1

Spatial Unmasking of Speech Based on Near-Field Distance Cues

Craig Jin[1], Virginia Best[2], Gaven Lin[2] and Simon Carlile[2]
[1]*School of Electrical and Information Engineering, The University of Sydney, Sydney NSW*
[2]*School of Medical Sciences and Bosch Institute, The University of Sydney, Sydney NSW*
Australia

1. Introduction

These days it is recognised that for bilateral hearing loss there is generally benefit in fitting two hearing aids, one for each ear (see Byrne, 1980 and Feuerstein, 1992 for clinical studies, see Byrne *et al.,* 1992, Durlach *et al*., 1981, and Zurek, 1981 for laboratory studies). Bilateral fitting is now standard practice for children with bilateral loss and as of 2005 bilateral fittings account for approximately 75% of all fittings (Libby, 2007). Nonetheless, it is only within the last half-decade that it has become possible to transfer audio signals between bilaterally-fitted hearing aids (Moore, 2007). This is primarily attributed to the technological advances in integrated circuit design, longer lasting batteries and also wireless inter-communication between the two hearing aids, e.g., using near-field magnetic induction (NFMI) communication. The possibility to exchange audio signals between bilaterally-fitted aids opens the door to new types of binaural signal processing algorithms to assist hearing-impaired listeners separate sounds of interest from background noise. In this chapter, we consider whether or not the manipulation of near-field distance cues may provide a viable binaural signal processing algorithm for hearing aids. More specifically, this chapter describes three experiments that explore the spatial unmasking of speech based on near-field distance cues.

In a typical cocktail party setting, listeners are faced with the challenging task of extracting information by sifting through a mixture of multiple talkers overlapping in frequency and time. This challenge arises as a result of interference in the form of energetic masking, where sounds are rendered inaudible due to frequency overlap, and informational masking, where sounds from different sources are confused with one another (Bronkhorst, 2000; Brungart *et al*., 2001; Kidd *et al*., 2008). Despite this, listeners are reasonably adept at parsing complex mixtures and attending to separate auditory events.

One factor that influences speech intelligibility in mixtures is perceived spatial location. Many studies have established that sounds originating from separate locations are easier to distinguish than sounds which are co-located (Hirsh, 1950; Bronkhorst and Plomp, 1988; Ebata, 2003). Separating sounds in space can result in an increase in the signal-to-noise ratio at one ear (the 'better ear'). Moreover, sounds that are spatially separated give rise to differences in binaural cues (interaural time and level differences, ITDs/ILDs) that can improve audibility by reducing energetic masking (Durlach and Colburn, 1978;

Zurek, 1993). Perceived differences in location can also be used as a basis for perceptual streaming, and this has been shown to be a particularly important factor in the segregation of talkers with similar voice characteristics, resulting in a significant reduction of informational masking (Kidd *et al*., 1998; Freyman *et al*., 1999; Arbogast *et al*., 2002; Drennan *et al*., 2003).

While many studies have established the role of spatial cues in the unmasking of speech mixtures, the majority of these have focused on sources at a fixed, relatively far distance, with spatial separation in the azimuthal plane. Very few studies have examined the perception of speech mixtures in the acoustic 'near field', defined as the region less than one meter from the listener's head. Unlike in the far field, spatial cues at the two ears vary substantially as a function of distance in the near field (Brungart and Rabinowitz, 1999). Listeners can use these cues to estimate the distance of sources in the immediate vicinity (Brungart *et al*., 1999). A primary distance cue is overall intensity, with near sounds being louder than far sounds. In addition, ILDs increase dramatically with decreasing distance in both high and low frequency regions. Most notably, low-frequency ILDs, which are negligible in the far field, can be as large as 20 dB in the near field (Brungart, 1999; Brungart and Rabinowitz, 1999). In contrast, ITDs in the near field are independent of distance and remain relatively constant. This study investigated whether the increased ILD cues that occur at different distances in this region can provide a basis for improving speech segregation. Understanding the effect of distance cues on speech segregation will also enable a more complete picture of how spatial perception influences behaviour in cocktail party settings.

Two previous studies have shown that spatial separation of sources in the near field can lead to benefits in speech intelligibility. Shinn-Cunningham *et al*. (2001) showed that separating speech and noise in the near field could lead to improvements in speech reception thresholds. When one sound was fixed at one meter and the other was moved in closer to the listener, an improved target to masker ratio (TMR) occurred at one ear. In this case, masking was energetic and performance benefits were well-predicted by improvements in audibility. A study by Brungart and Simpson (2002) showed that separation of two talkers in distance improved accuracy in a speech segregation task. After controlling for better ear effects they found that there was an additional perceptual benefit, particularly when talkers were acoustically similar (the same sex). This suggests that distance cues in the near field may provide a basis for release from informational masking.

The primary aim of the current study was to further investigate the effects of near field distance cues on speech segregation. The first experiment was an extension of the study by Brungart and Simpson (2002). The aim was to measure the benefit of separating two competing talkers in distance, where one was fixed at one meter and the other was moved closer to the head. While Brungart and Simpson examined only the case where the two talkers were equal in level (0-dB TMR) and most easily confused, the current study aimed to discover whether this benefit generalized to a larger range of TMR values. Experiment 2 was identical to Experiment 1, but assessed whether low-frequency (< 2 kHz) spatial cues alone could produce the effects seen in Experiment 1. Experiment 3 investigated the effect of moving a mixture of three talkers (separated in azimuth) closer to the head. It was predicted that this manipulation, which effectively exaggerates the spatial cues, would offer improved segregation of the competing talkers.

2. General methods

2.1 Subjects

Eight subjects (six males and two females, aged between 20 and 32) participated in the experiments. Only one subject had previous experience with auditory experiments involving similar stimuli.

2.2 Virtual auditory space

Individualized head-related transfer functions for the generation of virtual spatialized stimuli were recorded in an anechoic chamber, and details of the procedure can be found elsewhere (Pralong and Carlile, 1994, 1996). In brief, a movable loudspeaker (VIFA-D26TG-35) presented Golay codes from 393 locations on a sphere of radius 1 m around the subject's head. Binaural impulse responses were collected using a blocked-ear approach, with microphones (Sennheiser KE 4-211-2) placed in the subject's ear canals. Recordings were digitized at a sampling rate of 80 kHz, and converted to directional transfer functions (DTFs) by removing location-independent components. The DTFs were bandpass filtered between 300 Hz and 16 kHz, the range in which the measurement system is reliable, but then the energy below 300 Hz was interpolated based on the spherical head model (below) so that fundamental frequency energy in the speech stimuli would not be filtered out.

A distance variation function (DVF) as described by Kan *et al*. (2009) was used to convert the far-field DTFs (1-m distance) to near-field DTFs (0.25- and 0.12-m distances). The DVF approximates the frequency-dependent change in DTF magnitude as a function of distance. It is based on the rigid sphere model of acoustic scattering developed by Rabinowitz *et al*. (1993) and experimentally verified by Duda and Martens (1998). According to this model, the head can be approximated as a rigid sphere of radius a with ears toward the back of the head at 110° from the mid-sagittal plane. If a sinusoidal point source of sound of frequency 'ω' is presented at distance 'r' and angle θ from the centre of the head, the sound pressure 'p' at the ear can be expressed as:

$$p(a,\omega,\theta,r) = -kr\sum_{m=0}^{\infty}(2m+1)\frac{h_m(kr)}{h'_m(ka)}P_m(\cos\theta)e^{-ikr} \tag{1}$$

where h_m is the spherical Hankel function, k is the wave number, and P_m is the Legendre polynomial. DVFs were applied to each subject's individualized DTFs. The head radius, a, for each subject was determined using Kuhn's (1977) equation:

$$\text{ITD} = \frac{3a}{c}\sin\theta_{\text{inc}} \tag{2}$$

where c is the speed of sound in air, θ is the angle of incidence to the head, and ITD is the ITD measured from a pair of DTFs using cross-correlation. Individualized DTFs modified with the DVF in this way were recently verified psychophysically for their ability to give rise to accurate near-field localization estimates (Kan *et al*., 2009). Fig. 1 shows a set of example DVF gain functions (to be applied to 1-m DTFs) as a function of frequency and distance for three azimuthal locations that were used in the study.

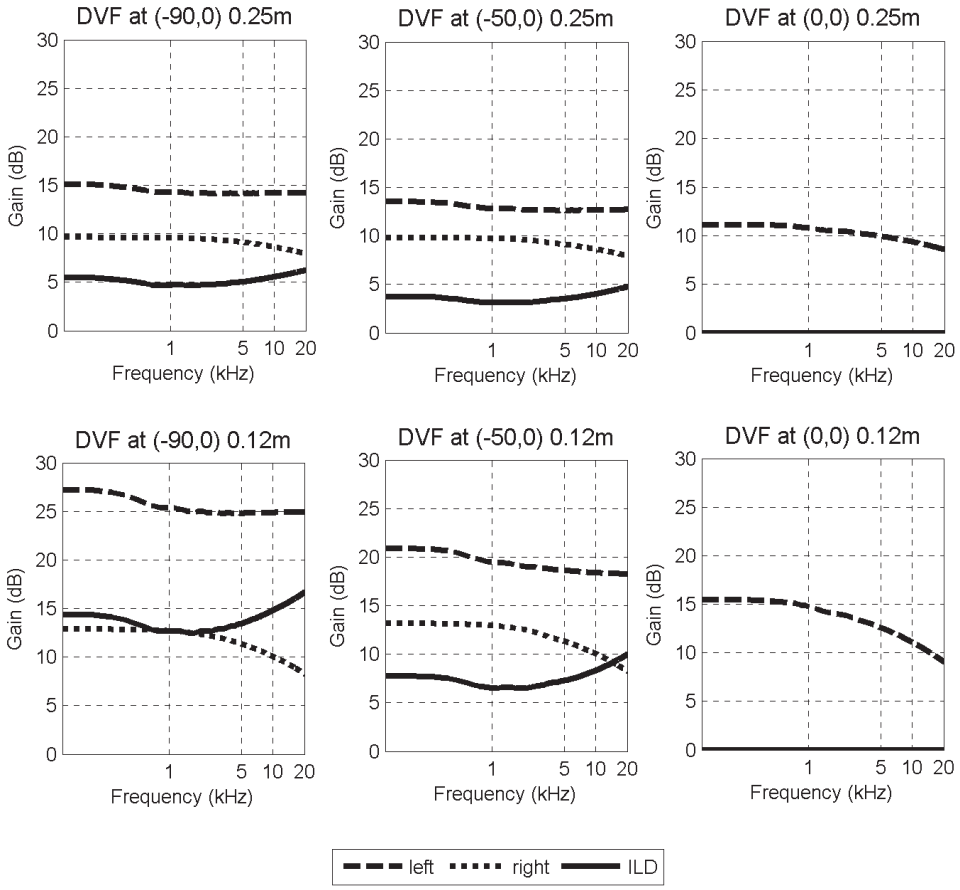

Fig. 1. The DVF for three locations and two near-field distances. The gain in dB is relative to the 1-m far-field case for each azimuth, and is shown for the left and right ears. Shown also is the induced ILD, which increases with increasing laterality (-90°>-50°>0°) and decreasing distance (0.12 m>0.25 m>1 m).

2.3 Speech stimuli

The speech stimuli used for this study were taken from the Coordinate Response Measure (CRM) corpus (Bolia *et al*., 2000). Each sentence is comprised of a call sign, color and number, spoken in the form "Ready (call sign) go to (color) (number) now". There are a total of 8 possible call signs ("arrow", "baron", "eagle", "hopper", "laker", "ringo", "tiger" and "charlie"), 4 possible colors ("red", "blue", "green" and "white") and 8 possible numbers (1-8). In total, there are 256 possible phrases, which are spoken by a total of 8 different talkers (4 male and 4 female), giving 2048 distinct phrases in the corpus.

In each experimental trial, the sentences were randomly selected without replacement and were chosen such that each sentence in a mixture had a unique talker, call sign, number and color. The same gender was used for each talker in a given trial. The call sign

"Charlie" was always assigned to the target. Sentences were normalized to the same RMS level and resampled from 40 kHz to 48 kHz for playback. The target sentence was then adjusted to achieve the desired TMR before all sentences were filtered through the relevant DTFs (also resampled to 48 kHz) and digitally added. There was no normalization of the stimulus level after the DTF filtering, thus the stimulus level would increase when presented nearer to the head. The stimuli were presented at a comfortable listening level that corresponded to a sensation level of approximately 40 dB for a source directly ahead at a distance of 1 m.
Experiments were conducted in a small audiometric booth. Stimuli were presented via an RME soundcard (48 kHz sampling rate) and delivered using insert earphones (Etymotic Research ER-1[1]). Subjects were seated in front of an LCD monitor, and registered their responses (a color and number combination for the target stimulus) by clicking with a mouse on a custom-made graphical user interface.

2.4 Analysis of results

The listener responses were scored as correct if both the color and number were reported correctly, and percent correct scores (over the 40 repetitions) were plotted as a function of TMR to give raw psychometric functions for each spatial configuration. However, a nominal TMR at the source gives rise to different TMRs at the listener's ears for different spatial configurations (according to the DVF). Thus, a normalization stage was applied to the data to factor out these changes in TMR at the ear. Of particular interest was whether there was still a perceptual benefit of the distance manipulations after taking into account any energetic advantages.
The RMS levels of the target and maskers at each ear were calculated during the experiment for each individual subject under the different spatial configurations. These values were then averaged and used to determine the TMR at the better ear for each condition. This better-ear TMR represented a consistent shift from the nominal TMR, and thus the psychometric functions could be re-plotted as a function of better-ear TMR by a simple shift along the TMR axis. The average normalization shifts for each condition are shown in Tables 1 and 2. A single mean value was appropriate (rather than individual normalization values for each listener) because the values varied very little (range across listeners < 1dB).
The perceptual benefit of separating/moving sources in the near field was defined as the remaining benefit (in percentage points) after taking into account energetic effects. To calculate these benefits, the normalized psychometric functions for the reference conditions were subtracted from the normalized psychometric functions for the various near-field conditions. Values were interpolated using a linear approximation where required.

3. Experiment 1

3.1 Experimental conditions

The spatial configurations used in Experiment 1 were essentially the same as those used by Brungart and Simpson (2002). One target and one masker talker were simulated at -90°

[1] Note that the ER-1 earphones reintroduce the ear-canal resonance that is removed by the DTF.

azimuth, directly to the left of the listener. This region was expected to be particularly important in the study of near field perception due to the large ILDs that occur. As illustrated in Fig. 2, there were a total of five different target or masker distances. One talker was always fixed at 1 m while the other was moved closer to the listener in the near field. In some conditions, the masker was fixed at 1 m while the target was presented at 0.25 m or 0.12 m from the head. Conversely, in other conditions, the target was fixed at 1 m while the masker was presented at 0.25 m or 0.12 m from the head. In the co-located condition, both talkers were located at 1 m. Five different TMR values were tested for each spatial configuration (see Table 1), resulting in a total of 25 unique conditions. Two 20-trial blocks for each condition were completed by each listener resulting in a total of 2x20x25=1000 trials per listener. The spatial configuration and TMR were kept constant within a block, but the ordering of the blocks was randomized.

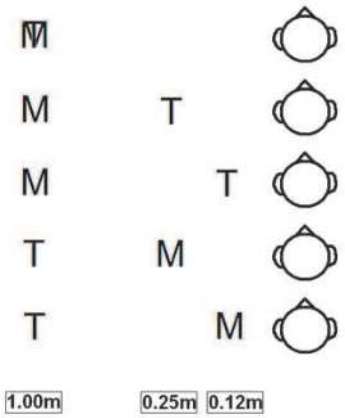

Fig. 2. The five spatial configurations used in Experiments 1 and 2. In one condition, both the target (T) and masker (M) were co-located at 1 m. In "target closer" conditions, the masker was fixed at 1 m while the target was located at 0.25 m or 0.12 m. In "masker closer" conditions, the target was fixed at 1 m while the masker was located at 0.25 m or 0.12 m.

Configuration	TMRs tested (dB)	Normalization shift (dB)
Target 1 m/Masker 1 m	[-30 -20 -10 0 10]	0
Target 0.25 m/Masker 1 m	[-40 -30 -20 -10 0]	+14
Target 0.12 m/Masker 1 m	[-40 -30 -20 -10 0]	+27
Target 1 m/Masker 0.25 m	[-20 -10 0 10 20]	-9
Target 1 m/Masker 0.12 m	[-20 -10 0 10 20]	-13

Table 1. The range of TMR values tested and normalization shifts for each spatial configuration in Experiments 1 and 2. The normalization shifts are the differences in the TMR at the better ear that resulted from variations in target or masker distance (relative to the co-located configuration).

3.2 Results

3.2.1 Masker fixed at 1 m and target near

The left column of Fig. 3 shows results (pooled across the eight listeners) from the conditions in which the masker was fixed at 1 m and the target was moved into the near field. Performance improved (Fig. 3, top left) when the target talker was moved closer (0.12 m>0.25 m>1 m). This trend was observed across all TMRs. Scores also increased with TMR as expected. A two-way repeated-measures ANOVA on the arcsine-transformed data[2] confirmed that there was a significant main effect of both target distance ($F_{2,14}$=266.5, p<.01) and TMR ($F_{3,21}$=58.2, p<.01). There was also a significant interaction ($F_{6,42}$=147.9, p<.01), implying that the effect of target distance differed depending on the TMR.

When the psychometric functions were re-plotted as a function of better-ear TMR, they looked almost identical (Fig. 3, middle left), except at 0-dB TMR. At this point, the co-located performance shows a characteristic plateau that is absent in the separated conditions, and this appears to drive the separation of the functions in this region. Fig. 3 (bottom left) shows the difference (in percentage points) between the separated conditions and the co-located condition as a function of TMR. The advantage is positive for the TMR range between -10 and 10 dB. T-tests confirmed that at 0-dB TMR, the advantages were significant for both the 0.25-m target (mean 23 percentage points, t_7=7.49, p<.01) and the 0.12-m target (mean 26 percentage points, t_7=8.29, p<.01).

3.2.2 Target fixed at 1 m and masker near

The right column of Fig. 3 shows results from the opposite conditions in which the target was fixed at 1 m and the masker was moved into the near field. The raw data (Fig. 3, top right) show that performance decreased as the masker was moved closer to the listener (1 m>0.25 m>0.12 m) for negative TMRs. However at higher TMRs, scores approached 100% for all distances. A two-way repeated-measures ANOVA on the arcsine-transformed data confirmed that there was a significant main effect of masker distance ($F_{2,14}$=37.4, p<.01) and TMR ($F_{3,21}$= 58.2, p<.01). The interaction did not reach significance ($F_{6,42}$=12.9, p=0.07).

When the psychometric functions were re-plotted as a function of better-ear TMR, there was a reversal in their ranking. Once the energetic disadvantage of moving a masker closer was compensated for, mean performance was slightly better when the masker was separated from the target compared to the co-located case. The benefit plots in Fig. 3 (bottom right) show that the spatial advantage was positive at all TMRs, but was particularly pronounced at 0-dB TMR. The advantage at 0-dB TMR was significant for both the 0.25-m masker (mean 26 percentage points, t_7= 7.71, p<.01) and the 0.12-m masker (mean 34 percentage points, t_7=8.44, p<.01). Again this benefit peaks in the region where the psychometric function for the co-located case is relatively flat.

The filled symbols in the middle and bottom rows of Fig. 3 show data from Brungart and Simpson (2002) under the analogous conditions of their study. Mean scores are higher overall in the current study (Fig. 3, middle row), however the benefit of separating talkers in distance is roughly the same across studies (Fig. 3, bottom row).

[2] The arcsine transformation converts binomially distributed data to an approximately normal distribution that is more suitable for statistical analysis (Studebaker, 1985).

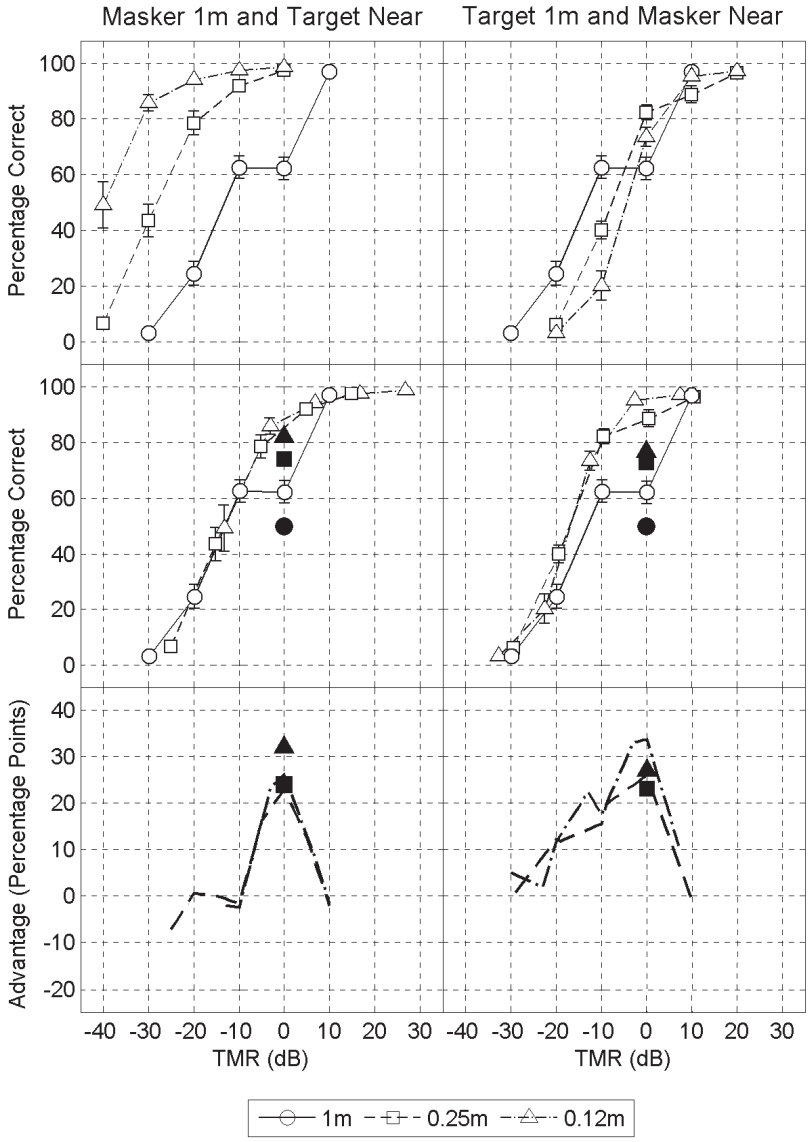

Fig. 3. Mean performance data averaged across all 8 subjects (error bars show standard errors of the means) in Experiment 1. The left panel displays the raw (top) and normalized (middle) data for the conditions where the masker was fixed at 1 m and the target was moved closer to the listener. The right panel displays the raw (top) and normalized (middle) data for the conditions where the target was fixed at 1 m and the masker was moved in closer to the listener. The bottom panels display the benefits of separation in distance, expressed as a difference in percentage points relative to the co-located case. The results obtained by Brungart and Simpson (2002) at 0-dB TMR are indicated by the black symbols.

3.3 Discussion

For a target and masker talker located at a fixed azimuth, target identification improved when the target was moved increasingly nearer to the head (relative to the case where both talkers were co-located at 1 m), but got worse when the masker moved closer. This basic pattern of results was likely driven by energetic effects: the closer source dominates the mixture and this either increases or reduces the effective TMR at the better ear depending on which source is moved.

The remaining benefit of spatial separation after the TMR changes were accounted for was restricted to a better-ear TMR region around 0 dB. This region is approximately where the psychometric function for the co-located case shows a clear plateau, which is no longer present in the separated cases. This plateau has been described previously (Egan *et al*., 1954; Dirks and Bower, 1969; Brungart *et al*., 2001), and is thought to represent the fact that listeners have the most difficulty segregating two co-located talkers when they are equal in level (0-dB TMR), but with differences in level listeners can attend to either the quieter or the louder talker. Apparently the perception of separation in distance also alleviates the particular difficulty of equal-level talkers, by providing a dimension along which to focus attention selectively. This finding adds to a growing body of evidence indicating that spatial differences can aid perceptual grouping and selective attention. Interestingly, the effect does not appear to be "all or nothing"; larger separations in distance gave rise to larger perceptual benefits. The lack of a spatial benefit at other TMRs, especially at highly negative TMRs, suggests that the main problem was audibility and not confusion between the target and the masker. Consistent with this idea, in the co-located condition, masker errors made up a larger proportion of the total errors as the TMR approached 0 dB. In Experiment 1, the proportion of masker errors was 38%, 45%, 62%, and 93% at -30, -20, -10, and 0-dB TMR.

Listeners in Experiment 1 performed around 10-20 percentage points better than Brungart and Simpson's (2002) listeners for the same stimulus configurations. This may be simply due to differences in the cohort of listeners, but there are two methodological factors that may have also played a role. Firstly, their study used HRTFs measured from an acoustic mannequin as opposed to individualized filters and thus the spatial percept may have been less realistic and thus less perceptually potent. Secondly, while the two studies used the same type of stimuli, Brungart and Simpson used a low-pass filtered version (upper cut-off of 8 kHz) and we used a broadband version (upper cut-off of 16 kHz). Despite the difference in overall scores, the mean benefit (in percentage points) obtained by separating talkers in distance was equivalent across the two studies.

4. Experiment 2

4.1 Experimental conditions

Experiment 2 was identical to Experiment 1 and used the same set of spatial configurations and TMRs (Fig. 2 and Table 1). The only difference was that the stimuli were all low-pass filtered (before RMS level equalization) at 2 kHz using an equiripple FIR filter with a stopband at 2.5 kHz that is 50 dB down from the passband.

4.2 Results

4.2.1 Masker fixed at 1 m and target near

The left column of Fig. 4 shows results from the conditions in which the masker was fixed at 1 m and the target was moved into the near field for the low-pass filtered stimuli of

Experiment 2. The raw data followed a similar trend to that observed in Experiment 1 (Fig. 4, top left). As the target was moved closer to the listener, performance improved, with best performance in the 0.12-m target case. A two-way repeated-measures ANOVA on the arcsine-transformed data revealed that there was a significant effect of target distance ($F_{2,14}$=332.9, p<.01) and TMR ($F_{3,21}$=120.6, p<.01) and a significant interaction ($F_{6,42}$=5.1, p<.05).

When the psychometric functions were plotted as a function of better-ear TMR, the results for all three distances were very similar (Fig. 4, middle left). After taking into account level changes with distance, there appears to be only a minor additional perceptual benefit of separating the low-pass filtered target and masker in distance. Fig. 4 (bottom left) shows that the advantage of separating the target from the masker was positive only for the small TMR range between -5 and +5 dB. The advantages across TMR were also smaller than those observed in Experiment 1. However, the advantages were still significant for both the 0.25-m target (mean 13 percentage points, t_7=4.20, p<.01) and the 0.12-m target (mean 17 percentage points, t_7=4.88, p<.01).

A three-way ANOVA with factors of bandwidth, distance, and TMR was conducted to compare performance in Experiments 1 and 2 in the target-near configuration (compare Fig. 3 and Fig. 4, top left). The main effect of bandwidth was significant ($F_{1,7}$=8.9, p<.05), indicating that performance was poorer for low-passed stimuli than for broadband stimuli overall. A separate two-way ANOVA on the benefits at 0 dB (compare Fig. 3 and Fig. 4, bottom left) found a significant main effect of distance ($F_{1,7}$=14.5, p<.01) but no significant effect of bandwidth ($F_{1,7}$=3.7, p=.10) and no interaction ($F_{1,7}$=0.7, p=.44).

4.2.2 Target fixed at 1 m and masker near

For the opposite configuration, where the masker was moved in closer (Fig. 4, right column), results were similar to those in Experiment 1. Listeners were less accurate at identifying the target when the masker was moved closer (Fig. 4, top right). A two-way repeated-measures ANOVA on the arcsine-transformed data revealed a significant effect of target distance ($F_{2,14}$=76.4, p<.01) and TMR ($F_{3,21}$=260.2, p<.01) and a significant interaction ($F_{6,42}$=5.1, p<.01).

Normalization of the curves based on better-ear TMR (Fig. 4, middle right) resulted in a reversal of the result, showing that there was indeed a perceptual benefit once the energetic disadvantage of a near masker was accounted for. Normalized scores were higher for maskers at 0.12 m and 0.25 m relative to 1 m, particularly around 0-dB TMR. This is reinforced by the benefit plots (Fig. 4, bottom right) which show that there was a positive advantage across all TMRs. Again, the largest advantage was observed at 0-dB TMR and was statistically significant for both the 0.25-m masker (mean 24 percentage points, t_7=7.31, p<.01) and the 0.12-m masker (mean 32 percentage points, t_7=7.51, p<.01).

A three-way ANOVA comparing the results from Experiments 1 and 2 in the masker-near configuration (compare Fig. 3 and Fig. 4, top right) revealed that performance was poorer for low-passed stimuli than for broadband stimuli overall ($F_{1,7}$=11.7, p<.05). A two-way ANOVA conducted on the benefits at 0 dB (compare Fig. 3 and Fig. 4, bottom right) found a significant main effect of distance ($F_{1,7}$=11.1, p<.05), but no significant effect of bandwidth ($F_{1,7}$=0.2, p=.66) and no interaction ($F_{1,7}$=0.6, p=.47).

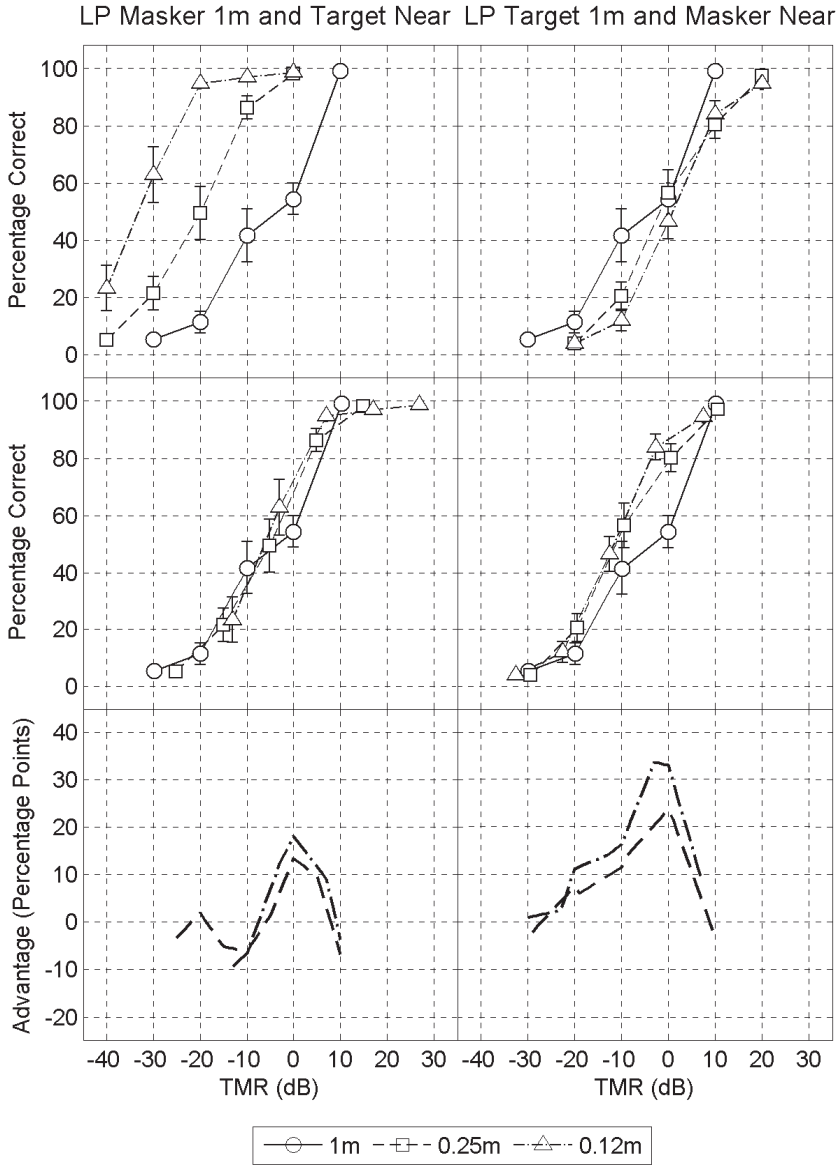

Fig. 4. Mean performance data averaged across all 8 subjects (error bars show standard errors of the means) in Experiment 2. The left panel displays the raw (top) and normalized (middle) data for the conditions where the masker was fixed at 1 m and the target was moved closer to the listener. The right panel displays the raw (top) and normalized (middle) data for the conditions where the target was fixed at 1 m and the masker was moved in closer to the listener. The bottom panels display the benefits of separation in distance, expressed as a difference in percentage points relative to the co-located case.

4.3 Discussion

The results from Experiment 2 in which the speech stimuli were low-pass filtered at 2 kHz were largely similar to those from Experiment 1. Performance across conditions was generally poorer, consistent with a more difficult segregation task, and subjects reported that voices appeared muffled and were more difficult to distinguish from each other in this condition. However, the perceptual benefit of separating talkers in distance condition was for broadband and low-pass filtered stimuli. This demonstrates that the low-frequency ILDs that are unique to this near field region of space are sufficient to provide a benefit for speech segregation.

5. Experiment 3

5.1 Experimental conditions

In Experiment 3, three talkers were used, and they were separated in azimuth at -50°, 0°, and 50° as illustrated in Fig. 5. For a given block, the distance of all talkers was set to either 1 m, 0.25 m or 0.12 m from the listener's head. Six different TMR values were tested for each spatial configuration (see Table 2), resulting in 18 unique conditions. The location of the target within the three-talker array was varied randomly within each block, such that half the trials had the target in the central position and the other half had the target in one of the side positions. Two 40-trial blocks were completed per condition by each listener resulting in a total of 2x40x18=1440 trials per listener. The distance and TMR were kept constant within a block, but the order of blocks was randomized.

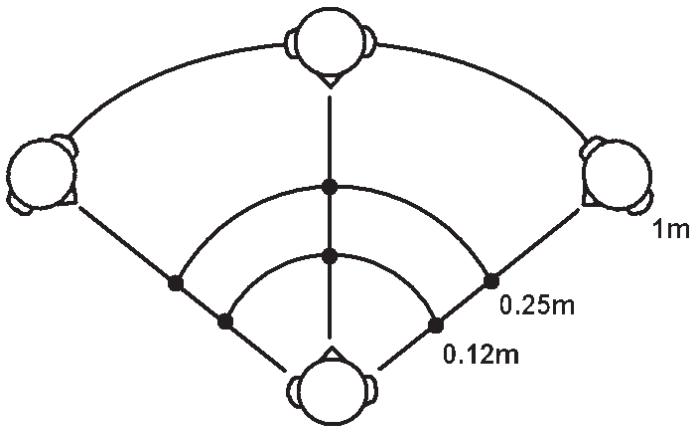

Fig. 5. The spatial configurations used in Experiment 3. Three talkers were spatially separated in azimuth at -50°, 0° and 50°and were either all located at 1 m, 0.25 m or 0.12 m from the listener's head. The location of the target talker was randomly varied (left, middle, right).

Configuration (target position/distance of mixture)		TMRs tested (dB)	Normalization shift (dB)
Central target	1 m	[-20 -15 -10 -5 0 5]	-3
	0.25 m	[-20 -15 -10 -5 0 5]	-5
	0.12 m	[-20 -15 -10 -5 0 5]	-8
Lateral target	1 m	[-20 -15 -10 -5 0 5]	0
	0.25 m	[-20 -15 -10 -5 0 5]	+3
	0.12 m	[-20 -15 -10 -5 0 5]	+6

Table 2. The range of TMR values tested and normalization values for each spatial configuration in Experiment 3. The normalization shifts are the differences in TMR at the better ear that resulted from variations in distance and configuration.

5.2 Results

5.2.1 Centrally positioned target

When the target was directly in front of the listener, with a masker on either side at ±50° azimuth, moving the whole mixture closer to the head had very little effect on raw performance scores (Fig. 6, top left). A two-way repeated-measures ANOVA on the arcsine-transformed data, however, showed that the effect of distance was statistically significant ($F_{2,14}$=7.7, p<.01), as was as the effect of TMR ($F_{5,35}$=159.4, p<.01). The interaction did not reach significance ($F_{10,70}$=1.4, p=0.2).

When the psychometric functions were re-plotted as a function of better-ear TMR, the distance effects were more pronounced (Fig. 6, middle left). This normalization compensates for the fact that the lateral maskers increase more in level than the central target when the mixture approaches the head. Mean performance was better for most TMRs when the mixture was moved into the near field. Fig. 6 (bottom left) shows the difference (in percentage points) between the near field conditions and the 1-m case, illustrating the advantage of moving sources closer to the head. The mean benefits were significant at all TMRs for both distances (p<.05).

5.2.2 Laterally positioned target

Raw results for the condition in which the target was located to the side of the three-talker mixture are shown in Fig. 6 (top right). Performance was better when the mixture was closer to the listener (0.12 m>0.25 m>1 m) particularly for low TMRs (below -5 dB). At higher TMRs, performance for all three distances appears to converge. Performance generally increased with increasing TMR but reached a plateau at around 80%. A two-way repeated-measures ANOVA on the arcsine-transformed data confirmed that there was a main effect of both distance ($F_{2,14}$=24.5, p<.01) and TMR ($F_{5,35}$=104.4, p<.01) and a significant interaction ($F_{10,70}$=17.4, p<.01).

When the psychometric functions were normalized to account for level changes at the better ear, the distinction between the different distances was reduced. An advantage of the near field mixtures over the 1-m mixture was found only at low TMRs (Fig. 6, middle right).

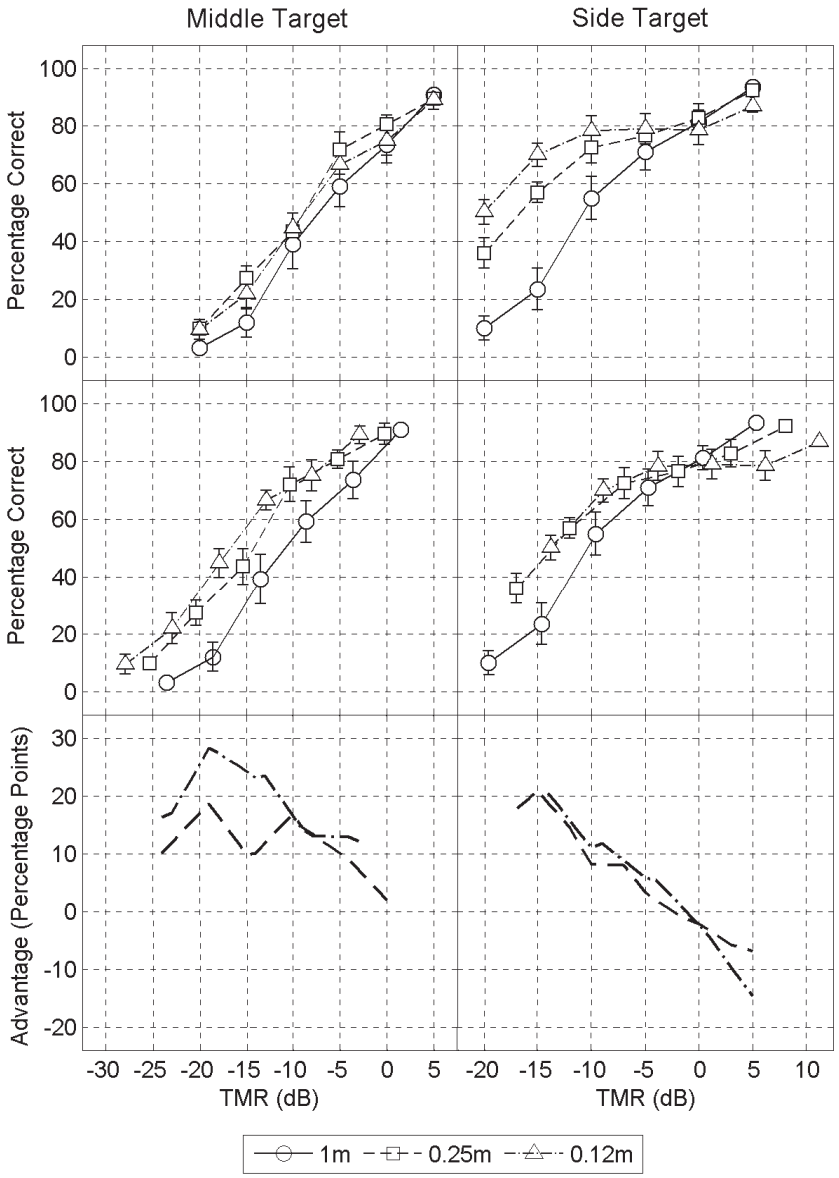

Fig. 6. Mean performance data averaged across all 8 subjects (error bars show standard errors of the means) in Experiment 3. The left panel displays the raw (top) and normalized (middle) data for the conditions where the target was located in the middle of three talkers. The right panel displays the raw (top) and normalized (middle) data for the conditions where the target was located to one side. The bottom panels display the benefits of decreasing the distance of the mixture, expressed as a difference in percentage points relative to the 1-m case.

At higher TMRs, the curves in fact reversed in order. These effects are reiterated in the benefit plots (Fig. 6, bottom right). The advantage was positive at negative TMRs but negative at positive TMRs. The mean benefits were significant at -15-dB TMR (t_7=4.30, $p<.01$) for the 0.25-m condition and at -10-dB TMR (t_7=2.78, $p<.05$) for the 0.12-m condition. A significant disadvantage was observed at 5-dB TMR for both distances ($p<.05$).

5.3 Discussion

Experiment 3 investigated the effect of moving a mixture of three talkers (separated in azimuth) closer to the head. Given that this manipulation essentially exaggerates the spatial differences between the competing sources, we were interested in whether it might improve segregation of the mixture. The manipulation had different effects depending on the location of the target. When the target was located in the middle, raw performance improved only very slightly with distance. However, this improvement occurred despite a decrease in TMR at the ear (both ears are equivalent given the symmetry) in this configuration (Table 2). In other words, performance improved despite an energetic disadvantage when the mixture was moved closer. Normalized performance thus revealed a perceptual benefit. When the target was located to the side, moving the mixture closer provided increases in better-ear TMR, and raw performance reflected this, but even after normalization there was a perceptual benefit of moving the mixture in closer. We attribute these benefits to an exaggeration of the spatial cues for the sources to the side, giving rise to a greater perceptual distance between the sources. It is not clear to us why this benefit was biased towards the lower TMRs in both cases, although the drop in benefit for high TMRs appears to be related to the flattening of the psychometric functions at high TMRs at the near field distances. It is possible that performance reaches a limit here due to the distracting effect of having three loud sources close to the head.

6. Conclusions

The results from these experiments provide insights into how the increase in ILDs that occurs in the auditory near field can influence the segregation of mixtures of speech. Spatial separation of competing sources in distance, as well as reducing the distance of an entire mixture of sources, led to improvements in terms of the intelligibility of a target source. These improvements were in some cases partly explained by changes in level that increased audibility, but in other cases occurred despite decreases in target audibility. The remaining benefits were attributed to salient spatial cues that aided perceptual streaming and lead to a release from informational masking.

In terms of binaural hearing-aids with the capability of exchanging audio signals, the experimental findings described here with normally-hearing listeners indicate that there may be value in investigating binaural signal processing algorithms that apply near-field sound transformations to sounds that are clearly lateralized. In other words, when the ITD or ILD cues strongly indicate a lateralized sound is present, a near-field sound transformation can be applied which artificially brings the sound perceptually closer to the head. We anticipate further experiments conducted with hearing-impaired listeners to investigate the value of such a binaural hearing-aid algorithm.

7. References

Arbogast, T. L., Mason, C. R., and Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *Journal of the Acoustical Society of America*, Vol. 112, pp. 2086-2098.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America*, Vol. 107, pp. 1065-1066.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica*, Vol. 86, pp. 117-128.

Bronkhorst, A. W., and Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *Journal of the Acoustical Society of America*, Vol. 83, pp. 1508-1516.

Brungart, D. S. (1999). Auditory localization of nearby sources. III. Stimulus effects. *Journal of the Acoustical Society of America*, Vol. 106, pp. 3589-3602.

Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (1999). Auditory localization of nearby sources. II. Localization of a broadband source. *Journal of the Acoustical Society of America*, Vol. 106, pp. 1956-1968.

Brungart, D. S., and Rabinowitz, W. R. (1999). Auditory localization of nearby sources. Head-related transfer functions. *Journal of the Acoustical Society of America*, Vol. 106, pp. 1465-1479.

Brungart, D. S., and Simpson, B. D. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *Journal of the Acoustical Society of America*, Vol. 112, pp. 664-676.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, Vol. 110, pp. 2527-2538.

Byrne, D. (1980). Binaural hearing aid fitting: research findings and clinical application, In *Binaural Hearing and Amplification: Vol 2*, E.R. Libby, pp. 1-21, Zenetron Inc., Chicago, IL

Byrne, D., Nobel, W., Lepage, B. W., (1992). Effects of long-term bilateral and unilateral fitting of different hearing aid types on the ability to locate sounds. *J. Am. Acad. Audiology*, Vol. 3, pp. 369-382.

Dirks, D. D., and Bower, D. R. (1969). Masking effects of speech competing messages. *Journal of Speech and Hearing Research*, Vol. 12, pp. 229-245.

Drennan, W. R., Gatehouse, S. G., and Lever, C. (2003). Perceptual segregation of competing speech sounds: The role of spatial location. *Journal of the Acoustical Society of America*, Vol. 114, pp. 2178-2189.

Duda, R. O., and Martens, W. L. (1998). Range dependence of the response of a spherical head model. *Journal of the Acoustical Society of America*, Vol. 104, pp. 3048-3058.

Durlach, N. I., and Colburn, H. S. (1978). Binaural phenomena, In *The Handbook of Perception*, E. C. Carterette and M. P. Friedman, Academic, New York.

Durlach, N. I., Thompson, C. L., and Colburn, H.A. (1981). Binaural interaction in impaired listeners - a review of past research. *Audiology*, Vol. 20, pp. 181-211.

Ebata, M. (2003). Spatial unmasking and attention related to the cocktail party problem. *Acoust. Sci and Tech.* , Vol. 24, pp. 208-219.

Egan, J., Carterette, E., and Thwing, E. (1954). Factors affecting multichannel listening. *Journal of the Acoustical Society of America*, Vol. 26, pp. 774-782.

Feuerstein, J. (1992). Monaural versus binaural hearing: ease of listening, word recognition, and attentional effort. *Ear and Hearing*, Vol. 13,, No. 2, pp. 80-86.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America*, Vol. 106, pp. 3578-3588.

Hirsh, I. J. (1950). The relation between localization and intelligibility. *Journal of the Acoustical Society of America*, Vol. 22, pp. 196-200.

Kan, A., Jin, C., and van Schaik, A. (2009). A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function. *Journal of the Acoustical Society of America*, Vol. 125, pp. 2233-2243.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). Informational masking, In *Auditory Perception of Sound Sources*, W. A. Yost, A. N. Popper, and R. R. Fay (Springer Handbook of Auditory Research, New York), pp. 143-190.

Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America*, Vol. 104, pp. 422-431.

Libby, E. R. (2007). The search for the binaural advantage revisited. *The Hearing Review*, Vol. 14, No. 12, pp. 22-31.

Moore, B.C.J. (2007). Binaural sharing of audio signals: Prospective benefits and limitations. *The Hearing Journal*, Vol. 40, No. 11, pp. 46-48.

Pralong, D., and Carlile, S. (1994). Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature "in-ear" recording system. *Journal of the Acoustical Society of America*, Vol. 95, pp. 3435-3444.

Pralong, D., and Carlile, S. (1996). The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *Journal of the Acoustical Society of America*, Vol. 100, pp. 3785-3793.

Rabinowitz, W. M., Maxwell, J., Shao, Y., and Wei, M. (1993). "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," Presence: Teleoperators and Virtual Environments 2.

Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. (2001). "Spatial unmasking of nearby speech sources in a simulated anechoic environment. *Journal of the Acoustical Society of America*, Vol. 110, pp. 1119-1129.

Studebaker, G. A. (1985). A rationalized arcsine transform. *Journal of Speech and Hearing Research*, Vol. 28, pp. 455-462.

Zurek, P. M. (1993). Binaural advantages and directional effects in speech intelligibility, In *Acoustical Factors Affecting Hearing Aid Performance*, G. A. Studebaker and I. Hochberg, pp. 255-276, Allyn and Bacon, Boston.

Advanced Biomedical Engineering

Edited by Dr. Gaetano Gargiulo

ISBN 978-953-307-555-6
Hard cover, 280 pages
Publisher InTech
Published online 23, August, 2011
Published in print edition August, 2011

This book presents a collection of recent and extended academic works in selected topics of biomedical signal processing, bio-imaging and biomedical ethics and legislation. This wide range of topics provide a valuable update to researchers in the multidisciplinary area of biomedical engineering and an interesting introduction for engineers new to the area. The techniques covered include modelling, experimentation and discussion with the application areas ranging from acoustics to oncology, health education and cardiovascular disease.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Craig Jin, Virginia Best, Gaven Lin and Simon Carlile (2011). Spatial Unmasking of Speech Based on Near-Field Distance Cues, Advanced Biomedical Engineering, Dr. Gaetano Gargiulo (Ed.), ISBN: 978-953-307-555-6, InTech, Available from: http://www.intechopen.com/books/advanced-biomedical-engineering/spatial-unmasking-of-speech-based-on-near-field-distance-cues

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821