JSLHR

Research Note

Does Time Compression Decrease Intelligibility for Female Talkers More Than for Male Talkers?

Eric M. Johnson,[a,b] Shae D. Morgan,[a,c] and Sarah Hargus Ferguson[a]

Purpose: This preliminary investigation compared effects of time compression on intelligibility for male versus female talkers. We hypothesized that time compression would have a greater effect for female talkers.
Method: Sentence materials from four talkers (two males) were time compressed, and original-speed and time-compressed speech materials were presented in a background of 12-talker babble to young adult listeners with normal hearing. Each talker/processing condition was heard by eight listeners (total $N = 64$). Generalized linear mixed-effects models were used to determine the effects of and interaction between processing condition and talker sex on keyword intelligibility. Additional post hoc analyses examined whether processing condition effects were related to talker vowel space and word frequency.
Results: For original-speed sentences, female and male talkers were essentially equally intelligible. Time compression reduced intelligibility for all talkers, but the effect was significantly greater for the female talkers. Supplementary analyses revealed that the effect of time compression depended on both talker vowel space and word frequency: The detrimental effect decreased significantly as word frequency and vowel space increased. Word frequency effects were also greater overall for talkers with larger vowel spaces.
Conclusions: While the small talker sample limits conclusions about the effects of talker sex, the secondary analyses suggest that intelligibility of talkers with larger vowel spaces is less susceptible to the negative effect of time compression, especially for high-frequency words.

Presenting speech at an accelerated rate by compressing it in time can have several useful scientific applications. Time-compressed (TC) speech has been used as a research tool for studying auditory processing in a wide variety of listeners, including older and younger adults with and without hearing loss (e.g., Gordon-Salant & Fitzgibbons, 1993; Sticht & Gray, 1969), patients with lesions of the auditory cortex (Kurdziel et al., 1976), children with auditory processing difficulties (Manning et al., 1977), listeners with cochlear implants (Fu et al., 2001), and children with specific language impairment (Guiraud et al., 2018). TC speech has also been used to study reading readiness (Riensche et al., 1986) and reading impairment (Abrams et al., 2009) in children. Other scientific uses for TC speech include increasing the difficulty of speech recognition tasks at positive signal-to-noise ratios (SNRs; Schlueter et al., 2015) and simulating the effects of natural, fast speech (Adams et al., 2012). Clinically, various TC speech tests have served to assess degraded speech understanding and aid in the diagnosis of auditory processing disorder (Emanuel et al., 2011). TC speech also has practical applications in education (Barabasz, 1968), radio advertising (LaBarbera & MacLachlan, 1979), and digital multimedia (Furini, 2008). Many individuals, including those who are blind, use TC speech when they listen to audiobooks, recorded lectures, and voicemail messages in order to increase processing efficiency (Gordon-Salant & Friedman, 2011).

Since almost all studies involving TC speech have focused on listener characteristics rather than talker characteristics, one factor that has received little attention has been the sex of the talker. This could be significant, considering the substantial acoustical differences between male and female speech, such as fundamental frequency, harmonic spacing, voice quality, vowel formant frequencies, vowel

[a]Department of Communication Sciences and Disorders, The University of Utah, Salt Lake City
[b]Department of Speech and Hearing Science, The Ohio State University, Columbus
[c]Department of Otolaryngology—Head and Neck Surgery and Communicative Disorders, University of Louisville, KY
Correspondence to Eric M. Johnson: johnson.7289@osu.edu

Editor-in-Chief: Bharath Chandrasekaran
Editor: Kate Bunton
Received March 25, 2019
Revision received October 23, 2019
Accepted January 2, 2020
https://doi.org/10.1044/2020_JSLHR-19-00301

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

space, fricative spectral peak location, voice onset time, and other acoustic characteristics (Clopper & Smiljanic, 2011; Jongman et al., 2000; Peterson & Barney, 1952; Simpson, 2009). According to several studies using young adult listeners with normal hearing, male and female talkers also differ in terms of their intelligibility, with females being more intelligible, on average, than males (Bradlow et al., 1996; Ferguson, 2004; Markham & Hazan, 2004; Wang & Humes, 2010; Yoho et al., 2019). Ferguson (2004) also found that the effect of speaking clearly on vowel intelligibility was greater for female talkers than male talkers. Given the considerable differences between male and female speech, it is reasonable to suppose that time compression could have a differential effect on intelligibility based on talker sex, which could have significant implications for the wide-ranging uses of TC speech listed above. However, virtually every study of TC speech and intelligibility has used materials from a single talker, usually a male (e.g., Adams et al., 2012; Gordon-Salant & Fitzgibbons, 1993; Sticht & Gray, 1969) and only occasionally a female (e.g., Stollman & Kapteyn, 1994). This lack of talker diversity in TC speech studies speaks to a more general problem in speech perception research (Yoho et al., 2019).

To our knowledge, only one study of TC speech (Versfeld & Dreschler, 2002) has used materials produced by more than a single talker. They found that listeners required a slower rate of speech to achieve 50% sentence intelligibility for a female versus a male talker, indicating that time compression may have a greater adverse effect on intelligibility for women's speech than for men's. We have heard anecdotal accounts from audiology patients who report particular difficulty understanding women, and the results of Versfeld and Dreschler (2002) suggest a possible source of decreased intelligibility for female talkers who use a naturally fast rate of speech, since listening to natural fast speech is even more difficult than listening to artificially TC speech (Gordon-Salant et al., 2014). Many studies using TC speech from a single talker have made inferences about listeners with diverse backgrounds and disorders, but the findings of Versfeld and Dreschler suggest that the sex of the talker matters. If women's speech is more susceptible to the negative effects of time compression than men's, this could complicate comparisons between TC speech studies and clinical tests that use different talkers. Talker sex could also have implications for other uses of TC speech, such as listening to recorded materials at accelerated speeds to reduce processing time.

In the present, preliminary study, we explored whether the deleterious effects of time compression on intelligibility differed between male and female talkers. In light of the acoustic differences observed between male and female speech, the findings of Versfeld and Dreschler (2002), and anecdotal evidence from audiology patients, we hypothesized that time compression would have a greater negative effect on the intelligibility of sentences produced by female talkers than it would on the intelligibility of sentences produced by male talkers.

Method and Initial Analysis

Materials

Sentence materials from two female talkers and two male talkers were selected from the Utah Speaking Style Corpus (USSC) based on their approximately equivalent speaking rates (see Table 1). The USSC contains recordings of three speech tasks (a read passage, a read list of 110 sentences, and a picture description task) produced by several speakers of General American English from the state of Utah. The set of talkers from which we chose the four used in this study repeated the complete task set 4 times in two speaking styles: conversational and clear. The list of 110 sentences included six lists (Lists 3, 5, 11, 19, 22, and 23) from the Hearing in Noise Test (HINT; Nilsson et al., 1994) intermixed with neutral sentences that had keywords targeting vowels in a fixed phonetic context; only the six HINT lists were used as stimuli in this study. Whereas each talker produced the sentences 8 times (2 speaking styles × 4 repetitions), we chose to use speech produced under instructions to be conversational as possible to provide greater ecological validity. In addition, we selected sentences from the third repetition of the task set to avoid learning effects in the first and second sets or fatigue effects in the fourth set.

After the original sentences were excised from the recordings of the full task set, a second TC version of each HINT sentence file was created in Adobe Audition Creative Cloud using the "time stretch/preserve pitch" function, which is based on the pitch-synchronous overlap-add method (Charpentier & Stella, 1986). Pitch-synchronous overlap-add techniques, which compress the speech signal in time while preserving pitch by deleting periods of voiced speech, are commonly used in TC speech research because they yield a high signal quality with relatively few artifacts that would unduly deteriorate intelligibility (e.g., Adank & Janse, 2009; Janse, 2004; Shibuya et al., 2012). The files were compressed to 66.7% of their original durations using a splicing frequency of 46 Hz and 27.1% overlap, yielding a playback speed 1.5 times faster than the unprocessed files.[1] Pilot listening indicated the TC sentences sounded like fast speech with the same pitch and timbre as the unprocessed sentences, with no other obvious distortions. MATLAB scripts were then used to add 50 ms of silence at the beginning and end of each file and to scale all files to the same average root-mean-square amplitude.

Speech rates for the four talkers' unprocessed and TC sentences are shown in Table 1. The speech rate for each sentence token was calculated by dividing its number of syllables by its duration in seconds. Duration was

[1]Many audio playback programs and virtually all audiobook applications (such as Audible, Google Play Books, Apple Books) offer a wide range of playback speeds, usually from "0.5 ×" (i.e., half as fast as the original recording) up to "3.0 ×" (i.e., 3 times as fast as the original recording). The time compression applied in this study is equivalent to a "narration speed" of "1.5 ×," a common playback speed for many audiobook listeners (Hu, 2017).

Table 1. Mean speech rate and standard deviation in syllables per second for the four talkers in both conditions.

Condition	Talker speech rate			
	Female 1	Female 2	Male 1	Male 2
Unprocessed sentences				
M	4.97	4.93	4.83	4.87
SD	0.67	0.68	0.78	0.61
Time-compressed sentences				
M	7.42	7.36	7.21	7.27
SD	1.00	1.02	1.16	0.91

measured manually in Cool Edit 2000 as the time between the onset of speech energy and the moment where speech energy was no longer detectable in the waveform. Mean speaking rate for each talker was then calculated by averaging across the 60 HINT sentences. The unprocessed and TC speech rates of the four talkers are consistent with "normal" and "fast" speaking rates, respectively, as measured in previous studies (e.g., Hargrave et al., 1994).

The 480 test stimuli (4 talkers × 60 sentences × 2 processing conditions) were arranged into eight test blocks. Each test block contained all 60 sentences from one talker in one processing condition. That is, there were two blocks of unprocessed sentences from female talkers, two blocks of TC female sentences, two blocks of unprocessed male sentences, and two blocks of TC male sentences. All sentences were presented at 0 dB SNR (see Procedure section for details). The masking noise was a 30-s sample of 12-talker babble low-pass filtered at 8500 Hz and digitized from the noise channel of a recording of the Speech Perception in Noise Test (Kalikow et al., 1977). Although the babble noise includes both male and female talkers, Kalikow et al. (1977) did not state how many talkers of each sex are present in the mixture. All materials were resampled to 24414 Hz prior to the perceptual experiment for presentation via Tucker-Davis Technologies (TDT) System III audio hardware.

Listeners

Sixty-four young adults (63 women, one man; ages 18–34 years; $\bar{x}$ = 22.3 years)[2] recruited from The University of Utah Department of Psychology Participant Pool and from undergraduate courses in the Department of Communication Sciences and Disorders (CSD) participated as listeners in this study. To participate, listeners from the Psychology Pool were required to be native speakers of American English aged 18–35 years with no history of speech, language, or hearing disorders, by self-report. All interested participants from CSD undergraduate courses were tested, but data from listeners not meeting these criteria were excluded from statistical analyses. In total, 75 participants were tested, but 11 participants' data were excluded for this reason. Listeners received research participation credit in Psychology or extra credit in their CSD undergraduate course.

Procedure

All test procedures were approved by The University of Utah Institutional Review Board. Before each listener was tested, their right ear was examined otoscopically. Individual listeners were tested in a quiet room, seated in front of a computer monitor, keyboard, and mouse. Target stimuli were presented at 70 dB SPL, a typical speech level in noisy environments (Pearsons et al., 1977), in a background of 12-talker babble. The babble was also presented at 70 dB SPL, resulting in an SNR of 0 dB. This SNR was identified via pilot testing as challenging enough to prevent ceiling effects. On each trial, a test sentence and a segment of the 12-talker babble were played from separate channels of a TDT RP2.1 real-time processor. The babble segment, which was 1 s longer than the sentence, was selected on each trial from a random location within the stored 30-s babble sample. The sentence and babble segment were centered temporally, with the babble starting 500 ms prior to the target sentence on each trial.

The sentence and babble segment were attenuated by separate TDT programmable attenuators (PA5) to achieve the desired presentation level and SNR. The speech and babble were then mixed (TDT SM5) and routed via a headphone buffer (TDT HB7) to an insert earphone (E-A-RTONE 3A) for monaural presentation. To identify each sentence, participants typed what they understood into a dialog box displayed on the monitor.[3] Partial answers and guessing were encouraged if listeners were unsure. After typing their response, listeners pressed "Enter" to advance to the next trial, giving them as much time as needed to respond to each sentence. Listeners were not permitted to replay the stimulus.

Each listener was tested in single session consisting of one test block (described above). Eight listeners heard each test block; the stimulus order was randomized for each listener. Due to the listeners' good hearing and comprehension, there was no familiarization task. The duration of the experiment was approximately 30 min for each participant.

Data Analysis

Prior to scoring, we identified 161 keywords (e.g., nouns, verbs, adjectives, and adverbs) in the set of 60 target sentences.[4] Each keyword was scored in a binary fashion as either correct (1) or incorrect (0). Correct morphological endings were required for a keyword to be scored as correct; homophones and responses containing spelling or

[2]Although the distribution of males and females is unbalanced here, several studies have found no effect of listener sex on speech intelligibility (e.g., Markham & Hazan, 2004; Yoho et al., 2019).

[3]Listener transcription is a commonly used methodology in speech perception studies, especially when subjects have normal hearing (Borrie et al., 2019). Compared to verbally repeating responses, typing also has the advantage of eliminating experimenter listening errors.

[4]See Supplemental Material S1 for the complete set of sentence materials and keywords used in this study.

typographical errors within one keystroke of the correct letter were scored as correct. For each listener, 161 keyword scores were collected. Therefore, a total of 10,304 binomial intelligibility scores (8 test conditions × 8 listeners per condition × 161 scores per listener) were collected and used as the outcome variable in the analysis.

To represent the categorical variables of processing condition and talker sex, deviation coding was used. Deviation coding facilitates model interpretation by specifying contrasts that are analogous to factors in traditional analyses of variance. The original-speed (unprocessed) condition was coded as −0.5 (baseline), and the TC condition was coded as 0.5 (comparison). The baseline and comparison levels for talker sex were arbitrarily assigned, with females coded as −0.5 and males coded as 0.5.

The analysis was performed using a generalized linear mixed-effects model (GLMM) for binomial data. The fixed effects for this model were processing condition, talker sex, and the interaction between them. The model also included random intercepts for talker[5] and keyword, as well as random slopes for processing condition by talker.[6] Listener was not included as a random factor in the model because listener identity was confounded with processing condition and talker in the between-subjects design. Visual inspection of residual plots did not reveal any clear violations of homoscedasticity or normality. All analyses were performed using R 3.5.1 (R Core Team, 2019) and the *lme4* package (Bates et al., 2015).

Results

Percent correct keyword intelligibility scores for the four talkers in the two processing conditions are shown in Figure 1. Apparent is the intelligibility difference between the two female talkers. In the unprocessed condition, Female 1 was the least intelligible of the four talkers (57.8% correct), while Female 2 was the most intelligible (93.6% correct). In contrast to the two female talkers, Males 1 and 2 were much more similar in their intelligibility in both the unprocessed (83.5% and 78.3% correct, respectively) and TC (52.9% and 52.4% correct, respectively) conditions. Also evident in Figure 1 is the reduced intelligibility in the TC processing condition. All talkers were less intelligible in the TC condition than they were in the unprocessed condition ($\bar{x}$ = 45.1% and 78.3% correct, respectively). GLMM analysis confirmed that the fixed effect of time compression was large and significant ($\beta = -1.92$, $z = -12.8$, $p < .001$). While the average scores for the female talkers are lower than those for the male talkers ($\bar{x}$ = 56.6% vs. 66.8% correct), the fixed effect of talker sex was not significant in the GLMM analysis ($\beta = 0.368$, $z = 0.489$, $p > .5$).

[5]Although there were only four talkers in this study, we chose to model this with a random effect, since they represent a pseudorandom sample from the population.

[6]The model in lme4 style is: SCORE ~ PROCESSING_CONDITION × TALKER_SEX + (1 + PROCESSING_CONDITION | TALKER) + (1 | KEYWORD)

Figure 1. Average percent correct sentence keyword identification scores for individual talkers in two conditions. The talkers are arranged from left to right in order of increasing vowel space perimeter. Error bars indicate 95% confidence intervals.

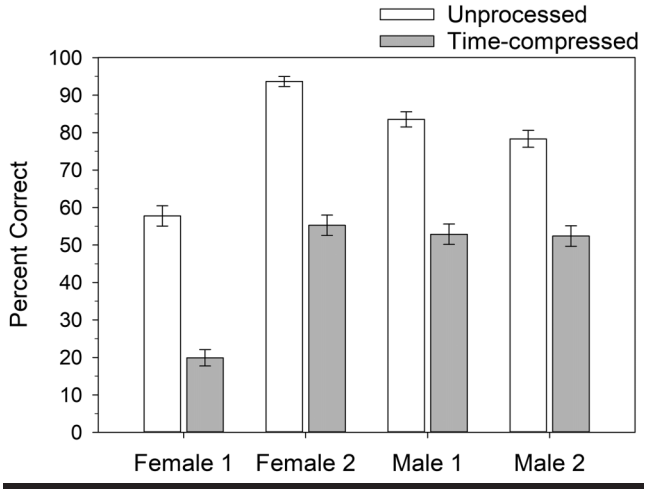


Figure 1 also seems to show a greater effect of time compression for the female talkers. Females 1 and 2 experienced greater intelligibility decrements due to time compression (−37.9 and −38.4 percentage points, respectively), compared to Males 1 and 2 (whose decrements equaled −30.7 and −25.9 percentage points, respectively). The GLMM analysis confirmed that the interaction between talker sex and processing condition was significant ($\beta = 0.822$, $z = 2.75$, $p < .01$).

Secondary Analysis

The significant interaction between talker sex and processing condition prompted a post hoc investigation seeking a mechanistic explanation for the differences in the effect of time compression based on talker sex. Two additional variables known to affect speech intelligibility were examined: vowel space and word frequency.

The study of vowel space (i.e., the two-dimensional area bounded by lines connecting the coordinates of first and second formant frequencies [F1 and F2, respectively] for vowels) has a long history in speech science (see Sandoval et al., 2013, for a review). Several studies have found larger vowel spaces to be correlated with increased speech intelligibility (e.g., Bradlow et al., 1996). Since time compression increases demands on the auditory processing of the listener, it could also change the effect of talker vowel space on intelligibility. Time compression largely maintains the spectral properties of the signal and thus could enhance the vowel space effect by making listeners rely more on steady-state vowel cues in the face of reduced temporal cues. Inversely, by giving listeners less time to process the spectral properties of the speech, time compression might eliminate the benefit of more distinct acoustic–articulatory vowel targets, thus reducing the effect of

vowel space. Differences in vowel space among talkers may therefore help account for differential intelligibility decrements for TC speech.

Another factor that influences speech intelligibility, especially under adverse listening conditions, is word frequency (e.g., how often a word occurs in daily speech). In a phenomenon known as the word frequency effect, frequently occurring words are more readily perceived than low-frequency words (e.g., Savin, 1963). This effect has been observed in a variety of listening conditions, including TC speech (e.g., Dupoux & Mehler, 1990). Like the vowel space effect, the word frequency effect could either be preserved, enhanced, or reduced in TC speech. Furthermore, if time compression does change the magnitude of the word frequency effect, this change may depend on talker characteristics, such as vowel space. Such findings could not only help account for the initial results of this study but also substantially impact our understanding of TC speech perception.

Data Analysis

The secondary analysis used the same 161 keyword scores as defined for the initial analysis. Each keyword was coded for word frequency, as recorded in the SUBTLEX-US corpus (Brysbaert & New, 2009) and expressed on the Zipf scale (van Heuven et al., 2014). Zipf values range from approximately 1 to 7 on a continuous logarithmic scale, with lower values indicating lower frequency words and higher values indicating higher frequency words. In this study, the lowest frequency keyword is "handstand," with a frequency of 2.19 Zipf (or 0.14 occurrences per million words), and the highest frequency keyword is "well," with a frequency of 6.48 Zipf (or 2,990 occurrences per million words).

Vowel space perimeter was also measured for these four talkers and is reported in Table 2. In a separate experiment, steady-state F1 and F2 frequencies were extracted using Praat from the vowels /i/, /æ/, /ɑ/, and /u/ spoken in a fixed phonetic context in the neutral sentences that were intermixed with the HINT sentences during recording of the USSC (see Materials section). F1 and F2 were converted from Hertz to Barks (Traunmüller, 1990), and perimeter was computed for each talker by summing the four Euclidean distances between the four values. For the present secondary analysis, each keyword token was coded for the vowel space perimeter, in Barks, of the talker that produced it.

Statistical analysis was performed using a GLMM for binomial data. To facilitate model interpretation, the variables of vowel space perimeter and word frequency were mean centered. The categorical variable of processing condition was deviation coded, as described in the initial analysis. The fixed effects for this model were processing condition, vowel space perimeter, and word frequency, as well as the two- and three-way interactions between these effects. The model also included random intercepts for talker and keyword, as well as random slopes for processing condition and word frequency by talker.[7] Again, listener was not included as a random factor because it is confounded with processing condition and talker.

Table 2. Vowel space perimeter in Barks for the four talkers.

Talker vowel space perimeter			
Female 1	**Female 2**	**Male 1**	**Male 2**
13.884	14.234	15.306	16.025

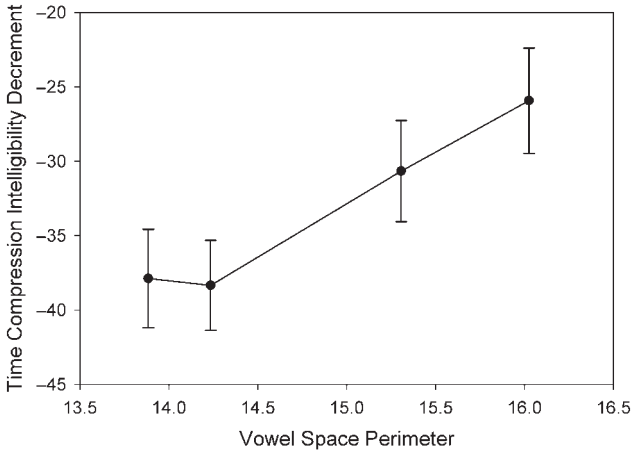
Results

As in the first analysis, the effect of time compression was large and significant in the post hoc model ($\beta = -1.92$, $z = -11.8$, $p < .001$). In Figure 1, the four talkers are arranged from left to right in order of increasing vowel space perimeter. Female 1, who had the smallest vowel space perimeter, was also the least intelligible talker. However, Female 2 was overall more intelligible than the two male talkers, even though their vowel space perimeters were larger, indicating an inconsistent relationship between vowel space perimeter and overall intelligibility. The fixed effect of vowel space perimeter was not significant in the GLMM ($\beta = 0.329$, $z = 0.738$, $p = .460$). Figure 1 also shows that time compression caused greater intelligibility decrements for talkers with smaller vowel spaces and, inversely, smaller intelligibility decrements for talkers with larger vowel spaces. The overall intelligibility decrements for Female 1, Female 2, Male 1, and Male 2 were −37.9, −38.4, −30.7, and −25.9 percentage points, respectively. Figure 2 plots these decrements as a function of vowel space perimeter. The GLMM analysis confirmed that the interaction between processing condition and vowel space perimeter was significant ($\beta = 0.486$, $z = 1.24$, $p = .214$).

The model also indicated that, overall, higher frequency words were somewhat more likely to be correctly identified than lower frequency words, but this fixed effect did not reach significance ($\beta = 0.0828$, $z = 0.821$, $p = .412$). Figure 3 shows average percent correct keyword intelligibility as a function of word frequency (on the Zipf scale) for the four talkers, again (as in Figure 1) arranged from left to right in order of increasing vowel space perimeter. Each point on the scatter plot represents a keyword in either the unprocessed or TC condition (filled or open circles, respectively). The solid and dashed regression lines represent the unprocessed and TC conditions, respectively. In general, the slopes of the regression lines increase (become more positive) as vowel space perimeter increases, indicating the effect of word frequency is stronger for talkers with larger vowel spaces, as confirmed by a significant interaction between vowel space perimeter and word frequency ($\beta = 0.0910$, $z = 2.40$, $p = .0162$). It also appears in Figure 3 that the slopes of the dashed lines (TC condition) tend to be more positive than the slopes of the solid lines (unprocessed condition), which would seem

[7]The model in lme4 style is: SCORE ~ PROCESSING_CONDITION × VOWEL_SPACE × WORD_FREQUENCY + (1 + PROCESSING_CONDITION + WORD_FREQUENCY | TALKER) + (1 | KEYWORD)

Figure 2. Average time compression intelligibility decrements (in percentage points) as a function of talker vowel space perimeter (on the Bark scale). Error bars indicate Welch's *t* intervals for 95% confidence. The talkers with larger vowel spaces have smaller intelligibility decrements when their speech is time compressed.



to indicate that the word frequency effect is stronger in the TC condition than in the unprocessed condition. However, the interaction between processing condition and word frequency fell just short of significance in the GLMM (β = 0.137, z = 1.96, p = .0506).

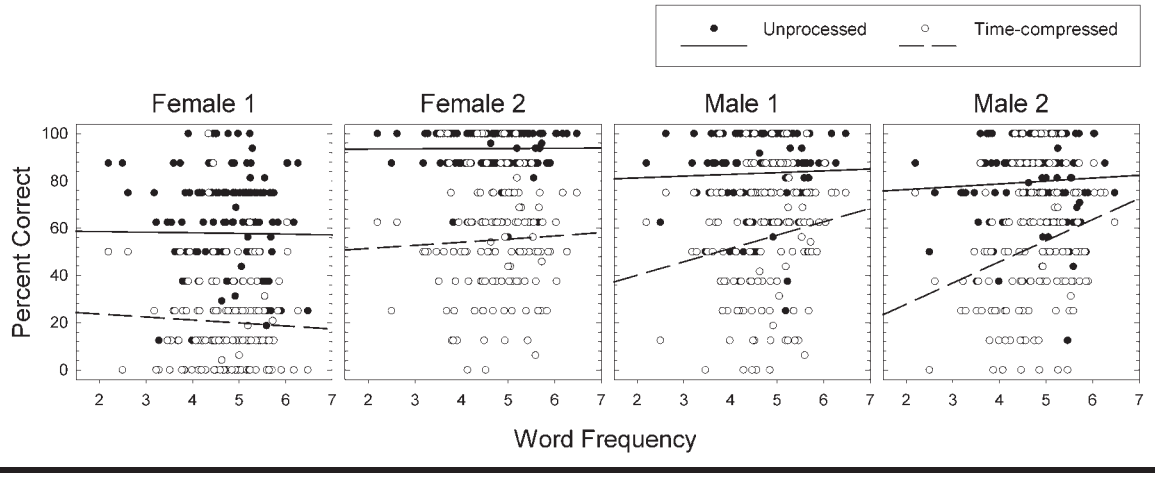
The interaction between vowel space perimeter and word frequency appears particularly strong in the TC condition, where Female 1 (who had the smallest vowel space) has an unusual negative word frequency effect; Female 2 (who had the second smallest vowel space) has a weak, positive word frequency effect; Male 1 (who had second largest vowel space) has a stronger, positive word frequency effect; and Male 2 (who had the largest vowel space) has the strongest word frequency effect. Thus, it appears that the size of the word frequency effect in TC speech is positively correlated with vowel space perimeter. In the GLMM, the three-way interaction between processing condition, vowel space perimeter, and word frequency was significant (β = 0.196, z = 2.54, p = .0111), indicating that the significant interaction between vowel space perimeter and word frequency depends on processing condition.

Discussion

The analyses support prior findings showing that TC speech is less intelligible than unprocessed speech (e.g., Adams et al., 2012; Gordon-Salant & Fitzgibbons, 1993). Furthermore, in this study, time compression had a greater negative effect on the intelligibility of female speech than it did male speech, with a significant interaction effect between talker sex and processing condition, suggesting that speech produced by female talkers may be more susceptible to time compression than that produced by male talkers. This finding is consistent with the results of Versfeld and Dreschler (2002), who found that subjects' average performance for female TC speech was poorer than their average performance for male TC speech.

We probed this finding in post hoc analyses by evaluating possible mechanisms underlying the differential loss of intelligibility between the female and male talkers under time compression. We found that the negative effect of processing on intelligibility was reduced in talkers with larger vowel spaces, with a significant interaction between vowel space perimeter and processing. This suggests that, because the temporal degradation caused by time compression leaves spectral cues relatively intact, acoustic–articulatory

Figure 3. Average percent correct keyword identification scores for individual talkers as a function of word frequency (on the Zipf scale). The filled circles and solid regression lines represent keywords in the unprocessed condition. The open circles and dashed regression lines represent keywords in the time-compressed processing condition. The talkers are arranged from left to right in order of increasing vowel space perimeter. As vowel space perimeter increases, the slopes of the regression lines increase, especially in the time-compressed processing condition.



vowel distinctiveness is even more important in TC speech than in unprocessed speech. In other words, accelerating the timing of the speech signal alters the temporal envelope cues that carry information about manner of articulation and voicing as well as the formant transition cues that carry information about place of articulation. Steady-state formant frequencies, however, are left unscathed, which means that vowel space remains instrumental in speech intelligibility even under time compression. In TC conditions, therefore, the speech of talkers with larger vowel spaces retains vowel space cues even as temporal cues are degraded. On the other hand, when the speech of talkers with smaller vowel spaces loses temporal cues, comparatively less information remains in the speech signal, leading to larger intelligibility decrements.

When talkers are instructed to speak clearly, they often increase the size of their vowel space (e.g., Picheny et al., 1986). Therefore, clear speech may be more robust to time compression than conversational speech. This could have significant implications for various scientific, clinical, and educational applications of TC speech. For example, children with specific language impairment, who have shown increased difficulty processing both TC and naturally fast speech relative to typically developing children (Guiraud et al., 2018), might show even greater difficulty processing accelerated conversational speech. Many listeners with cochlear implants also showed poorer recognition performance than listeners with normal hearing for TC speech (Fu et al., 2001), and this disadvantage could be highly dependent on the speaking style and vowel space of the talker. Individuals, blind or otherwise, who listen to recorded materials at an accelerated playback rate may have greater difficulty recognizing conversational speech or the speech of talkers with smaller vowel spaces who might otherwise be intelligible at their original speaking rate.

Another finding of interest was the significant interaction between vowel space perimeter and word frequency, suggesting that larger vowel spaces may "magnify" the word frequency effect. In general, studies have shown that larger vowel spaces are correlated with greater speech intelligibility (e.g., Bradlow et al., 1996). However, this intelligibility benefit may apply to high-frequency words more than to low-frequency words. Broadbent (1967) suggested that the word frequency effect arises because listeners are "biased in such a way as to accept a smaller amount of information before deciding in favor of a probable word rather than an improbable word." He therefore considered the word frequency effect to be a result of a decision-making bias, on the part of the listener, in setting the criterion for responding. Where listeners set their criterion for responding may therefore depend on the amount of information available to them, resulting in an increased inclination toward high-frequency words when vowel contrasts are more distinct.

In the TC processing condition, the effect of vowel space perimeter on the word frequency effect is particularly strong. We found a significant three-way interaction between processing condition, vowel space perimeter, and word frequency, which arose because, in the TC condition, the word frequency effect was stronger for the male talkers, who had larger vowel spaces than the female talkers. Under time compression, the talkers with smaller vowel space perimeters show weak or even negative word frequency effects. For the talkers with larger vowel spaces, in contrast, intelligibility for TC keywords approaches intelligibility for their corresponding unprocessed tokens as word frequency increases, and the negative effect of time compression diminishes in magnitude until nearly disappearing for the high-frequency keywords (see Figure 3). Indeed, high-frequency words appear to be nearly "immune" to the effect of time compression, but only for the talkers with the larger vowel spaces, suggesting that high-frequency words are more readily perceived than low-frequency words when temporal speech cues are degraded, as long as spectral cues remain intact.

This apparent differential effect of time compression on the intelligibility of low- versus high-frequency words for talkers with larger vowel spaces may also have substantial implications for research, clinical tests, and other recorded media that use TC speech. Listener populations who show increased TC speech processing deficits might have less difficulty processing TC speech when more familiar words are used, especially if the talker's vowel space is larger.

Limitations

We explicitly acknowledge the limitations in this study that arise from employing a sample size of only four talkers, two in each sex group. Even though we found a significant interaction between talker sex and processing condition on intelligibility in the initial model, the secondary analyses suggest that this finding could be due to the vowel space of the talkers rather than their sex. In this study, the two female talkers had smaller vowel spaces than the two male talkers; however, since male talkers generally have smaller vowel spaces than female talkers (Hillenbrand et al., 1995), it may be that, on a more general scale, time compression actually has a greater negative effect on the intelligibility of male talkers, especially for high-frequency words. Because of the limited sample size and the inherent possibility of sampling error, the findings presented here should be interpreted with caution until other studies replicate the results herein.

Furthermore, the listener sample consisted primarily of females. To our knowledge, no studies have found a significant effect of listener sex on speech intelligibility (or a significant interaction between listener sex and talker sex on intelligibility). However, Ellis et al. (1996) found that female listeners tended to indicate that a male talker was overall more intelligible while male listeners indicated that a female talker was more intelligible, suggesting a possible interaction between listener sex and talker sex on subjective intelligibility ratings. Regardless of the effects of listener sex on intelligibility measurements (or lack thereof), a more diverse and representative listener group would have been desirable.

Conclusion

The initial purpose of this preliminary investigation was to test the possible differential effect of time compression on speech intelligibility for male versus female talkers. The speech of female talkers was found to be more susceptible to time compression than the speech of the male talkers. Secondary analyses found that the intelligibility of the talkers with larger vowel spaces (the two male talkers) was less susceptible to the negative effect of time compression, especially for high-frequency words.

The results of research studies and clinical tests that use TC speech may be influenced by the vowel space of the talker, particularly when the materials include a large proportion of low-frequency words, since these words may be less robust to time compression for talkers with reduced vowel space perimeters. Careful attention should be paid to the word frequencies of items in TC speech tests that use clear speech (as most do), since word frequency may have a strong effect on the intelligibility of TC clear speech. This factor may be less critical if the test uses TC conversational speech. These results also imply that audiologists should counsel their patients' communication partners to avoid speaking excessively fast, since natural fast speech is even less intelligible than TC speech, partly because natural fast speech results in reduced vowel space (Turner et al., 1995), compounding the detrimental effects of increased speech rate and low-frequency words. Such counseling may be an effective way to help patients who continue to complain of difficulty understanding female speech even after the application of appropriate amplification and other well-known communication strategies.

Clear speech may be more robust to time compression, especially for high-frequency words. Many listeners who have difficulty processing TC speech may have less difficulty when the talker's vowel space is large or when the lexical content of the speech is more familiar, and these factors should be considered when conducting research or clinical tests with TC speech. Vowel space and word frequency may also affect processing speed when TC speech is used for practical purposes (listening to recorded lectures, audiobooks, etc.).

Acknowledgments

The development of the Utah Speaking Style Corpus was supported by National Institutes of Health Grant R01DC012315 to Eric Hunter. The experiment software was originally developed at Indiana University by Bill Mills; the most recent updates were made by Skyler Jennings and by the second author. Portions of these data were presented at the 171st meeting of the Acoustical Society of America Conference in Salt Lake City in May 2016.

References

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2009). Abnormal cortical processing of the syllable rate of speech in poor readers. *The Journal of Neuroscience, 29*(24), 7686–7693. https://doi.org/10.1523/JNEUROSCI.5242-08.2009

Adams, E. M., Gordon-Hickey, S., & Morlas, H. (2012). Effect of rate-alteration on speech perception in noise in older adults with normal hearing and hearing impairment. *American Journal of Audiology, 21*(1), 22–32. https://doi.org/10.1044/1059-0889(2011/10-0023)

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America, 126*(5), 2649–2659. https://doi.org/10.1121/1.3216914

Barabasz, A. F. (1968). A study of recall and retention of accelerated lecture presentation. *Journal of Communication, 18*(3), 283–287. https://doi.org/10.1111/j.1460-2466.1968.tb00077.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Borrie, S. A., Barrett, T. S., & Yoho, S. E. (2019). Autoscore: An open-source automated tool for scoring listener perception of speech. *The Journal of the Acoustical Society of America, 145*(1), 392–399. https://doi.org/10.1121/1.5087276

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*(3–4), 255–272. https://doi.org/10.1016/S0167-6393(96)00063-5

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74*(1), 1–15. https://doi.org/10.1037/h0024206

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Charpentier, F. J., & Stella, M. G. (1986). Diphone synthesis using an overlap-add technique for speech waveforms concatenation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 11, pp. 2015–2018). IEEE. https://doi.org/10.1109/ICASSP.1986.1168657

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics, 39*(2), 237–245. https://doi.org/10.1016/J.WOCN.2011.02.006

Dupoux, E., & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language, 29*(3), 316–335. https://doi.org/10.1016/0749-596X(90)90003-I

Ellis, L., Fucci, D., Reynolds, L., & Benjamin, B. (1996). Effects of gender on listeners' judgments of speech intelligibility. *Perceptual and Motor Skills, 83*(3), 771–775. https://doi.org/10.2466/pms.1996.83.3.771

Emanuel, D. C., Ficca, K. N., & Korczak, P. (2011). Survey of the diagnosis and management of auditory processing disorder. *American Journal of Audiology, 20*(1), 48–60. https://doi.org/10.1044/1059-0889(2011/10-0019)

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America, 116*(4), 2365–2373. https://doi.org/10.1121/1.1788730

Fu, Q.-J., Galvin, J. J., & Wang, X. (2001). Recognition of time-distorted sentences by normal-hearing and cochlear-implant listeners. *The Journal of the Acoustical Society of America, 109*(1), 379–384. https://doi.org/10.1121/1.1327578

Furini, M. (2008). Fast play: A novel feature for digital consumer video devices. *IEEE Transactions on Consumer Electronics, 54*(2), 513–520. https://doi.org/10.1109/TCE.2008.4560123

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research, 36*(6), 1276–1285. https://doi.org/10.1044/jshr.3606.1276

Gordon-Salant, S., & Friedman, S. A. (2011). Recognition of rapid speech by blind and sighted older adults. *Journal of Speech, Language, and Hearing Research, 54*(2), 622–631. https://doi.org/10.1044/1092-4388(2010/10-0052)

Gordon-Salant, S., Zion, D. J., & Espy-Wilson, C. (2014). Recognition of time-compressed speech does not predict recognition of natural fast-rate speech by older listeners. *The Journal of the Acoustical Society of America, 136*(4), EL268–EL274. https://doi.org/10.1121/1.4895014

Guiraud, H., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Caillot-Bascoul, A., Gonzalez-Monge, S., & Boulenger, V. (2018). Don't speak too fast! Processing of fast rate speech in children with specific language impairment. *PLOS ONE, 13*(1), e0191808. https://doi.org/10.1371/journal.pone.0191808

Hargrave, S., Kalinowski, J., Stuart, A., Armson, J., & Jones, K. (1994). Effect of frequency-altered feedback on stuttering frequency at normal and fast speech rates. *Journal of Speech and Hearing Research, 37*(6), 1313–1319. https://doi.org/10.1044/jshr.3706.1313

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Hu, J. (2017, June, 8). What's your 'x' rating? [Blog post]. https://www.audible.com/blog/the-listening-life/whats-your-x-rating/

Janse, E. (2004). Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech. *Speech Communication, 42*(2), 155–173. https://doi.org/10.1016/j.specom.2003.07.001

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America, 108*(3), 1252–1263. https://doi.org/10.1121/1.1288413

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337–1351. https://doi.org/10.1121/1.381436

Kurdziel, S., Noffsinger, D., & Olsen, W. (1976). Performance by cortical lesion patients on 40 and 60% time-compressed materials. *Journal of the American Audiology Society, 2*(1), 3–7.

LaBarbera, P., & MacLachlan, J. (1979). Time-compressed speech in radio advertising. *Journal of Marketing, 43*(1), 30–36. https://doi.org/10.1177/002224297904300103

Manning, W. H., Johnston, K. L., & Beasley, D. S. (1977). The performance of children with auditory perceptual disorders on a time-compressed speech discrimination measure. *Journal of Speech and Hearing Disorders, 42*(1), 77–84. https://doi.org/10.1044/jshd.4201.77

Markham, D., & Hazan, V. (2004). The effect of talker- and listener-related factors on intelligibility for a real-word, open-set perception test. *Journal of Speech, Language, and Hearing Research, 47*(4), 725–737. https://doi.org/10.1044/1092-4388(2004/055)

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). *Speech levels in various noise environments* (Report No. EPA-600/1-77-025). U.S. Environmental Protection Agency.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24*(2), 175–184. https://doi.org/10.1121/1.1906875

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 and Hearing Research, 29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Riensche, L. L., Curran, C. E., & Porch, B. E. (1986). The assessment of reading readiness using multidimensionally scored time-compressed speech. *Journal of Auditory Research, 26*(1), 1–4.

Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., & Spanias, A. (2013). Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America, 134*(5), EL477–EL483. https://doi.org/10.1121/1.4826150

Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *The Journal of the Acoustical Society of America, 35*(2), 200–206. https://doi.org/10.1121/1.1918432

Schlueter, A., Brand, T., Lemke, U., Nitzschner, S., Kollmeier, B., & Holube, I. (2015). Speech perception at positive signal-to-noise ratios using adaptive adjustment of time compression. *The Journal of the Acoustical Society of America, 138*(5), 3320–3331. https://doi.org/10.1121/1.4934629

Shibuya, T., Kobayashi, Y., Watanabe, H., & Kondo, K. (2012). Differences in the effect of time-expanded and time-contracted speech on intelligibility by phonetic feature. In *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 4489–4492). IEEE. https://doi.org/10.1109/ICASSP.2012.6288917

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass, 3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Sticht, T. G., & Gray, B. B. (1969). The intelligibility of time compressed words as a function of age and hearing loss. *Journal of Speech and Hearing Research, 12*(2), 443–448. https://doi.org/10.1044/jshr.1202.443

Stollman, M. H. P., & Kapteyn, T. S. (1994). Effect of time scale modification of speech on the speech recognition threshold in noise for elderly listeners. *Audiology, 33*(5), 280–290. https://doi.org/10.3109/00206099409071888

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America, 88*(1), 97–100. https://doi.org/10.1121/1.399849

Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 38*(5), 1001–1013. https://doi.org/10.1044/jshr.3805.1001

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Versfeld, N. J., & Dreschler, W. A. (2002). The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *The Journal of the Acoustical Society of America, 111*(1), 401–408. https://doi.org/10.1121/1.1426376

Wang, X., & Humes, L. E. (2010). Factors influencing recognition of interrupted speech. *The Journal of the Acoustical Society of America, 128*(4), 2100–2111. https://doi.org/10.1121/1.3483733

Yoho, S. E., Borrie, S. A., Barrett, T. S., & Whittaker, D. B. (2019). Are there sex effects for speech intelligibility in American English? Examining the influence of talker, listener, and methodology. *Attention, Perception, & Psychophysics, 81*(2), 558–570. https://doi.org/10.3758/s13414-018-1635-3