DESIGN CONSIDERATIONS FOR IMPROVING THE EFFECTIVENESS OF MULTITALKER SPEECH DISPLAYS

Douglas S. Brungart
Mark A. Ericson

Air Force Research Laboratory
Human Effectiveness Directorate
Wright-Patterson AFB, Ohio, USA
douglas.brungart@wpafb.af.mil

Brian D. Simpson

Veridian
Dayton, Ohio, USA
brian.simpson@wpafb.af.mil

ABSTRACT

Although many researchers have commented on the potential of audio display technology to improve intelligibility in multitalker speech communication tasks, no consensus has been reached on how to design an "optimal" multitalker speech display. This paper reviews a set of experiments that used a consistent procedure to evaluate the impact of six different parameters on overall intelligibility in multitalker speech displays: 1) the signal-to-noise ratio; 2) the number of competing talkers; 3) the voice characteristics of the talkers; 4) the relative levels of the talkers; 5) the apparent locations of the talkers; and 6) the listener's *a priori* information about the listening task. The results are discussed in terms of their implications in the design of more effective multitalker speech displays.

1. INTRODUCTION

Many critically important tasks require listeners to monitor and respond to speech messages originating from two or more competing talkers. Air traffic controllers, emergency service dispatchers, and military commanders are just a few examples of personnel who routinely make life and death decisions on the basis of information they receive through multichannel communications systems. Although a number of researchers have commented on the substantial benefits that audio display technology can provide in these multitalker communications tasks, no consensus has been reached on how to design an "optimal" multitalker speech display. In part, at least, this lack of consensus is a result of the extreme complexity of the multichannel listening problem— performance in such tasks depends on a wide variety of factors, including: 1) the ambient noise in the communications system; 2) the number of competing talkers; 3) the voice characteristics of the talkers; 4) the relative levels of the talkers; 5) the apparent locations of the talkers; and 6) the listener's *a priori* knowledge about the listening task. A further complicating issue is the variety of methodologies that have been used to examine these factors: procedural variations often make it difficult to compare the results of different multitalker listening experiments. In this paper, we review a series of experiments that used the same speech intelligiblity test (the Coordinate Response Measure or CRM) to examine the impact of six different audio display design parameters on overall intelligibility in a multichannel communications task. This allows a consistent comparison of the relative importance of each of these factors that can be used as a guide in the design of more effective multitalker speech displays.

2. EXPERIMENTAL METHODOLOGY: THE COORDINATE RESPONSE MEASURE

All of the experiments described in this paper were conducted using the Coordinate Response Measure (CRM). This speech intelligibility test was originally developed to provide greater operational validity for military communications tasks than standard speech intelligibility tests based on phonetically balanced words [1]. In the CRM task, a listener hears one or more simultaneous phrases of the form "Ready, (Call Sign), go to (color) (number) now" with one of eight call signs ("Baron," "Charlie," "Ringo," "Eagle," "Arrow," "Hopper," "Tiger," and "Laker"), one of four colors (red, blue, green, white), and one of eight numbers (1-8). The listener's task is to listen for the target sentence containing their pre-assigned call sign (usually "Baron") and respond by identifying the color and number coordinates contained in that target phrase.

Although the CRM was originally intended to measure speech intelligibility with a noise masker, its call-sign-based structure makes it ideal for use in multitalker listening tasks. The embedded call sign is the only feature that distinguishes the target phrase from the masking phrases, so the listener is forced to listen to the content of all the simultaneous phrases in order to determine which phrase contains the most relevant information. In this regard, it is similar to many command and control tasks where operators are required to monitor multiple simultaneous channels for important information that may originate from any channel in the system.

The experiments described in this paper were conducted using the corpus of CRM speech materials that has been made publically available in CD-ROM format by researchers at the Air Force Research Laboratory [2]. This CRM corpus contains all 256 possible CRM phrases (8 call signs X 4 colors X 8 numbers) spoken by eight different talkers (four male, four female). The experiments described in the following sections were conducted using this corpus. In all cases, the stimulus consisted of a combination of a target phrase, which was randomly selected from all of the phrases in the corpus with the call sign "Baron," and one or more masking phrases, which were randomly selected from the phrases in the corpus with call signs, colors, and numbers that differed from those used in the target phrase. These stimuli were presented over headphones at a comfortable listening level (approximately 70 dB SPL), and the listener's responses were collected either by using the computer mouse to select the appropriately colored number from a matrix of colored numbers on the CRT or by pressing an appropriately marked key on a standard computer keyboard. Each of the following sections discusses a different factor that influences

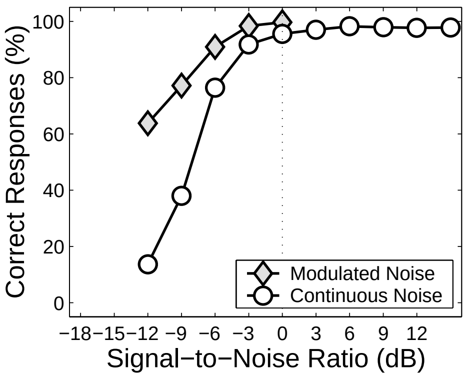

Figure 1: Percentage of correct color and number identifications for a CRM target phrase masked by a continuous or modulated speech-shaped noise signal. The error bars represent 95% confidence intervals. Adapted from Brungart, Simpson, Ericson, and Scott (2001).

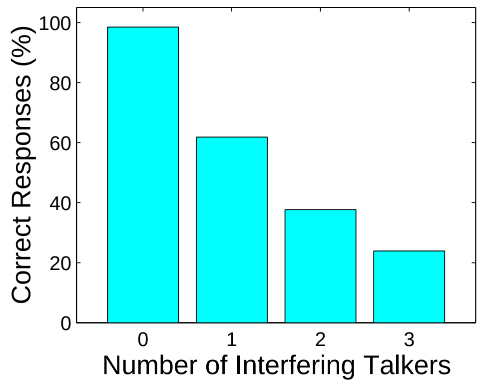

Figure 2: Percentage of correct color and number identifications for a CRM target phrase masked by 0, 1, 2, or 3 simultaneous same-sex masking phrases. All of the competing talkers were presented diotically at the same level. Adapted from Brungart, Simpson, Ericson, and Scott (2001).

speech intelligibility in a multitalker listening environment.

3. SIGNAL-TO-NOISE RATIO

One factor that influences the performance of any audio display is the signal-to-noise ratio of the output signal. In the case of a speech display based on radio communications, this noise could originate at any of three points in the transmission path: 1) ambient noise in the environment of the talker that is picked up by the microphone that records the talker's voice; 2) electronic noise or distortion in the transmission system (radio or wired); or 3) ambient noise in the environment of the listener.

The effects of signal-to-noise ratio (SNR) on speech perception are well documented, and, in many cases, it is possible to use the Articulation Index (AI) or the Speech Transmission Index (STI) to make a quantitative prediction of speech intelligibility directly from the acoustic properties of the noise and speech signals [3, 4]. In general, the sensitivity of speech intelligibility to the SNR depends on the phonetic structure, vocabulary size, and context of the speech signal. Although the CRM phrases provide no contextual information (it is impossible to predict the color or number in a CRM phrase from any of the other words in the phrase), they are limited to a small vocabulary of colors and numbers. This allows listeners to perform well in the CRM task even at very low SNRs. Figure 1 shows performance in the CRM as a function of SNR (calculated for each stimulus as the ratio of the RMS level measured across the entire individual speech utterance in the stimulus to the long-term RMS level of the individual noise sample in the stimulus) for a continuous speech-shaped noise (circles) and for a speech-shaped noise that has been modulated to match the envelope of a speech signal from the CRM corpus (diamonds). In each case, both the target speech and the noise were presented diotically.[1] The results show that performance in the CRM task is nearly perfect in continuous noise when the SNR is 0 dB or higher, and that performance with a noise masker that is modulated to match the amplitude variations that occur in speech is reasonably good ($> 80\%$) even at an SNR of -6 dB. It should be noted, however, that these surprisingly good results are a direct result of the small vocabulary size in the CRM corpus — the most demanding speech materials (nonsense syllables) require an SNR of approximately +20 dB in the speech band (200 Hz - 6100 Hz) to achieve $\geq 99\%$ performance [3]. Thus, an ideal multitalker speech display should be able to achieve an SNR of +20 dB in the frequency range from 200 Hz to 6100 Hz (measured from the overall RMS levels of the speech and noise signals).[2]

4. NUMBER OF COMPETING TALKERS

One of the more obvious factors that can affect the performance of a multitalker speech display is the number of competing talkers in the stimulus. As a general rule, performance in a multitalker listening task decreases when the number of talkers increases. Figure 2 shows how performance in the CRM task changes as the number of interfering talkers increases from 0 to 3. The data are shown for different same-sex talkers who were speaking at the same level in a single-channel signal that was presented diotically over headphones [5]. When no competing talkers were present in the stimulus, performance was near 100%. The first competing talker reduced performance by a factor of approximately 0.6, to 62% correct responses. Adding the second competing talker reduced performance by another factor of 0.6, to 38% correct responses. Adding the third competing talker reduced performance by another factor of 0.6, to 24% correct responses. Thus we see that CRM performance in a monaural or diotic multitalker speech display decreases by approximately 40% for each of the first three same-sex talkers added to the stimulus.

These results clearly show that it is advantageous to reduce the number of simultaneous talkers in a multitalker speech display whenever it is practical to do so. Possible ways to achieve

[1] In a diotic presentation, the same audio signal is presented simultaneously to both ears.

[2] It should be noted that the relative importance of each frequency range to speech intelligibility has been thoroughly documented in the literature on Articulation Theory. This information is invaluable when tradeoffs between bandwidth and SNR become necessary in the design of communications systems.

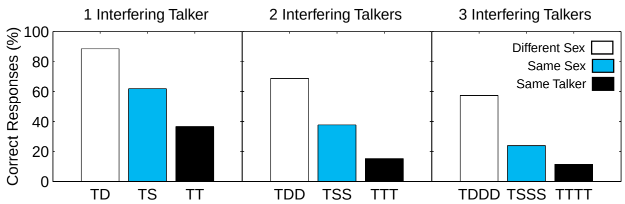

Figure 3: Percentage of correct color and number identifications for a CRM target phrase masked by 1, 2, or 3 simultaneous masking phrases. The white bars show performance with masking talkers who were different in sex than the target talker (the TD condition). The gray bars show performance with different masking talkers who were the same sex as the target talker (the TS condition). The black bars show performance when the target and masking phrases were all spoken by the same talker (the TT condition). All of the competing talkers were presented diotically at the same level. Adapted from Brungart, Simpson, Ericson, and Scott (2001).

this reduction range from simple protocols that reduce the chances of overlapping speech signals on a radio channel (such as marking the end of each transmission with a terminator like "over,") to systems that allow only one talker to speak on a radio channel at any given time, to sophisticated systems that queue incoming messages that overlap in time and play them back to the listener sequentially. However, none of these solutions is appropriate for complex listening situations where a single communication channel is in near-constant use by two or more simultaneous talkers or where a listener has to monitor two or more communications channels for time-critical information that might occur on any channel. For these situations, the designers of speech displays must rely on other cues to help users segregate the competing speech messages.

5. VOICE CHARACTERISTICS

Differences in voice characteristics provide one important cue that can be used to segregate competing speech signals. The voices of different talkers can vary in a wide variety of ways, including differences in fundamental frequency (F0), formant frequencies, speaking rate, accent, and intonation. Talkers who are different in sex are particularly easy to distinguish, because female talkers generally have substantially higher F0 frequencies and substantially shorter vocal tracts than male talkers.

Figure 3 illustrates the effect that differences in voice characteristics can have on a listener's ability to segregate a target speech signal from one, two, or three interfering talkers. The white bars show performance when the interfering talkers were different in sex than the target talker. The gray bars show performance when the masking phrases were spoken by different talkers who were the same sex as the target talker. The black bars show performance when the target and masking phrases were all spoken by the same talker. In all cases, performance was best when the interfering talkers were different in sex than the target talker, and worst when all the phrases were spoken by the same talker.

In situations where it is possible to control the voice characteristics of the competing talkers in a multitalker speech display, the characteristics of the competing voices should be made as different as possible. One example of a situation where this should be relatively easy to accomplish is in the use of computer-generated voice icons in an audio display. Consider, for example, a cockpit display where one voice icon might be used to indicate an engine fire and another might be used to indicate a terrain warning. Because the relative priority of these two warnings can vary with the situation, both of these warnings must be presented to the pilot as soon as they occur. If the two warnings are pre-recorded in both male and female voices, the display system can act to ensure that the two warnings are spoken by different-sex talkers. This would make it easier for the pilot to selectively attend to the warning with greater immediate relevance.

In audio displays that are designed to present externally generated voice communications rather than internally generated audio icons, it is much more difficult to control the vocal characteristics of the competing talkers. One possible option is to perform some kind of real-time or near-real-time audio processing on the different competing voice signals to make them more distinct. It may be possible to achieve this result by manipulating the parameters used to reconstruct the voice in communication systems that use low-bandwidth parametric vocoders. For example, the fundamental frequencies (F0s) of the two talkers could be manipulated to introduce a difference between the two competing talkers in real time. Assman and Summerfield [6] have shown that a difference in F0 of 1/6th of one octave is sufficient to produce a significant improvement in intelligiblity. However, this approach also has a major drawback: it may make it substantially more difficult (or impossible) for the listener to use voice characteristics to determine the identity of the talker. Thus the segregation efficiency that is gained by introducing differences in voice characteristics may be more than offset by the reduction in a listener's ability to correctly identify the target talker. A good rule of thumb might be to restrict the use of voice modification to situations in which speaker identification is not important and avoid the use of voice modification when accurate speaker identification is critical.

6. TARGET-TO-MASKER RATIO

Another factor that has a strong influence on a listener's ability to segregate competing speech signals is the level of the target talker relative to the competing talkers. In general, it is much easier to attend to the louder talker in a multitalker stimulus than to the quieter talker in a multitalker stimulus. This is illustrated in Figure 4, which shows performance as a function of the target-to-masker ratio (TMR) for 1, 2, or 3 interfering talkers. In this context, TMR is the ratio of the overall RMS level of the target talker to the overall RMS level of each of the interfering talkers in the stimulus. Thus, when the TMR is 0 dB, all of the talkers in the stimulus are presented at the same level. The results in Figure 4 show that performance is substantially improved when the target talker is the most intense talker in the stimulus (TMR$>$0 dB).

Clearly a substantial improvement in speech intelligibility can be achieved by increasing the level of the target talker relative to the levels of the other talkers in the stimulus. Unfortunately, this also degrades the intelligibility of the other talkers in the stimulus. Because it is usually difficult or impossible for the audio display designer to identify the target talker in the stimulus, there is no way to automatically determine which talker should be amplified relative the the others. One alternative approach is to allow the listener to adjust the relative levels of the talkers and thus increase the level of the talker who is believed to be the most important in the

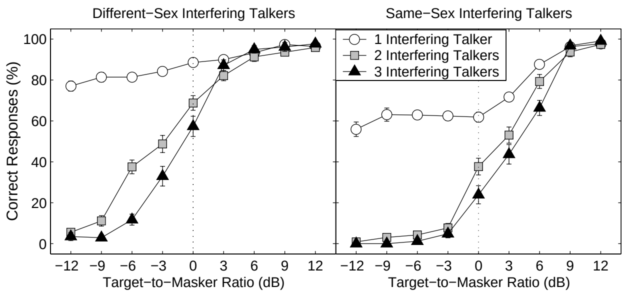

Figure 4: Percentage of correct color and number identifications for a CRM target phrase masked by 1, 2, or 3 interfering talkers. The results are shown as a function of the target-to-masker ratio (TMR), which is the ratio of the level of the target talker to the levels of the other interfering talkers in the stimulus (note that all the interfering talkers were presented at the same level). The left panel shows performance with different-sex interfering talkers; the right panel shows performance with same-sex interfering talkers. The error bars show the 95% confidence intervals of each data point. Adapted from Brungart, Simpson, Ericson, and Scott (2001).

current listening situation [7]. This ability is provided by current multichannel radio systems, which typically have adjustable level knobs for each radio channel. It should be noted, however, that a potential drawback of this approach is that the listener will miss crucial information that is spoken by one of the low-level talkers in the stimulus: the data in Figure 4 show that performance decreases rapidly with TMR when there are two or more interfering talkers and that listeners essentially receive no information from the low-level talkers when the TMR falls below -6 dB.

The data for the situation with one same-sex interfering talker (open circles in the right panel of Figure 4) have some interesting implications for the design of two-channel communications systems. In this condition, listeners were apparently able to selectively attend to the quieter talker in the stimulus. Consequently, performance in this condition did not decline when the TMR was reduced below 0 dB. Performance did, however, improve rapidly when the TMR was increased above 0 dB. Although one might expect that two equally important communications channels should be presented at the same level, the data in Figure 4 suggest that this is a poor strategy. When a level difference is introduced between the two channels, performance improves substantially when the target talker occurs on the louder channel, but is unaffected when the target talker occurs on the quieter channel. Thus, overall performance in the CRM task improves substantially when the speech stimuli are presented at levels that differ by 3-9 dB. These data are also consistent with results of a previous experiment that examined performance as a function of TMR with a different call-sign-based task [8]. Note, however, that this approach may not be a useful strategy in noisy environments, where the less-intense talker may be masked by ambient noise. Further investigation is needed to explore this in more detail.

7. SPATIAL SEPARATION

To this point, our discussion has been restricted to factors that influence the performance of monaural or diotic speech displays,

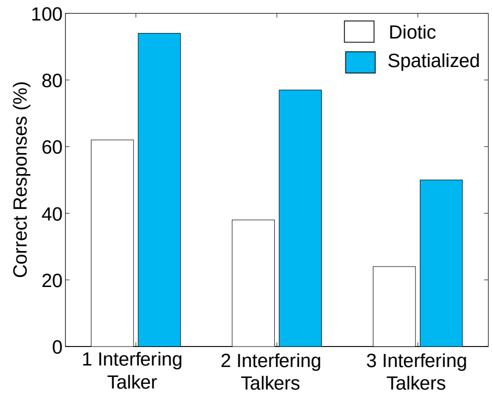

Figure 5: Percentage of correct color and number identifications for a CRM target phrase masked by 1, 2, or 3 same-sex interfering talkers. The white bars show results for a diotic condition where the competing talkers were not spatially separated (adapted from Brungart, Simpson, Ericson, and Scott (2001)). The gray bars show performance where the competing talkers were spatially separated by 45° (talkers at 0° and 45° with one interfering talker; -45°, 0° and 45° with two interfering talkers; and -45°, 0°, 45° and 90° with three interfering talkers).

where all of the competing talkers are mixed together into a single audio channel prior to presentation to the listener. When it is possible to use stereo headphones to present a binaural audio signal to the listener, substantial performance improvements can be achieved by using a virtual audio display to spatially separate the apparent locations of the competing sounds [9, 10, 11, 12]. Figure 5 shows the effect of spatial separation on performance with one, two, or three same-sex interfering talkers. In the case with one interfering talker, spatial separation increased performance by approximately 25 percentage points. In the cases with two or three interfering talkers, spatial separation of the talkers nearly doubled the percentage of correct responses. Clearly spatial separation in azimuth is a powerful tool for improving the performance of multitalker audio displays.

The advantages of spatial separation can be even more pronounced in a noisy environment. Figure 6 shows the effect of spatial separation with one same-sex competing talker as a function of the amount of ambient noise in the environment. These data are taken from an experiment where the CRM phrases were spoken by live talkers wearing oxygen masks and heard by listeners wearing ANR headsets [10]. The results show that the advantages of spatial separation were greatest when the listeners were subjected to an ambient noise field of 110 dB. This should be taken into consideration in the design of displays for use in noisy environments.

The advantages of spatial separation are not limited to direction. Recent experiments have shown that substantial improvements in performance can also be achieved by spatially separating nearby talkers in distance in the near-field region where listeners can use binaural cues to determine the distances of sound sources. In an experiment that used phrases from the CRM corpus that were spatially processed with near-field HRTFs measured at 90° in azimuth, the percentage of correct color and number identifications increased from 55% when the target and masking talkers were presented at the same distance (1 m) to more than 80% when one

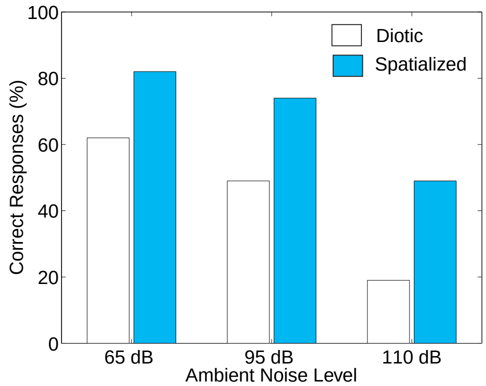

Figure 6: Percentage of correct color and number identifications for a CRM target phrase masked by one same-sex interfering talker. The white bars show results for a diotic condition where the competing talkers were not spatially separated. The gray bars show performance where the competing talkers were spatially separated by 45°. Note that the target and masking phrases were spoken by live talkers wearing an aviation helmet and oxygen mask, and that these signals were passed through a military intercom system before being spatially processed and presented to the listeners over Bose AH-A ANR headsets. Adapted from Ericson and McKinley (1997).

talker was presented at 1 m and the second talker was presented at 12 cm [13].

8. A PRIORI INFORMATION

A final factor that influences performance in multitalker speech displays is the amount of information the listener has about "where to listen" for the target information. In general, listeners do better when they know who the target talker is or where the target talker is than when they have no *a priori* information about the location or identity of the target talker. This effect is illustrated in Figure 7, which shows performance with one, two, or three spatially-separated same-sex interfering talkers in one of three different listening configurations. The white bars show a condition where the target talkers and target locations were chosen randomly and the listeners had no information about the location of the target talker. The gray bars show a condition where the listeners knew who the target talker was but the location of the target talker was randomized across trials. The black bars show a condition where the listeners knew who the target talker was and where the target talker was located prior to each block of trials. The results show that overall performance increased systematically as the listeners were provided with more information about who and where to listen for the target phrase. When the stimulus contained two or more interfering talkers, overall performance was approximately 20 percentage points higher when the listeners knew the identity and the location of the target talker in advance than when they had no information about the target talker.

Although these results demonstrate that listeners perform better in multichannel communications tasks when they know who the target talker is and where the target talker is, it is probably not practical to take advantage of this performance improvement in the

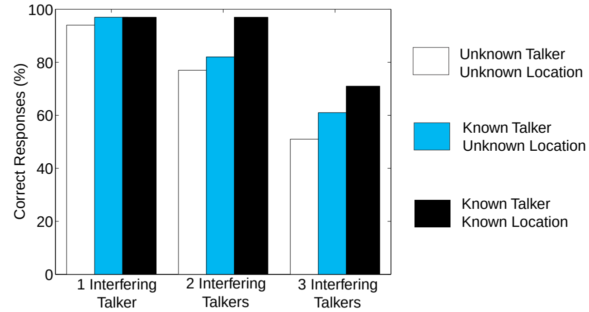

Figure 7: Percentage of correct color and number identifications for a CRM target phrase masked by 1, 2, or 3 same-sex interfering talkers. In each case, the talkers were spatially separated by 45° (same configurations as in Figure 5). The white bars show results for a condition where the listeners had no information about the location or identity of the target talker. The gray bars show a condition where they knew who the target talker was but not his or her location. The black bars show a condition where both the identity and location of the target talker were known in advance.

design of multitalker speech displays. Care should, however, be taken in the evaluation of multitalker speech displays to accurately model the amount of *a priori* information that will be available to the eventual end user of the system. In some operational tasks, the target talker will change frequently and the listener must always monitor all the channels vigilantly. In other tasks, the listener will engage in a conversation with a single talker for a long period of time before switching attention to one of the other channels of the system. Failure to account for the differences in these two situations may prevent an accurate assessment of the true operational effectiveness of the system.

9. CONCLUSIONS AND RECOMMENDATIONS

The most efficient way to improve the effectiveness of a multitalker speech display is to use virtual synthesis techniques to spatially separate the locations of the competing talkers. The data from Figure 5 show that spatially separating same-sex competing talkers by 45° produced a 25-35 percentage point increase in overall performance in the CRM task. In terms of the other factors examined in this paper, this is roughly equivalent to: 1) reducing the number of competing talkers in the stimulus by 1 to 1.5 talkers (Figure 2); 2) replacing the same-sex interfering talkers with different-sex interfering talkers (Figure 3); or 3) increasing the target-to-masker ratio by 3-9 dB (Figure 4).

However, spatial separation has substantial advantages over these other techniques. The biggest advantage is that spatial separation improves the intelligibility of all the talkers in the stimulus roughly equally, while the other techniques tend to increase the intelligibility of only one of a few selected talkers. Reducing the number of talkers in the stimulus increases the intelligibility of the remaining talkers at the expense of losing all the information from the eliminated talker. Replacing the same-sex interfering talkers with different-sex talkers provides a benefit only for the talker who is different in sex from the other talkers in the stimulus. Increasing the target-to-masker ratio increases the intelligibility of one talker but generally reduces the intelligibility of the other talkers in the stimulus. Only spatial separation is able to improve overall performance across all the talkers in a three- to four-talker stimulus.

Spatial separation is also relatively inexpensive to implement in multitalker speech displays. Many of the benefits of spatially separating speech signals can be obtained with low-cost digital signal processing techniques that simply introduce interaural time differences [14] and interaural level differences [15] into the different communications channels of the system. The listener-specific pinna-related spectral details that are required to produce realistic, localizable, externalized virtual sounds in non-speech virtual displays [16] provide little additional benefit to speech intelligibility in multitalker listening tasks for presentation in azimuth [17, 9]. Similarly, real-time head-tracking devices are not required to achieve good intelligibility in multitalker speech displays (the data shown in Figure 5 were collected without any head tracking). If a communications system or intercom is capable of processing audio signals in the digital domain, it may be possible to implement an effective speech segregation algorithm in software for little or no additional cost. The only restriction is that the system must be capable of producing a stereo output signal: no spatialization is possible in a system with only one analog output channel.

Although this paper has reviewed many of the factors that can influence the performance of a multitalker speech display, it has by no means explored all of these issues. Further investigation is needed to determine how the different display techniques outlined in this paper interact with one another. More research is needed to determine the optimal locations of the talkers in a spatialized speech display: most researchers have placed the competing talkers at evenly spaced locations in azimuth, but no systematic studies have been conducted to determine if this placement is ideal. Other factors, such as the effect of talker motion on speech segregation or the benefits that can be obtained by adding real-time head tracking to a multitalker speech display, also require further exploration. Finally, greater efforts must be made to determine how multitalker displays can be tailored for the specific communication tasks they are designed to address. Communication tasks can vary widely in terms of vocabulary size, speech syntax, and available contextual information. Communication tasks can also vary in terms of how frequently the listener is required to switch attention across the different competing talkers, and in terms of the non-speech tasks listeners are required to perform concurrently with the communication task [18]. At this point, most research in multitalker speech displays has been focused on "general-purpose" communications tasks. New techniques are needed to develop and test speech displays for more specific applications. Only when these issues are resolved will it be possible to begin converging on a series of protocols for designing truly "optimal" multitalker speech displays.

10. REFERENCES

[1] T.J. Moore, "Voice communication jamming research," in *AGARD Conference Proceedings 331: Aural Communication in Aviation*. 1981, pp. 2:1–2:6, Neuilly-Sur-Seine, France.

[2] R.S. Bolia, W.T. Nelson, M.A. Ericson, and B.D. Simpson, "A speech corpus for multitalker communications research," *Journal of the Acoustical Society of America*, vol. 107, pp. 1065–1066, 2000.

[3] K.D. Kryter, "Methods for calculation and use of the articulation index," *Journal of the Acoustical Society of America*, vol. 34, pp. 1689–1697, 1962.

[4] H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *Journal of the Acoustical Society of America*, vol. 67, pp. 318–326, 1980.

[5] D.S. Brungart, B.D. Simpson, M.A. Ericson, and K.R. Scott, "Informational and energetic masking effects in the perception of multiple simultaneous talkers," *Journal of the Acoustical Society of America*, vol. 110, pp. 2527–2538, 2001.

[6] P. F. Assman and Q. Summerfield, "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," *Journal of the Acoustical Society of America*, vol. 88, pp. 680–697, 1990.

[7] W. Spieth, J.F. Curtis, and J.C. Webster, "Responding to one of two simultaneous messages," *Journal of the Acoustical Society of America*, vol. 26, pp. 391–396, 1954.

[8] J. Egan, E. Carterette, and E. Thwing, "Factors affecting multi-channel listening," *Journal of the Acoustical Society of America*, vol. 26, pp. 774–782, 1954.

[9] W. T. Nelson, R. S. Bolia, M. A. Ericson, and R. L. McKinley, "Spatial audio displays for speech communication. a comparison of free-field and virtual sources," *Proceedings of the 43rd Meeting of the Human Factors and Ergonomics Society*, pp. 1202–1205, 1999.

[10] M.A. Ericson and R.L. McKinley, "The intelligibility of multiple talkers spatially separated in noise," *in Binaural and Spatial Hearing in Real and Virtual Environments*, 1997, Edited by R.H. Gilkey and T.R. Anderson (Erlbaum, Hillsdale N.J.), pp. 701-724.

[11] K. Crispien and T. Ehrenberg, "Evaluation of the 'cocktail party effect' for multiple speech stimuli within a spatial audio display," *Journal of the Audio Engineering Society*, vol. 43, pp. 932–940, 1995.

[12] K.S. Abouchacra, T.V. Tran, J. Besing, and J. Koehnke, "Performance on a selective attention task as a function of stimulus presentation mode," *in Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, Florida*, 1997.

[13] D.S. Brungart, "Optimizing multitalker speech displays with near-field hrtfs," in *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001*, 2001, pp. 169–174.

[14] R. Carhart, T.W. Tillman, and K.R. Johnson, "Release of masking for speech through interaural time delay," *Journal of the Acoustical Society of America*, vol. 42, pp. 124–138, 1967.

[15] A.W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level difference on speech intelligibility in noise," *Journal of the Acoustical Society of America*, vol. 83, pp. 1508–1516, 1988.

[16] E.M. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman, "Localization using non-individualized head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 94, pp. 111–123, 1993.

[17] R. Drullman and A.W. Bronkhorst, "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," *Journal of the Acoustical Society of America*, vol. 107, pp. 2224–2235, 2000.

[18] P. Tun and A. Wingfield, "Speech recall under heavy load conditions: Age, predictability and limits on dual-task interference," *Aging, Neuroscience, and Cognition*, vol. 1, pp. 29–44, 1994.