WORD SPOTTING IN A MULTICHANNEL VIRTUAL AUDITORY DISPLAY AT NORMAL AND ACCELERATED RATES OF SPEECH

Derek Brock, Christina Wasylyshyn, and Brian McClimens

U.S. Naval Research Laboratory,
4555 Overlook Ave., S.W.
Washington, DC 20375 USA
[derek.brock|christina.wasylyshyn|brian.mcclimens]@nrl.navy.mil

ABSTRACT

The demands of concurrent radio communications in Navy shipboard command centers contribute to the problem of operator information overload and impede personnel optimization goals for new platforms. Motivations for serializing this task and human performance research with virtual, multichannel, rate-accelerated speech in support of this idea are briefly reviewed, and the results of a recent listening study in which participants carried out a Navy-relevant word-spotting task in this context are reported.

1. INTRODUCTION

The broad operational range of radio circuits that require active attention in Navy command centers is a factor in operator information overload and an impediment to personnel optimization goals for new and existing platforms [1][2]. As part of an effort to address this and other performance issues related to multitasking in shipboard decision environments, our research group is exploring machine-mediated task serialization concepts.

Under one of our proposals, concurrent voice communications would be buffered and handled one at a time. To offset the extra time serialization would potentially impose, operators would monitor and/or interact with rate-accelerated speech as needed [3].

In a series of recent participant studies with news and story-based speech materials, measures of attention, comprehension, and effort were significantly improved by mediated serialization, in marked contrast to concurrent listening at normal speaking rates in the same span of time [4].

More recently, we have developed and vetted a corpus of simulated Navy voice communications based on a short set of fictitious tactical scenarios. In the present report, we describe the outcome of a preliminary listening study with these speech materials in which we simulated a mission-specific attentional concern for rate-accelerated voice communications in virtual auditory displays.

2. BACKGROUND

Navy watchstanders work in heavily loaded, multitask settings and must attend to and integrate a wide variety of auditory and visual information tasks. Specialists who have responsibility for particular tactical information domains, such as air or (ocean) surface defense, sit before a visual representation of entities being tracked in the operational theater, monitor and initiate relevant voice communications, and maintain an up-to-the-moment assessment of the tactical situation and their ship's capacity to act. Recent growth in shipboard information capacities has profoundly increased the human performance challenges of this work. Training in the expertise and skills these positions require is done at Navy facilities where teams learn current operational practices on legacy systems and coordinate with each other in highly realistic tactical simulations. Coverage of all requisite voice communications is ensured via team augmentation and redundant monitoring. The communications workload is limited to two active channels (operationally referred to as circuits) per operator, and critical circuits are assigned to two or more individuals as needed [4].

Fleet modernization has brought with it an ongoing opportunity to study how watchstanding can be reshaped to move beyond the constraints of its present operational framework. Increased task loads are already being supported by tactical information systems that have far more functionality and three times the visual display space of previous consoles.

With new ship classes coming online, machine-aided techniques for conveying and managing the display of competing information tasks are being investigated. The intent is to reduce the metacognitive effort associated with unassisted multitasking and to structure the presentation and/or modulation of information in accord with the operator's perceptual strengths. Team augmentation and redundant listening, for example, are not optimal uses of personnel, but this strategy does succeed in minimizing the operational risk of missing critical information in the process of having to attend to two voices at once. Ideally, operators should be able to focus on one message at a time. Mediated serialization of the communications task would allow operators to do this and would also afford a number of optimizations. As competing messages are enqueued, they could be rendered to text, analyzed for priority and duration, and their rates of speech accelerated as needed. The effort of divided listening would be alleviated, and because of varying distributions of idle time on competing channels, trained operators could reasonably cover an array of more than two voice communications circuits in a virtual auditory display [4].

Important listener performance questions are raised by the use of serialization and rate-accelerated speech, especially if the latter is to be rendered as virtual sound. These concerns include listeners' abilities to attend to and encode rapid messaging, adapt to imposed aural attention switching, and maintain and resume an understanding of multiple, aurally modulated, information contexts. In a

DOI: 10.21785/icad2016.031

sequence of studies with continuous speech materials developed from news commentaries and short narratives used in age-related cognition research, listeners in our lab were found to be markedly (and significantly) better at following and understanding serialized, rated-accelerated speech from a forward array of up to four spatially separate sources in a three-dimensional auditory display than they were when listening to two concurrent, opposing, and unaccelerated sources in the same virtual setting [5][6][4]. Attention and comprehension were measured, respectively, by the ability of participants to discriminate between actual and falsely sampled content while they were actively listening and, afterwards, to categorize queries derived from the spoken materials as being in agreement with, or not stated in, what they had heard. Listening performance was found to be equivalent for normal and accelerated listening up to an increase of at least 65% and then to exhibit an approximately linear decline that remains well above chance up to an increase of at least 125% [6][4]. Mixed costs were found for imposed attention switching in manipulations that compared serial listening at normal and 100% faster rates of speech. Attention and comprehension dropped significantly when the rate of speech was doubled but remained substantially above the same measures for current listening, whereas only attention exhibited an additional significant drop—albeit modest—when successive utterances in four serialized contexts were also randomly alternated to completion, as opposed to not being alternated [7].

These findings show that mediated serialization and rate-accelerated speech may in fact be a feasible technology for increasing performance capacities in Navy voice communications. Up to this point, however, only continuous speech has been evaluated, which is not representative of the radio communication patterns operators are actually exposed to. There are important reasons for having adopted this approach. Since continuous speech does not feature intermittent periods of idle time, it allows effectively equivalent performance comparisons to be made between serial and concurrent listening and between normal and accelerated listening. The next step is to explore these and other performance questions with more realistic speech corpora situated in the context of a Navy-relevant task.

As was noted earlier, watchstanders must attend to and integrate a wide variety of auditory and visual information. One of the voice communications skills they learn is the use of changing sets of code words, which are employed to disguise names and other references that would expose operational goals. In our previous work, we measured aural attention performance with ordered checklists of spoken phrases from each virtual source of speech, wherein listeners had to mark phrases they heard and pass over spurious phrases. In a more Navy-like setting involving several channels of speech communications, listeners would be expected to be aware of code words and this could also serve as a measure of aural attention. In the remainder of this paper, we outline and present the results of a limited study that was designed to explore the ramifications of this type of attentional measure.

3. EXPERIMENTAL DESIGN AND METHOD

As part of a related research effort studying the use of chat-based communications in watchstanding operations, we recently developed a corpus of interrelated voice communications with multiple talkers on four radio circuits that serve different operational functions. The scripted speech materials cover four fictional naval scenarios that run for about seven or eight minutes each. Each scenario involves an ongoing tactical operation in which the listener, in the role of a watchstander, is expected to monitor the actions of several radar-tracked air and surface entities/objects that are verbally identified and visually depicted on a corresponding tactical situation display, known as a "TACSIT." The scenarios are designed for use in a laboratory-based mock-watchstanding environment, and an experienced Navy reservist has vetted the visual and spoken content for operational realism and difficulty appropriate to non-specialists in an experimental setting.

The laboratory setup entails a multi-screen tactical information console and a head-tracked immersive auditory display with a fixed, real-world frame of reference corresponding to the console's center screen. Speech and/or cueing and other forms of auditory information are virtually positioned in the listening space using non-individualized head-related transfer functions (HRTFs) and are binaurally rendered with a stereo headset. The center screen displays the TACSIT, and other tactical information tasks are shown separately, as needed, on adjacent screens to the left and right.

For the present study, we identified a different set of eight "code words" in each of the communications scenarios and modified our TACSIT software to show a given set as a list in an onscreen box, together with an overhead depiction of each voice circuit's virtual location in the auditory display relative to the listener. A screen shot of the TACSIT and these additional elements is shown in Figure 1. Both the code word list and the four circuit positions, labeled "1," "2," "3," and "4," were implemented as interactive widgets. The list was only visible when its box was moused over with the computer's pointer, and the circuit labels functioned as clickable buttons. Interactions with the widgets were programmed to be logged as time-stamped performance data.

The code word lists were then used as the basis of an active listening task. Five volunteer listeners from our laboratory (three men and two women, with a mean age of

Figure 1. A screenshot of the visual display listeners used. The TACSIT showing a number of radar-tracked objects is positioned at the top. The list of code words can be seen in the tall box on the lower left, and the interactive depiction of each voice circuit's virtual location in the auditory display relative to the listener is positioned to the right of the list.

34) were given a short time to study and commit one of the lists to memory. Next, they listened to the corresponding scenario and were told they could follow radar tracks and attendant behaviors on the TACSIT as they were mentioned on any of the radio circuits. At the same time, they were asked to spot any spoken instances of the listed code words and to indicate where each instance came from by clicking on the corresponding circuit label as quickly as possible. Approximately equal numbers of code words were spoken on each circuit during the exercise, with half of the eight words distributed across the four circuits and occurring only once and the others occurring up to four times. The four circuits were virtually positioned in the listener's forward horizontal plane at 75° and 25° to the left and 25° and 75° to the right of the console's centerline; the spread between these positions was exaggerated in the visual display to make it easier to click on the circuit labels (see Fig. 1). All of the voice communications were serially interleaved, meaning that the temporal order of utterances across all circuits was preserved in the manner of a first-in-first-out queue, but only one circuit was sounded at a time. To determine how many times listeners needed to look at the list for verification, as well as the amount of time they spent looking at the list, the code words were intentionally hidden during the listening exercise, but could be revealed, if needed, by mousing over the list box. Participants were urged to refer to the list as little as possible.

Each participant performed four word-spotting exercises, each based on a different scenario and each corresponding to a different manipulation of the speech materials. The voice communications were unaccelerated in one of the exercises and were uniformly 50%, 65%, and 100% faster in the other three, respectively. Speech in the faster manipulations was accelerated with a speech analysis/synthesis technique known as pitch-synchronous segmentation developed at our facility in the early 1990s that preserves pitch and facets of the speech waveform associated with intelligibility [8]. To ensure that shorter utterances corresponded to what was being shown on the TACSIT, which was not accelerated in the faster manipulations, each accelerated utterance was played at the same point in time it had originally been scripted to occur prior to being accelerated. To be clear, the visual part of each of the "faster" scenarios ran for its original length of time, and each accelerated utterance $\boldsymbol{u}_{accel}$ started at the same time $\boldsymbol{t}$, relative to the start of the visual part of its scenario, as its source $\boldsymbol{u}_{unaccel}$ did in the original, "unaccelerated" version of the scenario. The unaccelerated exercise was given to all participants first, and the other three were given to each in a successively changed order. The distribution of spoken code words on each circuit in the four manipulations is described in Table 1.

Table 1. Distribution of code words spoken on each radio circuit in each manipulation. Four of the eight code words participants were asked to spot in each listening exercise occurred only once and were uniformly distributed; these occurrences are respectively indicated with the number "1" in each cell of the 4x4, manipulation-by-circuit matrix shown in the table. The remaining four code words in each exercise were spoken up to four times and were distributed so that the total number of code words spoken on each circuit was approximately equal; these occurrences are indicated with the parenthetical numbers in each cell. In the first cell, for example, three different code words were spoken, one being said three times, for a total of five occurrences.

	Circuit 1	Circuit 2	Circuit 3	Circuit 4
Normal speech	1+(3+1)=5	1+(3)=4	1+(3)=4	1+(1+2)=4
50% faster	1+(3+1)=5	1+(4)=5	1+(3)=4	1+(3)=4
65% faster	1+(3)=4	1+(2)=3	1+(3+1)=5	1+(1+1+1+2)=6
100% faster	1+(4)=5	1+(4)=5	1+(4)=5	1+(1+3)=5

In the following analysis, list visits, list look times, the timing of mouse clicks on circuit numbers in the visual display, the number of errors, and the proportions of correct responses were treated as dependent variables. Our expectations were a) that listeners would uncover the code lists two or three times during each exercise, b) that response times would be slower in the faster manipulations, c) that there would be few if any erroneous identifications of the circuit a given code word was spoken on, and d) that the mean proportion of correct responses across all manipulations would be above 50%, with the lowest scores occurring in the faster manipulations.

4. RESULTS

There were no significant differences in the number of list visits across the four speech rates, $F(3,9) = 0.434$, $p = 0.731$. However, the mean number of list visits per listening exercise was 10.35, which was much higher than anticipated. There were also no significant differences in the average amount of time participants looked at the list (and/or kept the list visible) across the four manipulations, $F(3,9) = 1.515$, $p = 0.276$. On average, participants spent roughly 5% of each exercise referring to the code words. In summary, increasing the speaking rate of the talkers on each of the radio circuits by up to 100% did not lead to meaningful changes in the number of times subjects looked at the list of code words nor in the amount of time the list was kept visible on the TACSIT.

In the response data, one listener's clicks were lost in the 65% and 100% faster manipulations due to a technical problem. The remaining data for this participant was included in the following analyses. There were no significant differences in the time listeners took to click on a circuit after spotting a code word, $F(3,9) = 2.234$, $p = 0.154$. A mixed criterion was used for this measure: a 4000 ms cutoff was applied unless the code word was embedded in an utterance that took more than this amount of time to complete. Contrary to what was expected, the mean values for this performance metric ranged from 3240 ms for normal speech to 1747 ms in the 65% faster manipulation; The next slowest mean response time (2598 ms), however, occurred in the 100% faster scenario. There were no significant differences in the number of errors listeners made across the four listening exercises, $F(3,9) = 1.0$, $p = 0.436$. An average of 2.06 errors were made in each manipulation, an error being defined as clicking on the wrong circuit when a code word was spoken. More notably, the total number of clicks listeners made decreased significantly as the rate of speech was accelerated. Thus, the proportion of code words listener's spotted and clicked the correct source of was

Table 2. Summary of mean performance measures in each manipulation. "*" indicates a main effect.

	Normal speech	50% faster	65% faster	100% faster
List visits	9.6	13.8	8.25	9.75
List look time (s)	15.3	27.5	17.5	23.4
Resp. time (ms)	3240	2133	1747	2598
Errors	1.5	3.25	1.75	1.75
Prop. correct*	0.4853	0.2639	0.3472	0.2000

significantly predicted by speech acceleration rate, $F(3,9) = 8.440$, $p = 0.006$, $\eta^2 = 0.738$. This proportion dropped from nearly 50% for unaccelerated speech to 20% when the rate of talker's speech was doubled. A summary of the performance measures discussed up to this point is given in Table 2.

To assess performance differences associated with the central and/or peripheral circuits, a 4x4 repeated measures ANOVA of the corresponding proportions of correct responses across the four rates of speech, showed there was no main effect of spatial position, $F(3,9) = 0.072$, $p = 0.974$. However, there was a significant circuit by speed interaction $F(9,27) = 5.542$, $p < 0.001$, $\eta_p^2 = 0.649$. These results are depicted in Figure 2.

To more closely examine the interaction, separate repeated measures ANOVAs of the word spotting responses were conducted for each rate of speech. As can be seen in Figure 2, listeners correctly responded to a greater proportion of code words spoken on the third circuit (0.80, right central position) in the normal speech manipulation (blue line) than in any other part of the study; the performance differences in this manipulation were marginally significant, $F(3,12) = 3.124$, $p = 0.066$, $\eta^2 = 0.439$. In addition, the comparatively high proportion of correct responses to code words on the fourth circuit (0.55, right peripheral position) was significant, relative to performance on the other circuits in the 100% faster manipulation (purple line), $F(3,9) = 10.500$, $p = 0.003$, $\eta^2 = 0.778$. There were no significant performance differences between the four circuit positions in the 50% and 65% faster manipulations (respectively, the red and green lines).

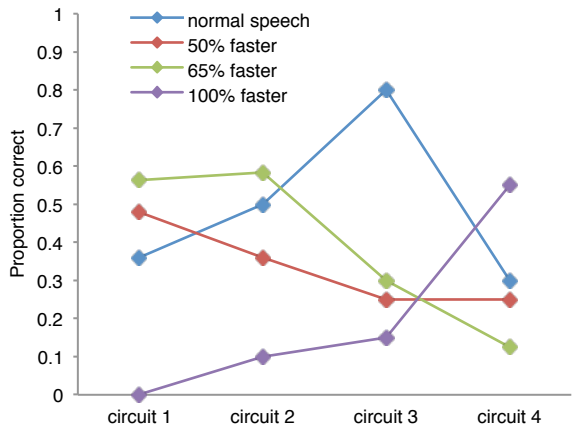

Figure 2. Correct identifications of circuits code words were spoken on, expressed as a proportion, in each of the four manipulations.

5. DISCUSSION

Several performance questions for mediated serialization of voice communications were explored in this preliminary study. In particular, the use of code words in Navy tactical communications was studied as a novel way to compare aural attention performance under normal and accelerated rates of speech as well as when rendered at different source positions in a virtual auditory display. While no test of scenario comprehension was included in the study and only a small number of participants were recruited, the findings showed that when listeners recognized code words, they could generally identify its virtual source shortly after it was spoken regardless of the rate of speech. Surprisingly, instead of being slower to respond to accelerated speech as we conjectured, listeners responded progressively faster in the 50% and 65% manipulations (see "Resp. time (ms)" in Table 2). The reasoning behind our expectation was that because making sense of speech can occasionally require a momentary mental review of what has been variously called "echoic memory" [9], "precategorical acoustic storage" [10], and "brief auditory storage" [11], it therefore seems plausible that listeners without exposure or practice might need to do this more often when processing rate-accelerated speech, and so, become progressively slower to respond. The decline in the mean proportion of correct responses shown in Table 2 as the rate of speech increases across manipulations—albeit, not fully linear and not significant until the rate of acceleration is is 100%—is consistent with this processing conjecture in the sense that attentional difficulties tend to correlate with missed information. Moreover, listeners' mean response time in the 100% faster condition was the second slowest in the study, suggesting that, as in our previous studies (e.g., [6]), at some point above an increase of 65% in the rate of normal speech, aural processing begins to require genuine attentional effort. Noting that "correct responses" required listeners to identify the circuit a code word was spoken on, rather than the word itself (thus, only implying that a specific code word was heard), the unexpected pattern of response times observed here may reflect a native attentional ability to adapt to the pace of auditory events up to a point, much as melodic and/or rhythmic information can generally be followed within a range of tempos and becomes difficult to process outside of this range (see, e.g., [12] and [13]).

That said, it is evident that the word-spotting part of the experimental task was not as easy as we had anticipated. While lower scores did occur in the faster manipulations as anticipated, the mean proportion of correct responses across all manipulations was below 50%, and despite the seemingly low numbers shown in Table 2, listeners made more errors than were expected. Put another way, listeners simply missed more than half of the code words in each of the manipulations, and although on the whole they were equally attentive to both the central and peripheral circuits (see Fig. 2: by circuit, the mean proportion correct responses ranges from .3125 for circuit 4 to .3858 for circuit 2), on average, listeners clicked on the wrong circuit 27% of the time (the error rate ranged from 15.4% for normal speech to 40.6% when the speech was 50% faster). Given the extent and pattern of missed code words across manipulations and the surprising rate of source identification errors, it is clear that listeners struggled to do well. Factors that may have contributed to their performance difficulties include distracted listening arising from looking at the code word list (note that these numbers in Table 2 are much higher than were expected), response completion errors arising from ongoing listening demands, attentional fatigue arising from varying distributions of code words in each manipulation and the relative sparsity of instances (just under 2.5 words per minute) and, to a lesser extent (because of the horizontal layout of source positions) poor display fidelity due to the use of non-individualized HRTFs. An aspect of the spoken material worth considering, that may have also influenced listeners' performance, is the relatively innocuous nature of the code words that were used in each manipulation. The incorporation of a range of Navy-like code words—words that are generally colorful and somewhat salient relative to ordinary speech—was not considered in the scripting and

recording of the radio communications part of the four tactical scenarios, which was done several months before the study this conceived. Thus, words that could plausibly function as "code words" had to be identified in each script. A range of nouns ("knots," "sensors," "queen," etc.) and, to a lesser extent, verbs ("proceed," "verify") and response words ("aye," "copy") were chosen within the fixed wording of the scripts so as to achieve a relatively uniform spread of words for listeners to spot across the four circuits in each scenario; the full selection of code words is given in Table 3. Additionally, in each manipulation, four of the eight code words were spoken only once and the remaining four were said multiple times in roughly equal numbers see Table 1). Performance very nearly at or above 50% correct occurred in only six cells of the 4x4 manipulation-by-circuit response matrix (these measures are plotted in Fig. 2). Listeners only needed to spot two words, "report" and "whiskey," in the best of these cells (circuit 3 under normal speech), and it could be argued that "whiskey" is a more readily spotted word than any or most of the more ordinary words participants were asked to listen for. In contrast, though, listeners also only had two words to spot in three of the four cells with the lowest performance (circuits 1, 2, and 3 under 100% faster speech and circuit 4 under 65% faster speech). Curiously, in the lowest of these cells (circuit 1, with no correct responses), one of the two words was "reconnaissance," which was chosen for its potential salience. In spite of these somewhat contradictory performance patterns, similar to a point made above, the predominant occurrence of the poorest performance in the study under the fastest rate of speech is consistent with our earlier finding that listeners perform at parity with normal speech only up to a 65% increase in the rate of speech. In the other lowest performing cell (circuit 4 under 65% faster speech), a different factor may have interacted with the listening task: here, instead of only two words, there were five separate words to spot—more than in any other cell in the study—giving listeners more work to do. Even so, this observation is somewhat countered by the best performance in the 100% faster manipulation (circuit 4) wherein, unlike circuits 1, 2, and 3 in this manipulation, there were three words to spot. The possibility that many of the code words were not memorable is also supported by the much higher than expected numbers of list looks participants resorted to in each manipulation in spite of having been given ample time to study each list before each of the listening exercises.

In addition to examining a structured set of performance questions, the broader intent of this study was to gain a preliminary sense of how carefully listeners are likely to listen in a somewhat "realistic" serialized multimodal framework wherein both auditory and visual information display components are referentially related. If mediated serialization of competing aural information tasks is to be adopted, it must be viable within a unifying operational context such as tactical situation monitoring, which was used here, or air traffic control. The counterpart of this study's measure of auditory attention will be an evaluation of listeners' attention to and knowledge of the tactical information content of the scenarios in a future experiment. Another aspect of managed task switching we have explored in the past and now plan to study in an integrated operational setting is the utility of virtual auditory cueing as a technique for guiding the operator's attention from one task to the next [14][15][16]. Although auditory cues significantly improved task performance in a series of prior studies with a cockpit-like dual task involving rapid decision making and continuous tracking, a range of additional questions are raised by their use in mixed auditory information settings. Among these are the development of an empirically based set of organizing principles for the presentation of competing sounds and performance-based evaluations of different cueing designs for cross-modal task switching involving prioritization and modulated information displays such as rate accelerated speech and visual augmentation to guide the operator's attentional focus.

Table 3. Lists of code words listeners were respectively asked to spot in each of the four manipulations. The first four in each list were only spoken once (each on a separate circuit) and the remaining four were said more than once in a given manipulation (see Table 1.)

Normal speech	50% faster	65% faster	100% faster
attention	inbound	supplies	reconnaissance
route	vessel	information	anchor
report	feet	neutral	direction
threat	mark	reflection	sensors
proceed	copy	knots	charlie
intentions	blockade	waters	queen
whiskey	values	aye	team
sea	status	stations	verify

6. ACKKNOWLEDGMENT

This research was supported by the Office of Naval Research under work order N0001412WX20879.

7. REFERENCES

[1] D. Wallace, C. Schlichting, 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, January, 2002.

[2] C. T. Bush, J. R. Bost, P. S. Hamburger, and T. B. Malone, "Opimizing manning on DD21," in *Proceedings of the Association of Scientists and Engineers (ASE) 36th Annual Technical Symposium*, April, 1999.

[3] B. McClimens, D. Brock, and F. E. Mintz, "Minimizing information overload in a communications system utilizing temporal scaling and serialization," in *Proceedings of the 12th International Conference on Auditory Display (ICAD)*, London, UK, June, 2006.

[4] D. Brock, C. Wasylyshyn, B. McClimens, and D. Perzanowski, "Facilitating the watchstander's voice communications task in future Navy operations," in *Proc. of the 2011 IEEE Military Communications Conference (MILCOM)*, Baltimore, MD, 2011.

[5] D. Brock, B. McClimens, J. 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," in *Proceedings of the 14th International Conference on Auditory Display (ICAD)*, Paris, France, June, 2008.

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

[7] D. Brock, S. Camille Peres, and B. McClimens, "Evaluating listeners' attention to and comprehension of serially interleaved, rate-accelerated speech," in *Proceedings of the 18th International Conference on Auditory Display (ICAD)*, Atlanta, GA, USA, June, 2012.

[8] 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, November, 1994.

[9] U. Neisser, *Cognitive Psychology*, New York: Appleton-Century-Crofts, 1967.

[10] R.G. Crowder and J. Morton, "Precategorical acoustic storage," *Perception and Psychophysics*, vol. 5, pp. 365-373, 1969.

[11] C.J. Darwin, M.T. Turvey, and R.G. Crowder, "An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage," *Cognitive Psychology*, vol. 3, pp. 255-267, 1972.

[12] M.M. Baese-Berk, C.C. Heffner, L.C. Dilley, M.A. Pitt, T.H. Morrill and J.D. McAuley, "Long-term temporal tracking of speech rate affects spoken-word recognition," *Psychological Science*, vol. 25, pp. 1546–1553, 2014.

[13] J.D. McAuley, "Tempo and rhythm," in M.R. Jones, R.R., Fay, and A.N. Popper (Eds.), *Music Perception*, Springer Handbook of Auditory Research, vol. 36, pp. 165-199, 2010.

[14] D. Brock, J. A. Ballas, J. L. Stroup, and B. McClimens, "The design of mixed-use, virtual auditory displays: Recent findings with a dual-task paradigm," in *Proc. of the 10th Int. Conf. on Auditory Display (ICAD)*, Sydney, Australia, July 6- 9, 2004.

[15] D. Brock, B. McClimens, and M. McCurry, "Virtual auditory cueing revisited," in *Proceedings of the 16th International Conference on Auditory Display*, Washington, DC, June 9-15, 2010.

[16] D. Brock and B. McClimens, "To what extent do listeners use aural information when it is present?" in *Proceedings of the 17th International Conference on Auditory Display*, Budapest, Hungary June 20-24, 2011.