DISTANCE-BASED SPEECH SEGREGATION IN NEAR-FIELD VIRTUAL AUDIO DISPLAYS

Douglas S. Brungart

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

In tasks that require listeners to monitor two or more simultaneous talkers, substantial performance benefits can be achieved by spatially separating the competing speech messages with a virtual audio display. Although the advantages of spatial separation in azimuth are well documented, little is known about the performance benefits that can be achieved when competing speech signals are presented at different distances in the near field. In this experiment, head-related transfer functions (HRTFs) measured with a KEMAR manikin were used to simulate competing sound sources at distances ranging from 12 cm to 1 m along the interaural axis of the listener. One of the sound sources (the target) was a phrase from the Coordinate Response Measure (CRM) speech corpus, and the other sound source (the masker) was either a competing speech phrase from the CRM speech corpus or a speech-shaped noise signal. When speech-shaped noise was used as the masker, the intelligibility of the target phrase increased substantially only when the spatial separation in distance resulted in an improvement in signal-to-noise ratio (SNR) at one of the two ears. When a competing speech phrase was used as the masker, spatial separation in distance resulted in substantial improvements in the intelligibility of the target phrase even when the overall levels of the signals were normalized to eliminate any SNR advantages in the better ear, suggesting that binaural processing plays an important role in the segregation of competing speech messages in the near field. The results have important implications for the design of audio displays with multiple speech communication channels.

1. INTRODUCTION

Many critically important occupational tasks require listeners to monitor and respond to speech messages originating from two or more simultaneous competing talkers. Examples of these tasks are commonplace in air traffic control towers, military command and control centers, and emergency-service radio dispatch centers. Previous research has shown that these multitalker listening tasks become much easier when a virtual audio display is used to spatially separate the apparent azimuthal locations of the competing talkers [1, 2].

The spatial separation cues provided by these virtual displays have two distinct advantages over conventional monaural communications systems. The first advantage is the additional information that is contained in the apparent spatial locations of the competing speech messages. This spatial information can be extremely useful for keeping track of the different competing talkers. For example, if all the pilots controlled by an air traffic controller are assigned unique angular positions in the audio display, the apparent locations of the incoming speech messages can be used to associate the messages with their originating aircraft. In more advanced audio displays, the apparent locations of the speech messages can be used to provide information about the actual location of the originating talkers. For example, speech signals from a pilot's wingman might originate from the location of the wingman's aircraft.

The second advantage provided by spatially separating the different voices in a multitalker audio display is a substantial improvement in the listener's ability to selectively attend to the most important talker in the stimulus. This improvement occurs because the display takes advantage of our natural ability to use binaural difference cues to segregate spatially separated talkers in real-world listening environments, which is often referred to as the "cocktail party" effect (see [3] for a recent review of the "cocktail party" phenomenon).

One aspect of the "cocktail party" phenomenon that has not yet been fully explored is the effect that spatial separation in distance has on a listener's ability to process multiple simultaneous sound sources. Nearly all previous studies that have examined the "cocktail party" phenomenon have focused on the angular separation of relatively distant speech signals, located 1 m or more away from the listener's head. In this far-field region, the interaural level difference (ILD) cues and interaural time difference (ITD) cues that listeners use to segregate speech signals depend only on the direction of the sound source, and not on its distance. Thus, there is no reason to believe that spatial separation in distance will have any meaningful effect on a listener's ability to selectively attend to competing speech signals when both sources are located at distances greater than 1 m. When a sound source is located near the head, however, the ILD is highly dependent on distance [4]. The ILD can increase by as much as 20-30 dB when the distance of a sound source at 90° azimuth is reduced from 1 m to 12 cm, while the ITD increases only slightly with decreasing distance in this region. Free-field localization experiments have shown that listeners are able to use these distant-dependent changes in the ILD to determine the distance of a nearby source when that source is located near the interaural axis [5]. Even when the overall levels of the stimuli were randomized, the distance judgments in these near-field experiments were highly correlated with the actual source distances ($r \approx 0.85$) when those sound sources were located near 90° in azimuth. However, little is known about the effects that these distant-dependent changes in ILD have on the segregation of speech signals presented at different distances in the near field.

To this point, the only study that has examined the effects of separation in distance on the processing of multiple sound sources has focused exclusively on the perception of spatially separated speech and noise sources in the near field [6]. The results of that experiment indicate that spectral differences at the ear with the better SNR can explain almost all of the benefits of spatially separating a nearby speech signal from a nearby noise masker. Binaural processing could only account for a 1-2 dB release from masking. However, recent studies have shown that binaural difference cues contribute substantially more to the spatial unmasking of speech when the masking sound is a competing speech signal than when the masking sound is a noise signal [7, 8]. It has been suggested that these differences occur because listeners derive a larger benefit from spatial separation in the apparent locations of sounds when “informational” masking forces them to disentangle two or more clearly audible but similar sounding signals than when traditional “energetic” masking overpowers the target sound and renders it inaudible at the periphery [7]. Since energetic and informational masking both play an important role in determining the intelligibility of competing speech messages, one would expect the informational component of speech on speech masking to produce a greater amount of spatial unmasking than would be predicted for a noise masker in the same target-masker configuration.

This paper describes an experiment that used near-field HRTFs measured with a KEMAR acoustic manikin to examine the perception of a speech signal masked by an interfering speech or noise signal when the two sources were located at different distances along the listener’s interaural axis. The results are discussed in terms of their applications in the design of multitalker speech displays.

2. METHODS

2.1. Listeners

A total of nine paid listeners, five male and four female, participated in the experiment. All had normal hearing (15 dB HL from 500 Hz to 6 kHz), and their ages ranged from 21-55. All of the listeners had participated in previous experiments that utilized the speech materials used in this study.

2.2. Stimuli

2.2.1. Speech Materials

The speech stimuli were taken from the publicly available Coordinate Response Measure (CRM) speech corpus for multitalker communications research [9]. This corpus consists of phrases of the form “Ready (call sign) go to (color) (number) now” spoken with all possible combinations of eight call signs (“arrow,” “baron,” “charlie,” “eagle,” “hopper,” “laker,” “ringo,” “tiger”), four colors (“blue,” “green,” “red,” “white”), and eight numbers (1-8). Thus, a typical utterance in the corpus would be “Ready baron go to blue five now.” Eight talkers (four male, four female) were used to record each of the 256 possible phrases, so a total of 2048 phrases are available in the corpus. The sentences in the corpus were resampled to 25 kHz prior to their use in this study.

2.2.2. Speech-Shaped Noise

In some trials, a speech-shaped noise signal was used as the masker. The spectrum of this noise masker was determined by averaging the log-magnitude spectra of all of the phrases in the CRM corpus.

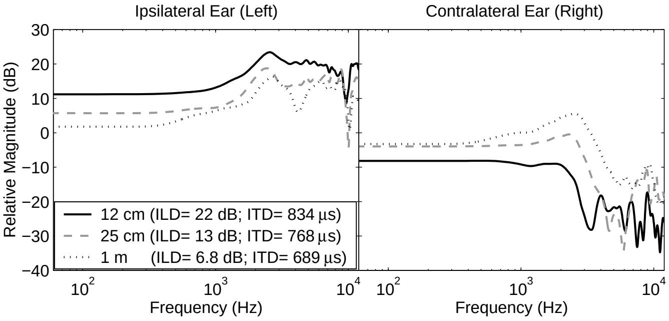

Figure 1: These curves show the frequency responses of the HRTF filters used to spatially process the stimuli used in the experiment. The headphone response corrections described in the text have been removed from these plots, so they represent the frequency responses of the raw HRTFs measured directly from the KEMAR manikin (as described in Brungart and Rabinowitz, 1999). The numbers in the legend show the average interaural level difference (ILD) (measured from overall RMS power for a speech-shaped noise stimulus) and the interaural time delay (ITD) (implemented with a linear phase delay in the HRTF for the contralateral ear) for each stimulus distance used in the experiment. Note that in each case the HRTF has been normalized to the sound pressure level that would occur at the location of the center of the head if the manikin’s head were removed.

This average spectrum was used to construct a 129-point Finite Impulse Response (FIR) filter that was used to shape Gaussian noise to match the average spectrum of the speech signals.

2.2.3. Stimulus Spatialization

Virtual synthesis techniques were used to control the locations of the stimuli in the experiment. The head-related transfer functions (HRTFs) used for this spatial processing were derived from an earlier set of HRTFs measured in the near field of a Knowles Electronics Manikin for Acoustic Research (KEMAR). These HRTFs, which are described in detail elsewhere [4], were measured at every degree in azimuth with an acoustic point source located 12 cm, 25 cm, and 1.0 m from the center of the manikin’s head. The overall effects of distance and the frequency characteristics of the point source were removed from these HRTFs by subtracting the free-field spectrum of the sound source (as measured by a single microphone placed at a location corresponding to the center of the manikin’s head) from the HRTFs measured at the manikin’s left and right ears. The HRTF measurements were made in the frequency domain and consisted of 600-point transfer functions with 32-Hz resolution from 100 Hz to 19.2 kHz. Figure 1 shows the frequency responses, ITD values, and ILD values of these HRTFs at the three source locations tested in this experiment.

The spatialization filters used in this experiment were derived directly from these HRTFs using the following procedure. First, the headphones used in the experiment (Sennheiser HD540) were placed on the KEMAR manikin and the same frequency-domain method used to measure the original HRTFs was used to measure the 600-point left- and right-ear transfer functions of the headphones. These transfer functions were subtracted from the raw HRTFs for the left and right ears in order to determine the desired transfer functions of the headphone-corrected HRTFs for each stimulus location. Then the MATLAB FIR2 command was used to generate 251-point, linear-phase FIR filters matching the frequency responses of the desired transfer functions over the frequency range from 100 Hz to 15 kHz at a 44.1 kHz sampling rate. These linear-phase filters were upsampled to a 1 MHz sampling rate in order to delay the contralateral-ear HRTF by the interaural time delay, which was determined from the average slope of the unwrapped phase of the original interaural HRTF over the frequency range from 160 to 1600 Hz. Finally, the HRTFs were downsampled to a 25 kHz sampling rate to efficiently accommodate the 8 kHz band-limited speech corpus used in this experiment. The resulting HRTFs were stored in a MATLAB file and directly convolved with the target and masker signals immediately prior to each stimulus presentation.

Although these KEMAR HRTFs do not capture the high - frequency, listener-specific detail that would be present in individualized HRTFs, they do accurately capture the distance- and direction-dependent changes in the low-frequency portion of the HRTF that are most likely to influence the spatial unmasking of speech. In the speech intelligibility tasks used in these experiments, it is reasonable to expect the manikin HRTFs to generate performance comparable to what would be achieved with individualized HRTFs or with free-field sound sources in an anechoic environment. Note that previous researchers who have compared multitalker speech intelligibility with virtual sources generated with non-individualized HRTFs to multitalker intelligibility with free-field sources [2] or virtual sources generated with individualized HRTFs [3] have reported no significant differences between the generic virtual sources and the more realistic source presentations.

2.2.4. Normalization

In real-world environments, the overall intensity of a stimulus varies with the distance of the source. Thus, if two equally intense speech signals were separated in distance, one would expect the closer speech signal to be substantially easier to comprehend simply because it would be more intense at the location of the listener; the contribution of binaural cues to the release from masking would be minimal relative to these distance-dependent intensity cues. Therefore, in order to examine the contribution of binaural cues and control for these distance-based intensity variations, the relative levels of the target and masker signals were adjusted in two different ways. In the *"center-of-the-head"* normalization condition (COH), the target and masker signals were convolved with the appropriate HRTFs and then scaled to make the SNR of the target signal 0 dB when measured from a microphone placed at a location corresponding to the center of the manikin's head (with the head removed from the sound field). In the *"better-ear"* normalization condition (BE), the target and masker were first convolved with the appropriate HRTFs. Then the SNR of the target signal was computed at each ear, and the filtered speech signals were scaled to make the SNR at the ear with the greater SNR (the *"better ear"*) equal to 0 dB. It is important to note that when the synthesized location of the target signal was far and that of the masker signal was near, the SNR at the ear on the opposite side of the head from the two signals was greater than the SNR at the ear on the same side of the head as the signals, and thus was the *"better ear."*

In the speech-shaped noise masker conditions, the masker level was increased by 9 dB after the normalization process in order to produce an SNR of -9 dB at the normalization point. This was done because previous speech-perception experiments in our laboratory have shown that performance with the CRM is most sensitive to changes in the relative level of a speech-shaped noise masker when the SNR of the target phrase is approximately -9 dB [10].

The signals were presented at a comfortable listening level (approximately 65 dB SPL on average) as measured at the output of the headphones, and the overall level of each stimulus presentation was randomly roved over a 6-dB range (in 1-dB steps).

2.2.5. Stimulus Configurations

All of the target and masker stimuli were presented along the interaural axis directly to the left of the listener. A total of three different target and masker configurations were used. The distance of the first signal, which was equally likely to be the target or the masker, was randomly selected from three possible distances: 12 cm, 25 cm, or 1 m. The distance of the second signal was always set to 1 m. Thus, the target could be at 12 cm or 25 cm with the masker at 1 m, the masker could be at 12 cm or 25 cm with the target at 1 m, or both the target and masker could be co-located at 1 m. Due to the logic used in the random selection process, twice as many trials were collected with the target and masker co-located at 1 m than in the other possible configurations.

2.3. Procedure

On each trial, the target phrase was selected randomly from the 256 phrases in the speech corpus with the call sign "Baron," with the restriction that each talker was used the same number of times in each listening session. In the trials with a speech masker, the masking phrase was selected randomly from the 1176 phrases in the speech corpus with a different call sign, a different color coordinate, and a different number coordinate than the target phrase. Note that the random selection of the target and masking phrases resulted in different-sex competing talkers in 50% of the trials, different competing talkers of the same sex in 37.5% of the trials, and competing phrases spoken by the same talker in 12.5% of the trials. In the trials with a noise masker, a random Gaussian noise was filtered with the speech-spectrum shaping noise filter and gated rectangularly to the beginning and end of each phrase. The normalization scheme (center-of-head or better-ear) was randomly chosen on each trial.

The data were collected with the listeners seated in front of the CRT of a Windows-based control computer. The stimuli for each trial were generated by an interactive MATLAB script, which selected the stimulus signals, processed the signals with the appropriate HRTFs, and presented the signals over headphones (Sennheiser HD- 540) through a Soundblaster AWE-64 sound card. The listeners were instructed to listen for the target phrase, which was always addressed to the call sign "Baron," and use the mouse to select the color and number contained in the target phrase from an array of colored digits displayed on the screen of the control computer. Each listener first participated in a total of 1560 trials with a speech masker. These trials were collected in 13 blocks of 120 trials each, with each block taking approximately 15 minutes to complete. Each listener then heard a total of 1000 trials with a speech-shaped noise masker. These trials were collected in 5 blocks of 200 trials each, with each block taking approximately 20 minutes to complete. One or two blocks were run per day for each listener over a period of several weeks. Note that some of the data were collected with normalization schemes or target-masker distance configurations that are not discussed in this paper, and that

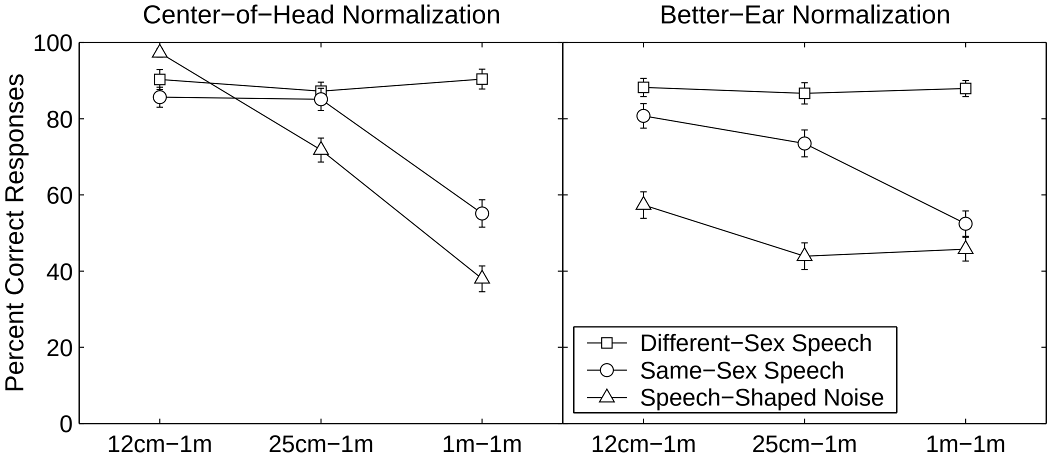

Figure 2: Percentage of correct color and number identifications for a CRM target phrase masked by a simultaneous speech or noise signal when both sources were presented directly to the left of the listener (90° azimuth). The left panel shows performance when the levels of the two signals were normalized to equalize the RMS power levels of the two signals at the location of the center of the listener's head. The right panel shows performance when the levels of the two signals were normalized to equalize the levels of the two signals at the ear with the greater SNR. Results are shown separately for each type of masker. The error bars represent 95% confidence intervals.

these points were excluded from the data analysis. Thus, the results that follow represent a total of 8,431 trials collected with the speech masker and 5,536 trials collected with the noise masker.

3. RESULTS

The overall results of the experiment indicate that the effects of distance separation on speech intelligibility are different for different masking signals. Figure 2 shows the percentage of trials where the listeners correctly identified both the color and the number used in the target phrase for three types of maskers: a speech phrase spoken by a talker who was different in sex than the target talker, a speech phrase spoken by a talker who was the same sex as the target talker, and a speech-shaped noise signal. When the target and masking phrases were spoken by different-sex talkers, spatial separation in distance had little or no impact on performance. The listeners correctly identified both the color and number coordinates in the target phrase in approximately 85% of the trials in all of the target and masker configurations tested. Apparently, the monaural cues that allow listeners to segregate different-sex talkers are so effective that no additional intelligibility advantage can be obtained by presenting the target and masking speech signals at different distances.

Spatial separation in distance had a much larger impact on performance when the masking phrase was spoken by a talker who was the same sex as the target talker. With both center-of-the-head (COH) and better-ear (BE) normalization, the percentage of correct identifications was approximately 30% greater in the 12 cm - 1 m configuration than in the 1 m - 1 m configuration. Overall performance in the spatially-separated same-sex conditions was approximately 8% better with COH normalization than with BE normalization (significant with p <0.001 in a two-tailed t-test). This improvement occurred because the signal-to-noise ratio at the better ear was, on average, 4.8 dB higher in the COH trials than it was in the BE trials.

When the target speech was masked by a speech-shaped noise masker, the benefits of spatial separation were much greater with COH normalization than with BE normalization. When the target and masker levels were normalized at the center of the head, the percentage of correct identifications increased from 40% in the 1 m - 1 m configuration to near 100% in the 12 cm - 1 m configuration. However, when the levels were normalized at the better ear, the percentage of correct identifications only improved to 60% in the 12 cm - 1 m configuration. These results suggest that speech intelligibility with a speech-shaped noise masker is more sensitive to monaural cues based on the SNR at the better ear and less sensitive to binaural processing based on distance-dependent changes in the ILD and ITD than speech intelligibility with a same-sex speech masker.

Figure 3 compares the better-ear normalized target-masker configurations measured in this experiment to the results of an earlier diotic experiment that used the same target and masker signals and the same panel of listeners [10]. As would be expected, performance in the co-located (1 m - 1 m) target-masker configurations of this experiment (the filled triangles in the figure) was essentially identical to performance in the corresponding diotic listening conditions with the same SNR at the better ear. Performance in the spatially separated conditions (the filled squares and circles in the figure) was consistently better than performance in the corresponding diotic listening conditions with the same SNRs at the better ear.

The data shown in Figure 3 can be used to quantify the contribution of binaural processing to the segregation of spatially-separated target and masking signals. A quantitative estimate of the "binaural advantage" can be obtained from the increase in SNR at the better ear that would be required to bring performance in a monaural or diotic presentation of the stimulus up to the level of performance achieved with a binaural presentation of the stimulus. By this definition, the binaural advantage of spatial separation in

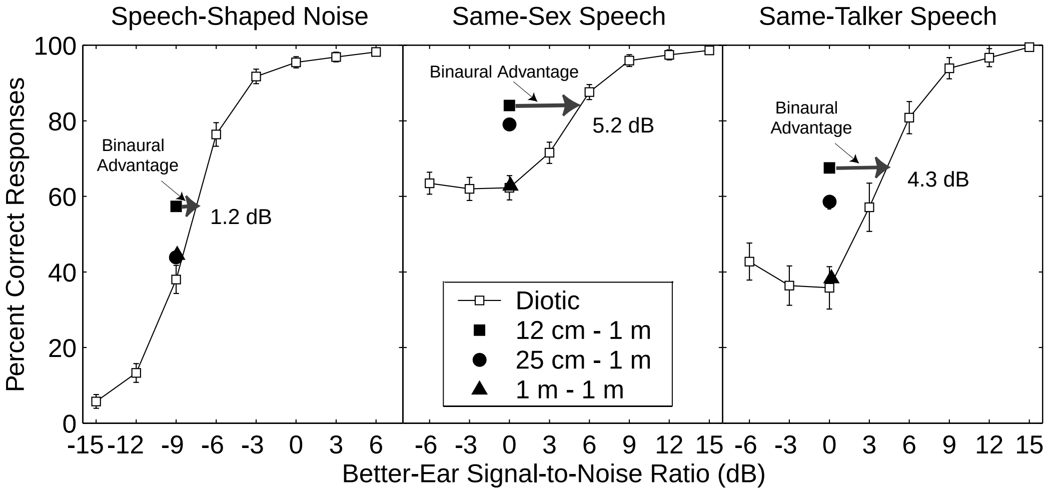

Figure 3: Binaural advantages of spatially separating the distances of the target and masking signals for a speech-shaped noise masker (left panel), a competing speech signal spoken by a talker who was the same sex as the target talker (center panel), and a competing speech signal spoken by the same talker used in the target phrase (right panel). The filled symbols show the percentage of correct color and number identifications for each target-masker configuration when the target and masking signals were normalized to equalize their levels at the ear with the better SNR. The open squares (diotic condition) show the percentages of correct responses for diotic presentations of the same target and masker signals as a function of overall SNR (Brungart, 2001). The arrows and numbers represent a decibel estimate of the binaural unmasking in the 12 cm - 1 m configuration for each type of masker. These estimates have been derived from the increase in SNR that would improve performance in the diotic condition to the same level that occurred in the 12 cm - 1 m condition with the same SNR at the better ear. The error bars represent 95% confidence intervals.

the 12 cm - 1 m configuration was 1.2 dB with a speech-shaped noise masker, 5.2 dB with a same-sex speech-masker, and 4.3 dB with a same-talker speech masker.

These results demonstrate that the binaural difference cues associated with spatial separation in distance produce a much larger improvement in intelligibility for speech-on-speech masking than they do for speech-on-noise masking. Note that the 1.2 dB binaural advantage in the speech-on-noise masking condition was roughly comparable to the 1-2 dB binaural advantage reported in an earlier study examining the binaural advantages of spatial separation in distance with a speech target and a noise masker in the near field [6]. The substantially larger binaural advantages that were found in the speech-on-speech masking conditions of this experiment may have occurred because differences in the perceived distances of sounds tend to produce a larger release in informational masking than in energetic masking [7]. Informational masking effects are known to dominate the 2-talker CRM response task used in the speech-on-speech masking conditions of this experiment [10], so the larger binaural advantages that occurred in these conditions are consistent with the hypothesis that informational masking is more susceptible than energetic masking to differences in the apparent locations of the target and masking sounds. Note that somewhat different results might be achieved with speech materials that are more susceptible to energetic masking than the CRT, such as the modified rhyme test (MRT) [11].

4. DISCUSSION AND CONCLUSIONS

The results of this experiment clearly show that listeners are able to use the distance-dependent changes that occur in the interaural time and level differences of nearby sounds to segregate competing speech signals presented at different distances along the interaural axis. When the target speech signal was masked by speech from a same-sex talker, spatially separating the target and masking sounds produced improvements in intelligibility that substantially exceed those that could be obtained by selectively attending to the ear with the highest SNR. The binaural advantage gained by spatially separating the competing speech signals in distance was approximately equivalent to a 5 dB increase in SNR in the 12 cm - 1 m configuration of the experiment. Indeed, binaural processing can account for most of the spatial unmasking that occurs when same-sex competing speech signals are separated in distance in the near field. A comparison of the COH and BE normalization conditions in Figure 2 shows that most of the intelligibility advantages that occurred when same-sex talkers were separated in distance in the near-field were retained when the signals were scaled to eliminate the advantages of selectively attending to the ear with the better SNR. Thus, it appears that spatial separation in distance can substantially improve the intelligibility of multiple same-sex talkers, and that binaural difference cues play a critical role in this improvement in intelligibility.

Spatial separation in distance can also produce a substantial improvement in speech intelligibility with a noise masker. When COH normalization was used, correct identifications improved from 40% in the 1 m - 1 m configuration to near 100% in the 12 cm -1 m configuration (Figure 2, left panel). However, nearly all of this benefit could be obtained by simply listening to the ear with the highest SNR. The binaural difference cues that are vitally important for the segregation of competing same-sex speech signals in the near field contribute little or nothing to our ability to segregate a nearby talker from a masking noise. Spatial separation never produced a binaural advantage much larger than 1 dB in any of the target-masker configurations tested in this experiment.

It is important to note that spatial separation of the target and masker in distance did not significantly improve the intelligibility of the target speech when the listening task was relatively easy to perform monaurally. This is apparent from the lack of any discernible differences between the spatially separated and co-located conditions when the target phrase and masker phrase were spoken by different-sex talkers (Figure 2).

The results of this experiment have important implications in the design of virtual audio displays that maximize the information processing capabilities of human listeners. In the past, audio display designers who wanted to present more than one simultaneous speech signal to a listener had to depend on angular separation to allow the listeners to segregate the competing messages. Near-field HRTFs offer an alternative way to segregate these messages— presenting them at different distances in the near field. This approach can have two major advantages over spatial separation based solely on differences in the azimuth locations of the sounds. One is that the different talkers can be presented in the same direction without compromising the listener's ability to segregate the two signals. In displays where the locations of the competing speech messages are used to convey information to the listener (i.e. air-traffic control locations that originate from the locations of the aircraft), there is always some danger that listeners will lose their ability to segregate the speech signals when they originate from the same direction relative to the listener. If the speech signals are presented at different distances, the listeners should be able to perceive them in the same azimuth locations without losing their ability to segregate the speech messages. Further investigation is needed to determine how well listeners are able to use differences in distance to segregate speech signals when they are presented at azimuth locations other than the 90° location examined in this experiment.

The second major advantage of using near-field HRTFs to segregate speech signals in a virtual audio display is that the nearby, lateral source locations tested in this experiment provide two additional speech channels that do not interfere with speech messages presented at any azimuth angle in the far field. For example, if one assumes that a 30° separation in azimuth is necessary to effectively segregate two competing speech signals, 7 independent speech channels would be available in a traditional far-field multitalker speech display (at azimuth locations of $-90°$, $-60°$, $-30°$, $0°$, $30°$, $60°$, and $90°$). Near-field HRTFs could provide at least two additional independent channels, at ±90° in azimuth and a distance of 12 cm, for a 29% increase in the total channel capacity of the system. Of course, listeners would never be able to process 9 simultaneous speech signals. But it may be useful to design a communications system where 9 different channels are assigned different spatial locations in order to allow the listener to use apparent position to keep track of the origins of the incoming speech messages. In most multitalker environments, incoming speech messages rarely arrive simultaneously from more than two or, possibly, three different communications channels. In these situations, a key design goal for a multitalker speech display is to allow the listener to segregate any pair of different talkers that may happen to speak at the same time while retaining their ability to determine who is talking from the fixed locations assigned to each communication channel. With spatial separation in azimuth only, there are only about 7 different locations that insure that no interference will occur for any pair of simultaneous speech signals. With the near-field lateral positions examined in this study, at least 9 of these non-interfering locations are available. Although more research is needed to find the optimal way to use near-field HRTFs in multitalker speech displays, the results of this experiment suggest that near-field cues will have important applications in the design of advanced speech interfaces for demanding communication environments.

5. REFERENCES

[1] 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.

[2] 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.

[3] A. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acustica*, vol. 86, pp. 117–128, 2000.

[4] D.S. Brungart and W.M. Rabinowitz, "Auditory localization of nearby sources. i: Head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 106, pp. 1465–1479, 1999.

[5] D.S. Brungart, N.I. Durlach, and W.M. Rabinowitz, "Auditory localization of nearby sources. ii: Localization of a broadband source," *Journal of the Acoustical Society of America*, vol. 106, pp. 1956–1968, 1999.

[6] B.G. Shinn-Cunningham, J. Schickler, N. Kopco, and R. Litovsky, "Spatial unmasking of nearby speech sources in a simulated anechoic environment," Unpublished manuscript under review at the Journal of the Acoustical Society of America, 2000.

[7] R.L. Freyman, K.S. Helfer, D.D. McCall, and R.K. Clifton, "The role of perceived spatial separation in the unmasking of speech," *Journal of the Acoustical Society of America*, vol. 106, pp. 3578–3587, 1999.

[8] M.L. Hawley, R.L. Litovsky, and J. Culling, "The 'cocktail party' effect with four kinds of maskers: Speech, time-reversed speech, speech-shaped noise, or modulated speech-shaped noise," *in Proceedings of the Midwinter Meeting of the Association for Research in Otolaryngology*, p. 31, 2000.

[9] 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.

[10] D.S. Brungart, "Informational and energetic masking effects in the perception of two simultaneous talkers," *Journal of the Acoustical Society of America*, vol. 109, pp. 1101–1109, 2001.

[11] D.S. Brungart, "Evaluation of speech intelligibility with the coordinate response measure," In press, Journal of the Acoustical Society of America, 2001.