COMPREHENDING SYNTHETICALLY ACCELERATED SPEECH: THE RELATIONSHIP BETWEEN PERFORMANCE AND SELF-CONFIDENCE

Christina Wasylyshyn, Derek Brock, and Brian McClimens

U.S. Naval Research Laboratory
Washington, DC 20375
christina.wasylyshyn@nrl.navy.mil, derek.brock@nrl.navy.mil, brian.mcclimens@nrl.navy.mil

ABSTRACT

The present study examines the ability to comprehend speech presented at synthetically accelerated rates in an auditory display environment. We sought to determine whether listeners could accurately predict their own performance when listening to accelerated speech. Comprehension performance and self-confidence ratings were compared at seven different rates of presentation. Self-confidence and accurate comprehension were related at slow to moderately accelerated rates of speech, however, listeners demonstrated an overconfidence effect at higher accelerated speech rates.

1. INTRODUCTION

The workload for an individual US Navy watchstander is attentionally demanding. It often involves the concurrent monitoring of various radio communications channels, as well as, actively monitoring and responding to events on several visual displays. Listening to multiple, concurrent speech communications has been shown to be difficult and detrimental to performance [1], [2], [3], even in the absence of a simultaneous visual display-monitoring task. Our research group at the Naval Research Laboratory is actively involved in developing new methods to improve the comprehension of communications messages in auditory display environments to support Naval watchstanders. One of these methods presents listeners with messages that are buffered, synthetically accelerated and presented serially which potentially removes the need to monitor concurrent talkers [3], [4]. While solutions like this one seem intuitive and thus instantly attractive, they can also introduce more problems than they solve. In the case of accelerated speech, we are concerned that watchstanders might overestimate their own cognitive and perceptual abilities, judging themselves to be more capable of understanding accelerated speech than they actually are. The present study examines the relationship between listeners' performance in accurately comprehending accelerated speech and their self-reported confidence about the accuracy of their responses.

Recent research has demonstrated the utility of listening to synthetically accelerated speech in auditory display environments [4]. Comprehension of auditory messages was compared at normal speed speech and at seven accelerated speech rates. The optimum acceleration rate for comprehension, or the fastest rate at which speech could be presented so that performance did not differ from that at a normal speech rate, was 65% faster-than-normal. Training was found to be effective at slower speech rates compared to higher rates, and it was also shown that practice did not aid comprehension performance. Listeners simply adapted quickly when listening to accelerated speech.

However, even if listeners are able to adapt quickly and can reasonably comprehend synthetically accelerated speech, do they know when they have made errors? Can listeners accurately predict their own behavior? The present study seeks to extend the results in [4] to determine if accurate performance and self-reported confidence are directly related to one another when evaluating information that has been presented at accelerated rates. If accelerated speech communications are to become a viable tool for the Naval warfighter, it is imperative that they be able to accurately judge the certainty of their listening comprehension abilities and act accordingly.

There is an extensive literature on the relationship between cognitive performance in various domains and self-reported confidence. In general, there are two phenomena often studied in this area of research. The first is the assumption that as one's confidence increases so does his/her level of accuracy; this is known as the *confidence-accuracy relationship*. The second is that self-assessment of accuracy often produces overconfidence; this is known as the *overconfidence effect*. That is, one's subjective confidence in the accuracy of his/her judgments is often reliably greater than his/her objective accuracy in performing the cognitive task. For example, overconfidence effects have been shown in investigations of text-based recall [5] and recognition memory [6]. Previous research has also shown that certain cognitive domains are more likely to bring about overconfidence than others and that more difficult cognitive items are often associated with greater overconfidence [7]. Overconfidence has also been related to *escalation of commitment* [8], or what is sometimes referred to as *sunk costs.* These phenomena manifest as decision-makers becoming overly committed to particular courses of action, even in the face of obviously diminishing returns. In the case of day-to-day decision-making, this may be harmless, but in the case of military decision-making, overconfidence may cost countless lives.

To our knowledge, there have been no investigations regarding self-confidence and comprehension of synthetically accelerated speech. The present study examined the relationship between comprehension performance and self-reported confidence of narratives presented at various accelerated speech rates.

The study sought to answer two primary research questions:

1. How is accelerated speech rate related to comprehension performance and self-reported confidence? That is, how accurate are listeners in comprehending synthetically accelerated speech and how accurate are listeners at predicting their own abilities to comprehend accelerated speech?
2. Do listeners become underconfident or overconfident in their comprehension performance at higher accelerated speech rates? That is, is the relationship between comprehension accuracy and self-reported confidence the same across speech rates, or do they diverge?

2. METHOD

2.1 Participants

Eighteen NRL employees participated (11 males, 7 females, mean age = 40.89, SD = 10.82). All participants were native English speakers and claimed to have normal (i.e., non-corrected) hearing.

All participants were presented with a baseline listening comprehension task. This entailed listening to two narratives at a normal speech rate and completing comprehension questions. The mean accuracy for comprehension performance at baseline was 77.66% (SD = 10.69).

2.2 Task and Apparatus

The main battery was composed of auditory narratives and associated comprehension questions [9]. These narratives were approximately 300 words in length and were equated for content difficulty. Each narrative was recorded in a female voice at a speaking rate of approximately 180 words per minute (normal speaking rates range between 130 to 200 words per minute).

After listening to each narrative, participants were asked to evaluate statements about ideas that were represented (or not) in the narrative. These consisted of 24 statements that included both main ideas and specific details about the narrative. Three different types of statements were included for comprehension evaluation:

1. *True* statements represented ideas that were included in the narrative
2. *False* statements represented ideas that were inconsistent with those heard in the narrative
3. *Distractor* statements represented ideas that were consistent with the content of the narrative, but were not actually part of it.

After evaluating each statement, participants were asked to rate their self-confidence about each response on a scale. The left end of the scale was labeled "low" to indicate low confidence ratings, and the right end of the scale was labeled "high" to indicate high confidence ratings. Participants were asked to move a sliding marker on the scale to denote the level of their confidence. The position of the sliding marker on the scale was converted to a percentage score (range 0-99).

The narratives were synthetically accelerated at rates ranging in 15% increments from 50% to 140% faster-than-normal. Participants listened to three narratives at each accelerated speech rate, for a total of 21 narratives. In order to create the accelerated test battery, the narratives were first recorded at a normal speaking rate. They were then subjected to a patented NRL speech-rate compression algorithm [10], known as "pitch synchronous segmentation" (PSS). PSS retains the fundamental frequency of speech signals and preserves a high degree of intelligibility by representing speech as a combination of individual pitch waveforms that do not destructively interfere with one another. PSS does not generate an electric analog of human speech; instead, the output speech is constructed from raw speech and pitch interference is absent from the representation.

The visual part of the study was displayed on a large flat-panel monitor, and the auditory component was rendered binanrally in Sony MDR-600 headphones. Brief auditory examples of what participants heard at each accelerated speech rate will be presented with the poster. They are given in the following sound files:

Speed50% [SPEED50.MP3]

Speed65% [SPEED65.MP3]

Speed80% [SPEED80.MP3]

Speed95% [SPEED95.MP3]

Speed110% [SPEED110.MP3]

Speed125% [SPEED125.MP3]

Speed140% [SPEED140.MP3]

2.3 Procedure

Participants provided informed consent, and then completed a short practice exercise that resembled the format of the experimental task and a baseline comprehension measure. Immediately after listening to each narrative, participants were visually presented with 24 statements (8 true, 8 false, and 8 distracter) and asked to evaluate whether or not the statement identified ideas heard in the narrative. Following the evaluation of each statement, participants provided a self-assessment of the accuracy of their responses on a confidence rating scale.

3. RESULTS

The percentage of correctly identified comprehension statements and the self-confidence ratings (expressed as percentages) served as the dependent measures. Speech rate (normal, 50%, 65%, 80%, 95%, 110%, 125%, 140%) served as

	Correlation	Significance
Overall	r = 0.39	p < 0.001
True Statements	r = 0.41	p < 0.001
False Statements	r = 0.47	p < 0.001
Distractor Statements	r = 0.29	p < 0.001

Table 1: The relationship between comprehension performance and self-confidence ratings across speech rates.

the independent measure. A repeated measures multivariate analysis of variance (MANOVA) was conducted to determine if the rate at which information was presented would have an effect on comprehension performance and confidence ratings.

The MANOVA procedure also takes into account the relationship among the dependent measures. The dependent measures should be related conceptually and correlated with one another at a low to moderate level. Otherwise, the dependent measures should not be analyzed together. Indeed, across speech rates, there is a significant positive relationship between accurate comprehension and self-reported confidence, $r = 0.39$, $p < 0.001$. That is, on average, as accurate comprehension performance increases, so does one's self-reported confidence. This linear relationship is also valid for each type of statement (true, false, and distractor) that participants were asked to evaluate. The correlations are shown in Table 1.

Means and standard errors for overall comprehension accuracy and self-rated confidence (both expressed as percentages) can be seen in Figure 1. Please note that Speech Rate 0 indicates the baseline measure (i.e., normal speed speech). All subsequent numbers on the x-axis indicate the percent speech rate faster-than-normal (e.g., 50 = 50% faster-than-normal, etc…).

Significant univariate main effects for speech rate were obtained for comprehension accuracy, $F(7, 119) = 16.82$, $p < 0.001$, partial eta square =.497, power = 1.0 and for confidence, $F(7, 119) = 14.06$, $p < 0.001$, partial eta square =.453, power = 1.0. These effects can be seen in Figure 1 in that both comprehension accuracy and confidence decrease as speech rate increases. Planned contrasts indicated that comprehension accuracy and confidence ratings were not significantly different from one another at the low to moderate speech rates (0, 50, 65, 80, and 95). That is, listeners were able to accurately predict their comprehension performance. However, at the higher speech rates (110, 125, and 140), confidence ratings were significantly greater than comprehension accuracy. That is, listeners exhibited an overconfidence effect; they expressed a greater degree of confidence than was justified by the accuracy of their comprehension scores.

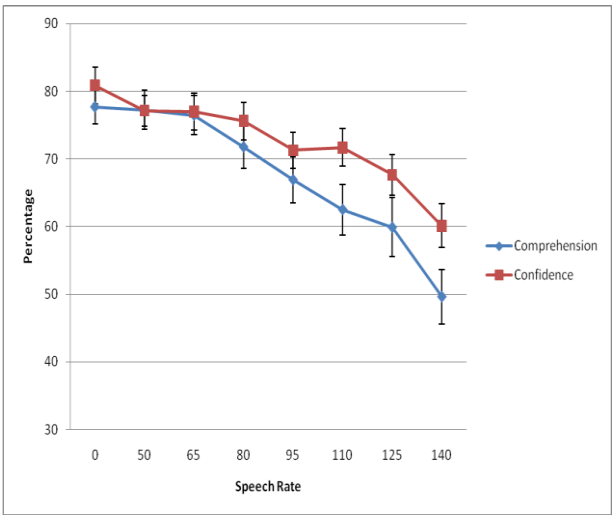

Figure 1: Means and standard errors of percentage of correctly identified comprehension statements and percentage self-confidence ratings as a function of speech rate. Note: 0 indicates normal speed speech and 50-140 denote % speech rate faster-than-normal.

4. DISCUSSION

The US Navy seeks to reduce the size of watchstanding teams. However, Navy radio communications and/or watchstanding workload will remain in its current state. The number and use of critical communications circuits in Navy operations has been steadily increasing, leading to the need for concurrent monitoring of multiple radio communications channels. While a strategy of buffering, accelerating and serializing these communications seems reasonable, we must make sure to present them to the warfighter in a way that doesn't hamper their abilities to make effective decisions.

The present study reports results from work in progress that examines the relationship between confidence and accuracy for comprehension of synthetically accelerated speech. Previous research conducted at NRL [3], [4] demonstrated that presenting audio communications in a serialized and accelerated fashion is a viable alternative to monitoring multiple, concurrent communications channels. The current study extends those previous results to determine if human operators can accurately judge their abilities to comprehend information that is presented at accelerated rates.

The confidence-accuracy relationship was confirmed for slow to moderate accelerated speech rates. That is, listeners were able to accurately predict their comprehension performance up to rates as fast as 95% faster-than-normal. Their confidence ratings were statistically equivalent to their comprehension scores. However, at high speech rates (110, 125, and 140% faster-than-normal), listeners confirmed the overconfidence effect. Listeners reported a greater degree of confidence than was justified by the accuracy of their comprehension performance. Their confidence ratings were statistically greater than their actual comprehension scores. The ability to make accurate confidence judgments regarding one's behavior is a central element of adaptive decision making. Knowing that one is certain or uncertain of his/her ability to carry out a task allows one to take appropriate actions that reflect that certainty or uncertainty. For example, if operators are reasonably good at predicting their own behavior when listening to accelerated speech, then they might be able to ask for clarification when needed or ask that messages be repeated if they were missed or misunderstood.

Several questions and research directions remain open for exploration. For example, we envision situations where time constraints motivate information to be presented at rates exceeding the empirically-derived thresholds we have explored in our experiments to date. In these situations, might it be possible to prompt warfighters to engage in serious metacognition or hypothetical thinking in order to counteract the potentially detrimental effects of the overconfidence effect? Interventions such as these remain unexplored at the present time, but are no doubt important directions to investigate in the design of future auditory display technology for the military.

5. ACKNOWLEDGMENT

This work was supported by the Office of Naval Research, N0001410WX20273.

6. REFERENCES

[1] D. Wallace, C. Schlicting, and U. Goff, *Report on the Communications Research Initiatives in Support of Integrated Command Environment (ICE) Systems*, Naval Surface Warfare Center, Dahlgren Division, TR-02/30, Jan. 2002.

[2] D. S. Brungart, M. A. Ericson, and B. D. Simpson, "Design considerations for improving the effectiveness of multitalker speech displays," *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, 2002.

[3] D. Brock, B. McClimens, G. Trafton, M. McCurry, and D. Perzanowski, "Evaluating listeners' attention to and comprehension of spatialized concurrent and serial talkers at normal and a synthetically faster rate of speech," *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, 2008.

[4] C. Wasylyshyn, B. McClimens, and D. Brock, "Comprehension of speech presented at synthetically accelerated rates: Evaluating training and practice effects," *Proceedings of the 16th International Conference on Auditory Display*, Washington, DC, USA, 2010.

[5] A. M. Glenberg, A. C. Wilkinson, and W. Epstein, "The illusion of knowing: Failure in the self-assessment of comprehension," *Memory & Cognition*, vol. 10, pp. 597-602, 1982.

[6] T. A. Busey, J. Tunnicliff, G. R. Loftus, and E. F. Loftus, "Accounts of the confidence-accuracy relation in recognition memory," *Psychonomic Bulletin & Review*, vol. 7, pp. 26-48, 2000.

[7] J. V. Baranski, and W. M. Petrusic, "Realism of confidence in sensory discrimination," *Perception & Psychophysics*, vol. 61, pp. 1369-1383, 1999.

[8] B. M. Staw, "Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action," *Organizational Behavior and Human Performance*, vol. 16(1), pp. 27-44, 1976.

[9] R. A. Dixon, and C. M. de Frias, "The Victoria Longitudinal Study: From characterizing cognitive aging to illustrating changes in memory compensation," *Aging Neuropsychology, and Cognition*, vol. 11, pp. 346-376, 2004.

[10] G. S. Kang, and L. J. Fransen, *Speech Analysis and Synthesis Based on Pitch-Synchronous Segmentation of the Speech Waveform*, Naval Research Laboratory, TR-9743, Nov. 1994.