View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Scholarly Materials And Research @ Georgia Tech

MINIMIZING INFORMATION OVERLOAD IN A COMMUNICATIONS SYSTEM UTILIZING TEMPORAL SCALING AND SERIALIZATION

Brian McClimens, Derek Brock, Farilee E. Mintz

Naval Research Laboratory
4555 Overlook Ave., SW Washington, DC 20375
mcclimen@aic.nrl.navy.mil
brock@itd.nrl.navy.mil

ITT Industries, Advanced Engineering and Sciences (AES)
2560 Huntington Ave., Alexandria, VA 22303
farilee.mintz@nrl.navy.mil

ABSTRACT

Recent Navy research has identified the monitoring of multiple communications streams as a performance bottleneck. We introduce a novel approach for improved monitoring of multi-channel voice communications which makes use of time-compression of speech in order to present communications serially, and discuss some of its potential benefits and pitfalls. An experiment currently being developed and piloted is detailed as part of a larger series of studies designed to examine the plausibility of this new approach. This study will explore the effects of sped up speech on listeners' comprehension of vocal radio transmissions.

1. INTRODUCTION

1.1. Motivation minimize

The US Navy is interested in reducing the number of watch-standers in command centers aboard naval vessels without making sacrifices in performance. Several technologies have been introduced that allow watchstanders to perform their tasks with increased efficiency. However, monitoring multiple channels of voice communications has proven to be a performance bottleneck.

A recent Navy report on communications research [1] examines the effectiveness of using spatialized audio and speech-to-text methods in an attempt to alleviate this problem. Results suggest that while spatialized audio provides an objective performance improvement and participants report improved performance with text-to-speech windows, concurrent presentation of communication streams still results in a significant performance drop.

In an attempt to address this issue, a method was developed at the Naval Research Laboratory [2] in which concurrent communications are buffered and sped up and then presented serially rather than in parallel. The serialization of speech introduces a possibility that time-critical transmissions will be delayed by an unreasonable amount, and temporal compression of audio communications may have adverse effects on the comprehension of transmissions. If a system were cleverly designed to minimize these problems, the benefits received from avoiding concurrent presentation of communication data may outweigh the costs. This potential has led to a large multi-disciplinary project within NRL to explore the feasibility of such a system.

1.2. Project Overview

The NRL Communications Overload project has the goal of reducing the workload required to monitor multiple channels of speech. This project will examine the effectiveness of spatialized audio, speech to text, language processing, temporal scaling of audio channels, serialization of speech, and a graphical interface in the realm of communications monitoring.

Using an algorithm developed by Kang and Fransen at NRL, the rate of speech can be compressed in time without affecting the frequency content or sacrificing much intelligibility [3]. With this temporal scaling of speech, it is believed that concurrent presentation of communications channels can be avoided, and that the ratio of information processed to time can be improved for Navy watch-standers. With an increased information to time ratio, it is believed that communications lines can be monitored effectively with a reduction in the amount of personnel required.

The concern that time sensitive material will be delayed by the serialization of incoming communications will be addressed by a combination of speech to text technology, language processing and the graphical interface. All incoming transmissions will be converted to text and be displayed on screen as soon as possible. The text will be passed to a language-processing component that will assess the priority of all incoming messages and interrupt the current transmission if necessary.

2. CURRENT WORK

A series of studies is planned to evaluate the effectiveness of temporally scaled speech and spatialization in reducing an operator's workload, as well as the effect of temporal scaling on the intelligibility of communications. These experiments will follow a common structure. In each of these studies, participants will be presented with multiple streams of audio, and comprehension will be measured. Throughout the experiments, different methods of presentation and a variety of measures of comprehension will be explored. The first of these experiments is currently being designed and piloted, and will be the primary focus of this paper.

The main goal of this first experiment is to compare comprehension of serialized audio streams sped up using Kang and Fransen's algorithm, to a baseline presentation of concurrent audio streams at their original speed. Work by Garvey on the intelligibility of speeded speech suggests that up to 60% of a word can be removed without strongly affecting the intelligibility of speech [4]. However, Garvey's work centered around recognition of a single word. The present research will explore listeners' ability to comprehend sped up speech presented continuously over extended periods of time.

2.1. Content of Audio Streams

Access to actual Navy communications is limited, so this study will use radio news stories acquired from National Public Radio (NPR) as sources for the audio streams. The use of radio news stories provides a few advantages. Typically newscasters on national radio stations are trained to have even rates of speech, clear enunciation and do not have strong local accents. These qualities allow the study to focus on the effects of temporal compression while minimizing the effects of irregularities found in speech. In addition, each of the stories features a single speaker, so performance differences will not be confounded by the number of speakers or transitions between speakers.

These streams differ from what watchstanders on naval ships listen to in a few important ways. For example, the news stories have a 100% rate of service. Generally Navy communications have lower rates of service that change over time. The constant speech found in news reports allows the study to examine listeners' ability to interpret speech presented at increased rates for sustained periods of time.

Another difference between audio streams used in this study and Naval communications is that each news story consists of exactly one transmission that has a clear beginning and end. Audio streams in a communications setting typically consist of a series of shorter transmissions that may come at unpredictable times. Future studies will explore the effects of attention shifts that result from several shorter transmissions.

2.2. Presentation of Audio Streams

Each of the conditions in this experiment will present four audio streams to the participant. These streams will be presented over headphones, and will be spatialized using a head related transfer function (HRTF). It has not yet been decided whether participants will use one generalized HRTF or whether individuals will use an HRTF selected from a set to best match their ears. In addition, head tracking may be used in order to obtain better results from the HRTF(s) used. Participants will hear a total of sixteen different news stories across four different conditions. In the first condition, the participant will hear four news stories presented concurrently at regular speed. In the other three conditions, participants will hear four stories sped up by a variety of degrees.

Originally, sounds were intended to be spatialized at equidistant points along a circle around the participant's head. These points were located at angles of 30, 70, 110, and 150 degrees. After listening to the audio streams positioned at these points, it was concluded that the positions at 30 and 70 degrees sounded very close together, as did the positions at 110 and 150. Interestingly, the points at 70 and 110 sounded much further apart in comparison. For the pilot study, positions have been changed to 30, 80, 100 and 150 degrees. A visual comparison of these two setups is shown in figure1, and audio samples are provided for an aural comparison between equidistant (wav1.wav) and uneven (wav2.wav) spacing of the sounds. For the actual experiment, the positions used will be derived from current research rather than a subjective determination.

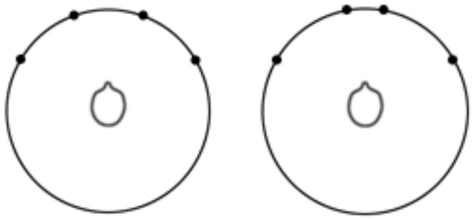

Figure 1. *The left image shows positions at 30, 70, 110 and 150 degrees. The right image shows the positions used in the pilot at 30, 80, 100 and 150 degrees.*

Three different rates of time-compression will be used in this experiment. Speech will be presented sped up by 50% (wav3.wav), 75% (wav4.wav) and 100% (wav5.wav). The rates of speed increase were chosen such that the slowest increase in speed is reasonably easy to understand, while still providing a significant increase in speed, and the fastest presentation is anticipated to be very difficult, but not impossible to understand. Audio clip three, provides a demonstration of speech presented at these three rates, and compares each to the original.

2.3. Measurement of Comprehension

For each of the audio streams presented, participants will be provided with a list of phrases. Half of the phrases in the list will occur in the audio stream, while the other half will be foils taken from news stories about similar issues. As participants are listening to the audio stream, they will be asked to mark all of the phrases that they hear. Comprehension of the audio stream will be measured in terms of the number of phrases that are marked correctly, and the number of foils that were marked. In the pilot, pen and paper will be used to mark the phrases heard. A program will be used to record this data in the final experiment, allowing access to reaction times, which may provide insight into the level of difficulty associated with interpreting the speech.

3. HYPOTHESES

When monitoring communications in an ideal setting, comprehension should be close to 100%. In practice, perfect comprehension can never be achieved and sustained over long periods of time. So, even in critical situations, there must be some level of acceptable error when measuring comprehension of communications. It is hoped that at least one of the rates of speech used in this experiment will provide a greater information to time ratio without dropping below accepted performance levels.

It is hypothesized that as the rate of speech is increased through use of temporal scaling, a drop off in comprehension will result. It is further hypothesized that this drop off will not be linear and that speech can be sped up by some amount without significantly affecting the levels of comprehension participants can achieve. In addition, it is expected that the fastest presentation of speech in this experiment will result in very poor comprehension levels.

4. ADDITIONAL TOPICS

There are many issues not addressed in the experiment described above which may be explored in similar experiments. Brief discussions about a couple of these topics follow.

4.1. Attentional Shifts

The described experiment switches between different audio streams, however each audio stream is presented in its entirety without interruption, and when the presentation switches to another audio stream, it never returns to one that was previously presented. In a realistic setting, watchstanders will have to divide their attention between several channels, and will frequently be shifting their attention back and forth between these channels.

It is expected that operators will have some difficulty comprehending the communications directly following an attention shift. One possible solution to this is that the communications will be presented at regular speed following a shift of attention, and slowly be ramped up to an optimal rate of speech.

Experiments will likely be conducted to explore the effects of attention shifts. Results from such experiments may help guide the behavior of a system that must switch between presentation of multiple audio streams.

4.2. Spatialization

In the task participants will be asked to do for the experiment, there is no need for participants to make distinctions about which audio stream they are listening to. For operators in real world settings, some information may have meaning that is dependant on the source of the transmission. It is expected that the ability to determine the source of a transmission will be crucial in realistic settings.

Spatialization has been used with success in other applications (e.g. air traffic control) where identifying the source of audio transmissions is crucial [5]. While research has been done concerning the effectiveness of spatialization in applications similar to naval communications, further experimentation may be required to study the effects of spatialization in this specific application.

5. CONCLUSION/SUMMARY

The conjunctive use of temporal scaling, serialization and spatialization of communication streams constitute a novel approach towards simplifying the task of monitoring multiple streams of communication. This approach carries inherent consequences including the delay of some transmissions and a distinct possibility of negative effects on the comprehension of presented audio. The authors of this paper believe that despite the problems introduced by this approach, the benefits will outweigh the costs, resulting in an increase in the amount of information a single operator can monitor to a satisfactory level.

A system using these techniques could be developed and performance on such a system directly compared to performance on current systems. However, there are many distinct issues involved in the use of such a complex system, each of which deserve scientific exploration. It is hoped that by performing basic research in these areas prior to developing a system to be used in communications settings, knowledge may be obtained that can help direct an efficient and effective system design.

The design of the first in a series of experiments was discussed in detail and several issues were discussed which may be the focus of future experimentation. It is clear that this line of research could follow a number of possible paths and careful thought will be required to plan a course of research that will make efficient use of resources and time, while answering as many questions as possible.

6. REFERENCES

[1] Daniel Wallace and Christine Schlichting, *Report on the Communications Research Initiatives in Support of Integrated Command Environment (ICE) Systems,* Jan. 2002.

[2] G. S. Kang and D. Brock, “Improved monitoring of multi-channel, time-overlapped, naval tactical voice messages” Naval Research Laboratory, Washington, DC, Tech. rep. (in preperation)

[3] George S. Kang and Lawrence J. Fransen, *Speech Analysis and Synthesis Based on Pitch-Synchronous Segmentation of the Speech Waveform,* Nov. 1994

[4] William David Garvey, “The Intelligibility of Speeded Speech”, *Journal of Experimental Psychology*, vol. 45, no. 2, pp. 102–108, 1953.

[5] Durand R. Begault, “Virtual Acoustics, Aeronautics and Communications”, Nov. 1996