



Second-Language Speakers and the Clear-Speech Benefit: Does Exaggerated Voice Pitch Increase Understanding?

Sarah L. Larson and Robert S. Schlauch

Department of Speech-Language-Hearing Sciences, University of Minnesota Twin Cities



I. INTRODUCTION

Background:

- Asking someone to speak clearly dramatically increases listeners' speech understanding in background noise (Miller et al., 2010; Shen & Souza, 2017a, 2017b)
- This 10-30% increase can be correlated with measured acoustic changes in the talker's speech. These include an emphasis in high-frequency energy (i.e., frequencies containing consonant sounds), more extreme formant frequencies in vowels, slowing down the speaking rate, and increasing variation in the voice pitch (Picheny et al., 1986; Uchanski et al., 1996; Bradlow et al., 2003; Hazan & Markham, 2004; Krause & Braida, 2004)
- The role that pitch plays has not been explored experimentally
- Second-language speakers are predominant workers in healthcare facilities, where aging populations who suffer from hearing loss are prevalent
- Koreans from Seoul believe that it is impolite to vary voice pitch, but a varied voice pitch may be the key to increasing understanding (Han 2019)

This experiment sought to assess whether there is a way for Native Korean talkers to modify their speech so that it can be more easily understood. Speakers were asked to vary their voice pitch more than it would be in conversational speech, and the number of correctly comprehended words heard by listeners was recorded

Research Questions:

- Is there a clear-speech benefit when second-language talkers produce exaggerated speech?
- Does an exaggerated voice pitch alone contribute significantly to the comprehension of speech?

Study Aim: To replicate the results of a previous study (Han 2019), only using one-third of the participants, and to see if an exaggerated pitch produced by second-language talkers produces an improvement in understanding.

II. METHODS

Participants (TALKERS):

- Five male native speakers from Seoul, Korea (T1-T5), with no speech, language, and hearing problems

Stimuli:

- Speech Materials: IEEE sentences
- Recording Equipment / Audio Quality: Marantz Professional PMD671 digital recorder with an AKG head-mounted condenser microphone / Sampling rate of 44,100 Hz with a 16-bit quantization
- Distance between microphone and participant's mouth (and a loudspeaker - BOSE Soundlink Color II Bluetooth Speaker - generating a 1-kHz reference tone): 10 cm

Recording Procedures:

- Individual talkers were seated in front of a computer monitor and wore a head-mounted condenser microphone.
- Individuals were assigned a different set of sentences with three practice sentences and instructed to produce the sentences under two different speaking styles: conversational and exaggerated pitches. For conversational speech, they were instructed to read aloud each sentence as they would in an everyday, normal conversation. For exaggerated speech, they were asked to read aloud each sentence by over-enunciating, as though they were speaking to an infant. Each sentence was repeated three times in each speaking style.
- Audio files with recording-mistakes were removed.
- A single session of the experiment lasted for approximately 1.5 hours including breaks.

Participants (LISTENERS):

- Five native speakers of American English, aged from 19 – 34, with normal hearing

Test Procedures:

- 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 speech spectrum noise were presented at 65 dB SPL binaurally through headphones (Sennheiser HD650) in a random order. Signal-to-noise ratios were fixed at -2.4, 0, -4, and -1.6 dB for T1, T2, T3, and T4, respectively.
- 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 first author, and the scores were verified by comparing the oral responses to the typed responses.
- The percentage of correct keywords was calculated for each talker for the exaggerated and conversational speech conditions.
- A single session of the experiment lasted for 1.5 hours, including breaks.

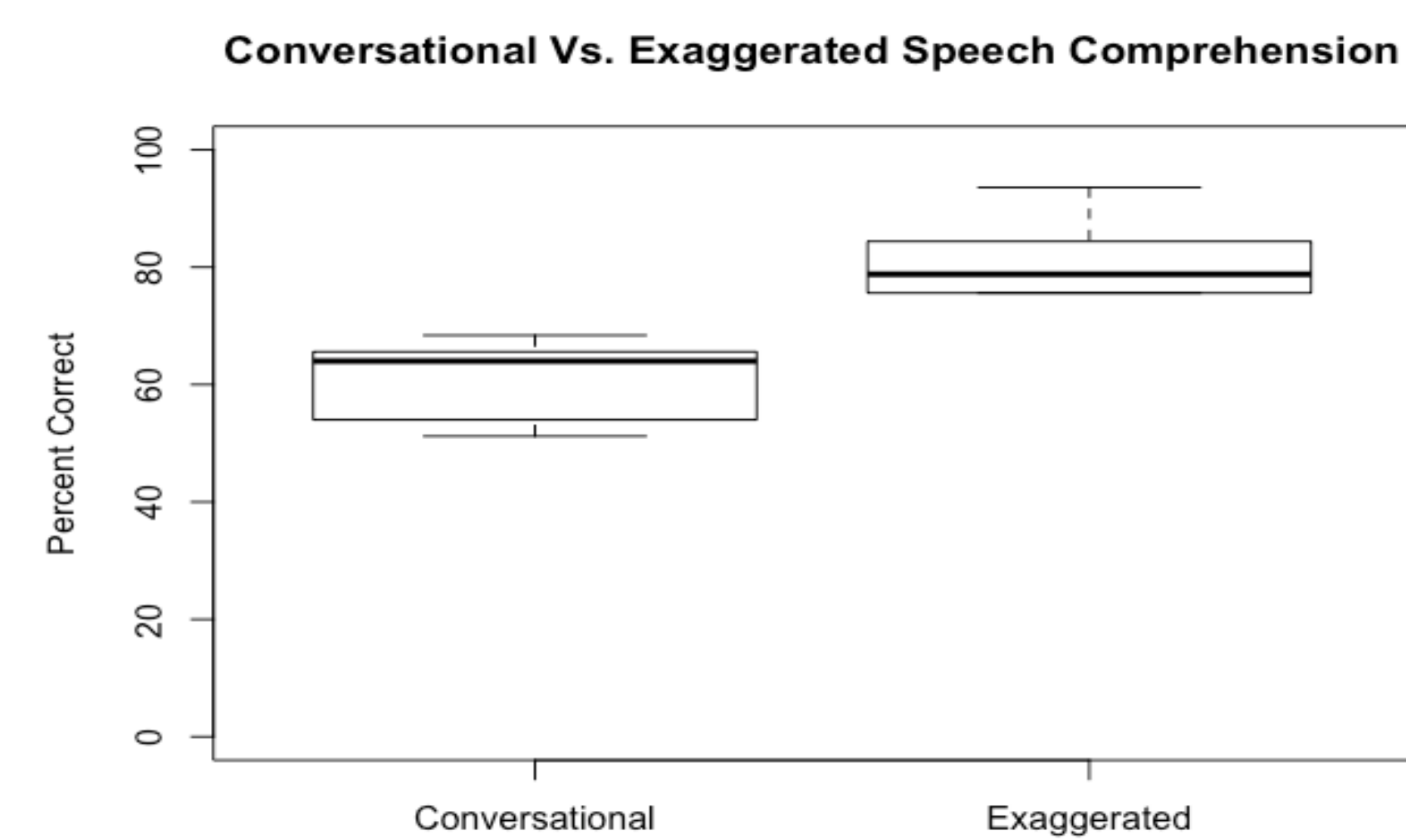
III. RESULTS

TABLE 1. Percentage of correctly comprehended words heard by 5 subjects (SXXX), spoken by 5 native Korean talkers (T1-T5), in a conversational pitch.

Talker	S004	S005	S009	S012	S014	Avg.	% Correct
T1	2.7	3	3.4	3.9	3	3.2	64
T2	3.7	3.3	3.5	3.6	3	3.42	68.4
T3	2.8	3.7	3.4	3.7	2.8	3.28	65.6
T4	3.4	3.6	2	2.8	1.7	2.7	54
T5	3.2	2.8	2.3	2.9	1.6	2.56	51.2
Average							60.64

TABLE 2. Percentage of correctly comprehended words heard by 5 subjects (SXXX), spoken by 5 native Korean talkers (T1-T5), in an exaggerated pitch.

Talker	S004	S005	S009	S012	S014	Avg.	% Correct
T1	4.1	3.5	4.1	4.1	3.9	3.94	78.8
T2	4.5	4.6	5	4.5	4.8	4.68	93.6
T3	3.2	3.4	4.1	4.5	3.7	3.78	75.6
T4	4.6	3.4	4.5	4.4	4.2	4.22	84.4
T5	4.5	3.4	3.2	4.1	3.7	3.78	75.6
Average							81.6



The *t*-value is 4.36576. The *p*-value is .002394

The result is significant at $p < .05$

Cohen's *d* = $(60.64 - 81.6) / 7.59105 = 2.761146$, revealing a large effect size

FIG. 1. Average percentage of words spoken by native Korean speakers in a conversational and exaggerated pitch correctly heard by 5 subjects.

IV. DISCUSSION

- The exaggerated speaking style resulted in a significant improvement in speech understanding over the conversational condition.** On average, listeners performed over 20% better in correctly identifying the sentences spoken by Koreans in an exaggerated pitch than they did for the conversational speakers. The result was robust and statistically significant even when the data from only five listeners was evaluated.
- Flattening the pitch would eliminate other variables, allowing for isolation of the effect of pitch on comprehension.** Although the percentage of words correctly heard by subjects increased when Korean talkers spoke with an exaggerated pitch, other elements may have generated this effect. For instance, the sentences with the exaggerated pitch were produced at a slower rate (Picheny et al., 1986; Bradlow et al., 2003).
- It is possible that the monotone results may have been better comprehended, which would reveal the adverse effects of an exaggerated pitch on speech comprehension.** If that were true, the improvement in the speech understanding of the exaggerated condition would be attributed to other acoustic changes.
- It is important to further expand upon this experiment to discover how speaking with an exaggerated pitch alone impacts speech comprehension.** This experiment lacks data to describe the role that the pitch itself had on comprehension, since it was unable to control for other factors.
- The original goal of the Dean's First Year Research and Creative Scholars Program, prior to the suspension of in-person instruction and the closing of research laboratories, was to examine the role of voice pitch in the improvement of speech understanding in the exaggerated condition.** In preparation for a follow-up study to (Han 2019), the pitch in the exaggerated sentence was flattened (i.e., made monotone) using an acoustic editing computer software program (PRAAT), and each sentence was analyzed individually.

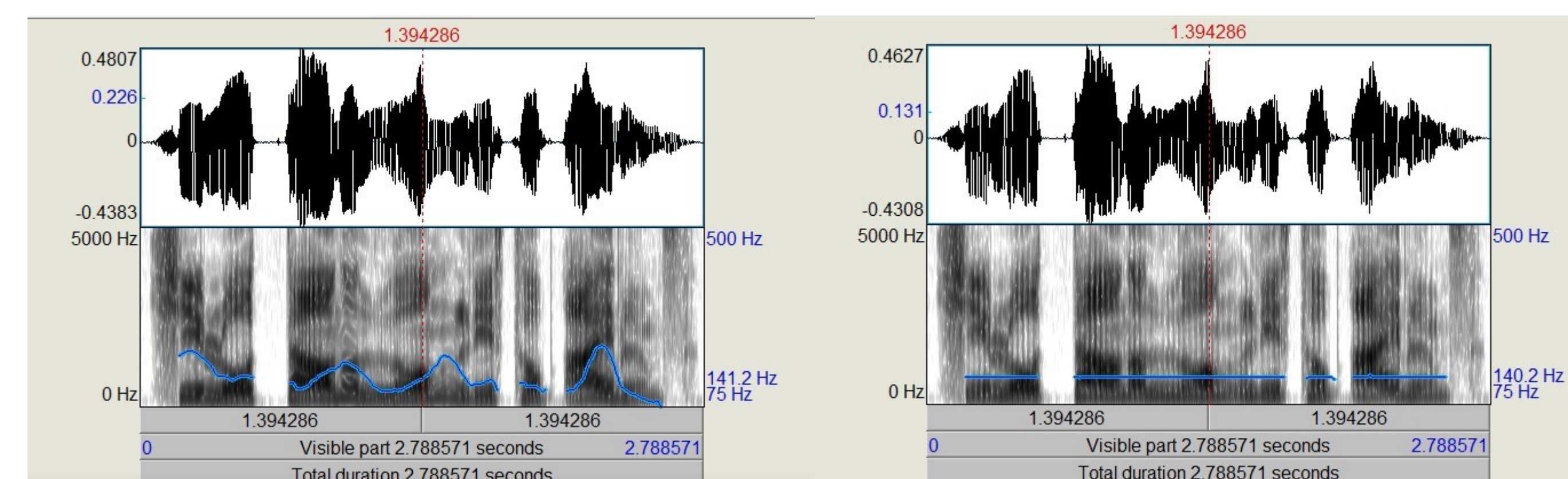


FIG. 2. Analysis of the sentence, "He wrote down a long list of items." Spoken in an exaggerated voice pitch (left) and flattened/made monotone (right) using the acoustic editing program PRAAT. [To hear the original exaggerated and flattened versions of this sentence, highlight and right click on the links below, then open the hyperlink]

[Exaggerated Sentence Example](#)

[Flattened Sentence Example](#)

- Aside from developing strong data analysis skills, I also learned how to set up and run the laboratory equipment to collect human data. The study is now properly prepared to proceed once the pandemic ends, when it is safe to return to campus.

V. CONCLUDING REMARKS

- Five subjects were enough to reveal that an exaggerated pitch yields better comprehension than a conversational pitch.** The results produced from this study, using only 5 participants, were nearly the same as those conducted in (Han 2019), where 15 subjects were used.
- There are uncertainties as to what role other acoustic factors may have played relative to the role of pitch.** A flattened (monotone) speech condition would be an important experiment to do to examine the role that exaggerated pitch plays in speech understanding.

ADDITIONAL DATA ANALYSES

The results from the analysis of 5 listeners generated nearly the same average values as the data from 15 listeners in (Han 2019). This can be seen by comparing the data in the far right of the figure below for the Exaggerated and Conversational conditions to that of Figure 1.

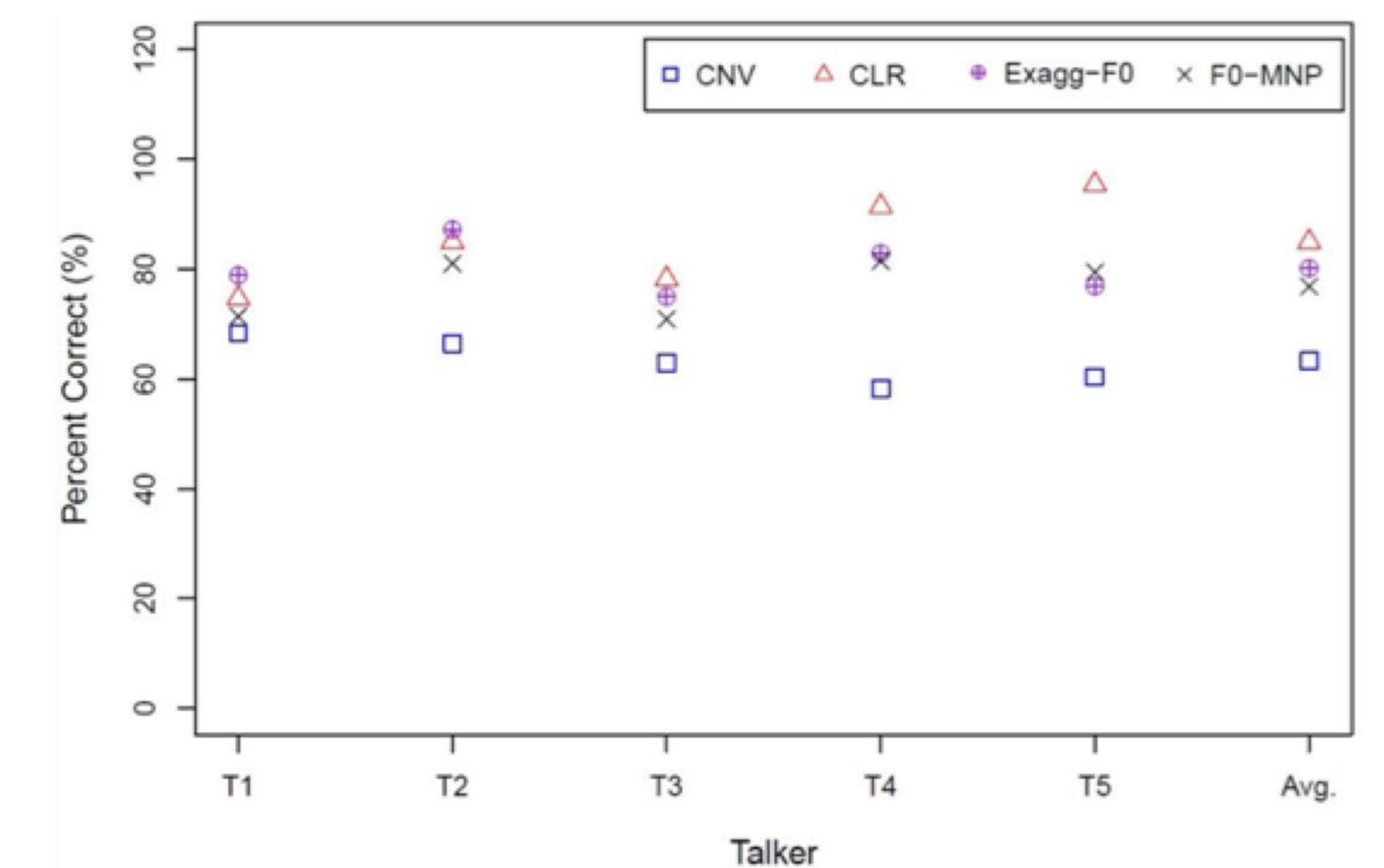


FIG. 3. Average intelligibility scores for 5 native Korean talkers (T1-T5) and grand mean scores in conversational (CNV), exaggerated-F0 (Exagg-F0), F0-manipulated (F0-MNP), and clear (CLR) speech conditions

ACKNOWLEDGEMENT

This study was supported by the Dean's First-Year Research and Creative Scholars program within the College of Liberal Arts at the University of Minnesota Twin Cities.

REFERENCES

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking Clearly for Children With Learning Disabilities. *Journal of Speech, Language, and Hearing Research*.

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America*, 116(5), 3108-3118.

Han, H. (2019). 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.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1), 362-378.

Miller, S. E., Schlauch, R. S., & Watson, P. J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *The Journal of the Acoustical Society of America*, 128(1), 435-443.

Picheny, M. A., Durlach, N. I., & BRAIDA, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 29(4), 434-446.

Shen, J., & Souza, P. E. (2017a). Do older listeners with hearing loss benefit from dynamic pitch for speech recognition in noise? *American Journal of Audiology*, 26(3S), 462-466.

Shen, J., & Souza, P. E. (2017b). The Effect of Dynamic Pitch on Speech Recognition in Temporally Modulated Noise. *Journal of Speech, Language, and Hearing Research*, 60(9), 2725-2739.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research*, 39(3), 494-509.