The Effects of Fundamental Frequency Contours on the Intelligibility Benefit of Clear Speech in Native Speakers of American English and Native Speakers of Seoul Korean

A Dissertation

SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

BY

Heekyung Han

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Robert S. Schlauch, Ph.D.

August 2019
Acknowledgements

I would like to express special thanks to my advisor, Dr. Robert Schlauch. You have not only been a tremendous mentor to me, but also you have shown me how to be a good mentor to other people. Thank you for your gracious patience, constant encouragement, and immense knowledge throughout the years of my Ph.D. program.

I would also like to thank the other members of my committee, Dr. Benjamin Munson, Dr. Peggy Nelson, and Dr. Peter Watson, for their invaluable feedback. Your insightful comments and helpful suggestions made this work markedly stronger.

I am also extremely thankful to my lab assistants, Emily Roznoski, Madison Murphy, Benjamin Morehead, and Emily Chu for their assistance with data analysis. Your assistance was crucial in completing this project.

My appreciation also extends to the Liberal Arts Technologies and Innovation Services (LATIS) and Research Methodology Consulting Center (RMCC) for their statistical consultation and to the Center for Applied and Translational Sensory Science (CATSS) for their technical support for my data collection.

A special thanks also goes to my friends and colleagues in Speech-Language-Hearing Sciences for their moral support and encouragement.

Finally, and most importantly, I would like to express immense gratitude to my family, my husband, my parents, and my parents-in-law for their tremendous love and support and also to Jesus Christ for His grace and mercy on me.
Table of Contents

List of Tables ........................................................................................................ v
List of Figures .......................................................................................................... vii
List of Appendices .................................................................................................... ix

Chapter 1: Overview ........................................................................................................ 1

Chapter 2: The Effects of Fundamental Frequency Contours on the Intelligibility Benefit of Clear Speech in Native English Speakers (Experiment 1) ....................... 3
I. Introduction ................................................................................................................ 3
II. Method ....................................................................................................................... 9
   A. Production Study ................................................................................................... 9
      a. Participants (Talkers) ....................................................................................... 9
      b. Stimuli ............................................................................................................. 9
   B. Perception Study .................................................................................................15
      a. Participants (Listeners) ...................................................................................15
      b. Speech Materials ............................................................................................16
      c. Data Collection ..............................................................................................16
   C. Data Analysis ......................................................................................................18
      a. Production Study ............................................................................................18
      b. Perception Study ............................................................................................18
III. Results .................................................................................................................. 19
   A. Production Study ................................................................................................. 19
      a. Speaking Rate .................................................................................................. 19
      b. F0 Features ..................................................................................................... 21
      c. Long-term spectra ......................................................................................... 23
      d. Vowel Space .................................................................................................. 25
      e. Vocal Intensity Level ..................................................................................... 31
   B. Perception Study .................................................................................................. 32

IV. Discussion .............................................................................................................. 36

V. Conclusion ............................................................................................................... 41

Chapter 3: The Effects of Fundamental Frequency Contours on the Intelligibility
Benefit of Clear Speech in Native Korean Speakers (Experiment 2) .................. 42
I. Introduction ............................................................................................................. 42
II. Method .................................................................................................................... 50
   A. Production Study ................................................................................................. 50
      a. Participants (Talkers) .................................................................................... 50
      b. Stimuli ........................................................................................................... 51
   B. Perception Study ................................................................................................ 57

Speech Intelligibility
   a. Participants (Listeners) ..................................................................................... 57
   b. Speech Materials ............................................................................................. 58
   c. Experimental Setup .......................................................................................... 58
   d. Experimental Procedure ................................................................................... 59
Speech Naturalness

a. Participants (Listeners) ................................................................. 60
b. Speech Materials ........................................................................ 60
c. Experimental Setup .................................................................... 61
d. Experimental Procedure ............................................................ 61
C. Data Analysis ............................................................................... 62
   a. Production Study ....................................................................... 62
   b. Perception Study ....................................................................... 62

III. Results ....................................................................................... 63
   A. Production Study ....................................................................... 63
      a. Speaking Rate ......................................................................... 63
      b. F0 Features ............................................................................. 65
      c. Long-term Spectra ................................................................. 69
      d. Vowel Space .......................................................................... 71
      e. Vocal Intensity Level ............................................................. 77
   B. Perception Study ...................................................................... 78
      a. Speech Intelligibility ............................................................... 78
      b. Speech Naturalness ............................................................... 81

IV. Discussion .................................................................................. 83
V. Conclusion ................................................................................... 89

Bibliography .................................................................................... 90

Appendices ...................................................................................... 96
List of Tables

Table 1. Words per minute (wpm) for each talker in conversational (CNV) and clear speech conditions ................................................................. 19

Table 2. Median, range, and variation (SD) of F0 values in Hz for each talker in clear and conversational speech conditions ......................................... 22

Table 3. Average band RMS level across the frequency range between 0.63- and 6-kHz for each talker in each speech condition. The third-octave bands centered at 0.63- and 0.8-kHz contributed a fractional amount so this average level is referred to in the text as a “high-frequency average.” ................................................................. 24

Table 4. Main effects of vowel and type and interaction between vowel and type for F1 values (reference vowel: /a/) .......................................................... 28

Table 5. Main effects of vowel and type and interaction between vowel and type for F2 values (reference vowel: /a/) .......................................................... 30

Table 6. Main effects of type for the average Euclidean distance ........................................ 31

Table 7. Individual talkers’ vocal intensity levels (SPL) in conversational and clear speech conditions ............................................................................. 32

Table 8. A total number of sentences used for intensity analysis over time within sentences (ms). Times listed represent midpoints of 300 ms analysis window. ........................................ 35

Table 9. Language experience and proficiency for each participant .................................... 50

Table 10. Words per minute (wpm) for each talker in conversational, clear, and exaggerated-F0 speech conditions ................................................. 63

Table 11. Median, range, and variation (SD) of F0 values in Hz for each talker for each speaking style in L1 (Korean) and L2 (English) production ........................................ 66

Table 12. Average band RMS level across the frequency range between 0.63- and 6-kHz for each talker in each speech condition. The third-octave bands centered at 0.63- and 0.8-kHz contributed a fractional amount so this average level is referred to in the text as a “high-frequency average.” ................................................................. 70
Table 13. Main effects of vowel and type and interaction between vowel and type for F1 values (reference vowel: /a/) ............................................................... 74

Table 14. Main effects of vowel and type and interaction between vowel and type for F2 values (reference vowel: /i/) ............................................................... 76

Table 15. Main effects of type for the average Euclidean distance ........................................ 76

Table 16. Individual talkers’ vocal intensity levels (dB SPL) in conversational, clear, and exaggerated-F0 speech conditions .............................................................................. 77

Table 17. A total number of sentences used for analyzing relative dB over time within sentences (ms). Times listed represent midpoints of 300 ms analysis window ...................... 81

Table 18. Mean speech naturalness scores for each talker group for each speaking style. (Scale 1=extremely unnatural to 7=extremely natural) ................................................................. 82
List of Figures

**Figure 1.** F0 contours for the sentence “The bill was paid every third week.” in three different speech conditions (T4 production) ........................................................................................................... 14

**Figure 2.** Long-term spectral distribution for conversational (Conv) and clear (Clear) speech in each talker.................................................................................................................................................. 23

**Figure 3.** F1/F2 values (Bark) of four corner vowels (/i/, /u/, /a/, and /æ/) for each talker and the average Euclidean distance (Bark) of each vowel from the grand mean F1/F2 value across talkers in clear (CLR) and conversational (CNV) speech ......................................................... 27

**Figure 4.** Type and vowel interaction for F1 (left panel) and F2 (right panel) .......................... 29

**Figure 5.** Average intelligibility scores for each talker and grand mean scores in conversational, clear, and F0-manipulated speech conditions...................................................... 33

**Figure 6.** Keyword recognition scores across temporal keyword position (left panel) vs. relative intensity levels over time within sentences (right panel). Times listed in the right panel represent midpoints of 300 ms analysis window............................................................... 34

**Figure 7.** F0 contours for the sentence “The bill was paid every third week.” in three different speech conditions (T4 production) ........................................................................................................... 55

**Figure 8.** Long-term spectral distribution for conversational (Conv), clear (Clear), and exaggerated-F0 (Exagg) speech in each talker .................................................................................................................. 69

**Figure 9.** F1/F2 values (Bark) of vowels for each talker and average Euclidean distance (Bark) of each vowel from the grand mean F1/F2 values across talkers in conversational (CNV), clear (CLR), and exaggerated-F0 (EXG) speech ................................................................................................................... 72

**Figure 10.** Type and vowel interaction for F1 (left panel) and F2 (right panel) ...................... 75

**Figure 11.** Average intelligibility scores for each talker and grand mean scores in conversational (CNV), clear (CLR), exaggerated-F0 (Exagg-F0), and F0-manipulated (F0-MNP) speech conditions ............................................................................................................. 78
Figure 12. Keyword recognition scores across temporal keyword position (left panel) vs. relative intensity levels over time within sentences (right panel). Times listed in the right panel represent midpoints of 300 ms analysis window.
List of Appendices

Appendix A ................................................................................................................................. 96
Appendix B ............................................................................................................................... 104
Appendix C ............................................................................................................................... 105
Chapter 1: Overview

Clear speech, which is characterized by specific acoustic changes that are distinguished from ordinary conversational speech, is a speaking strategy that enhances a talker’s intelligibility in adverse listening conditions. An increase in the range of a talker’s fundamental frequency (F0) is known as one of the several acoustic changes that is observed when clear speech is produced. Although an increase in F0 range is often seen in clear speech, its contribution to the clear speech benefit is unknown.

Experiment 1 in this dissertation examined whether an increase in F0 variation contributes significantly to the clear speech benefit in native speakers of American English. Experiment 2 evaluated clear speech effects in native speakers of Seoul Korean who started to learn English after the age of six and also examined the role of F0 variation.

The clear speech benefit was measured by having talkers produce sentences in a conversational and a clear speaking style. The stimuli for these experiments were produced by several talkers and were recorded digitally. At the time of recording, participants were instructed to read aloud low-context sentences in conversational (Experiments 1 & 2), clear (Experiments 1 & 2), and exaggerated-F0 (only Experiment 2) speaking styles. The exaggerated speaking style, which is similar to infant-directed speech with a wide range of F0, was a condition given to the native Korean talkers because Koreans do not typically vary their F0 much in various speaking styles. To characterize acoustic-phonetic changes at the sentence level in talkers’ productions, five acoustic changes were measured: speech rates, long-term spectra, F0 distribution, vowel formant frequencies, and vocal intensity levels.

Sentences from the talkers were presented to native listeners of American English in a perceptual study. F0-manipulated speech was synthesized from the clear speech (Experiment 1) and from the exaggerated-F0 speech (Experiment 2) to examine whether F0 variation is a contributing factor in the intelligibility benefits in native English speakers and in native
Korean speakers. This was accomplished by compressing the F0 contours of clear speech to match those of conversational speech in Experiment 1 and by compressing the F0 contours of exaggerated-F0 speech to match those of conversational speech in Experiment 2. Listeners were randomly presented with sentences in different speaking styles and asked to type in the sentence after orally repeating each sentence that they heard. The percentage of correct keywords was calculated for each speaking style.

The data revealed that F0 range did not contribute to the clear speech benefit. The exaggerated-F0 speech condition for the Korean talkers showed slightly poorer intelligibility benefit than the clear speech condition. A follow-up study of speech naturalness revealed that clear speech is more natural than exaggerated-F0 speech. However, a significant correlation between intelligibility and speech naturalness was not found.

Although the experiments were designed to examine directly the role of F0 range on the clear speech benefit, the recordings and perceptual data provided opportunities to study other perceptual correlates of this phenomenon. The primary acoustic factor contributing to the clear speech benefit for native English and native Korean talkers was an increase in the intensity of high-frequency speech sounds.
Chapter 2: The Effects of Fundamental Frequency Contours on the Intelligibility Benefit of Clear Speech in Native English Speakers (Experiment 1)

I. Introduction

Clear speech is a listener-oriented speaking style in which talkers increase articulatory effort and communication precision in various listening situations, such as when speaking with someone in a noisy environment, when talking to a person who learns a second language, or when talking to a person with hearing problems (e.g., Uchanski, 2005; Smiljanić & Bradlow, 2007, 2009; Searl & Evitts, 2013; Hazan et al., 2015). Clear speech is produced with specific acoustic changes that are distinguished from ordinary conversational speech. In this speaking style, talkers typically exhibit a decrease in speaking rates (Picheny et al., 1986; Uchanski et al., 1996; Bradlow et al., 2003; Liu & Zeng, 2006), insertion of pauses between words (Picheny et al., 1986; Bradlow et al., 2003; Liu & Zeng, 2006), an increase in vocal intensity levels (Picheny et al., 1986; Lam et al., 2012), a high-frequency emphasis in amplitude spectrum (Picheny et al., 1986; Krause & Braida, 2004, 2009), expansion of vowel space (Picheny et al., 1986; Ferguson & Kewley-Port, 2002; Bradlow et al., 2003; Ferguson & Quene, 2014), and an increase in fundamental frequency (F0) range (Picheny et al., 1986; Bradlow et al., 2003).

Numerous studies have shown that clear speech improves a talker’s intelligibility among native English speakers (e.g., Picheny et al., 1985; Payton et al., 1994; Uchanski et al., 1996; Krause & Braida, 2004). This intelligibility benefit is observed for nearly all talkers and is substantial (10 -34 percentage points) for most
talkers under a variety of conditions that include listeners with sensorineural hearing loss (Picheny et al., 1985; Payton et al., 1994; Schum, 1996; Uchanski et al., 1996; Liu et al., 2004), and for normally hearing listeners in noise (Uchanski et al., 1996; Bradlow et al., 2003; Liu et al., 2004) and in noise with reverberation (Payton et al., 1994). This indicates that talkers are able to enhance the intelligibility of their speech both for hearing-impaired listeners and for normal-hearing listeners in difficult listening situations.

There has been growing interest in perceptual correlates of the acoustic changes in clear speech. Uchanski et al. (1996) explored the relationship between pause structure and the clear speech benefit in listeners with and without hearing loss. The role of pause structure in clear speech was examined by removing the key words from conversational and clear sentences and measuring the intelligibility of keywords in isolation. The results showed that the pause structure barely affected the clear speech benefit in both listener groups. This outcome was supported by their second experiment in which they deleted and inserted pauses to match the pause duration of clear speech with that of conversational speech. The results showed that both insertion and deletion of pauses degraded the intelligibility scores in noise backgrounds for listeners with normal hearing. These outcomes suggest that the pause structure does not necessarily account for the clear speech benefit.

Krause and Braida (2002) investigated the relative importance of speech rates on the clear speech benefit. To examine whether slow speaking rates are essential for producing the clear speech advantage, five selected talkers produced conversational and clear speech at slow, normal conversation, and fast rates. The results showed that the intelligibility benefit can be achieved in clear speech produced at normal
speaking rates. Although the benefits were smaller (i.e., 14 percentage points) when compared with those at slow speaking rates (i.e., 18 percentage points), there were still substantial benefits from naturally produced clear speech at normal speaking rates. This outcome was consistent with the findings in other studies (Uchanski et al., 1996; Liu & Zeng, 2006). These results suggest that reduced speaking rates are not solely responsible for the clear speech benefit.

Krause and Braida (2004) investigated other acoustic properties of naturally produced clear speech at normal speaking rates. Sentence materials for the acoustical analysis were obtained from the five talkers in their previous study (Krause & Braida, 2002). They found that the energy in the 1- to 3-kHz range of long-term spectra increased in the naturally produced clear speech at normal speaking rates. This increase in the high frequencies appeared to be correlated with the improvement in speech understanding in speech-spectrum background noise. However, this finding was not fully supported in a follow-up study (Krause & Braida, 2009). They investigated the relative importance of the high-frequency energy (i.e., energy between 1- and 3-kHz) in the clear speech benefit by boosting the energy of voiced segments of conversational speech. The results showed that increased high-frequency energy improved intelligibility of the conversational speech for two talkers whereas it became worse for two others.

Although several studies examined the relative contributions of acoustic properties to the intelligibility benefit of clear speech, there has been little research on a direct effect of F0 range on the clear speech benefit. There has been a study that reported a tendency for a wider F0 range to correlate with more intelligibility benefit (Bradlow et al., 1996), but it has not yet been fully established whether the wider F0
range is a primary contributor to the intelligibility benefit of clear speech. Picheny et al. (1986) had three male talkers produce clear and conversational speech to explore the acoustic characteristics of clear speech. The results showed that all the three talkers increased their F0 range in clear speech. This outcome was consistent with the findings in Bradlow et al. (2003) where the average F0 range for a male and a female talker was increased in clear speech. In particular, the male talker showed the bigger change in F0 range for clear speech than the female talker. The same pattern was observed in the study by Krause and Braida (2004) where the male talker showed a considerable increase of F0 range for naturally produced clear speech at slow and normal rates, when compared with conversational speech at normal rates. The other four female talkers, however, showed no or relatively small increase in F0 range for the two types of clear speech, when compared with conversational speech at normal rates. These outcomes suggest that not everyone uses the increased F0 range as a clear speech strategy.

Previous studies have shown that manipulations of F0 contours can affect speech understanding (e.g., Binns & Cullings, 2007; Miller et al., 2010; Shen & Souza, 2017a, 2017b). Miller et al. (2010) modified F0 contours of low-predictability sentences that were produced by five female native English talkers to investigate whether the F0 contour manipulations affect speech intelligibility in speech-shaped noise for normal hearing listeners. They found that manipulations that reduced (flattened), exaggerated (increased F0 range by 1.75), or made the F0 contour unnatural (sinusoidal frequency modulated or inversed) reduced intelligibility. The same trend was found in the study of Binns and Cullings (2007) in which speech recognition thresholds (SRTs) were obtained at five different F0
manipulations, including inverse, flat (monotone), quarter, half, and standard F0 contours, under speech-shaped noise. The speech materials were low-context sentences (Rothauser et al., 1969) that were produced by an adult male talker. The results showed a trend for performance degradation with a decrease of F0 variability. Although there was no statistically significant difference in performance except for when contours were inversed, a trend for diminution in performance was found for the monotone and quarter (25% of full range) condition. In other studies, a reduction in speech intelligibility was also observed compared to the “natural” condition when the F0 range is increased by a factor of 1.75. This is true for listeners with and without hearing loss (Miller et al., 2010; Shen & Souza, 2017a, 2017b). The reduction observed might be a result of the large increase in the dynamic pitch range of female talkers in these studies. Watson and Schlauch (2008) state that the misalignment of source harmonics with vocal tract resonances is more likely for very high F0 values which would be consistent with poorer speech understanding. Support for this comes from a study by Clarke et al. (2017) who found that a male talker with an expanded F0 range did not show a reduction in speech understanding.

As shown above, most research on F0 and speech understanding done to date has focused on the effect of manipulations of F0 contours on speech understanding. Most previous studies reported that the speech intelligibility seems to be negatively affected when the original F0 contour of speech was flattened or made unnatural. These studies did not specify the speaking style of the recorded speech materials. It still remains an unsettled question whether an increase in F0 range, which is naturally produced in clear speech, is a significant contributor to the intelligibility benefit of clear speech in native English speakers. To examine the role
of F0 variation in the clear speech benefit four adult male speakers, who showed a significant difference in F0 range between conversational and clear speech, were selected. This experiment will answer the question of whether an increase of F0 range is a primary factor that contributes to the intelligibility benefit of clear speech or a secondary effect of other clear-speech acoustic changes (e.g., slow speaking rates).
II. Method

A. Production Study

a. Participants (Talkers)

Four adult male native speakers of Midwestern American English (T1 - T4), aged from 25 to 35 (mean: 30.3, SD=5.0), were recruited. All of the native English speakers had normal hearing at octave frequencies between 500 and 4000 Hz and self-reported no history of a speech-language disorder. A single session of the experiment lasted for approximately 1.5 hours including breaks. All participants were monetarily compensated for their time and effort at the end of the session.

Three other participants were recruited, but their speech production was not used for this study since the change in F0 range for clear speech did not meet a minimum requirement of a ratio of 1.75 to 1 for the range of clear to conversational speech. To test the idea that F0 range is an important factor contributing to the clear speech benefit, we set a criterion of a large change in this present study.

b. Stimuli

Selection of Speech Materials. Two-hundred forty low-predictability sentences were selected from the IEEE/Harvard sentence corpus (Rothauser et al., 1969). The sentences, in which each had five keywords, were selected based on word difficulty and frequency. The difficulty of words in the sentence corpus was evaluated by one adult native speaker of American English and five adult native speakers of Korean. All of the native Korean speakers started to learn
English after the age of six and lived in English-speaking countries for 5 years on average (range: 4 - 6 years). The inclusion of the Korean-speaking second language (L2) learners of English in the evaluation task was for selection of sentences that would be appropriate for use with native Korean-speaking populations in Experiment 2.

Based on their lexical knowledge, 431 sentences were removed from the original corpus of 720 sentences because these sentences contained a word or words that were unfamiliar to any of them. Of the remaining 289 sentences, 240 sentences were selected based on the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996). Their mean value of word frequency (mean: 182,767, range: 1,730 – 12,661,276) was higher than that of the Northwestern University Auditory Test No. 6 (NU-6) list 1A (mean: 70,905, range: 182 – 980,925) and lower than that of the Phonetically Balanced Kindergarten (PBK) test (mean: 346,551, range: 546 – 6,474,135). A higher number represents a more commonly occurring word. The examination of word difficulty and frequency was exclusively limited to the keywords in sentences.

**Recording Equipment.** Participants were recorded individually in a double-walled sound-isolated booth. All speech materials were recorded on a Marantz Professional PMD671 digital recorder with an AKG head-mounted condenser microphone. The microphone was placed approximately 10 cm from a participant’s mouth. Single-channel recordings were collected at a sampling rate of 44,100 Hz with 16-bit quantization, which is CD quality. In order to ensure enough gain without distortion or clipping, the author manually adjusted
the input level at the beginning of the recording session and maintained that level until each speech condition terminated.

The identical settings on the recorder input level were applied for recording a 1-kHz reference tone, which was used for estimating the level of vocal intensity in each speaking style. The reference tone was presented at 80 dB sound pressure level (SPL) by a loudspeaker (BOSE Soundlink Color II Bluetooth Speaker) and recorded on the Marantz Professional PMD671 digital recorder with AKG head-mounted condenser microphone. Microphones of the sound-level meter and the recorder were placed at a distance of 10 cm from the loudspeaker, the same distance as the talker’s mouth.

**Recording Sentence Productions.** Three types of speech materials were created: conversational, clear, and F0-manipulated speech. Conversational and clear speech were recorded by four adult male native speakers of Midwestern American English. The F0-manipulated speech was synthesized from the clear speech condition.

For conversational and clear speech tasks, individuals were seated in front of a computer monitor and wore a head-mounted condenser microphone. Participants were assigned 60 different sentences with three practice sentences and instructed to produce the sentences under two different speaking styles. First, they were recorded producing sentences in a conversational speaking style. Next, they were recorded producing the same sentences in a clear speaking style. All participants were provided with written instructions. Each target sentence was presented orthographically with a white background on a Dell 24-inch monitor in front of them during the recording
session. The font-family was Arial, and the height of the lettering was approximately 10 mm.

Prior to data collection, participants produced three practice sentences to familiarize themselves with producing sentences in each speaking style. In the conversational speech task participants were instructed to read aloud practice sentences as they would in an everyday, normal conversation. In the same speaking style, they were recorded producing 60 different sentences. After completing the conversational speech task, participants were instructed to read aloud three practice sentences in a clear speaking style. The instruction for this speech task contained two steps: on the first step, they were asked to read aloud the practice sentences while speaking clearly. On the second step, they were instructed to read aloud the same practice sentences clearly and naturally by over-enunciating. The second step instruction was what they used for producing actual sentences in a clear speaking style. This two-step instruction was developed to minimize unnatural speech production and maximize clear speech effects (Lam & Tjaden, 2013). Previous studies (Lam et al., 2012; Lam & Tjaden, 2013) reported that “over-enunciate” instruction produced the largest clear speech advantage among several different instructions. In an unpublished study prior to this experiment, the author instructed talkers to read aloud the sentences clearly by over-enunciating. This instruction produced unnatural speech with significant gaps (e.g., choppy speech). Therefore, the author modified the simple instruction to the two-step instruction in order to maximize the naturalness of clear speech. After the familiarization task, participants were
recorded producing 60 actual sentences in a clear speaking style. Each sentence was repeated three times during the recording session.

Throughout the recording session, the author continuously monitored the input level to ensure enough gain without exceeding the dynamic range of the recording system. No changes to the gain were required within each speaking condition. The recorded speech files were separately saved in WAV file format. Each speech file included three repeated sentences.

After the recording session, each speech file was segmented into sentence-length audio files with 50-millisecond silent leader and follower. After removing the audio files with recording-mistakes (e.g., peak clipping, narrow dynamic range, high level of noise), the mid-80 percentile ranges of F0 within the rest audio files were computed using the Praat program (Boersma & Weenink, 2008).

All of the four talkers showed wider F0 ranges (i.e., mid-80 F0 percentile ranges) in clear speech files, compared with those in conversational speech files (mean difference: 26.7 Hz, range: 11 - 49.6 Hz). From this larger set of sentences, which for most sentences included three repetitions, 48 conversational-clear speech pairs were selected that showed the biggest differences in F0 ranges for each talker. A total of 384 sentences were used for data analysis in the production study and for additional acoustic analysis (4 talkers x 48 sentences x 2 conditions = 384 sentences).

F0-manipulated speech was synthesized from the clear speech condition for use in the perception study. To create F0-manipulated speech, F0 contours
of clear speech were compressed to match those of conversational speech in the Praat program, using the following formula:

$$F0_i' = \left( \frac{F0_{i\text{conv}}}{F0_{i\text{clear}}} \right) \ast (F0_i - F0_{i\text{med}}) + F0_{i\text{med}}$$

where $F0_i'$ represents the new F0 of the frame for sentence i, $F0_{i\text{clear}}$ is the F0 range (i.e., mid-80 percentile ranges) of clear speech for sentence i, $F0_{i\text{conv}}$ is the F0 range (i.e., mid-80 percentile ranges) of conversational speech for sentence i, $F0_i$ is the F0 of clear speech for sentence i at a given time sample, and $F0_{i\text{med}}$ is the median F0 of clear speech for sentence i. From this formula 192 F0-manipulated speech files were created from the four talkers. In order to minimize potential differences between pre- and post-processing of speech materials, the 192 conversational-clear speech pairs were also processed by replacing the ratio between $F0_{i\text{conv}}$ and $F0_{i\text{clear}}$ with 1.

**Figure 1.** F0 contours for the sentence “The bill was paid every third week.” in three different speech conditions (T4 production)
Figure 1 shows representative F0 contours of a single sentence from T4 in three speech conditions. For a comparison of F0 contours in conversational and F0-manipulated speech (left panel of Figure 1), the duration of conversational speech was synthetically stretched to equal the duration of F0-manipulated speech because the sentence duration of the original conversational speech is shorter than that of clear and F0-manipulated speech. The left panel contains the contours for conversational and F0-manipulated speech conditions, whereas the right panel contains the contours for clear and F0-manipulated speech conditions. A total of 576 sentences including F0-manipulated sentences were used for the perceptual study (48 sentences x 4 talkers x 3 conditions = 576 sentences). The final sentence list across talkers for each speech condition is provided in Appendix A.

B. Perception Study

a. Participants (Listeners)

The participants were fifteen adults (12 females and 3 males), aged from 19 - 34 (mean: 23.2, SD: 3.6), for whom American English was their native language. All participants self-reported normal hearing with no history of any hearing problems. None of the listeners was familiar with the talkers’ voices or the speech materials.

The experimental protocols were approved by the Institutional Review Board of the University of Minnesota. Informed consent was obtained from each participant. A single session of the experiment lasted for 1.5 hours, including
breaks. All participants were monetarily compensated for their time and effort at the end of the experiment.

b. Speech Materials

A total of 576 IEEE sentences, from the production study, were used for speech intelligibility tasks. Forty-eight different sentences were obtained from each of the four talkers in three different speaking styles: conversational, clear, and F0-manipulated speech (48 sentences x 4 talkers x 3 conditions = 576 sentences). Each sentence-length audio file, with 50-millisecond silent leader and follower, was normalized to a root-mean-square (RMS) level of 65 dB SPL. Each stimulus sentence was then embedded in speech-shaped noise that was generated by obtaining the long-term average spectrum of all sentence files from four talkers. The speech-shaped noise began 150 ms before the onset of the sentence and terminated 150 ms after the offset of the sentence. An additional eight sentences in a conversational speaking style were used for practice trials, but these sentences were excluded from data analysis.

c. Data Collection

Experimental Setup. The experimental program was written in MATLAB®, version R2018a (The MathWorks, Inc., Natick, MA). The program randomly selected 192 from the 576 IEEE sentences for a single session of the experiment. The single session consisted of the same number of sentence trials from four talkers in the following three conditions: conversational, clear, and F0-manipulated speech conditions (16 sentences x 4 talkers x 3 conditions = 192 sentences). None of the 192 sentences was identical across the three speech
conditions in a single experimental session. The speech stimuli and speech-shaped noise were presented at 65 dB SPL binaurally through headphones (Sennheiser HD650) at fixed signal-to-noise ratios (SNRs). The SNRs were -2.4, 0, -4, and -1.6 dB for talkers T1, T2, T3, and T4, respectively. These SNRs were adjusted for each talker based on pilot data to avoid floor and ceiling effects in native English-speaking listeners whose hearing was within normal limits. During the experiment, each sentence was presented only one time. The pace of the experiment was manually determined by each participant who either pressed the “Enter” key or clicked a button to begin the presentation of a sentence.

**Experimental Procedure.** Participants were individually seated in a double-walled sound-isolated booth in front of a desktop computer with a video camera (Logitech HD Pro Webcam C920). Listeners were instructed to follow the prompts on the computer screen to proceed throughout the experimental tasks. After a familiarization session with eight practice trials in a conversational speech condition, 192 sentences embedded in noise were presented binaurally through headphones (Sennheiser HD650) in a random order. Listeners orally repeated each sentence that they heard while looking at a video camera mounted in the sound booth. Next, they typed in the sentence that they orally repeated. Throughout the experimental session listeners were videotaped and their typed responses were recorded into a spreadsheet by the MATLAB® program. Listeners’ oral responses were scored online by the author, and the scores were verified by comparing the oral responses to the typed responses.
C. Data Analysis

a. Production Study

Conversational-clear speech pairs were compared and analyzed for statistical significance in each acoustic characteristic, using an analysis of variance (ANOVA) and a linear-mixed effects regression.

b. Perception Study

The percentage of correct keywords was calculated for each talker for each speaking style. To obtain more reliable statistical results, the average scores were transformed to arcsine units (Studebaker, 1985) before performing a repeated measures ANOVA that was used for examining statistical significance among three speaking styles across four talkers. Other potential perceptual correlates that account for the clear speech benefit were also explored.
III. Results

A. Production Study

a. Speaking Rate

Average sentence length, average pause duration, and pause frequency were measured for each talker in conversational and clear speech using Praat. To derive a value of words per minute (WPM), the total number of words was divided by the number of minutes in 48 sentences for each talker in each speaking style. The results showed that all talkers produced fewer wpm in clear speech relative to conversational speech. The average and individual values are shown in Table 1.

Table 1. Words per minute (wpm) for each talker in conversational (CNV) and clear speech conditions

<table>
<thead>
<tr>
<th>Talker</th>
<th>CNV (wpm)</th>
<th>CLR (wpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>285</td>
<td>203</td>
</tr>
<tr>
<td>T2</td>
<td>259</td>
<td>224</td>
</tr>
<tr>
<td>T3</td>
<td>258</td>
<td>162</td>
</tr>
<tr>
<td>T4</td>
<td>284</td>
<td>194</td>
</tr>
<tr>
<td>Avg</td>
<td>272</td>
<td>196</td>
</tr>
</tbody>
</table>

Note: CNV = conversational speech, CLR = clear speech

The average conversational speech rate was 272 wpm, while clear speech rate was 196 wpm across talkers. This falls within the range of speaking rates reported by Krause and Braida (2002, 2004) for normal and quick productions. These values are much less than the “fast” rate of 372 wpm for a talker reported
in Valentini-Botinhao et al. (2019). The clear speech was on average 28% slower than the conversational speech (range: 14 - 32%).

The total length of sentences (TSL) for each talker was averaged across sentences in each speaking style to measure the average sentence length (ASL). The ASL of clear speech was 2,451 ms, while that of conversational speech was 1,745 ms across talkers. A repeated measures ANOVA revealed that the ASL of clear speech was significantly longer than that of conversational speech \[F(1, 3)=18.58, p<0.05\]. To estimate the extent to which an increase in the frequency and duration of pauses contributed to the increased sentence length in clear speech, average pause duration (APD) and pause frequency (PF) were calculated for each talker. The analysis included any period of silence at least 10 ms in duration (Picheny et al., 1986; Krause & Braida, 2004). The measures were conducted by the author and accuracy was verified by a native English speaker in 10% of the measures. The discrepancy for the durations assessed was less than 1%. The results showed that clear speech had approximately 1.5 times more pauses than conversational speech. The APD of clear speech was 81 ms, while that of conversational speech was 47 ms across talkers. A repeated measures ANOVA showed that clear speech had significantly longer pause duration \[F(1, 3)=14.85, p<0.05\] with a higher number of pauses than conversational speech across talkers, as reported in other studies (Pichney et al., 1986; Bradlow et al., 2003; Krause & Braida, 2004). The results revealed that the increased pause duration contributed to the increased sentence length in clear speech by on average 19% (range: 13 - 26%). Individual talkers’ quantitative measurements for speaking rate are provided in Appendix B.
b. F0 Features

The median, standard deviation (i.e., F0 variation, SD), and mid-80 percentile range of F0 values were computed for each talker in each speaking style, using the pitch detection algorithm provided in the Praat program. To reduce possible errors produced by the F0 tracker, the author removed any intervals with noticeable mis-tracked pitch points or intervals of glottal fry (i.e., creaky voice) before extracting pitch contours over the voiced segments from the speech files. The removed intervals were 0.8% and 3.5% of the total sentence duration (TSD) in clear and conversational speech, respectively, across talkers.

Pitch values for every 0.1 millisecond (i.e., 0.0001s time step) were estimated. The mid-80 percentile range of F0 values was calculated as a difference in Hertz (Hz) between the 10th and 90th percentile of the F0 distribution for each sentence-length file. The 10th percentile is the highest attested value for which at most 10% of all attested values are less or equal. In a similar vein, the 90th percentile is the highest value for which at most 90% of all attested values are less or equal. The value of F0 median is defined as the 50th percentile. F0 variation was estimated as dispersion of F0 values from the mean and summarized as the standard deviation (SD).

The F0 median, SD, and range were averaged across sentences in each speaking style for each talker. Table 2 shows the average median, mid-80 percentile range, and variation of F0 values for each talker in each speech condition. Differences in F0 features between conversational and clear speech are shown in semitones. The compression rate (%) represents the amount that the F0 range for clear speech was reduced to equal the range of F0 for
conversational speech. These values were used for synthesizing F0-manipulated clear speech, one of the listening conditions for the perception study.

Table 2. Median, range, and variation (SD) of F0 values in Hz for each talker in clear and conversational speech conditions

<table>
<thead>
<tr>
<th>Talker</th>
<th>F0</th>
<th>CNV (Hz)</th>
<th>CLR (Hz)</th>
<th>CLR-CNV (semitones)</th>
<th>Compression Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Median</td>
<td>93.6</td>
<td>115.8</td>
<td>3.7</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>40.0</td>
<td>108.0</td>
<td>15.2</td>
<td>62.9</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>15.9</td>
<td>38.8</td>
<td>15.5</td>
<td>-</td>
</tr>
<tr>
<td>T2</td>
<td>Median</td>
<td>129.4</td>
<td>132.7</td>
<td>0.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>37.6</td>
<td>67.3</td>
<td>10.1</td>
<td>44.1</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>18.0</td>
<td>26.3</td>
<td>6.6</td>
<td>-</td>
</tr>
<tr>
<td>T3</td>
<td>Median</td>
<td>129.7</td>
<td>138.0</td>
<td>1.1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>40.3</td>
<td>71.5</td>
<td>9.9</td>
<td>43.6</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>17.6</td>
<td>28.3</td>
<td>8.2</td>
<td>-</td>
</tr>
<tr>
<td>T4</td>
<td>Median</td>
<td>83.6</td>
<td>101.6</td>
<td>3.4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>20.7</td>
<td>53.8</td>
<td>16.5</td>
<td>61.6</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>8.6</td>
<td>20.9</td>
<td>15.4</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. CNV = conversational speech, CLR = clear speech

A repeated measures ANOVA was used to compare F0 features in conversational speech to those in clear speech across talkers. The result showed that F0 median values were not statistically different between two speaking styles across talkers [$F(1, 3)=7.9, p=0.07$]. For the range and SD of F0 values, clear speech had a wider range with greater variation, relative to the conversational speech across talkers [range: $F(1, 3)=19.5, p<0.05$, variation: $F(1, 3)=17.3, p<0.05$]. The mean F0 range and SD for clear speech were higher than those for conversational speech by 40.5 and 13.5 Hz, respectively. The F0 range values in this study are relatively high, when compared to those of male talkers in other studies (Pichney et al., 1986; Krause & Braida, 2004). It is notable that the
conversational-clear speech pairs, selected from the large set of sentences, have the biggest differences in F0 ranges. The selection criteria for the sentence pairs were for maximizing the effect of F0 range in the perception study.

c. Long-term spectra

Speech spectra of conversational and clear speech were computed after normalization for long-term rms level. Spectral components were obtained from average amplitude distributions of the speech over one-third octave intervals with center frequencies ranging from 62.5 to 8000 Hz, as described in a study by Krause and Braida (2004). Figure 2 shows the spectral distribution as a function of third-octave band frequency for conversational and clear speech in each talker.

![Figure 2](image)

**Figure 2.** Long-term spectral distribution for conversational (Conv) and clear (Clear) speech in each talker

As shown in the figure, clear speech contains more spectral energy above 0.63-kHz, relative to conversational speech in all talkers. Previous studies,
although they used slightly different criteria (i.e., spectral change in 1-3 kHz),
also reported increased high-frequency spectral energy for the clear speech
condition (Pitchney et al., 1986; Krause & Braida, 2004).

Table 3. Average band RMS level across the frequency range between 0.63- and 6-kHz for each talker in each speech condition. The third-octave bands centered at 0.63- and 0.8-kHz contributed a fractional amount so this average level is referred to in the text as a “high-frequency average.”

<table>
<thead>
<tr>
<th>Speech Condition</th>
<th>Average Band RMS Level (dB SPL)</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNV</td>
<td></td>
<td>65.7</td>
<td>63</td>
<td>67.7</td>
<td>64.5</td>
<td>65.2</td>
</tr>
<tr>
<td>CLR</td>
<td></td>
<td>67.0</td>
<td>63.9</td>
<td>69.3</td>
<td>65.3</td>
<td>66.4</td>
</tr>
<tr>
<td>CLR-CNv</td>
<td></td>
<td>13</td>
<td>0.9</td>
<td>16</td>
<td>0.8</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Note: CNV=conversational speech, CLR=clear speech, CLR-CNv=spectral change in clear speech

Table 3 shows the results of third-octave band analysis that summarizes in dB the high-frequency spectral change in clear speech. To quantify the high-frequency energy for each speaking style, band RMS levels were averaged across frequencies between 0.63- and 6-kHz where the articulation index (AI) accounts for between 77% and 87%, depending on the weighting function. There is no accepted weighting function for IEEE sentences. As shown in the table, clear speech had on average 1.2 dB more high-frequency energy than conversational speech across talkers, even when rms amplitude was equalized across speaking styles. The results show that T3 had the largest spectral change, while T4 showed the smallest change. These results suggest that clear speech has more favorable SNR than conversational speech for speech intelligibility, even when the stimuli from both speaking styles are equated for their rms levels.
d. Vowel Space

The analysis included four corner vowels (i.e., /i/, /u/, /ɑ/, and /æ/) from the keywords in 48 sentence stimuli for each talker in each speaking style. The first formant (F1) and second formant (F2) values were extracted at the midpoint of each vowel, using Linear Predictive Coding (LPC) formant tracking algorithm provided in the Praat program. Default settings were used for adult male talkers, in which the maximum formant was set for 5000 Hz, and five formants were estimated. The author manually segmented the vowel intervals to maximize the accuracy of the formant-tracking algorithm. When the formant values were apparently mis-tracked by the LPC tracking algorithm, the author corrected by manually measuring the formants using visual inspection. The accuracy of hand-measures was verified by a native English speaker in 10% of the measures, and no substantial discrepancies were found.

Before data analysis, the formant values (Hz) were converted to the auditory Bark scale that reflects vowels’ real position in oral cavity, using the formula of Zwicker and Terhardt (1980):

\[
\frac{Z_{c\text{Bark}}}{Bark} = 13 \arctan \left(0.76 \frac{f}{kHz}\right) + 3.5 \arctan \left(\frac{f}{7.5 kHz}\right)^2
\]

where \(\frac{Z_{c\text{Bark}}}{Bark}\) is a critical band rate in Bark and \(f\) is a formant value in kHz.
To estimate the expansion of vowel space in clear speech, the Euclidean distance of the individual vowels from the average F1/F2 was measured for each talker, using the following formula:

\[ D = \sqrt{(F1_v - F1_x)^2 + (F2_v - F2_x)^2} \]

where \( D \) is each vowel token’s distance from a center of vowel space, \( F1_v \) is the F1 value of each vowel token, \( F1_x \) is the average F1 of all vowels, \( F2_v \) is the F2 value of each vowel token, and \( F2_x \) is the average F2 of all vowels. All values in this formula are in Bark. After obtaining individual vowel token’s distance, the author calculated the mean of these distances for each vowel for each talker, as other studies did (Bradlow et al., 1996; Munson & Solomon, 2004). Figure 3 shows the vowel-space dispersion for each talker and the average values in clear and conversational speech.
Figure 3. F1/F2 values (Bark) of four corner vowels (/i/, /u/, /ɑ/, and /æ/) for each talker and the average Euclidean distance (Bark) of each vowel from the grand mean F1/F2 value across talkers in clear (CLR) and conversational (CNV) speech.

In the figure the individual X indicates the average F1/F2 value of all vowels for each talker. The square filled in red indicates the grand mean of the values. The individual bold vowel letters represent the average F1/F2 values of individual talker’s clear vowel tokens, while the others represent the values of conversational vowel tokens. The grand mean of the values is shown as the circle and triangle filled in red for clear and conversational speech, respectively. The average vowel-space dispersion across talkers is shown as the solid and dashed lines for clear and conversational speech, respectively. The figure shows clear speech has greater vowel-space dispersion with higher mean F1 values in the vowel /ɑ/ and /æ/ and higher mean F2 values in the vowel /i/ and /æ/, when compared with conversational speech.
A linear mixed-effects model was fitted to statistically analyze values of F1 and F2 and the Euclidean distance in each speech condition across talkers. When F1 values were the dependent variable, fixed-effect variables were vowel and type and random-effect variables were word and talker. Vowel and type were nested within individual talkers that allowed each of the talkers to have their own effect of vowel and type. The results showed that there were significant main effects of vowel and type. F1 values were statistically higher in vowel /æ/ relative to vowels /u/ and /i/ and in vowel /a/ relative to vowels /i/ and /u/ at a significance level of .05. On average, clear speech had higher F1 than conversational speech. However, it depended on the vowel. Table 4 summarizes the results of statistics in which the vowel /ɑ/ is the reference vowel.

Table 4. Main effects of vowel and type and interaction between vowel and type for F1 values (reference vowel: /ɑ/)

|                      | Estimate | Std. | Error | df  | t-value | Pr(>|t|) |
|----------------------|----------|------|-------|-----|---------|---------|
| (Intercept)          | 6.55     | 0.18 | 5.13  | 36.61| 0.00    | ***     |
| vowelæ               | 0.08     | 0.19 | 8.43  | 0.42 | 0.69    |         |
| vowelæ               | -3.63    | 0.18 | 7.76  | -20.55| 0.00    | ***     |
| vowelæ               | -3.56    | 0.21 | 7.46  | -17.16| 0.00    | ***     |
| typeconv             | -0.31    | 0.12 | 402.34| -2.47| 0.01    | *       |
| vowelæ:typeconv      | -0.24    | 0.16 | 398.55| -1.49| 0.14    |         |
| vowelæ:typeconv      | 0.39     | 0.15 | 399.78| 2.54 | 0.01    | *       |
| vowelæ:typeconv      | 0.43     | 0.18 | 395.97| 2.36 | 0.02    | *       |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Main effects were qualified by a number of significant interactions between vowel and type. The nature of these interactions is shown in the left panel of Figure 4 which plots the clear and conversational vowels across talkers.
In general, based on visual inspection, F1 is lower for /i/ and /u/, which are closed vowels, while F1 is higher for /a/ and /æ/, which are open vowels, in clear speech relative to conversational speech. These results are consistent with findings in other studies (e.g., Picheny et al., 1986; Ferguson & Kewley-Port, 2002; Krause & Braida, 2004).

Figure 4. Type and vowel interaction for F1 (left panel) and F2 (right panel)

F2 values were predicted with the same fixed-effect, random-effect, and nested random-effect variables as used for prediction of F1 values. The results showed that there were significant main effects of vowel and type. F2 values were statistically higher in vowel /i/ relative to vowels /a/, /æ/, and /u/ and in vowel /æ/ relative to vowels /a/ and /u/ at a significance level of .05. The association between clear and conversational speech with F2 depended on vowel. Table 5 summarizes the results of statistics in which the vowel /a/ is the reference vowel.
Table 5. Main effects of vowel and type and interaction between vowel and type for F2 values (reference vowel: /a/)

| Fixed Effects                          | Estimate | Std. Error | df  | t-value | Pr(>|t|) |
|----------------------------------------|----------|------------|-----|---------|----------|
| (Intercept)                            | 9.81     | 0.18       | 10.04 | 55.49   | 0.00     | ***      |
| vowel /æ/                              | 2.60     | 0.21       | 9.87  | 12.13   | 0.00     | ***      |
| vowel /ɪ/                              | 4.25     | 0.26       | 4.89  | 16.51   | 0.00     | ***      |
| vowel /u/                              | -0.03    | 0.22       | 52.62 | -0.12   | 0.90     |          |
| type conversational                    | 0.25     | 0.15       | 388.55 | 1.72   | 0.09     |          |
| vowel /æ/:type conversational          | -0.48    | 0.19       | 385.44 | -2.52   | 0.01     | *        |
| vowel /ɪ/:type conversational          | -0.51    | 0.18       | 386.56 | -2.80   | 0.01     | **       |
| vowel /u/:type conversational          | 0.51     | 0.22       | 383.21 | 2.34    | 0.02     | *        |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 1

Main effects were qualified by a number of significant interactions between vowel and type. The nature of these interactions is shown in the right panel of Figure 4. In general, based on visual inspection, F2 is lower for /a/ and /u/, which are back vowels, while F2 is higher for /i/ and /æ/, which are front vowels, in clear speech compared with conversational speech. These results are consistent with findings in the previous studies (e.g., Picheny et al., 1986; Ferguson & Kewley-Port, 2002; Krause & Braida, 2004).

To analyze the average Euclidean distance in different speaking styles, the linear-mixed effects model involved Euclidean distance as the dependent variable, type as the fixed-effect variable, and talker as the random-effect variable. Type was nested within individual talkers that allowed each of the talkers to have their own effect of type.
Table 6. Main effects of type for the average Euclidean distance

| Fixed Effects   | Estimate | Std.  | Error | df  | t-value | Pr(>|t|) |
|-----------------|----------|-------|-------|-----|---------|---------|
| (Intercept)     | 2.80     | 0.08  | 3.02  | 33.23 | 0.00    | ***     |
| type conversational | -0.37   | 0.09  | 4.24  | -4.10| 0.01    | *       |

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

As shown in Table 6, there was a significant main effect of type: overall, clear speech showed larger vowel-space dispersion than conversational speech. This outcome was consistent with the finding in previous studies (e.g., Picheny et al., 1986; Krause & Braida, 2004).

e. Vocal Intensity Level

Forty-eight sentence-length-files were concatenated to form a single, large file to estimate the average vocal intensity level for each talker in each speech condition. The average rms voltage of the concatenated file was converted to dB SPL using a digitized, 1-kHz, 80-dB SPL reference tone that was recorded with the same recorder settings used to digitize the speech. Since the microphone was placed 10 cm from a participant’s mouth (and the loudspeaker generating the reference tone) during the recording session, the intensity value was reduced according to the inverse-square law to correspond to the level at a distance of 1 m. One meter corresponds to the face-to-face conversation distance in a quiet room (Pearsons et al., 1977; Olsen, 1998).
Table 7. Individual talkers’ vocal intensity levels (SPL) in conversational and clear speech conditions

<table>
<thead>
<tr>
<th>Talker</th>
<th>CLR</th>
<th>CNV</th>
<th>CLR-CN</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>63.4</td>
<td>62.0</td>
<td>1.4</td>
</tr>
<tr>
<td>T2</td>
<td>64.3</td>
<td>62.5</td>
<td>1.8</td>
</tr>
<tr>
<td>T3</td>
<td>64.7</td>
<td>62.8</td>
<td>1.9</td>
</tr>
<tr>
<td>T4</td>
<td>64.3</td>
<td>60.3</td>
<td>4.1</td>
</tr>
<tr>
<td>Av.</td>
<td>64.2</td>
<td>61.9</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Note: Values are the vocal intensity levels in dB SPL for the face-to-face conversation distance, 1m. CLR=clear speech, CNV=conversational speech.

Table 7 shows vocal intensity levels for four talkers in each speech condition. The speech levels in a conversational speaking style fall within the range of the normal conversational speech level in a quiet room (i.e., 60-70 dB SPL). As reported in previous studies (Picheny et al., 1986; Lam et al., 2012), clear speech had greater intensity, compared with conversational speech. The mean vocal intensity level difference was 2.3 dB (range: 1.4 - 4.1 dB) across talkers in this study. The level differences were eliminated in the perception study, by equalizing rms amplitude across all sentences in different speaking styles.

B. Perception Study

Average intelligibility scores were obtained from a total of 225 sentences across 15 listeners in each speech condition for each talker. Figure 5 shows average percent-correct keyword scores for each talker and grand mean scores in each speech condition. All talkers showed substantial intelligibility benefits from clear and F0-manipulated speech, when compared with conversational speech. None of the
talkers showed a significantly poorer or better performance in the F0-manipulated speech over the clear speech.

![Figure 5. Average intelligibility scores for each talker and grand mean scores in conversational, clear, and F0-manipulated speech conditions](image)

To yield more reliable statistical analysis, the average scores were transformed to arcsine units (Studebaker, 1985) before performing a repeated measures ANOVA with speech condition as the within-subject factor. There was a significant main effect of speech condition \([F(2,28)=227.8, p<0.001]\). When Bonferroni-corrected \(\alpha=0.017\) level was applied, the repeated measures ANOVA revealed that conversational speech performed significantly poorer than clear \([F(1,14)=369.7, p<0.001]\) and F0-manipulated speech across talkers\([F(1,14)=353.3, p<0.0001]\). Clear speech did not statistically differ from F0-manipulated speech in performance across
talkers \[ F(1,14)=2.8, p=0.12 \]. This indicates that decreases in F0 range do not have detrimental effects on the speech intelligibility in noise backgrounds.

Keyword recognition scores were averaged across all listeners in each speech condition. This analysis was performed to examine whether the correct response rates are affected by the position of keywords within a sentence. The average scores show a decline in speech recognition performance from the beginning to the end of sentences. This pattern was shown in all three speech conditions, but conversational speech showed the most dramatic decrease in the performance over the temporal position (Left panel of Figure 6). Miller et al. (2010) showed the same pattern in their data which are also shown in this figure. Miller et al.’s results are most similar to our results for clear speech. The speaking style of the talkers was not specified in the study by Miller et al. (2010).

Figure 6. Keyword recognition scores across temporal keyword position (left panel) vs. relative intensity levels over time within sentences (right panel). Times listed in the right panel represent midpoints of 300 ms analysis window.
The right side of Figure 6 shows the relative intensity levels (dB) over time within sentences. The intensity level was estimated, based on the average rms values at different time points after normalization for long-term rms level. The values were obtained using 300 ms windows. Analysis included approximately half of the total number of sentences that had a duration equal to or longer than the average sentence duration in each speech condition. The 300 ms windows included for analysis were only calculated for intervals that were one window prior to the end of a sentence. Using this rule, fewer sentences remained for analysis for the longest duration sentences. When the pool of sentences was fewer than nine, those sentences were not included in the analysis. Table 8 shows a total number of sentences that were included for the intensity analysis over time within sentences. Since the intensity levels in F0-manipulated speech are the same as those in clear speech, F0-manipulated speech condition was not included in this analysis.

<table>
<thead>
<tr>
<th>Number of Sentences</th>
<th>150</th>
<th>450</th>
<th>750</th>
<th>1050</th>
<th>1350</th>
<th>1650</th>
<th>1950</th>
<th>2250</th>
<th>2550</th>
<th>2850</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNV</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>.20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CLR</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
</tr>
</tbody>
</table>

Table 8. A total number of sentences used for intensity analysis over time within sentences (ms). Times listed represent midpoints of 300 ms analysis window.

As shown in figure 6, the decline in speech understanding with time was accompanied by a decrease in intensity. This finding suggests that at least some of the decline in speech understanding can be accounted for based on audibility because the background noise was constant throughout the presentation of the sentences.
IV. Discussion

The clear speech production from this study showed acoustic changes that have been reported in numerous previous studies. These modifications include slower speaking rates with longer and more inter-word pauses, high-frequency emphasis, higher vocal intensity level, expanded vowel space, and wider F0 range.

Among these acoustic changes, the contribution of F0 range to the clear speech benefit was the focus of this study. All four talkers in the present study expanded the F0 range as one of their clear-speech modifications, and the amount of change was relatively larger, compared with values typically seen in other clear speech studies (Pichney et al., 1986; Krause & Braida, 2004). The participants were selected with this in mind. The mean difference between conversational and clear speech in the present study is 13.4 semitones (range: 9.9 - 17.2 semitones), which is more than an octave. The F0-range compression rate for clear speech required to yield the same range as conversational speech was on average 53% (range: 44 - 63%).

Despite the significant expansion of the range of F0 in clear speech, a reduction in F0 range did not have a statistically significant impact on the intelligibility benefit of clear speech in sustained noise. The F0-manipulated speech had nearly the same intelligibility scores as the clear speech that had significantly higher scores than the conversational speech across talkers. The overall clear speech benefit was 32 percentage points (range: 22 - 48 percentage points), which falls within the range of values reported by previous studies (e.g., Uchanski et al., 1996; Bradlow et al., 2003; Liu et al., 2004).

In contrast to the present study, Bradlow et al. (1996) reported a tendency for F0 range to have a positive correlation with sentence-level intelligibility.
They found a correlation that approached significance in a mixed gender talker group that consisted of 20 talkers of General American English. The present study provides experimental evidence that the positive correlation is not observed when holding acoustic changes other than F0 range constant in four adult male speakers of American English. This finding is consistent with the results of a study by Tjaden et al. (2014) that used a different approach to F0-manipulation. By resynthesizing F0 contours of conversational speech to have the characteristics of clear speech, they examined whether the increased F0 range in the resynthesized sentences enhances sentence-level intelligibility scores in two talkers with Parkinson’s disease. The results showed that this intonation hybrid did not improve talker intelligibility in multi-talker babble noise. Despite the major differences between Tjaden et al.’s study and this present study, both outcomes support that F0 contour is not a primary contributor to the clear-speech benefit.

The finding that F0 range does not affect the intelligibility benefit of clear speech has three possible explanations. First, audible changes in F0 may be all that are required to provide a cue for speech understanding. Binns and Cullings (2007) showed that a reduction of the F0 range by 75% of the original speech in a single talker with an unspecified speaking style was not significantly detrimental to speech understanding in a speech-shaped noise background. The reduction in F0 range in the present study was between 44% and 63%. By contrast, other studies (Miller et al., 2010; Shen & Souza, 2017a, 2017b) report that when F0 contours were flattened (i.e., monotone speech), speech intelligibility was significantly poorer. Flattening the contour entirely removed F0 cues which are important for signaling breaks.
between syllables which would make it harder to process speech (Cutler & Foss, 1977; Liss et al., 2000; Spitzer et al., 2007).

The second possible explanation is that a simple compression of F0 contours in clear speech does not remove complex concomitant changes in speech acoustics. Uchanski (2005) states that hyper-articulated clear speech increases spectral amplitude at high frequencies that is a result of increased vocal effort. In this process, the average F0 of clear speech would be increased, a finding reported in previous studies (Picheny et al., 1986; Krause & Braida, 2004). In a similar vein, over-enunciated clear speech generally reduces the speaking rate relative to conversational speech. This slower production rate would provide more time for F0 to vary within a word which provides segmental cues related to, for example, lexical stress.

Lastly, the clear-speech benefit may be partially attributed to the “micro-variation” in F0 contours of clear speech that exist separately from the F0 range. The left panel of figure 1 shows that the F0 contour of conversational speech does not line up with the F0 manipulated clear speech contour even after compressing the clear-speech sample in time. The latter half of the time within sentence, in particular, shows that F0 contours of F0-manipulated speech are qualitatively different from those of conversational speech. This leaves open the possibility that the intelligibility benefit of clear speech may be associated with the fine-grained changes in F0 contours rather than the F0 range per se.

In addition to the absence of an effect of F0 compression on speech understanding, another notable finding in the present study was that keyword recognition scores declined from the beginning to the end of sentences in all speech
conditions. In particular, sentences spoken in a conversational speaking style fell off more by the fifth keyword than that of a clear speaking style. One possible explanation for the decline involves uneven audibility across the temporal position of keywords within a sentence. Intensity levels measured at different time points within sentences show a reduction in level over time in both conversational and clear speech conditions (see the right panel of Figure 6). The drop is much larger for conversational than for clear speech, which is consistent with the intelligibility findings associated with keyword position. This is an expected finding because the speech noise stimulus maintained the same level throughout a sentence and a drop in the speech level would then represent a reduction in the signal-to-noise ratio (SNR). The SNR is one of the most important factors determining speech intelligibility.

Other additional perceptual correlates that can be translated into SNR differences were explored with the speech production and perception data obtained from the present study. These include pause durations and high-frequency emphasis. MacPherson and Akeroyd (2014) found, in a survey of archival speech understanding studies, that speech intelligibility increases on average 4.5% per dB when IEEE sentences are presented in speech maskers. By converting the production differences between clear and conversational speech into SNR, the percentage contribution to the clear speech advantage can be assessed.

Clear speech has more pauses than conversational speech. As a consequence, when these speaking styles are equalized in their rms levels, longer pauses result in a higher gain during this equalization process. A significant increase in the gain of clear speech resulting from this rms equalization process would result in better performance that is an artifact of calibration and not a consequence of speaking
clearly (Liu & Zeng, 2006). To quantify the contribution of pause extension to the clear speech benefit, a dB change that is attributed to the pause extension was calculated by subtracting the value of log ratio between the total sentence length with and without pauses in conversational speech from the value of log ratio in clear speech. The overall dB change was 0.2 dB (range: 0.1-0.3 dB), which predicts approximately 1% increase in the intelligibility score. This value accounts for approximately 3% of the total clear speech benefit observed in this study. This result is not unexpected because Krause and Braida found only a 4% reduction in the clear speech benefit when talkers produced clear speech at a conversational speaking rate.

To quantify the contribution of high-frequency emphasis to the clear speech benefit, the decibel difference in the high-frequency region was converted to the predicted increase value in percentage, using the relation of 4.5% per dB of gain in the high-frequency SNR. Details about the process were described in the results section and summarized in Table 3. The overall dB change was 1.2 dB (range: 0.8-1.6 dB), which predicts approximately 5% increase in the intelligibility score. This value accounts for approximately 16% of a total clear speech benefit observed in this study. Such a substantial contribution of high-frequency emphasis supports the findings in previous studies (e.g., Hazan & Markham, 2004; Krause & Braida, 2004). These analyses suggest that approximately 20% of a total clear speech benefit observed in the present study can be accounted for by the two acoustic modifications but mostly by the high-frequency emphasis.
V. Conclusion

The present study provides experimental evidence that the change in F0 range of clear speech is not a primary factor that directly contributes to the clear speech benefit. Rather, it may be a secondary effect that is accompanied by other acoustic changes that convey clear speech benefits, such as slower speaking rate. The contribution of high-frequency emphasis to the clear speech benefit was salient, whereas the contribution of pause extension to the clear speech benefit was minimal.
Chapter 3: The Effects of Fundamental Frequency Contours on the Intelligibility Benefit of Clear Speech in Native Korean Speakers (Experiment 2)

I. Introduction

The United States has become more linguistically and culturally diverse. Both the number and the proportion of foreign-born individuals have increased over the past few decades (Grieco et al., 2012). One of the biggest challenges that these individuals with limited English-language proficiency face is successfully communicating with others in the workplace. Communication becomes more challenging in situations where they work with individuals who have hearing problems, such as in health care facilities, nursing homes, and senior living communities. Considering the high proportion of foreign workers in the healthcare industry (Lowell, 2013), strategies are needed for ensuring effective communication between healthcare providers and patients.

One strategy to reduce barriers to communication is to use clear speech that is a listener-oriented speaking style. Clear speech is distinguished from typical, conversational speech, which is characterized by a wide range of acoustic-phonetic changes. Clear-speech modifications can be divided into two categories: global and local modifications (Bradlow et al., 1996; Bradlow & Bent, 2002; Krause & Braida, 2004; Smiljanić & Bradlow, 2008; Granlund et al., 2011; Cook et al., 2014; Hazan et al., 2015). Local modifications generally include segmental-level adjustments, such as longer segmental durations (Krause & Braida, 2004; Furgerson & Kewley-Port, 2007), increased short-term vowel spectra (Krause & Braida, 2004), more frequent
stop burst releases (Picheny et al., 1986; Krause & Braida, 2004), and a wider vowel space (Picheny et al., 1986; Bradlow et al., 2003; Furgerson & Kewley-Port, 2002, 2007; Smiljanić & Bradlow, 2007; Ferguson & Quene, 2014). In contrast, global modifications include signal enhancing adjustments, such as slower speaking rate (Picheny et al., 1986; Uchanski et al., 1996; Bradlow et al., 2003; Liu & Zeng, 2006), more and longer inter-word pause durations (Picheny et al., 1986; Bradlow et al., 2003; Liu & Zeng, 2006), high-frequency emphasis (Picheny et al., 1986; Krause & Braida, 2004, 2009), higher vocal intensity levels (Picheny et al., 1985; Lam et al., 2012), and wider fundamental frequency (F0) range (Picheny et al., 1986; Bradlow et al., 1996, 2003).

Numerous studies done to date have focused on speech production and perception in native English-speaking populations to investigate the clear speech effects in English. The majority of the studies showed reliable and robust clear-speech benefits (10-34 percentage points) under a variety of conditions that include listeners with sensorineural hearing loss (Picheny et al., 1985; Payton et al., 1994; Uchanski et al., 1996; Schum, 1996; Liu et al., 2004), and for normally hearing listeners in noise (Uchanski et al., 1996; Liu et al., 2004) and in noise with reverberation (Payton et al., 1994).

Previous cross-linguistic studies on clear speech effects have found that the intelligibility benefit of English clear speech is not exclusively observed in native populations (e.g., Bradlow & Bent, 2002; Smiljanić & Bradlow, 2005, 2011; Li & So, 2006; Rogers et al., 2010; Granlund et al., 2011, 2012; Luque & Bradlow, 2011). Several studies have demonstrated that non-native populations listening to native English talkers have clear-speech benefit but the degree of the benefit is relatively
small (Bradlow & Bent, 2002; Smiljanić & Bradlow, 2005, 2007; Bradlow & Alexander, 2007). One of the explanations for the smaller benefit is that non-native listeners mostly benefit from global acoustic modifications but not from all segmental-level modifications of clear speech production (Bradlow & Bent, 2002). Bradlow and Bent (2002) suggested that non-native listeners have difficulties with using the linguistic code at a segmental level when their experience with the target language is limited. Smiljanić and Bradlow (2011) supported this idea by showing that high-proficiency non-native listeners achieved a similar amount of clear-speech benefit from their native (L1) and second language (L2) clear speech productions. The evidence that the degree of clear-speech benefit is partially associated with listeners’ language experience is also shown in the study by Bradlow and Alexander (2007) in which clear speech effects were examined in both native and non-native listeners. All listeners were presented with low- and high-context English sentences in clear and conversational speaking styles under speech-shaped noise. The results of scoring by keyword showed that clear speech was more intelligible than conversational speech in both native and non-native listener groups, regardless of the level of sentence context. However, the clear-speech benefit in the low predictability context was smaller in the non-native listener group. This finding suggests that non-native listeners are better able to achieve clear-speech benefit when context cues are available.

Several cross-linguistic studies on clear speech have explored the extent to which acoustic-phonetic properties have cross-language similarities in clear speech production modifications. Smiljanić and Bradlow (2005) investigated whether Croatian-English bilinguals modify speaking rate, F0 range, and vowel space from
conversational to clear speech production in their L1 (Croatian) and L2 (English).
The results showed that Croatian-English bilinguals used all of the three clear-speech modifications in both languages and showed native-like enhancement in L2 production. These native-like clear speech adjustments in F0 range and vowel space were also found in native Cantonese speakers who lived in an English speaking country for about 2.5 years (Li & So, 2006). Cross-language similarities in global acoustic modifications were also reported in Finish-English bilinguals (Granlund et al., 2011, 2012). The Finish-English bilinguals modified high-frequency spectral energy and F0 features (e.g., F0 median and range) from conversational to clear speech production in their L1 (Finish) and L2 (English). The modifications in L2 production were comparable with those of native English speakers. Segmental-level modifications, by contrast, were not produced consistently in the two languages. For example, the voice onset time (VOT) of a stop voiced consonant was significantly shorter in L2, compared with L1 clear speech productions. These results suggest that global acoustic changes are generally shown in L1 and L2, whereas local changes are not consistently shown in both languages. In other words, non-native talkers do not seem to always incorporate the local acoustic features that are observed in a native talker’s clear speech. The production of the local change may depend on whether the phonological features of L2 are part of the talker’s native language. Although several studies demonstrated that global clear-speech modifications (e.g., F0 features, speaking rate, and high-frequency emphasis) were generally shown in L1 and L2 productions, not many languages have been assessed.

Korean presents an interesting case for examining F0 features that have been known to be global acoustic characteristics. Korean is identified as a syllable-timed
language, which is different from English that is a stress-timed language. In a syllable-timed language, each syllable is given equal stress. In a stress-timed language, by contrast, stressed vowels are given more emphasis with longer duration, higher F0, and greater intensity (e.g., Lieberman, 1960; Abercrombie, 1967; Beckman, 1986; Lee et al., 2006). Since Seoul Korean lacks significant word-level stress and pitch accent, English spoken by native Korean speakers often sounds to native English speakers quite flat (Gerlach, 2013). Cho et al (2011) examined acoustic-phonetic modifications in hyper-articulated speech produced by native Koreans in their native language. The results showed that a change in F0 features, including 0.6 expansion of F0 range and higher F0, was not observed in eight male native Korean talkers. There remains the question of whether Korean-speaking L2 learners of English do not show these F0 changes in their L2 clear speech production, like in their L1 speech production.

In English, an expanded F0 range is not necessary for a clear-speech benefit because some talkers show minimal or negligible differences whereas others show large differences (Krause & Braida, 2004). A recent study (Experiment 1) made an empirical test of the role of an expanded F0 range of the clear-speech benefit. That study found that the intelligibility benefit of clear speech did not change when the expanded F0 range produced by a clear-speaking style was compressed to equal the range of F0 for a conversational speaking style. This negative result is consistent with F0 range being a secondary acoustic factor when considering the clear-speech benefit.

The finding that F0 range is a secondary factor in the clear-speech benefit has three possible explanations. First, as long as variations in F0 are audible, the
increased F0 range would be a redundant cue for speech understanding. Binns and Cullings (2007) showed that there was no significant reduction in speech intelligibility until the F0 range was reduced by 75% of the original speech. By contrast, when F0 contours were flattened, speech intelligibility was significantly poorer. This outcome was consistently found in other studies (Miller et al., 2010; Shen & Souza, 2017a, 2017b). Since F0 cues are important for signaling breaks between syllables, it is hard to process speech when these cues are not available to listeners (Cutler & Foss, 1977; Liss et al., 2000; Spitzer et al., 2007).

Second, fine-grained changes in F0 contours of clear speech may play an important role in the clear-speech benefit. The simple compression of the F0 range in clear speech does not change the pattern for the F0 contour. The “micro-variation” in F0 contours, which still remain after decreasing the F0 range, may not be present in the F0 contours of conversational speech. These qualitative differences in F0 contours of the two speech styles may partially account for the intelligibility benefit of clear speech even when the range of F0 in clear speech is compressed to match the range in conversational speech.

The third possible explanation is that complex concomitant changes in speech acoustics that are part of a clear-speech speaking style would not be removed by the simple compression of F0 contours. For example, clear speech is usually produced at a slower rate than conversational speech, and this slower production rate offers more time for F0 to vary. Infant-directed speech is another example to show concomitant acoustic changes in hyper-articulated speech. Infant-directed speech with its wide range of F0 values has some clear-speech-like effects. This hyper-articulated speaking style is known for its wide ranging F0 that results in an expanded
vowel space (Uther et al., 2007). It is believed that the hyper-articulation increases speech contrasts making infant-directed speech more intelligible in noise backgrounds and easier to learn for infants. One goal of this present study is to investigate a change in F0 features in English clear speech produced by Korean-speaking L2 learners of English. As described above, Korean does not use lexical stress or pitch accent. One study, however, reported that late Korean-English bilinguals were able to produce English stressed and unstressed vowel production like native English talkers by using F0 differences at a word-level (Lee et al., 2006). This outcome does not guarantee that native Koreans are able to achieve a native-like F0 contour in clear speech production of sentences.

Because of the possible narrow F0 range in native Koreans, another goal of this study is to investigate whether native Koreans instructed to expand the F0 range in English improves speech understanding, relative to conversational and clear-speaking styles. It is generally known that infant-directed speech with its wide range of F0 values expands the vowel space, which leads to clear-speech-like improvements in speech understanding. However, the widely held belief that infant-directed speech aids speech understanding in noise is not always supported by empirical studies of synthetically expanded F0 range and adult listeners. (Miller et al., 2010; Shen & Souza, 2017a, 2017b). Miller et al. (2010) found that an exaggerated F0 contour reduced talker intelligibility in speech-shaped noise for normal hearing listeners. It was observed that the exaggerated F0 contour reduced the intelligibility by a similar amount as when the natural F0 range was exaggerated by a factor of 1.75 synthetically. A possible explanation for the reduction involves high F0 values of female talkers producing harmonics that are misaligned with the resonances in the
vocal tract that produce formant frequencies (Watson & Schlauch, 2008). Further, any change to an F0 contour that does not follow the rules of the language results in a reduction of speech understanding ability. For instance, when F0 contours were sinusoidally frequency-modulated (Miller et al., 2010) or inversely manipulated (Binns & Cullings, 2007; Miller et al., 2010), intelligibility was poorer than for a flattened condition. Miller et al. (2010) suggested that this outcome could be a result of the flattened F0 condition providing neutral cues for parsing syllables whereas the sinusoidal FM condition providing the wrong cues.

Finally, this present study will explore potential acoustic correlates of clear-speech benefit. A recent study (Experiment 1) revealed that high-frequency emphasis substantially contributes to the clear-speech benefit, relative to pause duration. Given that non-native populations benefit mostly from global, signal enhancements of clear speech production, a similar pattern should be shown in the L2 speech production.

In sum, this present study explores the effect of F0 range on the intelligibility benefits of hyper-articulated speech in Korean-speaking L2 learners of English. Further, F0 features were examined in L1 and L2 productions to explore whether these acoustic features have cross-language similarities. The potential acoustic correlates of clear-speech benefit were investigated using the same analysis technique in Experiment 1.
II. Method

A. Production Study

a. Participants (Talkers)

Five adult male native speakers of Seoul Korean (T1-T5), aged from 20 to 34 (Mean: 25.6, SD: 5.86), were recruited. All participants started to learn English after the age of six and lived in an English-speaking country or English-speaking countries for 69.4 months on average (range: 60 - 76 months). English language experience and proficiency were self-reported by each participant (Table 9). All participants had normal hearing at octave frequencies between 500 and 4000 Hz and self-reported no history of a speech-language disorder.

Table 9. Language experience and proficiency for each participant

<table>
<thead>
<tr>
<th>Variable</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOAE (yr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>8.0</td>
<td>8.0</td>
<td>13.0</td>
<td>11.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Listening</td>
<td>8.0</td>
<td>8.0</td>
<td>13.0</td>
<td>11.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Speaking</td>
<td>8.0</td>
<td>8.0</td>
<td>20.0</td>
<td>13.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Writing</td>
<td>8.0</td>
<td>8.0</td>
<td>20.0</td>
<td>13.0</td>
<td>10.0</td>
</tr>
<tr>
<td>County</td>
<td>73.0</td>
<td>72.0</td>
<td>76.0</td>
<td>60.0</td>
<td>66.0</td>
</tr>
<tr>
<td>Family</td>
<td>0.0</td>
<td>48.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>School</td>
<td>72.0</td>
<td>72.0</td>
<td>76.0</td>
<td>60.0</td>
<td>75.0</td>
</tr>
<tr>
<td>Exposure to English per day (hr)</td>
<td>5.9</td>
<td>10.0</td>
<td>7.0</td>
<td>2.0</td>
<td>5.0</td>
</tr>
<tr>
<td>English Proficiency&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>6.0</td>
<td>8.3</td>
<td>5.7</td>
<td>5.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Native Language Proficiency&lt;sup&gt;c&lt;/sup&gt;</td>
<td>9.0</td>
<td>9.0</td>
<td>9.7</td>
<td>8.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Difference between English and Native Language Proficiency&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.0</td>
<td>0.7</td>
<td>4.0</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Self-reported Foreign Accent perceived by Self&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7.0</td>
<td>3.0</td>
<td>5.0</td>
<td>8.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Self-reported Foreign Accent identified by others&lt;sup&gt;a&lt;/sup&gt;</td>
<td>8.0</td>
<td>3.0</td>
<td>7.0</td>
<td>10.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Note: Proficiency ratings were obtained using an 11-point scale for English and Korean. Ratings were obtained separately for the four language skills and averaged to yield the tabled, composite proficiency ratings.

<sup>a</sup>Ratings: 1=none; 2=very low; 3=low; 4=fair; 5=slightly less than adequate; 6=adequate; 7=slightly more than adequate; 8=good; 9=excellent; 10=perfect.

<sup>b</sup>Mean score of speaking, reading, and listening.

AOAE, Age of English Acquisition; ED, Immersion Duration.
The experiment took place over two sessions to avoid fatigue. The first session was for production of sentences in English. It lasted for approximately 1.5 hours including breaks. Within a week after the first session, participants returned for production of sentences in Korean. The second experimental session lasted for approximately an hour including breaks. All participants were monetarily compensated for their time and effort at the end of each session.

b. Stimuli

Selection of Speech Materials. Two-hundred and forty IEEE sentences, which were used in Experiment 1, were used in this experiment. The criteria for the selection of sentence stimuli were described in Experiment 1.

Recording Equipment. Each talker’s speech production was recorded, using the same recording equipment as in Experiment 1. For estimating the level of speech in each speaking style, 1-kHz reference tone was recorded at 80 dB SPL. Details were described in Experiment 1.

Recording Sentence Productions. Four types of speech materials were created: conversational, clear, exaggerated-F0, and F0-manipulated speech. Conversational, clear, and exaggerated-F0 speech were recorded by the five native Korean talkers. The F0-manipulated speech was synthesized from the exaggerated-F0 speech for perception study.

In the first session, participants were recorded producing English sentences in conversational, clear, and exaggerated-F0 speaking styles. For the English sentence production, participants were individually seated in front of a computer monitor and wore a head-mounted condenser.
Participants were assigned 48 different sentences with three practice sentences and instructed to produce sentences under three different speaking styles. First, they were recorded producing sentences in a conversational speaking style. Next, they were recorded producing the same sentences in a clear-speaking style. After completing tasks for these two speaking styles, they were recorded producing the same sentences in an exaggerated-F0 speaking style. All participants were provided with instructions that were written in Korean and English. Each target sentence was presented orthographically with a white background on a Dell 24-inch monitor in front of them during the recording session. The font-family was Arial, and the height of the lettering was approximately 10 mm.

Prior to data collection, participants produced three practice sentences based on written instructions to familiarize themselves with producing sentences in each speaking style. In the conversational speech task participants were instructed to read aloud practice sentences as they would in an everyday, normal conversation. In the same speaking style, they were recorded producing 48 different sentences. After completing the conversational speech task, participants were instructed to read aloud three practice sentences in a clear-speaking style. All participants followed a two-step instruction to minimize unnatural speech production and maximize the clear speech effect (Lam & Tjaden, 2013) as in Experiment 1. Details of the two-step instruction were described in Experiment 1. After the familiarization task, participants were recorded producing 48 different sentences clearly and naturally by over-enunciating. For the speech with exaggerated intonations, participants were
instructed to read aloud practice sentences while exaggerating their intonation by making their voice pitch cover a wider range as if they were talking to babies or dogs. In the same speaking style, they were recorded producing 48 different sentences. Each sentence was repeated three times during the recording session.

Throughout the recording session, the author continuously monitored the input level to ensure enough gain without exceeding the dynamic range of the recording system. No changes to the gain were required within each speaking condition. The recorded speech files were separately saved in WAV file format. Each speech file included three repeated sentences.

After the recording session, each audio file was segmented into sentence-length files with 50-millisecond silent leader and follower. After removing the audio files with recording-mistakes (e.g., peak clipping, narrow dynamic range, high noise level), the mid-80 percentile ranges of F0 within the rest audio files were computed using Praat. All of the five talkers showed wider F0 ranges in clear speech relative to conversational speech files, but some of them showed relatively small amount of changes in clear speech (mean difference: 7 Hz, range: 2 – 13.3 Hz). Exaggerated-F0 speech files had substantially wider F0 ranges relative to conversational speech files in all of the five talkers (mean difference: 31.7 Hz, range: 13 – 56.9 Hz). From the larger set of sentences, which for most sentences included three repetitions, 40 sentences were selected in each speaking style from each talker. For the file selection task, three native speakers of American English were recruited. For the selection of conversational and clear speech sentences, the process began by picking the middle of the three identical
sentences that were produced. If there were any recording-mistakes in the production, the sentence production that sounded most natural was selected by one of the three evaluators. For the selection of exaggerated-F0 speech, the production that sounded most natural was selected. The naturalness of speech was rated by the other two evaluators, and the speech files with highest average rating scores were selected (0: extremely unnatural to 5: extremely natural). In order to maximize the effect of F0 range in this experiment, 40 conversational-exaggerated-F0 speech pairs were selected that showed the biggest differences in F0 ranges for each talker. Later, forty clear speech files that corresponded to the pre-selected speech pairs were selected for each talker. A total of 600 speech materials were used for data analysis in production study and for additional acoustic analysis (40 sentences x 5 talkers x 3 speech conditions = 600 sentences).

F0-manipulated speech was synthesized from the exaggerated-F0 speech for use in the perception study. For the F0-manipulated speech, F0 contours of exaggerated-F0 speech were compressed to match those of conversational speech in the Praat program, using the following formula:

\[
F0_i' = \left( \frac{F0_{i\text{conv}}}{F0_{i\text{rexagg}}} \right) \ast (F0_i - F0_{i\text{med}}) + F0_{i\text{med}}
\]

where \( F0_i' \) represents the new F0 of the frame for sentence i, \( F0_{i\text{conv}} \) is the F0 range (i.e., mid-80 percentile range) of conversational speech for sentence i,
F₀ᵢ_{rexagg} is the F₀ range (i.e., mid-80 percentile range) of exaggerated-F₀ speech for sentence i, F₀ᵢ is the F₀ of exaggerated-F₀ speech for sentence i at a given time sample, and F₀ᵢ_{med} is the median F₀ of exaggerated-F₀ speech for sentence i. From this formula, 200 F₀-manipulated speech files were created from the five Korean talkers. In order to minimize potential differences between pre- and post-processing of speech materials, the sentence files in conversational, clear, and exaggerated-F₀ speech conditions were also processed by replacing the ratio between F₀ᵢ_{rconv} and F₀ᵢ_{rexagg} with 1.

Figure 7. F₀ contours for the sentence “The bill was paid every third week.” in three different speech conditions (T4 production)

Figure 7 shows representative F₀ contours of a single sentence from T4 in three speech conditions. For a comparison of F₀ contours in conversational and F₀-manipulated speech (left panel of Figure 7), the duration of conversational speech was synthetically stretched to equal the duration of F₀-manipulated speech because the sentence duration of the original conversational speech is shorter than that of exaggerated-F₀ and F₀-manipulated speech. The left panels contain the contours of conversational and F₀-manipulated speech, whereas the right panels
contain the contours of exaggerated-F0 and F0-manipulated speech. A total of 800 sentences including F0-manipulated sentences, were used for the perceptual study (40 sentences x 5 talkers x 4 conditions = 800 sentences). The final sentence list across talkers for each speech condition is provided in Appendix A.

In the second experimental session, sentence productions in Korean were obtained from the five talkers. Participants were individually seated in front of the Dell 24-inch monitor and wore a head-mounted condenser microphone as in the first experimental session. Participants were assigned the same sentences as they were assigned in the first session. All sentences were translated into Korean by the author, whose native language is Korean, for this experimental session. Each target sentence was presented orthographically with a white background on the Dell 24-inch monitor in front of them throughout the recording session. The font-family was Malgun Gothic, which is one of the most popular and widely used in Korea, and the height of the lettering was approximately 15 mm.

Participants were instructed to produce the assigned Korean-translated sentences in two different speaking styles: conversational and clear speech. First, they were recorded producing sentences in a conversational speaking style. Next, they were recorded producing the same sentences in a clear-speaking style. All of the participants were provided with the same instructions that they had in the first experimental session. To familiarize themselves with producing sentences in each speaking style, they produced three practice sentences prior to data collection.
During the recording session, the author continuously monitored the input level to ensure enough gain without exceeding the dynamic range of the recording system. No changes to the gain were required within each speaking condition. The recorded speech files were separately saved in WAV file format. Each speech file included three repeated sentences.

After the recording session, each audio file was segmented into sentence-length files with 50-millisecond silent leader and follower. After excluding files with recording-errors or translation-errors, the author obtained 147 Korean-translated conversational-clear speech pairs that were culturally appropriate across talkers: talkers T1, T2, T3, T4, and T5 had a total number of 35, 27, 31, 28, and 26 conversational-clear speech pairs, respectively. These speech materials were only used for F0 analyses in production study to explore whether a change in F0 features is consistently shown in their L1 and L2 clear speech productions.

B. Perception Study

Speech Intelligibility

a. Participants (Listeners)

Fifteen adult speakers of American English (6 males and 9 females), aged from 19 - 35 (Mean: 21.3, SD: 4.1), were recruited for speech intelligibility tasks. None of the participants were familiar with the talkers’ voices or the speech materials. All participants had no knowledge of the Korean language and self-reported normal hearing with no history of any hearing problems.
The experimental protocols were approved by the Institutional Review Board of the University of Minnesota. Informed consent was obtained from each participant. A single session of the experiment lasted for 1.5 hours including breaks. All participants were monetarily compensated for their time and effort at the end of the experiment.

b. Speech Materials

A total of 800 IEEE sentences in English, from the production study, were used for speech intelligibility tasks. Forty different sentences were obtained from each of the five native Korean speakers in four different speech conditions: conversational, clear, F0-manipulated, and exaggerated-F0 speech conditions (40 sentences x 5 talkers x 4 conditions = 800 sentences). Each sentence-length audio file, with 50-millisecond silent leader and follower, was normalized to a root-mean-square (RMS) level of 65 dB SPL. Each stimulus sentence was then embedded in speech-shaped noise that was generated by obtaining the long-term average spectrum of all sentence files from five talkers. The speech-shaped noise began 150 ms before the onset of the sentence and terminated 150 ms after the offset of the sentence. An additional ten sentences in a conversational speaking style were used for practice, but these sentences were excluded from data analysis.

c. Experimental Setup

The experimental program was written in MATLAB®. The program randomly selected 200 from the 800 IEEE sentences for a single session of the experiment. The single experimental session consisted of the same number of
sentence trials from five talkers in the following four conditions: conversational, clear, F0-manipulated, and exaggerated-F0 speech conditions (10 sentences x 5 talkers x 4 conditions = 200 sentences). None of the 200 sentences was identical across the four speech conditions in a single experimental session. The speech stimuli and speech-shaped noise were presented at 65 dB SPL binaurally through headphones (Sennheiser HD650) at 2 dB SNR across all talkers. The SNR was selected based on pilot data to avoid floor and ceiling effects in native English-speaking listeners whose hearing was within normal limits. During the experiment, each sentence was presented only one time. The pace of the experiment was manually determined by each participant who either pressed the “Enter” key or clicked a button to begin the presentation of a sentence.

d. Experimental Procedure

Participants were individually seated in a double-walled sound-isolated booth in front of a desktop computer with a video camera (Logitech HD Pro Webcam C920). Listeners were instructed to follow the prompts on the computer screen to proceed throughout the experimental tasks. After a familiarization session with ten practice trials in a conversational speech condition, 200 sentences embedded in noise were presented binaurally through headphones (Sennheiser HD650) in a random order. Listeners orally repeated each sentence that they heard while looking at a video camera mounted in the sound booth. Next, they typed in the sentence that they orally repeated. During the experimental session, listeners were videotaped and their typed responses were recorded into a spreadsheet by the MATLAB®
program. Listeners’ oral responses were scored online by the author, and the scores were verified by comparing the oral responses to the typed responses.

**Speech Naturalness**

a. Participants (Listeners)

In addition to the 15 participants for the speech intelligibility task, ten adult native speakers of American English (8 females and 2 males), aged from 21 - 49 (Mean: 27.9, SD: 10.5), were recruited for a speech naturalness task. None of the listeners was familiar with the talkers’ voices or the speech materials. All participants had no knowledge of the Korean language and self-reported normal hearing with no history of any hearing problems.

The experimental protocols were approved by the Institutional Review Board of the University of Minnesota. Informed consent was obtained from each participant. Listeners completed the task in a single half-hour session. All participants were monetarily compensated for their time and effort at the end of the experiment.

b. Speech Materials

A total of 100 IEEE sentences were randomly selected from Experiment 1 and 2. Fifty sentences were obtained from the native Korean talker group in this present study, while the remaining fifty were obtained from the native English talker group in Experiment 1. Sentences from the native English talker group were included to provide a baseline reference for the Korean talkers. Five different sentences were obtained from each of the five native Korean talkers for clear and exaggerated-F0 speaking styles (5 sentences x 5 talkers x 2 conditions =
These two speaking styles were selected to investigate whether the exaggerated-F0 speech is as (un)natural as clear speech. In the native English talker group, five different sentences were obtained from each of the four native English talkers for conversational and clear-speaking styles. To match a total number of sentences in both talker groups, an additional five sentences were obtained from one of the native English talkers for each speaking style. Each sentence-length file, with 50-millisecond silent leader and follower, was normalized to a rms level of 65 dB SPL.

c. Experimental Setup

The experimental program was written in MATLAB®. The program was used to randomize the 100 IEEE sentences for a single session of the experiment. The speech stimuli were presented at 65 dB SPL binaurally through headphones (Sennheiser HD650) in a quiet background. During the experiment, each sentence was presented only one time. The pace of the experiment was manually determined by each participant who either pressed the “Enter” key or clicked a button to begin the presentation of a sentence.

d. Experimental Procedure

Participants were individually seated in a double-walled sound-isolated booth in front of a desktop computer. Listeners were instructed to follow the prompts on the computer screen to proceed throughout the experimental tasks. A total of 100 sentences were randomly presented binaurally through headphones (Sennheiser HD650) in a quiet background. Listeners rated the naturalness of speech, by moving the slider on a scale of 1 to 7, with 1 being extremely
unnatural and 7 being extremely natural. Throughout the experimental session, listeners’ responses were recorded into a spreadsheet by the MATLAB® program. There was no familiarization session prior to the actual test.

C. Data Analysis

a. Production Study

Conversational, clear, and exaggerated-F0 speech were compared and analyzed for statistical significance in each acoustic characteristic, using an analysis of variance (ANOVA) and a linear-mixed effects regression.

b. Perception Study

The percentage of correct keywords and the speech naturalness were calculated for each talker for each speaking style. To obtain more reliable statistical results, the average percent-correct scores were transformed to arcsine units (Studebaker, 1985) before performing ANOVA that was used for examining statistical significance among four speaking styles across five talkers. Other potential perceptual correlates that account for the clear-speech benefit were also explored. The Spearman’s correlation was performed to examine the strength of a linear relationship between two measures.
III. Results

A. Production Study

Speaking rate, F0 features, long-term spectra, vowel space, and vocal intensity levels were measured in L2 (English) production for each talker for conversational, clear, and exaggerated-F0 speaking styles. The analysis included 40 sentences for each talker in each speech condition. Each property was evaluated, using the same measuring process and settings that were used in Experiment 1.

a. Speaking Rate

Words per minute (WPM), average sentence length (ASL), average pause durations (APD), and pause frequency (PF) were measured for each talker for each speaking style using Praat. Table 10 shows values of wpm for conversational, clear, and exaggerated-F0 speech. All talkers produced fewer wpm in clear speech relative to conversational speech. Exaggerated-F0 speech had nearly the same wpm as clear speech.

<table>
<thead>
<tr>
<th>Talker</th>
<th>CONV (wpm)</th>
<th>CLR (wpm)</th>
<th>EXAGG (wpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>225</td>
<td>189</td>
<td>199</td>
</tr>
<tr>
<td>T2</td>
<td>256</td>
<td>205</td>
<td>193</td>
</tr>
<tr>
<td>T3</td>
<td>216</td>
<td>159</td>
<td>161</td>
</tr>
<tr>
<td>T4</td>
<td>218</td>
<td>109</td>
<td>199</td>
</tr>
<tr>
<td>T5</td>
<td>252</td>
<td>181</td>
<td>185</td>
</tr>
<tr>
<td>Avg.</td>
<td>233</td>
<td>181</td>
<td>182</td>
</tr>
</tbody>
</table>

*Note. CONV = conversational speech, CLR = clear speech, EXAGG = exaggerated-F0 speech*
The average speech rate was 233, 181, and 182 wpm for the conversational, clear, and exaggerated-F0 speech, respectively, across talkers. Both clear and exaggerated-F0 speech were approximately 23% slower than conversational speech overall. This value was slightly smaller than that for native English talkers (i.e., 28%) in Experiment 1.

The total length of sentence (TSL) for each talker was averaged across sentences in each speaking style to measure the average sentence length (ASL). The ASL of conversational, clear, and exaggerated-F0 speech was 2,022, 2,617, and 2,594 ms, respectively, across talkers. These values were relatively longer than those for native English talkers in Experiment 1, overall. A repeated measures ANOVA revealed that there was a significant main effect of speaking style \([F(2, 8)=22.7, p<0.001]\). When Bonferroni-corrected \(\alpha=0.017\) level was applied, the repeated measures ANOVA revealed that clear speech was significantly longer than conversational speech across talkers \([F(1, 4)=68.5, p<0.001]\). Exaggerated-F0 speech did not significantly differ from clear speech across talkers \([F(1, 4)=0.26, p=0.64]\).

To estimate the extent to which an increase in the frequency and duration of pauses contributed to the increased sentence length in clear and exaggerated-F0 speech, APD and PF were calculated for each talker. The analysis included any period of silence at least 10 ms in duration (Picheny et al., 1986; Krause & Braida, 2004). The measures were conducted by the author, and no substantial discrepancies were found when a native English speaker evaluated 10% of them. The results showed that clear and exaggerated-F0 speech had more pauses than conversational speech by approximately 34% and 33%, respectively, across
talkers. These values were smaller than that for native English talkers (i.e., 50\%) in Experiment 1. The APD of conversational, clear, and exaggerated-F0 speech was 54, 77, and 74 ms, respectively, across talkers. A repeated measures ANOVA showed a significant main effect of speaking style \([F(2, 8)=23.9, p<0.001]\). When the Bonferroni-corrected \(\alpha=0.017\) level was applied, the repeated measures ANOVA revealed that clear speech had significantly longer pause durations relative to conversational speech across talkers \([F(1, 4)=30.5, p<0.01]\). There was no statistically significant difference between clear and exaggerated-F0 speech across talkers \([F(1, 4)=1, p=0.37]\).

The increased pause duration contributed to the increased sentence length in clear speech by on average 19\% (range: 13 - 32\%), which is comparable with that for native English talkers in Experiment 1. For exaggerated-F0 speech, the increased pause duration accounted for on average 17\% of the increased sentence length (range: 12 - 28\%). Individual talkers’ quantitative measurements for speaking rate are provided in Appendix C.

b. F0 Features

To explore whether native Korean talkers modify F0 features from conversational to clear speech production in their L1 and L2, the analysis included Korean sentence production for each talker in conversational and clear speech conditions.

The median, variation, (i.e., standard deviation), and mid-80 percentile range of F0 values were computed for each talker in each speaking style. Since measurement errors by the pitch detection algorithm in Praat were negligible, the
F0 features were analyzed without corrections: mis-tracked pitch points or intervals of glottal fry (i.e., creaky voice) were not found in L1 production and were less than 0.5% of the total sentence duration in L2 production for each speaking style.

Table 11 summarizes the median, range, and variation of F0 values in L1 and L2 productions for each talker for each speech condition. Differences in F0 features are shown in semitones. The compression rate (%) in L2 is the amount that the F0 range for F0-exaggerated speech was reduced to equal the range of F0 for conversational speech. These values were used for synthesizing F0-manipulated exaggerated-F0 speech, one of the listening conditions for the perception study.

**Table 11.** Median, range, and variation (SD) of F0 values in Hz for each talker for each speaking style in L1 (Korean) and L2 (English) production

<table>
<thead>
<tr>
<th>Talker</th>
<th>F0</th>
<th>L1 CNV</th>
<th>L1 CLR</th>
<th>L1 CLR-CN</th>
<th>L2 CNV</th>
<th>L2 CLR</th>
<th>L2 EXAGG</th>
<th>L2 CLR-CN</th>
<th>L2 EXAGG-CN</th>
<th>CR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Median</td>
<td>109.4</td>
<td>101.1</td>
<td>0.1</td>
<td>90.7</td>
<td>91.0</td>
<td>94.4</td>
<td>0.1</td>
<td>0.5</td>
<td>37.6</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>23.8</td>
<td>33.8</td>
<td>2.8</td>
<td>26.2</td>
<td>27.8</td>
<td>42.9</td>
<td>1.0</td>
<td>0.2</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>10.7</td>
<td>12.5</td>
<td>2.7</td>
<td>9.1</td>
<td>10.9</td>
<td>14.3</td>
<td>1.8</td>
<td>1.9</td>
<td>8.8</td>
</tr>
<tr>
<td>T2</td>
<td>Median</td>
<td>109.9</td>
<td>108.4</td>
<td>1.5</td>
<td>109.8</td>
<td>104.5</td>
<td>123.3</td>
<td>0.7</td>
<td>2.0</td>
<td>40.8</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>23.1</td>
<td>36.3</td>
<td>5.1</td>
<td>28.8</td>
<td>32.6</td>
<td>48.7</td>
<td>2.1</td>
<td>9.1</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>10.8</td>
<td>13.2</td>
<td>3.5</td>
<td>12.6</td>
<td>12.4</td>
<td>16.1</td>
<td>1.2</td>
<td>1.2</td>
<td>7.7</td>
</tr>
<tr>
<td>T3</td>
<td>Median</td>
<td>105.7</td>
<td>108.5</td>
<td>0.8</td>
<td>102.3</td>
<td>106.5</td>
<td>137.0</td>
<td>0.7</td>
<td>2.3</td>
<td>56.4</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>25.1</td>
<td>27.8</td>
<td>1.8</td>
<td>20.0</td>
<td>26.6</td>
<td>45.9</td>
<td>4.9</td>
<td>14.4</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>9.5</td>
<td>10.5</td>
<td>1.7</td>
<td>8.8</td>
<td>10.7</td>
<td>17.4</td>
<td>5.0</td>
<td>13.7</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>Median</td>
<td>97.7</td>
<td>111.8</td>
<td>2.3</td>
<td>101.3</td>
<td>119.9</td>
<td>142.6</td>
<td>2.9</td>
<td>5.8</td>
<td>65.7</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>35.4</td>
<td>92.1</td>
<td>6.7</td>
<td>34.2</td>
<td>47.7</td>
<td>69.4</td>
<td>5.6</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>13.8</td>
<td>10.9</td>
<td>6.3</td>
<td>15.2</td>
<td>18.0</td>
<td>36.5</td>
<td>6.3</td>
<td>15.7</td>
<td></td>
</tr>
<tr>
<td>T5</td>
<td>Median</td>
<td>122.4</td>
<td>135.2</td>
<td>1.7</td>
<td>104.6</td>
<td>110.8</td>
<td>125.8</td>
<td>1.0</td>
<td>3.2</td>
<td>60.8</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>46.7</td>
<td>51.6</td>
<td>1.7</td>
<td>33.9</td>
<td>44.2</td>
<td>66.4</td>
<td>4.6</td>
<td>16.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>17.6</td>
<td>19.2</td>
<td>1.5</td>
<td>12.6</td>
<td>16.9</td>
<td>32.9</td>
<td>5.1</td>
<td>16.6</td>
<td></td>
</tr>
</tbody>
</table>

Note: CLR=clear speech, CNV=conversational speech, EXAGG=Exaggerated speech, CR=compression ratio
The values for differences between two speaking styles are shown in semitones. All other values are shown in Hz.
**F0 Features in L1 (Korean) production.** The average median F0 value for clear speech was 115.1 Hz, while that for conversational speech was 108.5 Hz. A repeated measures ANOVA showed that the median F0 value was not significantly different between conversational and clear speech across talkers \([F(1, 4)=4.1, p=0.11]\).

The average F0 range (i.e., mid-80 percentile range) and variation for clear speech were 40.3 Hz and 15.1 Hz, respectively, while those for conversational speech were 32.6 Hz and 12.5 Hz, respectively. When the average F0 range was the dependent variable, a repeated measures ANOVA showed that clear speech had a significantly wider F0 range than conversational speech \([F(1, 4)=9.3, p<0.05]\). When F0 variation was the dependent variable, significantly greater variation was found in clear speech \([F(1, 4)=8.3, p<0.05]\).

The average F0 range and variation for clear speech were higher than those for conversational speech by 7.7 and 2.6 Hz, respectively.

**F0 Features in L2 (English) production.** The average values of F0 median were 101.8, 108.5, 120.2 Hz for conversational, clear, and exaggerated-F0 speech, respectively. A repeated measures ANOVA showed a significant main effect of speaking style when F0 median was the dependent variable \([F(2, 8)=8.9, p<0.01]\). When Bonferroni-corrected \(\alpha=0.025\) level was applied, the repeated measures ANOVA revealed that exaggerated-F0 speech had a significantly higher F0 value than clear speech across talkers \([F(1, 4)=13.5, p<0.025]\). The median value was not statistically different between conversational and clear speech across talkers \([F(1, 4)=4.6, p=0.1]\). However, when the analysis was performed within talker, clear speech had a significantly
higher median F0 value than conversational speech in all but T1 at a significance level of \( p=0.01 \) [T1: \( F(1, 78)=2.1, p=0.2 \)].

The average values of F0 range were 28.6, 35.8, 64.6 Hz for conversational, clear, and exaggerated-F0 speech, respectively. A repeated measures ANOVA showed a significant main effect of speaking style \([F(2, 8)=13.6, p<0.05]\). When Bonferroni-corrected \( \alpha=0.025 \) level was applied, the repeated measures ANOVA revealed that exaggerated-F0 speech had a significantly wider F0 range than clear speech across talkers \([F(1, 4)=14, p<0.025]\). The F0 range was not statistically different between conversational and clear speech across talkers \([F(1, 4)=11, p=0.029]\). However, when the analysis was performed within talker, clear speech had a significantly wider F0 range than conversational speech in all but T1 at a significance level of \( p=0.01 \) [T1: \( F(1, 78)=1.6, p=0.5 \)].

The average values of F0 variation were 11.1, 14, 24.3 Hz for conversational, clear, and exaggerated-F0 speech, respectively. A repeated measures ANOVA showed a significant main effect of speaking style \([F(2, 8)=13.6, p<0.05]\). When Bonferroni-corrected \( \alpha=0.025 \) level was applied, the repeated measures ANOVA revealed that exaggerated-F0 speech had significantly greater F0 variation than clear speech across talkers \([F(1, 4)=14.8, p<0.025]\). F0 variation was not statistically different between conversational and clear speech across talkers \([F(1, 4)=9.5, p=0.037]\). However, when the analysis was performed within talker, clear speech had a significantly greater F0 variation than conversational speech in all but T2 at a significance level of \( p=0.05 \) [T2: \( F(1, 78)=2.1, p=0.2 \)].
Cross-language similarities in F0 Features. Overall, a wider F0 range with greater F0 variation was found in L1 and L2 clear speech productions. These results are consistent with findings in other studies that examined cross-language similarities in L2 learners of English whose native language is Croatian or Finish (Smiljanić & Bradlow, 2005; Granlund et al., 2011, 2012). A higher F0 value was found in L2 but not in L1 clear speech production.

c. Long-term Spectra

Speech spectra of conversational, clear, and exaggerated-F0 speech were computed after normalization for long-term rms level. Figure 8 shows the spectral distribution as a function of third-octave band frequency for conversational, clear, and exaggerated-F0 speech in each talker.

Figure 8. Long-term spectral distribution for conversational (Conv), clear (Clear), and exaggerated-F0 (Exagg) speech in each talker.
Figure 8 reveals that clear speech contains more spectral energy above 0.63-kHz, relative to conversational speech in each talker. Talkers T4 and T5 show relatively large spectral enhancements, whereas talkers T1 and T3 show relatively small enhancements in clear speech. Exaggerated-F0 speech shows nearly the same or larger spectral enhancements than clear speech in all talkers. Table 12 shows the results of third-octave band analysis that summarizes in dB the high-frequency spectral change in clear and exaggerated-F0 speech.

**Table 12.** Average band RMS level across the frequency range between 0.63- and 6-kHz for each talker in each speech condition. The third-octave bands centered at 0.63- and 0.8-kHz contributed a fractional amount so this average level is referred to in the text as a “high-frequency average.”

<table>
<thead>
<tr>
<th>Speech Condition</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNV</td>
<td>61.4</td>
<td>61.2</td>
<td>64.7</td>
<td>62.3</td>
<td>61.8</td>
<td>62.3</td>
</tr>
<tr>
<td>CLR</td>
<td>61.9</td>
<td>62.4</td>
<td>65.6</td>
<td>66.8</td>
<td>64.4</td>
<td>64.2</td>
</tr>
<tr>
<td>EXAGG</td>
<td>62.5</td>
<td>62.7</td>
<td>65.1</td>
<td>68.5</td>
<td>65.2</td>
<td>64.8</td>
</tr>
<tr>
<td>CLR-CN V</td>
<td>0.3</td>
<td>1.2</td>
<td>0.9</td>
<td>4.5</td>
<td>2.6</td>
<td>1.9</td>
</tr>
<tr>
<td>EXAGG-CN V</td>
<td>0.9</td>
<td>1.4</td>
<td>0.4</td>
<td>6.2</td>
<td>3.4</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Note. CNV=conversational speech, CLR=clear speech, EXAGG=exaggerated speech, CLR-CN V=spectral change in clear speech, EXAGG-CN V=spectral change in exaggerated speech.

To quantify the high-frequency energy for each speaking style, band RMS levels were averaged across frequencies between 0.63- and 6-kHz as in Experiment 1. As shown in the table, clear speech had on average 1.9 dB more spectral energy than conversational speech. Exaggerated-F0 speech had on average 2.5 dB more spectral energy than conversational speech. In both clear and exaggerated-F0 speech, talkers T4 and T5 showed relatively larger spectral enhancements, compared with the other talkers. Overall, these results demonstrate that clear and exaggerated-F0 speech has more favorable SNR than
conversational speech even when rms amplitude was equalized across speaking styles. The clear-speech enhancement is larger than that for native English talkers in Experiment 1 and participant dependent.

d. Vowel Space

The analysis included four corner vowels (i.e., /i/, /u/, /a/, and /æ/) from the keywords in 40 sentence stimuli for each talker in each speech condition. Since Korean does not have the vowel /æ/, the present study did not assess whether the productions matched the target vowels. To investigate vowel space in each speaking style, first (F1) and second formant (F2) values were estimated, using the same measurement settings and process as described in Experiment 1. The author manually segmented vowel intervals to maximize the accuracy of the Linear Predictive Coding (LPC) formant-tracking algorithm embedded in Praat. When the algorithm apparently mis-tracked the formant values, the author corrected by manually measuring the formants using visual inspection. The accuracy of hand-measures was verified by a native English speaker in 10% of the measures, and no substantial discrepancies were found.

Prior to data analysis, formant values (Hz) were converted to the auditory Bark scale, a perceptually realistic scale of frequency (Zwicker & Terhardt, 1980). The Euclidean distance (Bark) of the individual vowels from the average F1/F2 (Bark) was calculated to estimate the vowel-space dispersion for each talker, as described in Experiment 1. Figure 9 shows the vowel-space dispersion for each talker and the average values in clear, conversational, and exaggerated-F0 speech.
Figure 9. F1/F2 values (Bark) of vowels for each talker and average Euclidean distance (Bark) of each vowel from the grand mean F1/F2 values across talkers in conversational (CNV), clear (CLR), and exaggerated-F0 (EXG) speech.

In the figure, the individual X indicates the average F1/F2 value of all vowels for each talker. The square filled in red indicates the grand mean of the values. Individual bold vowel letters represent the average F1/F2 values of each talker’s clear vowel tokens while individual italic vowel letters represent the values of each talker’s exaggerated-F0 vowel tokens. Individual small vowel letters represent the values of each talker’s conversational vowel tokens. The grand mean of the values is shown as the triangle, circle, and diamond filled in red for conversational, clear, and exaggerated-F0 speech, respectively. The average vowel-space dispersion across talkers is shown as the dashed, solid, and dotted lines for conversational, clear, and exaggerated-F0 speech, respectively.

Figure 9 reveals that clear speech has greater vowel-space dispersion with higher mean F1 values in the vowel /ɑ/ and /æ/ and higher mean F2 values in the
vowel /i/ and /æ/, when compared with conversational speech. This pattern is similar to that for native English talkers, but native Korean talkers show relatively lower F1 and F2 values than native English talkers in Experiment 1. Exaggerated-F0 speech shows larger vowel-space dispersion, relative to conversational speech. Overall, F1 values of exaggerated-F0 vowels are higher than those of conversational vowels except for the vowel /i/, whereas F2 values of exaggerated-F0 vowels are nearly the same as those of conversational vowels except for the vowel /i/. For the vowel /i/, F1 is nearly the same in both speaking styles while F2 is higher in the exaggerated-F0 speech condition relative to the conversational speech condition.

The linear mixed-effects models fitted in Experiment 1 were used to analyze F1 and F2 values and the Euclidean distance for each speaking style across talkers. Details about the models can be found in Experiment 1. When F1 values were the dependent measure, there were significant main effects of vowel and type. F1 values were statistically higher in the vowel /ɑ/ relative to vowels /i/ and /u/ and in the vowel /æ/ relative to vowels /u/ and /i/ at a significance level of 0.05. Overall, the association between clear and conversational speech with F1 depended on the vowel. In particular, for the vowels /ɑ/ and /i/, there was an interaction between vowel and speech type. F1 values in clear speech were comparable with those in exaggerated-F0 speech. Table 13 summarizes the results of statistics in which the vowel /ɑ/ is the reference vowel.
Main effects were qualified by a number of significant interactions between vowel and type. The nature of these interactions is shown in the left panel of Figure 10 which plots the clear, conversational, and exaggerated-F0 vowels across talkers. Based on visual inspection, F1 is slightly lower for the vowel /i/, which is a closed vowel, while F1 is higher for vowels /α/ and /æ/, which are open vowels, in clear speech relative to conversational speech. There is no notable difference for the vowel /u/. Exaggerated-F0 speech shows relatively higher F1 values for vowel tokens /α/, /æ/, and /u/, compared with conversational speech. The vowel /i/ is nearly the same in both speaking styles.
When F2 values were the dependent measure, there were significant main effects of vowel and type. F2 values were statistically higher in the vowel /i/ relative to vowels /ɑ/, /æ/, and /u/ and in the vowel /æ/ relative to vowels /ɑ/ and /u/ at a significance level of 0.05. Overall, the association between clear and conversational speech with F2 depended on the vowel. In particular, for the vowels /i/ and /u/, there was an interaction between vowel and speech type. F2 values in clear speech were comparable those in exaggerated-F0 speech in general. Table 14 summarizes the results of statistics in which the vowel /i/ is the reference vowel.

Figure 10. Type and vowel interaction for F1 (left panel) and F2 (right panel)
Table 14. Main effects of vowel and type and interaction between vowel and type for F2 values (reference vowel: /i/)

| Fixed Effects | Estimate | Std. Error | df | t-value | Pr(>|t|) | Signif. Codes |
|---------------|----------|------------|----|---------|----------|---------------|
| (Intercept)   | -13.12   | 0.17       | 3.37 | 7.97    | 0.00     | ***           |
| vowel /a/     | -4.07    | 0.22       | 10.26 | -18.11  | 0.00     | ***           |
| vowel /æ/     | -2.47    | 0.19       | 10.75 | -7.71   | 0.00     | ***           |
| vowel /u/     | -3.59    | 0.38       | 5.40  | -8.99   | 0.00     | ***           |
| type conversational | 0.26 | 0.15 | 24.77 | -1.47 | 0.11 | * |
| type exaggerated-F0 | 0.01 | 0.14 | 20.73 | 0.24 | 0.60 | |
| vowel /a/ | type conversational | 0.34 | 0.26 | 983.38 | 1.43 | 0.15 | |
| vowel /æ/ | type conversational | 0.06 | 0.21 | 985.80 | 0.44 | 0.66 | |
| vowel /u/ | type conversational | 0.34 | 0.22 | 985.35 | 2.10 | 0.03 | |
| vowel /æ/ | type exaggerated-F0 | 0.02 | 0.22 | 987.32 | 0.10 | 0.91 | |
| vowel /u/ | type exaggerated-F0 | 0.20 | 0.22 | 984.86 | 0.79 | 0.43 | |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Main effects were qualified by a number of significant interactions between vowel and type. The nature of these interactions is shown in the right panel of Figure 10 which plots the clear, conversational, and exaggerated-F0 vowels across talkers. Based on visual inspection, F2 is lower for vowels /a/ and /u/, which are back vowels, while F2 is higher for vowels /i/ and /æ/, which are front vowels, in clear speech, relative to conversational speech. Exaggerated-F0 speech shows nearly the same F2 values for all but the vowel /i/ as conversational speech.

Table 15. Main effects of type for the average Euclidean distance

| Fixed Effects | Estimate | Std. Error | df | t-value | Pr(>|t|) |
|---------------|----------|------------|----|---------|---------|
| (Intercept)   | 2.28     | 0.10       | 4.14 | 23.63   | 0.00    | ***      |
| type conversational | -0.20 | 0.07 | 9.31 | -2.65 | 0.01 | * |
| type exaggerated-F0 | -0.11 | 0.07 | 5.43 | -1.53 | 0.18 | |

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
When Euclidean distance was the dependent measure, there was a significant main effect of type. As shown in Table 15, clear speech showed larger vowel dispersion, compared with conversational speech. However, there was no significant difference in the vowel-space dispersion between clear and exaggerated-F0 speech.

e. Vocal Intensity Level

Table 16 shows vocal intensity levels for five talkers in each speech condition. The given values correspond to the level at a distance of 1 m (method described in Experiment 1).

<table>
<thead>
<tr>
<th>Talker</th>
<th>CLR</th>
<th>CNV</th>
<th>EXAGG</th>
<th>CLR-CN</th>
<th>EXAGG-CLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>64.2</td>
<td>64.9</td>
<td>66.3</td>
<td>-0.7</td>
<td>2.1</td>
</tr>
<tr>
<td>T2</td>
<td>67.7</td>
<td>66.1</td>
<td>68.3</td>
<td>1.5</td>
<td>0.6</td>
</tr>
<tr>
<td>T3</td>
<td>64.2</td>
<td>64.3</td>
<td>67.2</td>
<td>-0.1</td>
<td>3.0</td>
</tr>
<tr>
<td>T4</td>
<td>71.5</td>
<td>67.0</td>
<td>77.5</td>
<td>4.5</td>
<td>6.0</td>
</tr>
<tr>
<td>T5</td>
<td>65.5</td>
<td>64.3</td>
<td>67.0</td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Avg.</td>
<td>66.6</td>
<td>65.3</td>
<td>69.3</td>
<td>1.3</td>
<td>2.6</td>
</tr>
</tbody>
</table>

*Note: Values are the vocal intensity levels in dB SPL for the face-to-face conversation distance, 1m. CLR=clear speech, CNV=conversational speech, EXAGG=exaggerated-F0 speech*

As shown in the table, the speech levels in a conversational speaking style fall within the range of the normal conversational speech level in a quiet room (i.e., 60-70 dB SPL). In general, the intensity of clear speech was greater than that of conversational speech, which is consistent with the results for native English talkers in Experiment 1 and other studies (Picheny et al., 1986; Lam et al., 2012). However, individual variability was found: talkers T1 and T3
showed nearly the same or lower intensity level in clear speech relative to conversational speech. Exaggerated-F0 speech was on average 2.6 dB higher than clear speech across talkers. This pattern was found in all talkers. The level differences were eliminated in the perception study, by equalizing rms amplitude across all sentences in different speech conditions.

B. Perception Study

a. Speech Intelligibility

Average intelligibility scores were obtained from a total of 150 sentences across 15 listeners in each speech condition for each talker.

*Figure 11.* Average intelligibility scores for each talker and grand mean scores in conversational (CNV), clear (CLR), exaggerated-F0 (Exagg-F0), and F0-manipulated (F0-MNP) speech conditions
Figure 11 shows average Keyword percent-correct scores for each talker and grand mean scores in each speech condition. All talkers showed intelligibility benefits from clear and exaggerated-F0 speech, compared with conversational speech. Clear speech showed better performance relative to exaggerated-F0 speech across talkers, but individual differences were found. Talkers T4 and T5 showed substantially better performance in clear speech, whereas talkers T1-T3 showed nearly the same performance in the two speech conditions. Exaggerated-F0 speech showed nearly the same performance as F0-manipulated speech across talkers.

Prior to data analysis with a repeated measures ANOVA, the average scores were transformed to arcsine units that yield more reliable statistical results (Studebaker, 1985). The results showed that there was a significant main effect of speech condition \(F(3,42)=67.8, p<0.001\). When Bonferroni-corrected \(\alpha=0.0125\) level was applied, the repeated measures ANOVA revealed that clear speech performed significantly better than exaggerated-F0 speech \(F(1,14)=9.95, p<0.01\) that showed significantly better performance than conversational speech across talkers \(F(1,14)=302.7, p<0.001\). Manipulated-F0 speech did not statistically differ from exaggerated-F0 speech across talkers \(F(1,14)=4.26, p=0.06\).

To examine whether the correct response rates are affected by the temporal position of keywords within a sentence, keyword recognition scores were averaged across all listeners in each keyword position (position 1-5) for each speaking style. The average recognition scores were nearly the same from the beginning to the end of sentences, which is distinguished from the patterns for
native English talkers in Experiment 1. The left panel of Figure 12 includes the results for each condition in this study and, for comparison, the results for the sinusoidally frequency-modulated speech condition from Miller et al. (2010). The percent correct values for the frequency-modulated speech are the mean values for two frequency-modulated speech conditions at rates of 2.5 and 5.0 Hz.

![Figure 12. Keyword recognition scores across temporal keyword position (left panel) vs. relative intensity levels over time within sentences (right panel). Times listed in the right panel represent midpoints of 300 ms analysis window.](image)

The right side of Figure 12 shows the relative intensity levels (dB) over time within sentences. The relative dB level was estimated, based on the average rms values at different time points after equating speech from different conditions for the overall rms level. The values were obtained using 300 ms windows. Analysis included approximately half of the total number of sentences that had a duration equal to or longer than the average sentence duration in each speech condition. The 300 ms windows included for analysis were only calculated for intervals that were one window prior to the end of a sentence. Using this rule, fewer sentences remained for analysis for the longest duration
sentences. When the pool of sentences was fewer than 17, those sentences were not included in the analysis. Table 17 shows a total number of sentences that were included for the intensity analysis over time within sentences. Since the intensity levels in F0-manipulated speech are the same as those in exaggerated-F0 speech, F0-manipulated speech condition was not included in this analysis.

Table 17. A total number of sentences used for analyzing relative dB over time within sentences (ms). Times listed represent midpoints of 300 ms analysis window.

<table>
<thead>
<tr>
<th>Number of Sentences</th>
<th>Time within Sentences (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>150</td>
</tr>
<tr>
<td>CNV</td>
<td>93</td>
</tr>
<tr>
<td>CLR</td>
<td>93</td>
</tr>
<tr>
<td>EXAGG</td>
<td>92</td>
</tr>
</tbody>
</table>

Note: CNV=conversational speech, CLR=clear speech, EXAGG=exaggerated-F0 speech

As shown in the left panel of figure 12, the intensity level decreases over time within sentences and the degree of reduction in intensity across time are nearly the same in all three speech conditions. This decreasing trend is similar to that for native English talkers in Experiment 1. For the conversational speech, native Korean talkers had a smaller amount of reduction in intensity level across time, compared with native English talkers in Experiment 1 (3 dB vs 6.3 dB). For the clear speech condition, by contrast, a similar amount of reduction was observed for both groups (3.8 dB for native Korean talkers vs. 4.5 dB for native English talkers).

b. Speech Naturalness
Speech naturalness (scale 1=extremely unnatural to 7=extremely natural) was rated by ten native English-speaking listeners in a quiet background. Table 18 shows the average values of speech naturalness for two speaking conditions in the native Korean and the native English talker group.

**Table 18.** Mean speech naturalness scores for each talker group for each speaking style. (Scale 1=extremely unnatural to 7=extremely natural)

<table>
<thead>
<tr>
<th>Talker group &amp; Condition</th>
<th>NKT</th>
<th>NET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Scores</td>
<td>CLR</td>
<td>EXAGG</td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>3.9</td>
</tr>
</tbody>
</table>

*Note.* NKT=native Korean talker group, NET=native English talker group, CLR=clear speech, CNV=conversational speech, EXAGG=exaggerated-F0 speech

In the native Korean talker group, a repeated measures ANOVA revealed that clear speech was more natural than exaggerated-F0 speech across talkers [$F(1,9)=10.4, p=0.01$]. In the native English talker group, the analysis revealed that the naturalness for clear speech did not statistically differ from that for conversational speech across talkers [$F(1,9)=2.3, p=0.16$]. A significant correlation between speech naturalness and keyword performance was not found in the native Korean talker group across two speaking styles when the Spearman’s rank-order correlation was conducted ($r_{[128]}=0.2, p=0.5$).
IV. Discussion

English (L2) clear speech produced by five male native speakers of Seoul Korean, who started to learn English after the age of six, showed acoustic changes that have been reported in other cross-linguistic studies (Smiljanić & Bradlow, 2005; Li & So, 2006; Granlund et al., 2011, 2012). These modifications include slower speaking rates with longer and more pauses, wider vowel space, higher vocal intensity level, expanded F0 range, and high-frequency emphasis. An increase in high-frequency energy was more prominent while other modifications were relatively smaller in native Korean talkers, compared with gender-matched native talkers of English in Experiment 1 and other studies (Picheny et al., 1986; Bradlow et al., 2003; Krause & Braida, 2004). In general, exaggerated-F0 speech had clear-speech-like acoustic modifications across talkers.

Among those clear-speech modifications, the contribution of F0 range to the English clear-speech benefit was the primary focus of this study. Since the expanded F0 range was relatively small in clear speech (mean: 3.7 semitones, range: 1.0 - 5.8 semitones), exaggerated-F0 speech, which had clear-speech-like acoustic changes, was alternatively used to investigate the role of F0 range in the intelligibility benefit. The increased F0 range was on average 13.3 semitones (range: 8.2 - 18.5 semitones) in exaggerated-F0 speech across talkers. The F0 range of exaggerated-F0 speech was compressed by on average 52.3% (range: 37.6 – 65.7%) to yield the same range as conversational speech. This compression rate was comparable with that for native English talkers (mean: 53.1%, range: 44.1 – 62.9%) in Experiment 1. The native English talkers in Experiment 1 were selected to have a wide range of F0 in their clear speech. The results showed that a reduction in F0 range did not
significantly affect the intelligibility benefit in sustained noise, which is consistent with the finding for native English talkers in Experiment 1. The overall clear-speech benefit was 19 percentage points (range: 5.7 – 29.6 percentage points). Nearly the same benefit was seen for the exaggerated-F0 speech condition (mean: 16 percentage points, range: 10.3 – 23.3 percentage points). These intelligibility benefits were smaller, compared with the clear speech benefit for native English talkers in Experiment 1 (mean: 32 percentage points, range: 22 - 48 percentage points).

This present study extends the finding that F0 range does not affect the clear speech advantage when the range is compressed to match that of conversational speech. In one interpretation, this outcome supports the idea that concomitant acoustic-phonetic changes in naturally produced hyper-articulated speech are not removed by a simple compression of F0 contours. This could be a result of perceptually identical F0 cues remaining before and after F0-manipulation. F0 movements are thought to be crucial for segmenting speech signals (Cutler & Foss, 1977; Binns & Cullings, 2007; Spitzer et al., 2007; Liss et al., 2000; Miller et al., 2010). Since F0 contours were not completely flattened in the F0-manipulated speech, native English listeners could have taken advantage of the F0 cues to facilitate speech segmentation in noise.

A second interpretation questions whether the Korean talkers produced accurate prosodic cues in any of the speech conditions. If incorrect prosodic cues were audible in all three speech conditions (i.e., conversation, clear, and exaggerated-F0 speech) then compressing the F0 range in the exaggerated condition (i.e., F0-manipulated speech) would not be expected to reduce speech understanding. The evidence for this explanation is that native Korean talkers showed a different pattern
for the correct response rates by the position of keywords within a sentence from native English talkers, regardless of speaking styles (see the left panel of Figure 6 for native English talkers and left panel of Figure 12 for native Korean talkers). As shown in the left panel of Figure 12, sinusoidally frequency-modulated speech in the data of Miller et al. (2010) has the same pattern as those for native Korean talkers in this present study. This frequency-modulated speech conveys incorrect prosodic information by distorting the normal stress pattern of the words, which consequently results in poorer intelligibility scores than flattened-F0 speech which is considered a “neutral” cue condition (Miller et al., 2010). Lexical and sub-lexical segmentation cues are essential for speech segmentation. When these cues are incorrectly provided, listeners have difficulty in analyzing speech signals in an efficient way (Mattys & Samuel, 1997; Spitzer et al., 2007).

Unnaturally produced speech may be another indicator to show that Korean-speaking L2 learners of English do not use native-like prosodic cues when they produce an English sentence. Anand and Stepp (2015) suggested that speech naturalness can be negatively affected by the use of incorrect prosodic cues. In this present study, native Korean talkers’ speech productions were less natural relative to native English talkers’ in general. This result indicates that the highly efficient native Korean talkers in the present study did not have native-like language competence. It, however, cannot be ruled out that other aspects of speech (e.g., phonological differences) could also affect the naturalness judgment.

The idea that the prosodic cues used by the Korean talkers may have been incorrect has its basis in the different prosodic systems in Korean and English. Korean (Seoul dialect) has phrase-based prominence while English has lexical-based prominence (Tremblay et al., 2016). In Korean, a high phrase accent is observed in
the word- or phrase-final position, which is based on the tonal pattern of the accentual phrase. An F0 rise, therefore, signals word- or phrase-final boundaries in Korean (Jun, 1998). By contrast, in English pitch accents are placed on stressed syllables. In other words, they are not necessarily phrase-final like in Korean. Instead, an F0 rise statistically tends to appear more in word-initial rather than word-final boundaries (Cutler & Carter, 1987; Clopper, 2002). As shown in Figure 6 and 12, both native Korean and native English talkers have the same results (i.e., decreasing trend) for the intensity levels over time within a sentence. However, these two talker groups have different patterns for the correct response rates by the position of keywords within a sentence. Such different patterns may be partially attributed to the different prosodic systems in Korean and English. Native Korean talkers, who have incomplete mastery of English, may follow the rules of L1 for some prosodic features when they produce a sentence in English.

In this present study, all but T1 expanded F0 range in both L1 and L2 clear speech productions, as shown for Croatian-English and Finish-English bilinguals in previous studies (Smiljanić & Bradlow, 2005; Granlund et al., 2011, 2012). This finding contradicts the result in a previous study (Cho et al., 2011) in which native Korean talkers did not modify F0 range from conversational to clear speech production in their native language. Native Korean talkers in this present study had a high level of English proficiency as Croatian-English and Finish-English bilinguals had in previous studies (Smiljanić & Bradlow, 2005; Granlund et al., 2011, 2012). Their high-L2-proficiency may partially account for the cross-language similarities in the clear-speech modification. Cook (2003) suggested that L1 speech production can be influenced by L2 (i.e., backward transfer), and vice versa. This bidirectional
inter-lingual transfer has been reported in several studies (Pavlenko & Jarvis, 2002; Brown & Gullberg, 2008; Jarvis & Pavlenko, 2008). Nevertheless, since the study of Cho et al. (2011) and this present study had different clear speech instructions and small sample size, it is hard to make direct comparisons.

In addition to F0, other acoustic correlates of the clear-speech benefit were examined. Experiment 1 revealed that a high-frequency emphasis seen in clear speech contributes substantially to the clear-speech benefit, whereas extended pause duration contributes minimally to the clear-speech benefit in native English talkers. Since these two global modifications were also found in native Korean talkers, the perceptual correlates were explored with the speech production and perception data obtained from the present study. To quantify the contribution of each acoustic change to the clear-speech benefit, the relation of 4.5% per dB in IEEE sentences (MacPherson & Akeroyd, 2014) was used after converting the production difference between clear and conversational speech into the signal-to-noise ratio (SNR).

To quantify the contribution of pause extension to the clear-speech benefit, a dB change that is attributed to the pause extension was calculated by subtracting the value of log ratio between the total sentence length with and without pauses in conversational speech from the value of log ratio in clear speech. The overall dB change was 0.13 dB (range: 0.1-0.2 dB), which predicts approximately 0.6% increase in the intelligibility score when the relation of 4.5% per dB is used. This value accounts for approximately 5.2% of a total clear-speech benefit observed in this study. The degree of contribution was negligible.

To quantify the contribution of high-frequency emphasis to the clear-speech benefit, the decibel difference in the high-frequency region was converted to the
predicted increase in performance, using the relation of 4.5% per dB of gain in the high-frequency SNR. Details about the process are described in the results section. The overall dB change was 1.9 dB (range: 0.3-4.5 dB), which predicts approximately 8.6% increase in the intelligibility score. This value accounts for approximately 45% of a total clear-speech benefit observed in this present study. The degree of contribution is markedly higher than that for native English talkers in Experiment 1. In particular, substantially higher percentage contribution was found in T4 and T5 who showed biggest clear-speech benefit (average 29.6 percent points for both talkers). By contrast, a relatively small percentage contribution was found in other talkers who showed smaller clear-speech benefits. The Spearman’s rank-order correlation coefficient showed a significant, positive correlation between high-frequency emphasis and intelligibility benefit in native Korean talkers ($r_s[0.5]=0.97$, $p<0.01$). This outcome supports the finding that a high-frequency emphasis is an effective strategy for improving speech intelligibility in noise for native Korean talkers speaking English. When Spearman’s correlation was calculated using the data for native English talkers in Experiment 1 ($n=4$) and in the study of Krause and Braida (2009) ($n=4$), a significant correlation was not found ($r_s[75]=0.1$, $p=0.8$).

These analyses for native Korean talkers suggest that approximately a half of the total clear-speech benefit observed in the present study can be accounted for by the two acoustic modifications but mostly by the high-frequency emphasis. Further, compared with native English talkers, native Korean talkers tend to more rely on these global modifications to improve their intelligibility in clear speech.
V. Conclusion

The present study provides the extended finding that F0 range does not affect the clear-speech-like benefit. This outcome supports the conclusion of Experiment 1 that the change in F0 range of clear speech is a secondary effect, rather than a primary factor that directly contributes to the clear-speech benefit.

The contribution of high-frequency emphasis to the clear-speech benefit was substantial, whereas the contribution of pause extension to the clear-speech benefit was much smaller. In particular, the enhancement of high-frequency spectral energy can be an important indicator of determining the degree of clear-speech benefits in native Koreans speaking English.
Bibliography


Gerlach, A. (2013). *Psychoanalysis in Asia: China, India, Japan, South Korea, Taiwan.* Karnac Books.


Appendix A

Selected IEEE sentences for Experiments 1 and 2

Although many sentences are the same for both experiments, different selection criteria and recording errors resulted in some differences that are listed in this appendix.

One hundred ninety-two sentences were used for each speaking style in Experiment 1, while 200 sentences were used for each speaking style in Experiment 2.

Five keywords are in bold in each sentence.

**Sentences selected that are common to Experiments 1 and 2**

A BIG WET STAIN was on the ROUND CARPET.
A BREAK in the DAM ALMOST CAUSED a FLOOD.
A CASTLE BUILT from SAND FAILS to ENDURE.
A CLEAN NECK MEANS a NEAT COLLAR.
A KING RULED the STATE in the EARLY DAYS.
A POUND of SUGAR COSTS MORE than EGGS.
A RICH FARM is RARE in this SANDY WASTE.
A ROD is USED to CATCH PINK SALMON.
A SAW is a TOOL USED for MAKING BOARDS.
A SIX COMES up MORE OFTEN than a TEN.
A THIN BOOK FITS in the SIDE POCKET.
ADD SALT BEFORE you FRY the EGG.
ADDING FAST LEADS to WRONG SUMS.
AFTER the DANCE, they WENT STRAIGHT HOME.
ALL SAT FROZEN and WATCHED the SCREEN.
AT that HIGH LEVEL the AIR is PURE.
Be SURE to SET the LAMP FIRMLY in the HOLE.
BOTH BROTHERS WEAR the SAME SIZE.
BRING YOUR PROBLEMS to the WISE CHIEF.
CATS and DOGS EACH HATE the OTHER.
CHOOSE BETWEEN the HIGH ROAD and the LOW.
CRACK the WALNUT with your SHARP SIDE TEETH.
DRAW the CHART with HEAVY BLACK LINES.
DROP the ASHES on the WORN OLD RUG.
DULL STORIES MAKE HER LAUGH.
EVERY WORD and PHRASE he SPEAKS is TRUE.
FAIRY TALES SHOULD be FUN to WRITE.
FAKE STONES SHINE but COST LITTLE.
FEED the WHITE MOUSE some FLOWER SEEDS.
FINE SOAP SAVES TENDER SKIN.
FLOAT the SOAP on TOP of the BATH WATER.
FLY by NIGHT, and you WASTE LITTLE TIME.
GLUE the SHEET to the DARK BLUE BACKGROUND.
He ORDERED PEACH PIE with ICE CREAM.
He SAID the SAME PHRASE THIRTY TIMES.
He WROTE DOWN a LONG LIST of ITEMS.
He WROTE his LAST NOVEL there AT the INN.
His HIP STRUCK the KNEE of the NEXT PLAYER.
His SHIRT was CLEAN but ONE BUTTON was GONE.
It is HARD to ERASE BLUE or RED INK.
It is LATE MORNING on the OLD WALL CLOCK.
It was DONE BEFORE the BOY could SEE IT.
IT'S EASY to TELL the DEPTH of a WELL.
JAZZ and SWING FANS like FAST MUSIC.
KICK the BALL STRAIGHT and FOLLOW THROUGH.
LEAVE NOW and YOU will ARRIVE on TIME.
LEAVES TURN BROWN and YELLOW in the FALL.
LET'S all JOIN as we SING the LAST CHORUS.
LIFT the SQUARE STONE OVER the FENCE.
MARK the SPOT with a SIGN PAINTED RED.
MEN THINK and PLAN and SOMETIMES ACT.
MOST of the NEWS is EASY for US to HEAR.
MUCH of the STORY MAKES GOOD SENSE.
NEAT PLANS FAIL WITHOUT LUCK.
NINE MEN were HIRED to DIG the RUINS.
NINE ROWS of SOLDIERS STOOD in LINE
OAK is STRONG and ALSO GIVES SHADE.
On the ISLANDS the SEA BREEZE is SOFT and MILD.
OPEN YOUR BOOK to the FIRST PAGE.
PACK the KITS and DON'T FORGET the SALT.
PACK the RECORDS in a NEAT THIN CASE.
PAPER will DRY OUT WHEN WET.
PICK a CARD and SLIP it UNDER the PACK.
PILE the COAL HIGH in the SHED CORNER.
POST NO BILLS on this OFFICE WALL.
PRESS the PEDAL WITH your LEFT FOOT.
READ VERSE OUT LOUD for PLEASURE.
SEED is NEEDED to PLANT the SPRING CORN.
SEND the STUFF in a THICK PAPER BAG.
SHAKE HANDS WITH this FRIENDLY CHILD.
SHAKE the DUST from YOUR SHOES, STRANGER.
She HAS a SMART WAY of WEARING CLOTHES.
SHE SAW a CAT in the NEIGHBOR'S HOUSE.
SHE was KIND to SICK OLD PEOPLE.
SICKNESS KEPT him HOME the THIRD WEEK.
SLIDE the BOX INTO that EMPTY SPACE.
SLIDE the TRAY ACROSS the GLASS TOP.
SOAP can WASH MOST DIRT AWAY.
STOP and STARE at the HARD WORKING MAN.
SWEET WORDS WORK BETTER than FIERCE.
TAKE the MATCH and STRIKE it AGAINST your SHOE.
TEN PINS were SET IN ORDER.
That GUY is the WRITER of a FEW BANNED BOOKS.
THAT MOVE MEANS the GAME is OVER.
The AIM of the CONTEST is to RAISE a GREAT FUND.
The BABY PUTS his RIGHT FOOT in his MOUTH.
The BANK PRESSED FOR PAYEMENT of the DEBT.
The BEST METHOD is to FIX it in PLACE with CLIPS.
The BIG RED APPLE FELL to the GROUND.
The BILL was PAID EVERY THIRD WEEK.
The BILLS were MAILED PROMPTLY on the TENTH of the MONTH.
The BLIND MAN COUNTED his OLD COINS.
The BOX was THROWN BESIDE the PARKED TRUCK.
The BOY was THERE WHEN the SUN ROSE.
The CHILD ALMOST HURT the SMALL DOG.
The CLOTHES DRIED on a THIN WOODEN RACK.
The COFFEE STAND is TOO HIGH for the COUCH.
The CURTAIN ROSE and the SHOW WAS ON.
The DARK POT HUNG in the FRONT CLOSET.
The DRY WAX PROTECTS the DEEP SCRATCH.
The FIGHT will END in JUST SIX MINUTES.
The FLY MADE its WAY ALONG the WALL.
The LAZY COW LAY in the COOL GRASS.
The LEASE RAN OUT in SIXTEEN WEEKS.
The LONG JOURNEY HOME TOOK a YEAR.
The MEAL was COOKED BEFORE the BELL RANG.
The MUSIC PLAYED ON WHILE they TALKED.
The NEW GIRL was FIRED TODAY at NOON.
The PEARL was WORN in a THIN SILVER RING.
The PENCIL was CUT to the SHARP at BOTH ENDS.
The PLANT GREW LARGE and GREEN in the WINDOW.
The PLAY BEGAN as SOON as we SAT DOWN.
The PLAY SEEMS DULL and QUITE STUPID.
The POOR BOY MISSED the BOAT AGAIN.
The PRICE is FAIR for a GOOD ANTIQUE CLOCK.
The ROPE will BIND the SEVEN BOOKS at ONCE.
The RUDE LAUGH FILLED the EMPTY ROOM.
The SALT BREEZE CAME ACROSS from the SEA.
The SENSE of SMELL is BETTER THAN that of TOUCH.
The SINK is the THING in WHICH we PILE DISHES.
The SKY that MORNING was CLEAR and BRIGHT BLUE.
The SMALL RED NEON LAMP went OUT.
The SOURCE of the HUGE RIVER is the CLEAR SPRING.
The TEAM with the BEST TIMING LOOKS GOOD.
The TERM ENDED in LATE JUNE that YEAR.
The THREE STORY HOUSE was BUILT of STONE.
The TINY GIRL TOOK OFF her HAT.
The WALL PHONE RANG LOUD and OFTEN.
The WATER in this WELL is a SOURCE of GOOD HEALTH.
The WEIGHT of the PACKAGE was SEEN on the HIGH SCALE.
The WORK of the TAILOR is SEEN on EACH SIDE.
The YOUNG GIRL GAVE no CLEAR RESPONSE.
The YOUNG KID JUMPED the RUSTY GATE.
There are MANY WAYS to DO THESE THINGS.
THERE are MORE than TWO FACTORS HERE.
THERE is a LAG BETWEEN THOUGHT and ACT.
THERE the FLOOD MARK is TEN INCHES.
THERE was a SOUND of DRY LEAVES OUTSIDE.
THESE COINS will be NEEDED to PAY his DEBT.
These PILLS DO LESS GOOD than OTHERS.
They are PUSHED BACK EACH TIME they ATTACK.
They COULD LAUGH ALTHOUGH they WERE SAD.
They SANG the SAME TUNES at EACH PARTY.
THIEVES who ROB FRIENDS DESERVE JAIL.
Those LAST WORDS WERE a STRONG STATEMENT.
THOSE WORDS were the CUE for the ACTOR to LEAVE.
THROW out the USED PAPER CUP and PLATE.
TIME BRINGS US MANY CHANGES.
To REACH the END he NEEDS MUCH COURAGE.
To SEND it NOW in LARGE AMOUNTS is BAD.
TRY to HAVE the COURT DECIDE the CASE.
TWIST the VALVE and RELEASE HOT STEAM.
TWO PLUS SEVEN is LESS than TEN.
USE a PENCIL to WRITE the FIRST DRAFT.
WAKE and RISE and STEP into the GREEN OUTDOORS
WE are SURE that ONE WAR is ENOUGH.
We DON’T GET much MONEY but we HAVE FUN.
WE LIKE to SEE CLEAR WEATHER.
We TRIED to REPLACE the COIN BUT_failed.
WOMEN FORM LESS than HALF of the GROUP
WOOD is BEST for MAKING TOYS and BLOCKS.
WRITE at ONCE or you MAY FORGET IT.
You CANNOT BREW TEA in a COLD POT.

Sentences selected only for Experiment 1

A PINK SHELL was FOUND on the SANDY BEACH.
A THICK COAT of BLACK PAINT COVERED all.
BIRTH and DEATH MARK the LIMITS of LIFE.
DO THAT WITH a WOODEN STICK.
DRIVE the SCREW STRAIGHT INTO the WOOD.
EITHER MUD or DUST are FOUND at all TIMES.
GREEN MOSS GROWS on the NORTHERN SIDE.
GUESS the RESULTS FROM the FIRST SCORES.
He KNEW the SKILL of the GREAT YOUNG ACTRESS.
NO CEMENT will HOLD HARD WOOD.
SCHOOLS for LADIES TEACH CHARM and GRACE.
SET the PIECE HERE and SAY NOTHING.
The BOMBS LEFT MOST of the TOWN in RUINS.
The DOCTOR CURED HIM with THESE PILLS.
The **DUSTY BENCH** STOOD by the **STONE WALL**.
The **EARLY PHASE** of **LIFE MOVES FAST**.
The **FIRST PART** of the **PLAN NEEDS CHANGING**.
The **HOUSES** are **BUILT** of **RED CLAY BRICKS**.
The **INK STAIN** **DRIED** on the **FINISHED PAGE**.
The **KITE FLEW WILDLY** in the **HIGH WIND**.
The **MAN WENT** to the **WOODS** to **GATHER STICKS**.
The **PLEASANT HOURS** **FLY** by much **TOO SOON**.
The **PURPLE TIE** was **TEN YEARS OLD**.
The **STREETS** are **NARROW** and **FULL** of **SHARP TURNS**.
The **SUN CAME** up to **LIGHT** the **EASTERN SKY**.
The **WAY** to **SAVE MONEY** is **NOT** to **SPEND** much.
**WHEN** you **HEAR** the **BELL**, **COME QUICKLY**.
They **TOOK** the **AXE** and the **SAW** to the **FOREST**.
They **TOOK** their **KIDS** from the **PUBLIC SCHOOL**.

**Sentences selected only for Experiment 2**

A **GOLD RING** will **PLEASE** most **ANY GIRL**.
A **WHITE SILK JACKET** goes with **ANY SHOES**.
**ACID BURNS HOLES** in **WOOL CLOTH**.
**CODE** is **USED WHEN SECRETS** are **SENT**
**CUT** the **PIE INTO LARGE PARTS**.
**FILL** the **INK JAR** with **STICKY GLUE**.
**GO NOW** and **COME HERE LATER**.
He **BROKE** his **TIES** with **GROUPS** of **FORMER FRIENDS**.
He **RAN HALF WAY** to the **HARDWARE STORE**.
**HER PURSE** was **FULL** of **USELESS TRASH**.
It **TAKES** a **LOT** of **HELP** to **FINISH THESE**.
**NEXT TUESDAY WE MUST VOTE**.
**ONCE WE STOOD BESIDE** the **SHORE**.
**READ JUST WHAT** the **METER** **SAYS**.
SHE CALLED his NAME MANY TIMES.
SMALL CHILDREN CAME to SEE HIM.
The BEACH is DRY and SHALLOW at LOW TIDE.
The BLOOM of the ROSE LASTS a FEW DAYS.
The CEMENT had DRIED WHEN he MOVED IT.
The FACTS DON'T ALWAYS SHOW who is RIGHT.
The HORN of the CAR WOKE the SLEEPING COP.
The LAST SWITCH CANNOT be TURNED OFF.
The LAWYER TRIED to LOSE HIS CASE.
The LITTLE TALES THEY TELL are FALSE.
The MAP HAD an X that MEANT NOTHING.
The PIPE BEGAN to RUST WHILE NEW.
The QUICK FOX JUMPED on the SLEEPING CAT.
The TRAIN BROUGHT our HERO to the BIG TOWN.
The TWO MET WHILE PLAYING on the SAND.
There is a STRONG CHANCE it will HAPPEN ONCE MORE.
They are MEN WHO WALK the MIDDLE of the ROAD.
TRY to TRACE the FINE LINES of the PAINTING
WATCH the LOG FLOAT in the WIDE RIVER.
WE ADMIRE and LOVE a GOOD COOK.
We DRESS to SUIT the WEATHER of MOST DAYS.
WE FIND JOY in the SIMPLEST THINGS.
WIPE the GREASE OFF his DIRTY FACE.
Appendix B

Table of individual talkers’ quantitative measurements for speaking rate in Experiment 1

<table>
<thead>
<tr>
<th>Talker</th>
<th>Clear Speech</th>
<th>Conversational Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSL (s)</td>
<td>TPD (s)</td>
</tr>
<tr>
<td>T1</td>
<td>117.0</td>
<td>13.4</td>
</tr>
<tr>
<td>T2</td>
<td>97.3</td>
<td>5.2</td>
</tr>
<tr>
<td>T3</td>
<td>139.2</td>
<td>14.2</td>
</tr>
<tr>
<td>T4</td>
<td>117.1</td>
<td>8.4</td>
</tr>
<tr>
<td>Average</td>
<td>117.7</td>
<td>10.3</td>
</tr>
</tbody>
</table>

Note. TSL=Total Sentence Length, TPD=Total Pause Duration, ASL=Average Sentence Length, APD=Average Pause Duration, PF=Pause Frequency, SR=Speaking Rate
Appendix C

Table of individual talkers’ quantitative measurements for speaking rate in Experiment 2

<table>
<thead>
<tr>
<th>Talker</th>
<th>Clear Speech</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Conversational Speech</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Exaggerated-F0 Speech</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TSL (s)</td>
<td>TPD (s)</td>
<td>ASL (ms)</td>
<td>APD (ms)</td>
<td>PF</td>
<td>SR (rpm)</td>
<td>TSL (s)</td>
<td>TPD (s)</td>
<td>ASL (ms)</td>
<td>APD (ms)</td>
<td>PF</td>
<td>SR (rpm)</td>
<td>TSL (s)</td>
</tr>
<tr>
<td>T1</td>
<td>104.7</td>
<td>11.3</td>
<td>2617</td>
<td>70</td>
<td>143</td>
<td>180</td>
<td>88.2</td>
<td>6.1</td>
<td>2204</td>
<td>63</td>
<td>97</td>
<td>221</td>
<td>99.3</td>
</tr>
<tr>
<td>T2</td>
<td>87.5</td>
<td>5.7</td>
<td>2187</td>
<td>57</td>
<td>100</td>
<td>205</td>
<td>70.2</td>
<td>3.5</td>
<td>1754</td>
<td>42</td>
<td>82</td>
<td>226</td>
<td>92.8</td>
</tr>
<tr>
<td>T3</td>
<td>114.7</td>
<td>9.9</td>
<td>2888</td>
<td>88</td>
<td>113</td>
<td>120</td>
<td>84.4</td>
<td>4.0</td>
<td>2111</td>
<td>50</td>
<td>70</td>
<td>216</td>
<td>113.1</td>
</tr>
<tr>
<td>T4</td>
<td>115.6</td>
<td>8.2</td>
<td>2889</td>
<td>78</td>
<td>108</td>
<td>160</td>
<td>80.4</td>
<td>4.8</td>
<td>2234</td>
<td>55</td>
<td>87</td>
<td>215</td>
<td>115.2</td>
</tr>
<tr>
<td>T5</td>
<td>106.9</td>
<td>10.8</td>
<td>2522</td>
<td>87</td>
<td>124</td>
<td>181</td>
<td>72.3</td>
<td>5.5</td>
<td>1809</td>
<td>59</td>
<td>93</td>
<td>222</td>
<td>98.4</td>
</tr>
<tr>
<td>Average</td>
<td>104.7</td>
<td>9.2</td>
<td>2617</td>
<td>77</td>
<td>118</td>
<td>181</td>
<td>80.9</td>
<td>4.8</td>
<td>2022</td>
<td>54</td>
<td>88</td>
<td>233</td>
<td>103.8</td>
</tr>
</tbody>
</table>

Note: TSL=Total Sentence Length; TPD=Total Pause Duration; ASL=Average Sentence Length; APD=Average Pause Duration; PF=Pause Frequency; SR=Speaking Rate