Comparison of Speech Rate and Long-Term Average Speech Spectrum between Korean Clear Speech and Conversational Speech

Jeeun Yoo[1], Hongyeop Oh[1], Seungyeop Jeong[1], and In-Ki Jin[2]

[1]Department of Speech Pathology and Audiology, Graduate School, Hallym University, Chuncheon, Korea
[2]Division of Speech Pathology and Audiology, Research Institute of Audiology and Speech Pathology, College of Natural Sciences, Hallym University, Chuncheon, Korea

Background and Objectives: Clear speech is an effective communication strategy used in difficult listening situations that draws on techniques such as accurate articulation, a slow speech rate, and the inclusion of pauses. Although too slow speech and improperly amplified spectral information can deteriorate overall speech intelligibility, certain amplitude of increments of the mid-frequency bands (1 to 3 dB) and around 50% slower speech rates of clear speech, when compared to those in conversational speech, were reported as factors that can improve speech intelligibility positively. The purpose of this study was to identify whether amplitude increments of mid-frequency areas and slower speech rates were evident in Korean clear speech as they were in English clear speech. **Subjects and Methods**: To compare the acoustic characteristics of the two methods of speech production, the voices of 60 participants were recorded during conversational speech and then again during clear speech using a standardized sentence material. **Results**: The speech rate and long-term average speech spectrum (LTASS) were analyzed and compared. Speech rates for clear speech were slower than those for conversational speech. Increased amplitudes in the mid-frequency bands were evident for the LTASS of clear speech. **Conclusions**: The observed differences in the acoustic characteristics between the two types of speech production suggest that Korean clear speech can be an effective communication strategy to improve speech intelligibility.
J Audiol Otol 2019;23(4):187-192

KEY WORDS: Speech acoustics · Speech perception · Speech intelligibility · Communication · Rehabilitation.

Introduction

Clear speech is a speaking style that is used for effective communication in difficult listening situations and draws on techniques such as accurate articulation, a slow speech rate, and the inclusion of pauses [1,2]. Generally, clear speech includes acoustic articulation modifications, such as reduction in the speech rate, expansion of pitch, and an increase in the intensity of core vocabulary [3]. Sometimes, people use naturally produced clear speech to accurately communicate their intent.

Several published studies have demonstrated that clear speech is more intelligible than conversational speech (natural speech that is used in daily life) [1-4]. Clear speech has been reported to be significantly more intelligible than conversational speech for both children and the elderly with normal hearing [2]. Intelligibility scores in listeners with hearing loss were roughly 17% higher for clear speech than for conversational speech [1]. Further, a significant 16% increase in intelligibility scores for non-native listeners was observed when clear speech was used instead of conversational speech [3]. Another study showed that people with normal hearing had 15% improvement in speech perception on a background of noise when clear speech was used [4]. Clearly, these findings

indicate that clear speech is effective in improving speech intelligibility.

Compared with conversational speech, clear speech has better acoustic characteristics that are associated with improved intelligibility. Clear speech is normally spoken more slowly than conversational speech [5,6]. Picheny, et al. [6] reported that the speech rate for clear speech was around 100 words per minute (wpm), while that for conversational speech was roughly 200 wpm. To investigate the role of speech rate in intelligibility, Krause and Braida [7] tested the intelligibility of clear and conversational speech at normal and slow speech rates and found that both types of speech were more intelligible at a slower than the normal speech rate. The difference in intelligibility between normal and slower than normal speech rates was around 5% for both speech conditions. The long-term spectra also differ between the two speech conditions. Clear speech showed an increased level (1 to 3 dB) between 1,000 Hz and 3,000 Hz in the long-term averaged speech spectra (LTASS) compared with conversational speech [8]. Krause and Braida [5] investigated the contribution of increased LTASS in the mid-frequency range (1−3 kHz) to speech intelligibility using signal processing while controlling for other speech factors such as speech rate. In their study, modification of the LTASS improved intelligibility by approximately 6%. Other characteristics associated with clear speech include vowel expansion [9], increased variations in fundamental frequency [10], and increased temporal envelope modulations [8].

This study considered the Korean language, which has acoustic characteristics that are different from those of the English language. Specifically, the Korean language has different spectral characteristics. For example, averaged Korean LTASS levels have been shown to be significantly lower than English LTASS levels at frequencies above 2,000 Hz [11]. Although English clear speech has been identified to have several acoustic characteristics of benefit for intelligibility, the same has not been demonstrated for the Korean language. Because the English and Korean languages have different acoustic characteristics, the current study tried to identify what characteristics of English clear speech might also apply to Korean clear speech.

Based on English clear speech studies, increased amplitudes (1 to 3 dB) of mid-frequency area (1−3 kHz) and slow speech rates (50% rate of conversational speech) were the most prominent characteristics to improve speech intelligibility. Therefore, the purpose of this study was to compare the speech rate and LTASS between Korean clear speech and Korean conversational speech. The hypothesis was that if these acoustic characteristics of Korean clear speech are simi-

lar to those for English, Korean clear speech can be an effective communication strategy to improve speech intelligibility.

Subjects and Methods

Subjects

Sixty native Korean speakers (30 males, 30 females; mean age 21.5 years, range 20−28 years) participated in the study. All participants had no speech disorder history. For the screening of voice disorder, the Multi-Dimensional Voice Program was used (KayPENTAX™, CSL, Montvale, NJ, USA). All participants had normal ranges of fundamental frequency, pitch perturbation (Jitter), amplitude perturbation (Shimmer), and noise-to-harmonics ratio [12]. Each study participant received approximately $10 (10,000 won, Korean currency) as compensation for each hour spent participating in the study. The study protocol was approved by the institutional review board of Hallym University (approval number HIRB-2016-001 and each participant received a written explanation of the study aims, protocol, and procedures and provided written informed consent before participating). Each participant was provided a written informed consent form after receiving a written explanation of the study aims, protocol, and procedures.

Recording and analysis procedure

Korean clear speech instruction was undertaken in the same way as that reported for instructions of English clear speech [13]. Speakers were instructed to read the sentences "while speaking clearly" [13]. Each participant recorded Korean speech perception in noise (K-SPIN) test sentences, which is one of the standardized Korean sentence materials [14]. The K-SPIN sentences were developed according to a similar development principle to that for the English SPIN sentences [15]. K-SPIN sentences were balanced for intelligibility of speech, predictability and familiarity of key words, phonetic content, and length [14]. The K-SPIN test sentences consisted of 240 sentences, with half of the sentences being high-predictability sentences and the other half, low-predictability sentences.

Both conversational and clear speech examples were recorded using a computerized speech laboratory equipment (KayPENTAX™, CSL, Montvale, NJ, USA) and an e-835s microphone (Sennheiser, Wedemark, Germany). All recordings were made in a double-walled, sound-attenuated booth. The sampling mode was set to 44,100 Hz and the quantization was set to 16 bits. Before recording the example sentences, a practice session was held to relax each study participant. The recording was made at a distance of 10 cm from the mouth of the microphone while the subject was sitting in a chair.

The recording session for conversational speech lasted 2 hours and that for clear speech lasted 3 hours. First, each participant recorded K-SPIN sentences as conversational speech. Speakers were instructed to read sentences naturally, as they would usually speak to a friend. Second, the same participant recorded K-SPIN sentences as clear speech with the instruction. During both recording sessions, two audiologists (HO and SJ, both trained in clear speech) monitored the quality of the recorded speech. During the recording process, if one or both audiologists determined that a participant's recorded voice or quality of speech sounds was not appropriate, the recording was repeated. The subjects were allowed to take breaks if they wished.

The speech rate and LTASS for clear speech and conversational speech were analyzed using the recorded sentences. All data were analyzed according to the production method (30 datasets for clear speech and 30 datasets for conversational speech) and sex. For the speech rate, the periods of silence at the beginning and end of each recorded sentence were removed. Subsequently, each participant's recorded sentences were concatenated and digitized at a sampling rate of 44,100 Hz using Adobe Audition CS6 (Adobe Systems, San Jose, CA, USA). Each subject's number of syllables per minute (spm) was calculated (60 seconds×3,790 syllables÷total duration in seconds) and the averaged values were calculated for male and female participants. In English, it is possible to count the number of words on the basis of spacing; however, in Korean, there are cases wherein words are pasted without spacing. Therefore, spm, which was also used in a previous study to analyze speech rate for Korean, was considered in this study [16].

To calculate the LTASS, the sentences for each participant were concatenated without any separating pauses and digitized at a sampling rate of 44,100 Hz using Adobe Audition CS6. Subsequently, each concatenated sentence was normalized to a root mean square sound pressure level (SPL) of 65 dB, and then an averaged long-term root mean square speech spectrum of 21 bands for frequencies ranging from 100 Hz to 10,000 Hz was calculated using Matlab (version R2018a; MathWorks Inc., Natick, MA, USA). The band divisions were based on the critical band speech intelligibility index calculation procedure (American National Standards Institute, 1997/R2012). A 25.6 ms non-overlapping Hamming window with a 1024 fast Fourier transform size was used.

Statistical analysis

A paired t-test was used to identify any significant effects of speech rate and LTASS according to speech production type and the sex of the speaker. Statistical analysis was performed using SPSS software ver. 18.0 (SPSS Inc., Chicago, IL, USA), and the significance level was set at $p<0.05$.

Results

Speech rate

On average, the speech rates for clear speech were slower than those for conversational speech in both male and female subjects. For male speakers, the average (standard deviation, SD) speech rates for clear and conversational speech were 263.4 (SD=29.4) spm and 356.4 (SD=27.6) spm, respectively. The speech rates for the two different methods of speech production were significantly different ($t=7.512$, $df=58$, $p<0.05$). For female speakers, the average speech rates for clear and conversational speech were 227.9 (SD=25.8) spm and 330.2 (SD=22.8) spm, respectively. The speech rates for the two different methods of speech production were also significantly different ($t=12.726$, $df=58$, $p<0.05$).

LTASS

Compared with LTASS for conversational speech, increased amplitudes in mid-frequency bands were evident in LTASS for clear speech in both male and female speakers (Fig. 1). For male speakers, LTASS for clear speech was higher than that for conversational speech at center frequencies (CFs) between 700 Hz and 5,700 Hz. Differences between the two methods of speech production were found between 2.12 dB and 1.33 dB. Specifically, LTASS for clear speech at CFs of 700, 840, 1,000, 1,170, 1,600, 1,850, and 4,800 Hz was significantly higher than that for conversational speech ($p<0.05$). For female speakers, LTASS for clear speech was higher than that for conversational speech at CFs between 840 Hz and 3,400 Hz. Differences between the two methods of speech production were found between 2.26 dB and 1.22 dB. Specifically, the LTASS for clear speech at CFs of 1,000, 1,170, 1,600, 2,500, and 3,400 Hz was significantly greater than that for conversational speech ($p<0.05$).

Discussion

In this study, the speech rate and LTASS were compared between Korean clear speech and Korean conversational speech. Speech rates for clear speech were significantly slower than those for conversational speech. Clear speech was spoken at rates as slow as 35.3% by male speakers and 44.8% by female speakers. For LTASS, significantly increased amplitudes (1.22 dB to 2.26 dB) in mid-frequency bands were evident for clear speech compared to those for conversational speech. Overall, amplitudes of clear speech were higher than conversational speech at CFs between 700 Hz and 5,700
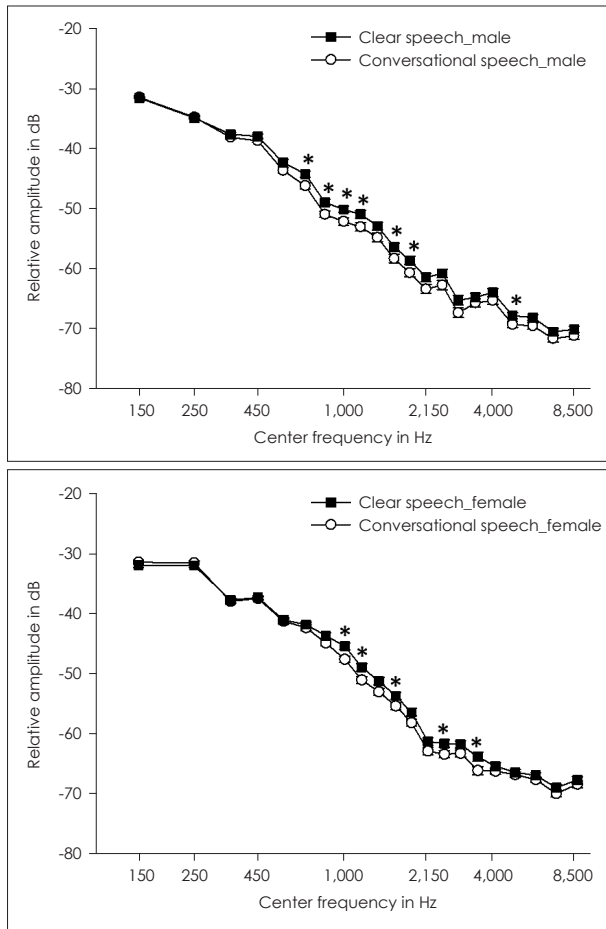
Fig. 1. Long-term average speech spectra for clear (square) and conversational (circle) speech as a function of center frequency (in Hz). The higher graph shows data obtained in the male group and the lower graph shows data obtained in the female group. Asterisks indicate statistical significance.

Hz for male speakers and between 840 Hz and 3,400 Hz for female speakers.

We found slower speech rates (35% to 45%) and increased levels of LTASS (1.22 dB to 2.26 dB) in the mid-frequency bands for Korean clear speech than for Korean conversational speech at the same normalized level (65 dB SPL). These findings are similar to those in reports on English clear speech [6,8]. English clear speech spoken at approximately 50% the rate of English conversational speech [6]. Moreover, increased levels of LTASS (1 to 3 dB) are evident in frequency areas between 1,000 Hz and 3,000 Hz in English clear speech [8]. Slow speech rates and increased LTASS in mid-frequency bands were obvious factors that contributed to improved intelligibility of clear speech in previous English studies [6,8]. Thus, results of this present study indicate that Korean clear speech has acoustical characteristics similar to those of English clear speech to improve speech intelligibility, such as speech rates and LTASS.

A comparison between the English and Korean speech rates indicated that Korean speech is faster than English speech in both clear and conversational speech [6]. This could be due to various reasons including a different scale (wpm vs. spm). Another possible reason is that the Korean syllable structure is simpler than the English syllable structure. The English syllable structure allows for more than two consonants to appear before and after a syllable nucleus, whereas the Korean syllable structure allows for only one consonant to appear before and after the syllable nucleus [16]. In other words, the number of phonemes that can be included in one syllable can be more in English than in Korean. Thus, the Korean speech rate can be faster than the English speech rate.

To be able to compare the rate between English and Korean that slows when using clear speech between the two, speech rates based on a similar unit were compared. The speech rate in English was measured by the wpm and counting the number of words based on the spacing in sentences [6]. '*Eogeol*' in Korean is a spacing unit that is composed of a content word (noun, verb, adjective or adverb) and a sequence of functional elements [17]. Although the structures of words between English and Korean are still different, *eogeol* in Korean seems to be a similar unit with the English word because it is based on spacing. For male speakers, the average speech rates for clear and conversational speech were 103.1 (SD=14.1) *eogeol* per miniute (epm) and 134.6 (SD=13.8) epm, respectively. The speech rates for the two different methods of speech production were significantly different (t=9.443, df=58, p<0.05). For female speakers, the average speech rates for clear and conversational speech were 89.2 (SD=11.13) epm and 124.7 (SD=9.83) epm. The speech rates for the two different methods of speech production were significantly different (t=14.152, df=58, p<0.05). Korean clear speech was spoken at rates that were as slow as 30.5% for male speakers and 39.8% for female speakers. Because English clear speech is spoken at approximately 50% of the rate of English conversational speech [6], the decrement rates of English clear speech seemed larger than the rates for Korean clear speech when compared to each language's conversational speech in terms of units based on spacing. Although the significant decrement rates of clear speech in both languages were evident, the degree of decrement rates did seem different. Thus, a further study is required to investigate the precise effects of decrement rate in Korean clear speech on speech intelligibility.

To record conversational speech, speakers were instructed to read sentences naturally as they would usually speak to a friend. To identify whether speech rates of this study are similar to those of normal conversational speech, results of this study were compared to those of a previous study that ana-

lyzed speech rates for Korean language [16]. Lee, et al. [16] investigated how speech rate in Korean is affected by socio-linguistic factors such as region, sex, and generation. The instruction of speech asked to be read sentences in his normal speed and tone. As the results for young adults who speak Korean standard language (Seoul speech), averaged speech rates for male speakers were 334.2 spm and those for female speakers were 314.4 spm. In this study, the averaged conversational speech rate for male speakers was 356.4 spm and that for female speakers was 330.2 spm. Because raw data from Lee, et al. [16] was not available, statistical comparison was not possible. In terms of quantitative comparison, the speech rates of this study and the rates of Lee, et al. [16] seem quite similar in both sexes. Results of this comparison indicate that speech stimuli for conversational speech by the instruction of this study show general speech rates of Korean conversational speech.

Although the overall trends of LTASS were similar between Korean and English clear speech, Korean clear speech seems to bring a broader boost of frequency ranges. In previous studies, clear speech was considered a production method involving efforts to speak clearly [1,2]. In terms of acoustic articulation modifications of clear speech, the speaker seems to make efforts to increase the amplitude of important frequency areas or areas that need more amplification for better intelligibility. If this assumption is reasonable, results of LTASS for Korean clear speech might be related to characteristics of LTASS and band-importance function of Korean. In the case of high frequency areas above 2 kHz, the LTASS was significantly lower for Korean than for English in reading passage conditions [11]. During clear speech, if the speaker's effort to include high frequencies that have less energy in normal vocalization is involved, a wider boost of amplitude compared to that in English clear speech may be one possible reason. With regard to low frequency areas below 1 kHz, low frequencies below 800 Hz were more important in Korean than in English because of the semi-tonal language characteristics of Korean [18]. During clear speech, if the speaker's effort to include low frequencies that are more important areas for speech intelligibility of Korean is involved, a wider boost of amplitude may be another possible reason. These reasons may explain why frequency ranges of the increased LTASS was wider for Korean clear speech than for English clear speech, but further analysis is required to confirm this.

In this study, speakers were instructed to read the sentences "while speaking clearly" when recording the clear speech. However, effects of improvement of speech intelligibility seem to vary depending on the instruction of clear speech. Lam and Tjaden [13] investigated effects of the improvement of

intelligibility for various instructions of English clear speech. They used four instructions: "speak to over-enunciate", "speak to a hearing loss person", "speak as clearly as possible", and "speak normally." The study found that the over-enunciated condition (when each word is over-enunciated) had the greatest intelligibility benefit. Although the present study identified different acoustic characteristics of Korean clear speech compared to those of conversational speech, only one instruction was applied. Therefore, further studies are required to determine which instruction provides the most beneficial effect on speech intelligibility in Korean clear speech. In this study, we demonstrated the possibility that Korean clear speech can contribute to improved intelligibility by analyzing the acoustic characteristics such as speech rate and LTASS. However, the extent to which Korean clear speech improves intelligibility remains unclear. Further studies on the intelligibility of Korean clear speech in hearing-impaired listeners in various environments are needed.

There are several limitations to this study. Although the results of this study provide a possibility that Korean clear speech provides intelligibility benefits because of its slower speech rates and higher mid-frequency amplitudes than those of conversational speech, the actual benefits remain unclear. Therefore, further studies are required to identify the actual intelligibility benefits of Korean clear speech as a next step. In addition, the degree of benefits may depend on the hearing sensitivity of listeners or background noise conditions [1,2]. Thus, various factors should be considered when studies assessing the intelligibility benefits of Korean clear speech are conducted. Also, there were sex differences in the LTASS of Korean clear speech (Fig. 1). Because sex differences of acoustical characteristics are not a new issue related to LTASS [11] and fundamental frequencies [19], LTASS can be different between male and female speakers in clear speech. However, the exact reason why different LTASS of clear speech was found in different sexes seems unclear. Therefore, further studies are required to evaluate the reasons for sex differences in LTASS of Korean clear speech.

This study showed that Korean clear speech has acoustical characteristics to improve speech intelligibility compared with conversational speech. The observed differences in the acoustic characteristics between the two types of speech production suggest that Korean clear speech can be an effective communication strategy to improve speech intelligibility in listeners with hearing loss or in noisy conditions.

A1A01052458).

Conflicts of interest

The authors have no financial conflicts of interest.

Author Contributions

Conceptualization: In-Ki Jin. Data curation: Jeeun Yoo, Hongyeop Oh, Seungyeop Jeong. Formal analysis: Jeeun Yoo. Funding acquisition: In-Ki Jin. Investigation: Jeeun Yoo and In-Ki Jin. Methodology: In-Ki Jin. Project administration: In-Ki Jin. Supervision: In-Ki Jin. Visualization: Jeeun Yoo. Writing—original draft: Jeeun Yoo and In-Ki Jin. Writing—review & editing: In-Ki Jin.

ORCID iDs

Jeeun Yoo   https://orcid.org/0000-0002-0217-8115
Hongyeop Oh   https://orcid.org/0000-0002-5493-4648
Seungyeop Jeong   https://orcid.org/0000-0002-9966-5736
In-Ki Jin   https://orcid.org/0000-0002-0834-5981

REFERENCES

1) Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. J Speech Hear Res 1985;28:96-103.
2) Schum DJ. Intelligibility of clear and conversational speech of young and elderly talkers. J Am Acad Audiol 1996;7:212-8.
3) Smiljanic R, Bradlow AR. Clear speech intelligibility: listener and talker effects [Internet]. Saarbrucken: ICPhS XVIth, Institute of Phonetics Saarland University; 2007 [cited 2019 Jan 7]. Available from: URL: https://www.researchgate.net/publication/228943878_Clear_speech_intelligibility_listener_and_talker_effects.
4) Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI. Speaking clearly for the hard of hearing IV: further studies of the role of speaking rate. J Speech Hear Res 1996;39:494-509.
5) Krause JC, Braida LD. Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. J Acoust Soc Am 2009;125:3346-57.
6) Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing. II: acoustic characteristics of clear and conversational speech. J Speech Hear Res 1986;29:434-46.
7) Krause JC, Braida LD. Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility. J Acoust Soc Am 2002;112(5 Pt 1):2165-72.
8) Krause JC, Braida LD. Acoustic properties of naturally produced clear speech at normal speaking rates. J Acoust Soc Am 2004;115:362-78.
9) Ferguson SH, Kewley-Port D. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. J Acoust Soc Am 2002;112:259-71.
10) Bradlow AR, Kraus N, Hayes E. Speaking clearly for children with learning disabilities: sentence perception in noise. J Speech Lang Hear Res 2003;46:80-97.
11) Noh H, Lee DH. How does speaking clearly influence acoustic measures? A speech clarity study using long-term average speech spectra in Korean language. Clin Exp Otorhinolaryngol 2012;5:68-73.
12) Kim JO. Acoustic characteristics of the voices of Korean normal adults by gender on MDVP. Phonetics and Speech Sciences 2009;1:147-57.
13) Lam J, Tjaden K. Intelligibility of clear speech: effect of instruction. J Speech Lang Hear Res 2013;56:1429-40.
14) Kim JS, Pae SY, Lee JH. Development of a test of Korean speech intelligibility in noise (KSPIN) using sentence materials with controlled word predictability. Speech Sciences 2000;7:37-50.
15) Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am 1977;61:1337-51.
16) Lee NR, Shin JY, Yoo DY, Kim KW. Speech rate in Korean across region, gender and generation. Phonetics and Speech Science 2017;9:27-39.
17) Choi KS, Isahara H, Kanzaki K, Kim H, Pak SM, Sun M. Word segmentation standard in Chinese, Japanese and Korean. In: Hammam Riza, editor. Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009; 2009 Aug 6-7; Suntec, Singapore: Association for Computational Linguistics; 2009. p.179-86.
18) Jin IK, Kates JM, Lee K, Arehart KH. Derivations of the band-importance function: a cross-procedure comparison. J Acoust Soc Am 2015;138:938-41.
19) Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. J Acoust Soc Am 1990;87:820-57.