Comparison of Long-Term Average Speech Spectra in Reading Context and Spontaneous Speech

Kyungju Lee[1], In-Ki Jin[2]

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

Purpose: The purpose of this study was to compare the long-term average speech spectra (LTASS) between reading context and spontaneous speech.

Methods: A continuous discourse for reading context and answers of an interview for the spontaneous speech were recorded by twenty speakers using a recording system in random order. Recorded stimuli were normalized to a root-mean-square (RMS) level of 65 dB SPL. Then, LTASS were analyzed as a function of frequency.

Results: There were no statistical differences between the two different styles in both male and female participants. Differences between the two styles were less than 2.1 dB in male participants and less than 2.9 dB in female participants across all frequency areas.

Conclusions: Results of this study indicate that reading context can be considered as a form to reflect the speech understanding in real life like spontaneous speech.

Keywords: LTASS, Acoustic characteristics, Korean stimuli

INTRODUCTION

Speech audiometry is an important test battery for measuring hearing ability [1,2]. Speech audiometry tests can be used to identify the reliability with pure-tone average thresholds or evaluate speech understanding in real life. When the speech audiometry is conducted, there are two stimulus presentation modes: One is a recorded voice mode and the other is a monitored live-voice mode [3]. Although the monitored live-voice mode is able to reflect better speech understanding in real life, a recorded voice mode is recommended in general because of the test reliability and consistency [4].

To test hearing aid (HA) outcomes for people with hearing loss, speech audiometry like hearing-in-noise test (HINT) is also one of important test battery [5]. The HINT is a test to measure speech understanding in noise conditions using reading context stimuli. Not only the HINT but also other standardized speech test materials like the Korean standard sentence lists for adults (KS-SL-A) have been tested by the recorded reading context stimuli [6]. The reading context refers to a speaking style like a person reads sentences. Thus, reading context is distinguished from spontaneous speech which refers to a speaking style like a person talks with other person naturally in terms of the naturalness of the vocalization [7].

This study considered acoustic characteristics between reading context and sponta-

neous speech. Most speech tests have used reading context stimuli [5,6]. Because the style of vocalization is different between reading context and spontaneous speech, however, we are unsure whether test results using reading context stimuli can reflect speech understanding in real life.

The long-term average speech spectrum (LTASS) is one of the important factors that determine the acoustic characteristics of speech [8]. The LTASS represents speech energy across the frequency in decibel (dB). Picheny et al. [9] reported that intelligibility scores measured by clear speech which is a method to speak as clear as possible were higher than the scores measured by normal speech for people with hearing loss. According to the result of acoustical analysis, the characteristics of the LTASS was different between two speaking styles [10]. Specifically, increased mid-frequency speech energy was apparent in the clear speech compared to the normal speech [10]. Thus, the LTASS can be an important factor to compare acoustic features between different speaking styles.

The purpose of this study was to compare the LTASS between reading context and the spontaneous speech. If different acoustic characteristics are found between the two speaking styles, the differences should be considered when test results by the reading context are interpreted. If there is no difference between the two different styles, the reading context can be considered as a form that reflects the speech understanding in real life.

METHODS

Participants

Ten male and ten female native Korean speakers aged between twenty and twenty-eight years participated in this study (mean=22.98 years). Pure-tone audiometry was conducted prior to the study, using a GSI 61 (Grason-Stadler, Eden Prairie, MN, USA). All participants had normal hearing sensitivity and none had any speech problems such as articulation disorders. Air-conduction thresholds and bone-conduction thresholds were within normal range (<20 dB) measured at 0.25, 0.5, 1, 2, 4, 8 kHz. Also, tympanometry test results were within normal range showing 'A' type. All participants recorded their voices in a double-walled sound booth for quality of recording. All participants were given written informed consent to participate in the study and researchers provided a sufficient explanation such as the purpose and method of study. All participants were offered compensation for their

participation (#HIRB-2016-006).

Test stimuli

To compare LTASS between reading context and spontaneous speech, a continuous discourse was used for reading context and a short interview was conducted for spontaneous speech. For reading context, all participants read a continuous discourse and their voices were recorded. A short story, 'The wind and the sun' (Borim, Kyeonggi-do, Paju), which is easy to read, was selected as the continuous discourse. For spontaneous speech, a short interview (15 questions) was conducted and participants' answers were recorded. Questions from 'Seoul Corpus' were chosen [11]. Questions such as 'What do you usually do when you are with your friends?', 'What do you think about the merits of smartphones?' and 'Could you tell me about your family background?' were included. Table 1 shows 15 interview questions for spontaneous speech.

Procedure

The recording was conducted using a computerized speech lab (CSL, KayPECTAX™, Montvale, NJ, USA) in a sound booth. All participants' voices were recorded at the same distance (10 cm) from the recording microphone (Sennheiser e-835s, Wedemark, Germany). The recording files were digitized at a 44,100 Hz with 16 bit. When recording reading context, all participants were instructed that they should read comfortably at normal speed. When an error occurred during record-

Table 1. Interview questions for spontaneous speech. Questions from 'Seoul Corpus' were chosen (Yun et al., 2015)

1	Could you tell me about your family background?
2	How do you spend time with your family when you are at home?
3	What do you do when your family gets together at a holiday?
4	Do you help out with chores around the house?
5	Who do you talk to more, your mother or father?
6	Who is your best friend?
7	What do you usually do when you are with your friends?
8	What are your interests when you meet your friends these days?
9	What is your hobby?
10	What are you studying now?
11	Have you ever thought about your future career?
12	Which season do you like best?
13	Tell me about your recent trip
14	What is your favorite drama or movie?
15	What do you think about the merits of smartphones?

ing such as mispronunciation or hesitation of speaking, those sentences were rerecorded. Before recording spontaneous speech, participants received an instruction that they should answer the questions as usual. After the researcher asked each question, the participants answered freely for 40 seconds. If participant went over their allotted time, the recording files were cut to 40 seconds.

To analyze the LTASS, each participant's recording file was concatenated without silence using Adobe Audition (adobe systems, version 3.0). Then, the recording files were normalized to a root-mean–square (RMS) level of 65 dB SPL using Praat v6.0.19 [12]. The LTASS of reading context and spontaneous speech were analyzed using the CSL. The results of LTASS ranging from 100 Hz to 10,000 Hz were used for analysis.

Statistical Analysis

A univariate analysis of variance (ANOVA) was used to identify whether there are significant differences in across-speaking style. The statistical analysis was conducted using SPSS-22 (IBM SPSS Inc., Armonk, NY, USA).

RESULTS

Results of LTASS for comparison of spontaneous speech and reading context in each gender are shown in Figure 1. The LTASS of spontaneous speech were marked with solid lines and the LTASS of reading context were marked with dotted lines. In both male and female speakers, there were no statistical differences between spontaneous speech and reading context in all frequency areas ($p > 0.05$). For male speakers,
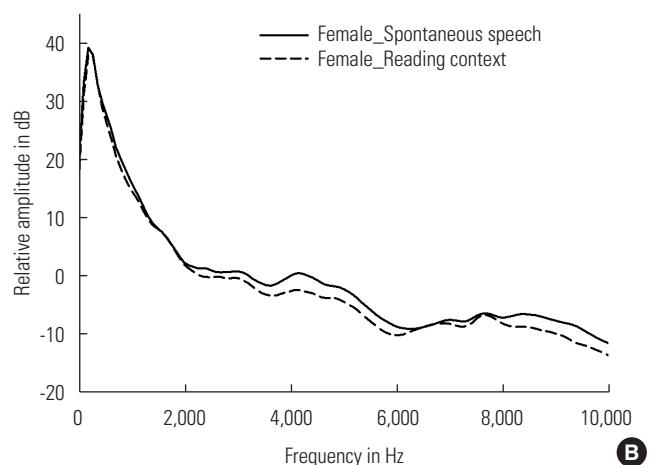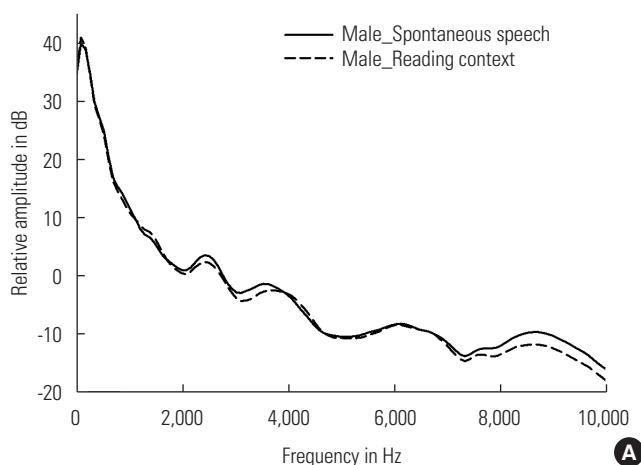

Figure 1. Averaged LTASS for spontaneous speech (solid line) and reading context (dotted line): (A) male and (B) female participants.
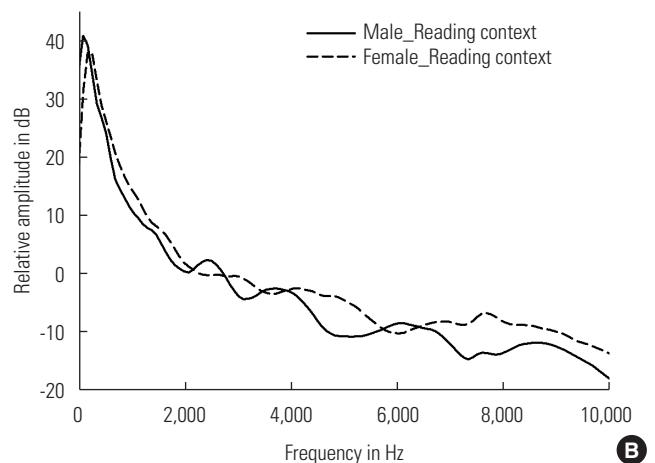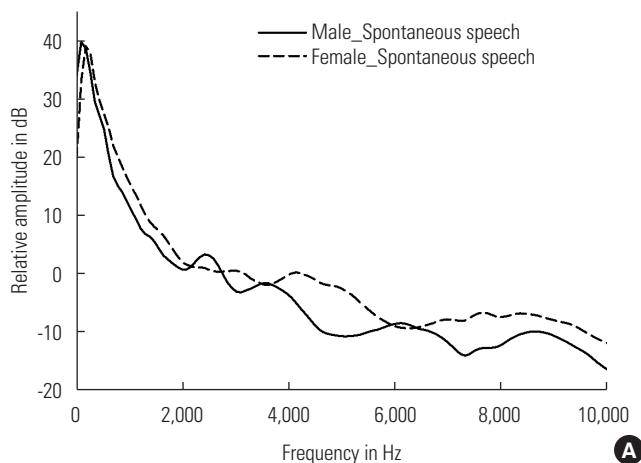

Figure 2. Averaged LTASS for male (solid line) and female (dotted line) participants: (A) spontaneous speech and (B) reading context.

the difference of the LTASS below 8,000 Hz was less than 1.6 dB and the difference of the LTASS above 8,000 Hz was less than 2.1 dB between the two styles. For female speakers, the difference of the LTASS was less than 2.9 dB between the two styles.

As a secondary comparison, results of LTASS for gender difference between spontaneous speech and reading context are shown in Figure 2. In both speaking styles, there were no statistical differences ($p > 0.05$) except for two frequency areas. In spontaneous speech for male and female speakers, there were significant differences in ranges between 4,200 Hz to 5,400 Hz and between 7,200 Hz to 8,100 Hz, respectively. Mean difference values of LTASS were 7.2 dB and 5.6 dB in male and female speakers, respectively ($p < 0.05$). In reading context for male and female speakers, there were significant differences in ranges between 4,500 Hz and 5,200 Hz and between 7,200 Hz and 8,000 Hz, respectively ($p < 0.05$). The mean differences were 6 dB and 6.2 dB in male and female speakers, respectively.

DISCUSSION

The current study compared the LTASS between spontaneous speech and reading context. As a result, there were no statistical differences between the two different styles in both male and female participants. Differences between the two styles were less than 2.1 dB in male participants and less than 2.9 dB in female participants across all frequency areas.

In the current study, different LTASS were found in several frequency areas between male and female participants. The result of the current study correlated to results from other studies that reported the LTASS. For example, Byrne et al. [8] reported that there were gender differences of the LTASS in few frequency areas above 4,000 Hz in several languages. In the study from Noh and Lee [13], different LTASS were also found between Korean male and female speakers in several frequency areas above 4,000 Hz. Because the importance of frequency to speech intelligibility accounts for over 80% in frequency areas below 4,000 Hz [14], however, we cannot conclude that results of the gender difference of the LTASS above 4,000 Hz may lead different performances for the speech understanding. Thus, further studies may require to identify the impact of gender differences of the LTASS in frequency areas above 4,000 Hz on the speech understanding.

There are a few limitations for the current study. Although LTASS were similar between the two styles, other factors like

formant characteristics may affect speech [10]. For example, Godoy et al. [10] compared formant characteristics in English vowels for different speaking styles such as clear speech and conversational speech. As a result, different formant characteristics (f1 and f2) were found between different speaking styles. Because the intelligibility performance of clear speech was different from the performance of conversational speech, they concluded that formant characteristics caused by different speaking styles may affect speech understanding performance [10]. The result of the current study provide evidence that the acoustic characteristics for the two styles are similar. However, impact of other factors that may affect the speech understanding may be considered for further study.

In the current study, LTASS for spontaneous speech and reading context were similar in male and female speakers. Although there is a possibility that other factors that affect the speech understanding may be different between the two styles, results of this study indicate that reading context can be considered as a form to reflect the speech understanding in real life like spontaneous speech in terms of the LTASS.

REFERENCES

1. Martin FN, Sides DG. Survey of current audiometric practices. ASAH. 1985 Feb;27(2):29.
2. Martin FN, Armstrong TW, Champlin CA. A survey of audiological practices in the United States. Am J Audiol. 1994 Jul;3(2):20-6.
3. Carhart R. Speech reception in relation to pattern of pure tone loss. J SpeechHear Dis. 1946 Jun;11(2):97-108.
4. Roeser RJ, Clark JL. Live Voice Speech Recognition Audiometry-Stop the Madness!. Audiology Today. 2008;20(1):32.
5. Moon SK, Kim SH, Mun HA, Jung HK, Lee JH, Choung YH, et al. The Korean hearing in noise test. Int J Audiol. 2008;47(6):375-6.
6. Jang HS, Lee JH, Lim DH, Lee KW, Jeon AR, Jung EJ. Development of Korean Standard Sentence Lists for Sentence Recognition Tests. Audiology and Speech Research. 2008; 4(2):161-77.
7. Llisterri J. Speaking styles in speech research. In: ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language, Dublin, Ireland. 1992 Jul.
8. Byrne D, Dillon H, Tran K, Arlinger S, Wilbraham K, Cox R, et al. An international comparison of long-term average speech spectra. J Acoust Soc Am. 1994 Jul;96(4):2108-20.
9. Picheny MA, Durlach NI, Braida LD. Speaking Clearly for the Hard of Hearing Intelligibility Differences between Clear and Conversational Speech. J Speech Lang Hear Res. 1985 Mar;28(1):96-103.
10. Godoy E, Koutsogiannaki M, Stylianou Y. Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles. Comput Speech Lang. 2014;28(2): 629-47.
11. Yun WH, Yoon KC, Park SW, Lee JH, Cho SM, Kang DS, et al. The Korean Corpus of Spontaneous Speech. Daegu: Industry-Academ-

ic Cooperation Foundation, Keimyung University (Distributor).

12. Boersma P, Weenink D. Praat software. University of Amsterdam. Im Internet: http://www. fon. hum. uva. nl/praat. 2016.

13. 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(2):68-73.

14. Jin IK, Kates JM, Lee KW, Arehart KH. Derivations of the band-importance function: A cross-procedure comparison. J Acoust Soc Am. 2015 Aug;138(2):938-41.