

3pSC10: Gender and Rate Effects on Speech Intelligibility

Eric M. Johnson, *Department of Communication Sciences and Disorders, University of Utah, Salt Lake City, Utah*

Sarah Hargus Ferguson, *Department of Communication Sciences and Disorders, University of Utah, Salt Lake City, Utah*



Background

Older adults seeking hearing help often complain of particular difficulty understanding female voices. This contrasts with studies using young listeners with normal hearing in which female talkers have been found to be generally more intelligible than male talkers for sentences presented in quiet (Bradlow et al., 1996) and for vowels presented in noise (Ferguson, 2004). Only two studies have examined gender effects on speech intelligibility for listeners with hearing loss. For vowel intelligibility in noise, Ferguson (2012) found that male and female talkers were equally intelligible for older adults with hearing loss (although the clear speech effect was larger for female talkers). In contrast, Larsby et al. (2015) found that groups of listeners with high-frequency hearing loss, low-frequency hearing loss, or normal hearing all required a more favorable signal-to-noise ratio (SNR) to recognize sentences produced by a female talker than sentences produced by a male talker.

These mixed results may well reflect talker or intelligibility task differences. The Larsby et al. data suggest that female speech might be more affected by distortion than male speech. Speech that has been time-compressed has been shown to be less intelligible than unprocessed speech (e.g. Gordon-Salant & Friedman, 2011), but no studies have explored whether an time compression causes an equal loss of intelligibility for male and female talkers.

The present study tested the following hypothesis:

- Compared to the intelligibility of unprocessed stimuli, the intelligibility of time-compressed stimuli will decrease for female talkers more than it will for male talkers.

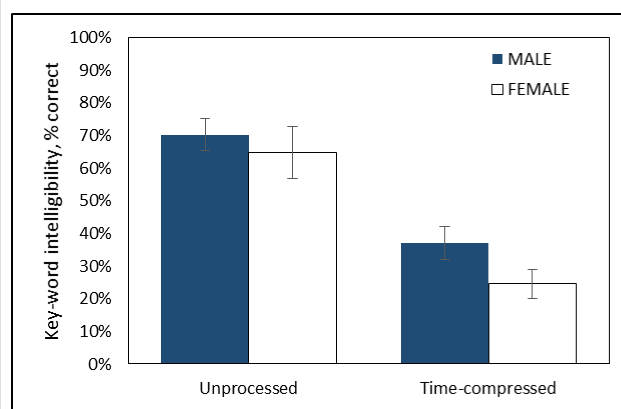
Materials

- **4 talkers** (2 males and 2 females) were selected from the Utah Speaking Style Corpus based on their approximately equivalent baseline speaking rates in conversational speech.
- For each talker, **6 sentence lists** (10 sentences per list) from the Hearing in Noise Test (HINT) were used, all of which were spoken in a conversational style and scaled to the same RMS amplitude.
- The speech rate of all audio files was digitally manipulated via the "time stretch/preserve pitch" function in Adobe Audition CC (2015). The audio files were time-compressed to 66.7% of their original durations so their playback speed would be 1.5 times as fast.
- There were 2 rate conditions for each sentence list: unprocessed and time-compressed. Thus, there were 8 blocks of 60 sentences (4 talkers x 2 rate conditions).
- Test stimuli were presented in a background of 12-talker babble via Tucker-Davis Technologies (TDT) System III audio hardware.

Procedures

Listeners were tested using **insert earphones** and seated in a quiet room in front of a computer monitor, keyboard, and mouse. **Tucker-Davis System 3** hardware and a custom **MATLAB** script were used to present sentences monaurally at **70 dB SPL** in the presence of **12-talker babble** at **0 dB SNR**. After hearing each sentence, listeners typed out what they heard using the computer keyboard. Each of the 8 blocks of 60 sentences was identified by **4 listeners**.

Lists of 10 sentences were **keyword** scored; correct morphological affixes were required for a word to be scored as correct. Homophones and responses containing minor spelling errors were scored as correct.



Linear mixed-effects model (LMM) analyses were used to analyze the data. LMMs account for random variance introduced by using stimuli from multiple talkers (Quené & van den Bergh, 2004). For the purposes of analysis, percent correct key-word identification scores for each list were converted to rationalized arcsine units (**RAU**; Studebaker, 1985).

Main effects of **processing condition** and **talker gender** as well as the interaction between them were analyzed. **Talker** was included in each model as a random factor.

Figure 1: Overall percent correct key-word identification performance for male and female talkers in unprocessed and time-compressed conditions. Error bars indicate 95% confidence intervals.

Significant effects for processing condition were observed for both male and female talkers, but there was a significant interaction between talker gender and processing condition. The effect of processing condition was much greater for the female talkers than for the male talkers.

The effect of talker gender was not significant for either the unprocessed or time-compressed speech

Analyses

Results

- 1) The interaction between talker gender and processing condition was significant ($z = -2.4, p < .02$).
- 2) The effect of rate condition was significant for both male ($z = -9.24, p < .001$) and female ($z = -12.68, p < .001$) talkers, but the effect of rate was much greater for the female talkers than for the male talkers.
- 3) The effect of talker gender was not significant for either the normal or the fast speech ($|z| < 2.0, p > .1$).

Discussion and Conclusion

The analyses support prior findings showing that time-compressed speech is less intelligible than unprocessed speech. Furthermore, in this study, time-compression had a greater effect on the intelligibility of female speech than it did male speech, with a significant interaction effect between talker gender and processing condition. This finding agrees with anecdotal reports in which people with hearing loss complain of a particular difficulty understanding women who speak quickly. It should be noted that all participants in this study were young, normal-hearing listeners. Nevertheless, if the intelligibility of female talkers is more subject to the effects of signal degradation than male talkers, it may also be more subject to the effects of hearing loss.

It should also be noted that making gender comparisons is difficult because the variability among talkers is so large; all talkers, male and female, vary tremendously in their speaking clarity, which makes it difficult to pinpoint whether any differences in intelligibility among talkers are actually caused by gender differences.

These results imply that audiologists should counsel communication partners of patients to avoid speaking excessively fast, especially in noise and especially if the patient complains of difficulty understanding women's voices. This may especially important for patients who complain of particular difficulty understanding speech in noise.

Development of Utah Speaking Style Corpus supported by NIH grant R01DC012315 to Eric Hunter. Experiment software originally developed at Indiana University by Bill Mills; the most recent updates were made by Shae Morgan and Skyler Jennings.

Listeners

- 32 listeners
- 31 females and 1 male
- 18-34 years old (mean: 22.3; SD: 4.3)
- No history of speech, language, or hearing disorders by self-report
- Native speakers of American English
- Each identified just 1 block of 60 sentences (1 talker and 1 rate condition)
- Received either research participation credit in the Department of Psychology or extra credit in Introduction to Audiology

References

1. Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Commun.* 20, 255-272.
2. Ferguson, S. H. (2004). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," *J. Acoust. Soc. Am.* 116, 2365-2373.
3. Ferguson, S. H. (2012). "Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss," *J. Speech Lang. Hear. Res.* 55, 779-790.
4. Gordon-Salant, S., and Friedman, S. A. (2011). "Recognition of rapid speech by blind and sighted older adults," *J. Speech Lang. Hear. Res.* 54, 622-631.
5. Larsby, B., Hallgren, M., Nilsson, L., and McAllister, A. (2015). "The influence of female versus male speakers' voice on speech recognition thresholds in noise: Effects of low- and high-frequency hearing impairment," *Speech Lang. Hear.* 18, 83-90.
6. Quené, H., and van den Bergh, H. (2004). "On multi-level modeling of data from repeated measures designs: A tutorial," *Speech Commun.* 43, 103-121.
7. Studebaker, G. A. (1985). "A rationalized arcsine transform," *J. Speech Lang. Hear. Res.* 28, 455-462.