GREATER BENEFIT FOR FAMILIAR TALKERS UNDER COGNITIVE LOAD

Erin M. Ingvalson[1,2]* & Trevor L. Stoimenoff[2]

[1]Florida State University
[2]Northwestern University
*erin.ingvalson@cci.fsu.edu

ABSTRACT

Earlier work has demonstrated that cognitive resources are expended when processing the speech of an unfamiliar talker. As such, processing the speech of a familiar talker is a more efficient, automated process. Similarly, data have shown that separating a speech signal from noise also uses cognitive resources and listeners with larger working memory capacities are better able to perceive speech in noise. Given the inverse relationship between perceiving familiar speech and perceiving speech in noise, and presuming that these processes tapped the same cognitive store, we tested the hypothesis that listeners would show a greater talker familiarity benefit when perceiving speech in noise while under cognitive load than when perceiving speech in noise with no load. Our hypothesis was confirmed. We discuss our results in terms of their implications for listeners for whom everyday listening is challenging.

Keywords: working memory, speech perception in noise, talker familiarity

1. INTRODUCTION

It is well established that listeners are more accurate when perceiving speech spoken by a familiar talker than by an unknown talker [6,8-11,14-17]. Adjusting to the speech patterns of unfamiliar talkers requires a period of normalization to compensate for talker differences while maintaining phonetic constancy [16]. Initially, it was suggested that processing talker characteristics was independent of the processing of phonemes [10] but that talker information was stored in an episodic trace that could be accessed during perception to facilitate recognition [11]. However, others have claimed that talker and phonetic information are jointly processed during speech perception [8]. In this second view, speech perception, including the processing of talker features, is an active process that requires computational capacity and greater cognitive capacity is required to perceive speech by multiple unfamiliar talkers [8]. Consistent with this perspective, listeners show a greater deficit for perceiving speech under a working memory load when that speech is produced by multiple talkers than by a single talker [9]. Perceiving speech by a familiar talker, then, is a more automatic and efficient process than perceiving speech by an unfamiliar talker, which places fewer computational demands on the perceptual system [16].

More recently, it has been suggested that the ability to separate signal from noise may also tax computational capacity [13]. Under ideal listening conditions, the ability to map the incoming speech signal to stored representations is an automatic process; but as listening conditions degrade, such as in the presence of noise, more cognitive resources are required to perceive speech [2]. Supporting this hypothesis, we have seen data indicating that cognitive skills such as working memory are linked to the ability to perceive speech in degraded listening conditions. Specifically, listeners who have larger verbal working memory capacities are more accurate at identifying speech in noise [3]. Indeed, training auditory working memory leads to significant gains in speech perception in noise [1].

Thus, perceiving speech by a familiar talker is hypothesized to require fewer cognitive resources whereas perceiving speech in noise is hypothesized to require more. Beyond introducing noise, the perceptual system can be further taxed by introducing a working memory load in addition to the speech recognition task [9]. If talker normalization, segmentation of signal from noise, and speech perception utilize a common pool of resources we would expect to see a greater benefit of hearing a familiar over an unfamiliar talker in noise under working memory load—when resources are greatly reduced—than when hearing a familiar over an unfamiliar talker in noise when no load is present.

2. METHODS

Following Souza et al. [15], we recruited participants in pairs rather than attempting to induce talker familiarity via training [10,11]. Multi-talker babble has been shown to result in greater interference than steady-state noise [14], and we therefore opted to use four-talker babble for the noise files. Noise was mixed with the signal at a moderate SNR level (0 dB) previously shown to interact with working memory load in young, normal hearing listeners [12].

2.1. Participants

Twenty-two participants (11 pairs, 16 females) were recruited from the Northwestern University community. Participants were recruited in pairs, with the constraint that the two members of the pair had to be living together currently, and had to have lived together continuously for at least two years prior to participation. Pairs included roommates, cohabitating couples, and married students (age M = 22.23 years, SD = 3.85 years). All participants reported no known hearing or cognitive deficits and were native speakers of American English.

2.3. Materials

Sentences were 200 low-context sentences drawn from the IEEE corpus from Souza et al. [15] Participants spoke into a Shure 58 microphone and the sentences were recorded directly to disk using a MOTU Ultralite external audio interface at 44.1 kHz and 16 bit accuracy. Recordings were monitored in Audacity to ensure sufficient output levels without clipping. Participants were instructed to speak at a natural pace, without any particular effort to speak loudly or clearly. The sentences were recorded twice by each participant, using two different list randomizations. The "best" production of each sentence was used for testing, as judged by the second author.

Each sentence was placed in a separate file. Within a talker, all sentences were RMS matched in amplitude. The noise file was the four-talker babble from the QuickSIN test [4]. Sentences were padded with 100 ms of silence on either side, then mixed with the noise at 0 dB SNR in Praat.

Recognition materials were presented over Sennheiser HD 280 headphones at 70 dB SPL. The experiment was controlled by E-Prime (Psychological Software Tools, Pittsburgh, PA).

2.4. Procedure

All recordings and speech perception in noise tests were conducted in a sound-attenuated booth. Recordings and speech recognition tests were conducted on separate days, spaced approximately one week apart.

Two practice lists from the QuickSIN [4] were used as practice to orient the listeners to the task. The practice lists were administered at a constant SNR of 0 dB, consistent with the speech recognition task. Listeners were instructed to repeat the sentences. Scoring was done live by an experimenter seated outside the booth. Keywords were marked as correct in any presentation order and without grammatical markers.

Ninety sentences were randomly selected for testing. Selections were unique for each listener. The no-load and working-memory-load conditions were blocked to avoid task switching costs. Further, the working-memory-load condition was always administered after the no-load condition to ensure participants were fully familiar with the speech recognition task prior to the introduction of a cognitive load. In both blocks, participants were instructed to repeat the sentences. Five talkers were heard by each listener: one talker was the listener's partner (the familiar talker) and the remaining four were unknown to the listener. Talkers were randomized within a block. Listeners were not informed that one of the talkers would be familiar.

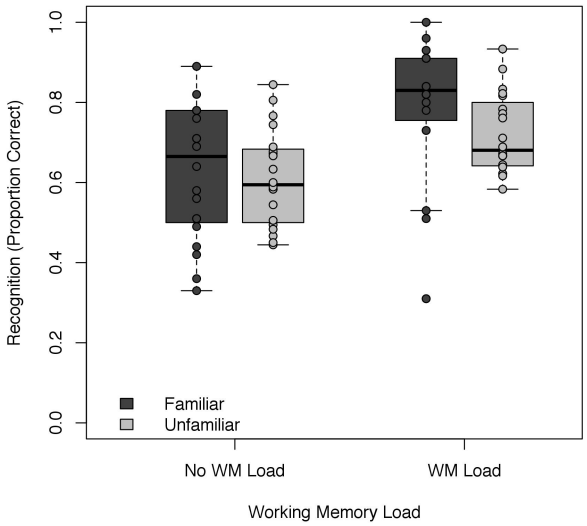
The 45 sentences selected for the working-memory-load condition were randomly blocked into three groups, with three sentences by each talker per group. Pilot testing revealed that listeners in this population had backward digit spans averaging eight digits. We therefore set the working memory task at eight digits to make it moderately challenging for all listeners. At the beginning of each group, the computer generated a random list of eight digits, presented visually. The listener was instructed to rehearse these digits while simultaneously performing the speech recognition task then, following a visual prompt, say the digits in the *reverse* order than they were presented. Digits were scored correct only if they were in the correct order. Listeners were told they would receive a bonus both for each keyword identified correctly and for each digit recalled correctly in the correct position, ensuring they devoted equal effort to both tasks.

3. RESULTS

Two participants were excluded from the perceptual analyses due to an equipment failure during the recognition task; excluded participants were not from the same pair. Raw data were arcsine transformed to normalize score variance for analysis. Figure 1 shows the effect of talker familiarity for all listeners in both the no load condition and the working memory load condition. There was a universal increase in performance in the working-memory-load condition $t(39) = 7.59$, $p < 0.001$, likely due to increased practice with the speech recognition task or to increased experience with the unfamiliar talkers [10,11].

Listeners were more accurate when recognizing words spoken by a familiar talker (familiar $M = 0.72$ proportion recognized; unfamiliar $M = 0.67$ proportion recognized), but the variability in recognition performance was too extensive for the difference between familiar and unfamiliar talkers to

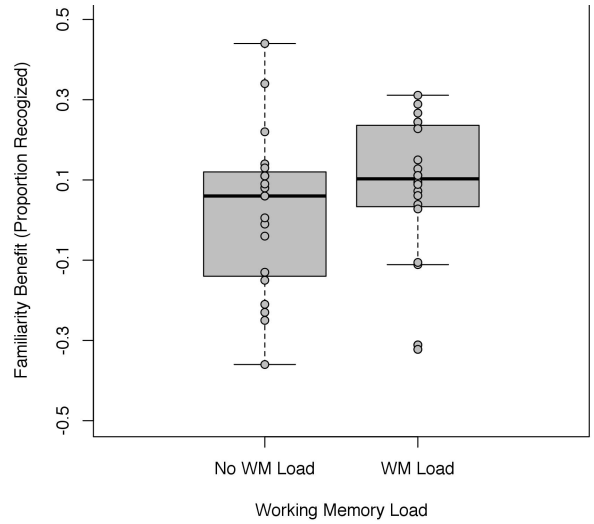
Figure 1: Recognition accuracy (proportion correct words identified) for familiar and unfamiliar talkers in both the no-load and working-memory-load conditions.



be significant (familiar SD = 0.19; unfamiliar SD = 0.12, $t(39) = 1.98$, $p = 0.05$). This variability in the magnitude of the talker familiarity benefit is consistent with earlier work [15], however, our data differ somewhat in that some listeners performed better with unfamiliar than with familiar talkers. To account for this magnitude difference, we calculated a familiarity benefit for each listener by calculating the difference in performance between their familiar talker and each unfamiliar talker [15]. Comparing the familiarity benefit in the no-load condition to the familiarity benefit in the working-memory-load condition revealed that listeners received a greater benefit from hearing a familiar talker when recognizing speech in noise under an additional cognitive load, $t(19) = 2.32$, $p = 0.03$, Figure 2.

To verify that the talker familiarity benefit under cognitive load was not a residual effect of the general increase in performance in the working-memory-load condition, accuracy data for both load conditions for familiar and unfamiliar talkers were entered into a 2 x 2 within-subjects ANOVA. The interaction was significant, $F(1,19) = 5.36$, $p = 0.03$, indicating that the difference in gain was significant.

Figure 2: Familiarity benefit (proportion correct familiar– proportion correct unfamiliar talker) in both the no-load and working-memory-load conditions.



4. DISCUSSION

Though both the talker familiarity benefit and working memory contributions to speech perception in noise are hypothesized to depend on available cognitive resources, surprisingly, to date their potential interaction has not yet been investigated. We hypothesized that listeners would show a greater benefit for recognizing speech in noise by a familiar talker when performing a simultaneous auditory working memory task than when recognizing speech when no working memory task was present. Our hypothesis was confirmed.

The task used in the current study—sentence repetition—is not dissimilar to the delayed repetition task of McLennan and Luce [7]. To ensure the entire sentence is perceived, listeners will wait to begin speaking until after stimulus presentation is complete, leading to a delay between the auditory presentation of the keyword and its repetition. This task is itself slow, and speech processing is further slowed by the addition of verbal rehearsal in the working-memory-load condition. It could therefore be argued that the effects seen here stem from increasing access to talker features due to increased slowing of speech processing and not a freeing of computational resources due to hearing a familiar talker [7]. However, we note that the hypothesis that processing talker information requires cognitive demands and the hypothesis that it takes time for talker information to influence speech processing are not easily separable, as increasing computation processing costs lead to increased processing time [7]. Further, we note that perceiving speech by multiple talkers requires more neural resources than perceiving single-talker speech, supporting the cognitive resource hypothesis [16].

These data add to the growing body of evidence that speech perception is not an isolated process. We have previously seen evidence that adjusting to the speech of an unfamiliar talker requires computational resources, leading to the talker familiarity benefit where hearing the speech of a familiar talker allows for more processing automaticity. Similarly, there are data demonstrating that separating speech from noise also demands computational resources and listeners with more cognitive capacity are more successful at this task. Though one may presume that these tasks tap the

same pool of resources, the data presented here confirm that presumption, and demonstrate that listeners receive a greater benefit from listening to a familiar talker when there are multiple demands on the system relative to when there are relatively few.

Our results perhaps have the largest implications for listeners for whom every day speech perception is degraded, possibly due to hearing loss or to being a non-native speaker [6]. Speech perception is more computationally intensive for these listeners than for native speakers with normal hearing [2,13]. It should therefore some as no surprise that both older adults [17] and adults with hearing loss [15] show a greater talker familiarity effect than normal hearing listeners and that working memory capacity is correlated with listeners success using hearing aids [2,3].

The current fascination with working memory training [5] provides an interesting opportunity for these listeners. It may be that increasing their working memory capacities, and thereby providing more resources that can be used to perceive speech, could lead to more successful speech perception [1]. Increasing working memory also has the potential to produce greater degrees of generalization than have been seen in earlier speech learning studies, as the increase in capacity would be available for all speech tasks, not only those talkers or contexts that had been trained [10,11].

Conversely, there may be learning situations in which it is more beneficial to reduce processing costs, particularly for listeners for whom everyday listening is challenging. Large variability in talkers could make it more difficult for these listeners to attend to relevant phonetic cues, segment speech from noise, or comprehend a message. Limiting learning to a single talker could lead to greater retention of the to-be-learned material by freeing up resources that would otherwise be devoted to talker normalization. In situations where multiple talkers are desirable, such as phonological training, a period of talker familiarization prior to training may allow listeners to better attend to the relevant features [11].

We look forward to future investigations into interactions between computational resources and speech perception. Not only do these studies inform our understanding of the speech perception mechanism, they have important implications for listeners for whom speech perception is challenging.

5. REFERENCES

[1] Ingvalson, E. M., Dhar, S., Wong, P. C. M., Liu, H. Working memory training to improve speech perception in noise across languages. Submitted. *J. Acoust. Soc. Am.*

[2] Arehart, K. H., Souza, P., Baca, R., Kates, J. M. 2013. Working memory, age and hearing loss: susceptibility to hearing aid distortion. *Ear Hear.* 34 251-260.

[3] Foo, C., Rudner, M. Rönnberg, J., Lunner. T. 2007. Recognition of speech in noise with new hearing instrument compression release setting requires explicit cognitive storage and processing capacity. *J. Am. Acad. Audiol.* 18, 618-631.

[4] Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit., L. J., Banerjee., S. 2004. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 116, 2395-2405.

[5] Klingberg, T. 2010. Training and plasticity of working memory. *Trends Cog. Sci.*14, 317-324.

7] Mattys, S. L. Davis, M. H., Bradlow, A. R., Scott, S. K. 2012. Speech recognition in adverse conditions: A review. *Lang. Cognitive Proc.* 27, 953-978.

[8] McLennan, C. T., Luce, P. A. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *J. Exp. Psychol. Learn.* 31, 306-321.

[9] Nusbaum, H. C. Magnuson, J. 1997. Talker normalization: Phonetic constancy as a cognitive process. In Johnson, K. A., Mullennix, J. W. (eds), *Talker Variability and Speech Processing*. New York: Academic Press, 109-132.

[10] Nusbaum, H. C., Morin, T. M. 1992. Paying attention to differences among talkers. In: Tohkura, Y., Sagisaka, Y., Vatikiotis-Bateson, E. (eds), *Speech Perception, Production, and Linguistic Structure*. Tokyo: OHM, 113-134.

[11] Nygaard, L. C., Pisoni, D. B. 1998. Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355-376.

[12] Palmeri, T., Goldinger, S., Pisoni, D. B. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *J. Exp. Psychol.* 19, 309-328.

[14] Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. 1995. How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593-608.

[15] Rönnberg, J., Rudner, M. Foo, C., Lunner, T. 2008. Cognition counts: A working memory system for ease of language understanding (ELU). *Int. J. Audiol.* 47, S99-S105.

[16] Rosen, S., Souza, P., Ekelund, C., Majeed, A. A. 2013. Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *J. Acoust. Soc. Am.* 133, 2431-2443.

[17] Souza, P., Gehani, N., Wright, R., McCloy, D. 2013. The advantage of knowing the talker. *J. Am. Acad. Audiol.* 24, 689-700.

[19] Wong, P. C. M., Nusbaum, H. C., Small, S. L. 2004. Neural bases of talker normalization. *J. Cog. Neuro.* 16, 1173-1184.

[20] Yonan, C. A., Sommers, M. S., The effects of talker familiarity on spoken word identification in younger and older listeners. *Psycol. Aging* 15, 88-99.