level (sensation level or loudness level) will maximize the perception of acoustic cues. Although this assumption may be valid for patients who communicate primarily with English speakers, it is not known whether this may be applied to patients who communicate primarily in other languages.

One language that is dissimilar to English both in phonetic content and pitch contour is Mandarin, the official and most commonly used Chinese dialect. According to Svantesson (1986), the phonetic structure of Mandarin includes several retroflex and palatal fricatives and affricates (/zh/, /ch/, /j/, /q/, and /x/) with corresponding acoustical energy spread across a wider frequency range (from 1 kHz to 9 kHz) than fricatives and affricates in English (3 kHz to 8 kHz). Moreover, Mandarin is a tonal language, which refers to the fact that each syllable is associated with one of four contrasting pitch patterns, high (–), high rising (∕), dipping-rising (\), and high falling (\). For example, the syllable /yi/ may mean "clothes" (yi –), "to suspect" (yi ∕), "chair" (yi \), or "meaning" (yi \). Acoustically, Mandarin tones are represented by changes in the fundamental frequency and/or low-frequency formants of the vocalic portion of a syllable (Gandour, 1978). Changes in vowel formants may alter the amount of low- and mid-frequency energy in Mandarin as compared to English. The purpose of the present study was to determine whether the differences in acoustic/phonetic structure and pitch contour between English and Mandarin could be revealed by comparing the long-term speech spectra of the two languages.
METHOD

Subjects

Twenty English speakers (10 males and 10 females; mean age = 28 years, age range = 19–48 years) and 20 Mandarin speakers (10 males and 10 females; mean age = 29 years, age range = 20–37 years) served as subjects.

Procedures

The speech materials consisted of (a) a 680-word short story selected from a magazine with a 6th grade reading level (English), and (b) a 750-word passage from a popular anthology of Chinese literature with a 6th to 8th grade reading level (Mandarin). Subjects were instructed to read the text at a normal conversational level, at a normal speed, and to omit occasional words or phrases that could not be pronounced. The length of time for reading each of the passages aloud was approximately two and a half minutes. Any subject who could not read aloud with reasonable fluency was excluded from the study.

To record the 2-minute speech samples, each subject was seated individually in a recording studio (ambient noise level = 40 dBA) approximately 18 inches from a dynamic univocal microphone (ATM 63, smooth peak-free frequency response from 50 to 17,000 Hz). The output of the microphone was sent to a mixing board connected to a digital audio tape recorder (Tascam DA-30). The vocal output at the microphone was monitored via the VU meter of the digital tape recorder. Samples with large variations in vocal output (± 5 dB) were eliminated from the study.

Data Analysis

Data analysis was conducted in the audiology laboratory at the University of California at Los Angeles. The recorded speech samples were replayed using a digital audio tape deck (Sharp R-DAT SX-0200). The output of the tape player was delivered to a real-time analyzer (Brüel and Kjaer, Type 2033), which integrated the output level in 25-Hz wide bands every 100 msec over the frequency range from 100 to 10,000 Hz. To obtain the long-term rms levels in each one-third octave band, the output of the real-time analyzer was interfaced with a computer (DEC, LSI II) that integrated the energy levels at 25-Hz points within specific standard one-third octave bands to determine the individual speech spectra for the 40 subjects.

In order to obtain long-term average speech spectra that (a) were independent of individual speaker differences in intensity, and (b) could be compared to previously reported results, a “normalizing” process was performed in which 64 spectra were arbitrarily sliced from highly vocalized portions of each individual speech sample. These spectra were averaged across subjects to produce a group speech spectrum for each language. Then, the relative rms level of speech in one-third octave bands was calculated by subtracting the group rms level of each bandwidth from the group overall rms level.

RESULTS AND DISCUSSION

The group speech spectra generated from the normalizing procedure are presented in Figure 1, where the relative rms level (in dB re: overall level) is shown as a function of one-third octave band center frequency for English (open circles) and Mandarin (filled circles). As seen in this figure, the speech spectrum of English contained the most energy around 200 Hz and gradually decreased in level as frequency increased. This is in agreement with previous reports of long-term average speech spectra (Pascoe, 1978; Berger et al, 1982; Cox and Moore, 1988). The speech spectrum of Mandarin was similar to the speech spectrum of English. In fact, the two speech spectra over-

![Figure 1](image-url)
lapped throughout the entire frequency range. Analysis of variance (two-way ANOVA, language x center frequency) showed no main effect for language (p < .05), and no language by frequency interaction (p < .05). These results imply that despite the phonetic structure and pitch contour differences, English and Mandarin generated similar long-term speech spectra.

Long-term average speech spectra usually are not considered in isolation but in the context of the relative importance of each frequency region. Although the English and Mandarin speech spectra were similar, the “load” that each frequency region carries for intelligibility probably is not. Information pertaining to the relative importance of each frequency region currently is unavailable for Mandarin. This information would be useful for calculating desired hearing-aid gain and determining the acceptability of deviations from target gain in the hearing-aid fitting.

Measurements derived from long-term speech spectra are only part of a hearing-aid evaluation or fitting protocol. Other factors, such as preferred loudness level of speech and preferred sound quality, also are important and may be related to cultural or ethnic background. It is recommended that future studies explore the relationships among language, culture, prescribed hearing-aid characteristics, and user satisfaction to benefit patients of all ethnicities.

Acknowledgment. The authors are grateful to Don Dirks for his time, expertise, and kindness during the data analysis process. The authors also thank Peter Ladefoged for his contribution to this manuscript. Funded in part by the San José State University Foundation.

REFERENCES


