# The Role of Reverberation in Release from Masking Due to Spatial Separation of Sources for Speech Identification

Gerald Kidd, Jr., Christine R. Mason, Andrew Brughera
Hearing Research Center and Communication Disorders, Sargent College, Boston University, 635 Commonwealth Avenue, Boston, MA 02215, USA. gkidd@bu.edu

William M. Hartmann
Dept. of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA

**Summary**
Arbogast *et al.* [1] found a large release from masking obtained by the spatial separation of a target talker and a competing speech masker. Both signal and masker were sentences from the Coordinate Response Measure corpus processed by extracting the envelopes of 15 narrow frequency bands and using the envelopes to modulate carrier tones at the center of each band. By playing nonoverlapping subsets (6–8) of bands from the signal and masker, the energetic component of masking was minimized while the informational component was maximized. The current study extended that work to determine the interaction between reverberation, masker type and spatial release from masking. The stimuli were processed in the same way and were presented in the same spatial configuration as the earlier study. The target sentence was presented at 0-deg azimuth while the masker sentence was played at either 0- or 90-deg azimuth. Noise-masker controls, comprised of overlapping or nonoverlapping frequency bands, were also tested. The listening environment was an IAC booth having interior dimensions of 12’4” × 13’ × 7’6”. Acoustic extremes were achieved by covering all surfaces with materials that were either highly reflective (Plexiglas® panels) or highly absorbent of sound (Silent Source® foam wedges). The results indicated that the amount of masking and the spatial release from masking depended both on the characteristics of the room and the masker type. When the masker was primarily energetic, spatial release from masking decreased from a maximum of about 8 dB in the least reverberant room to about 2 dB in the most reverberant room. For the informational masker, a larger advantage of 15–17 dB was found that was not affected by reverberation. Our interpretation of these findings was that spatial separation of sources could improve speech identification through acoustic filtering by the head, binaural interaction, and the strengthening of perceptual segregation of sound images. However, only the latter effect appears to be relatively insensitive to reverberation.

**PACS no. 43.66.Dc, 43.66.Lj, 43.66.Pn**

## 1. Introduction

Most everyday listening situations consist of multiple sources of sound and multiple paths (reflections) the sounds can take to the listener. When the listener wishes to attend to sounds from a particular source (the “signal” or “target”) and ignore sounds from other sources (“maskers”), the relative location of the sound sources and the reverberation characteristics of the listening environment may be crucial to accomplishing the task. Unwanted sounds from nonsignal sources, and the acoustic reflections of both signal and masker(s), adversely affect signal reception in somewhat different ways. Masking is thought to consist of two separate components – termed “energetic” and “informational” [2] – that roughly correspond to peripheral and central sites of origin within the auditory pathway, respectively. Unwanted sounds from nonsignal sources can produce either or, perhaps more typically, both types of masking.

Acoustic reflections of the signal can sometimes be beneficial in that they increase the signal energy reaching the listener relative to the case where signal energy is largely absorbed by nearby surfaces, but they can also be harmful in that they superimpose on the direct sound altering the waveform. Masking sounds produce reflections too, of course, which can complicate the task of the listener even more. Because the reflections may arrive at the two ears at different times and levels, they may interfere with the binaural advantage found in anechoic or mildly reverberant environments by reducing the extent to which the head effectively attenuates high-frequency sounds and by disrupting the fine timing cues used in binaural analysis. Although there have been studies indicating that the normal

Received 5 February 2004, Revised 11 June 2004,
accepted 1 September 2004.

process of echo suppression, or “precedence,” can interact with informational masking by playing a role in the perceptual segregation of auditory images [3, 4] it is not yet well-understood how this generalizes to the segregation and processing of sounds in actual reverberant environments.

Past work has shown that spatial separation of a signal and masker can provide a significant listening advantage in multisource environments for a variety of tasks including detection (e.g. [5, 6, 7, 8], nonspeech pattern identification (e.g. [9, 10]) and speech recognition (e.g. [11, 12, 13, 14, 15, 16, 3, 4, 1]. When the masking from unwanted sources is primarily energetic in nature, the listening advantage may result from attending to the ear with the more favorable signal-to-noise ratio (the “better ear advantage”), or from binaural analysis (e.g., see recent reviews by Yost, [17], Bronkhorst [18] and Ebata [19]). The principal binaural cues that afford a listening advantage due to spatial separation of sources arise from differences in interaural time-of-arrival at the two ears and the frequency-dependent differences in level at the two ears. Zurek [20] has reviewed much of the relevant work related to speech masked by noise as a function of spatial separation of sources in a free-field environment. He developed a quantitative model to predict masked speech intelligibility that was based on articulation index theory, the acoustical filtering of the head, and the function relating the masking level difference (MLD) to frequency. Further, he considered the performance that would be obtained in a “better ear” case and the differences that might occur for speech recognition materials presented in high vs. low context formats. Although it is an oversimplification to give any single numerical estimate of “the binaural advantage” (the reduction in signal-to-noise ratio at a criterion level of performance due to listening with two ears rather than one), Zurek’s model predicts a maximum benefit of approximately 8–9 dB in an anechoic environment by spatial separation of a target speech signal and speech-spectrum shaped noise masker. This prediction provides a close approximation to the 10 dB maximum advantage found empirically by Plomp ([21]; also [22]).

When the signal is speech from one talker and the masker is speech from a different talker, the situation is more complicated. This is due, in part, to the type of masking that is produced. Because running speech fluctuates in frequency content and amplitude over time, the energetic masking of one talker by another varies from moment to moment and may be difficult to predict. In addition, the informational masking value of the masker may be an important factor in determining the overall amount of masking and may be even more difficult to predict. In the classic “cocktail party problem” [23, 24, 25] the task of the listener is to focus attention on a particular talker while ignoring the speech of other talkers. This requires the listener to perceptually segregate the target speech stream from other speech streams and direct attention to it, which may be challenging if the other speech is difficult to ignore, highly uncertain, or there is a high degree of similarity between signal and masker and/or the signal-masker source characteristics. Dating at least from the work of Carhart *et al.* [26], it has been apparent that factors other than energetic masking are important in speech recognition in multisource environments, and that speech- on-speech masking contains a significant central component (which Carhart *et al.* referred to as “perceptual masking”).

In recent years, efforts to understand the contribution of informational masking to the cocktail party problem have intensified. Freyman *et al.* [3, 4] have demonstrated that the perceptual segregation of one talker from another can significantly reduce the informational masking produced by the unwanted talker. In their experiments, perceptual segregation was achieved by presenting the masker from two spatially separated loudspeakers with a brief time disparity between the sounds. When the temporally-leading masker was in the loudspeaker separated from the target, so that the masker image was pulled away from the target, a significant performance advantage was found relative to the case when the masker was only presented in the same loudspeaker as the target. However, this advantage was only found when the masker was speech not noise, suggesting that a significant amount of the masking observed in the reference condition was informational masking. Brungart [27] and Brungart and Simpson [28] have reported results from a series of experiments in which one or more unwanted talkers interfered with the intelligibility of a target talker in ways that were also attributed to informational masking. One piece of evidence supporting informational masking in many of the conditions they tested was the finding that the preponderance of errors in recognition came from the test words in the speech masker rather than being randomly distributed between test-item alternatives as would be the general expectation for energetic masking. In some cases the interference produced by a masking talker on a target talker presented to the same ear was so great that it apparently disrupted the ability of the listener to ignore unwanted information in the ear contralateral to the target. This caused the contralateral masker to degrade identification performance significantly more than for the ipsilateral masker alone [29]. A breakdown in binaural channel separation in some complex dichotic listening conditions has also been observed in nonspeech tasks as well (e.g. [30, 31]).

Recently, Arbogast *et al.* [1] studied the role of spatial separation of sources for the task of speech identification for maskers that varied along the energetic-informational masking continuum in a mildly reverberant sound field (same as one condition used in this study and described in detail below). The stimuli were derived from the Coordinate Response Measure corpus [32] processed through a modified version of software intended to produce speech that simulates the speech received by listeners with cochlear implants (cochlear implant simulation speech or CIS; [33]). The resulting speech signal was comprised of a set of narrow bands that were sine-wave carriers modulated by the envelopes extracted from the corresponding bands of the speech signal. The way that the energetic-informational distinction was accomplished was by choosing subsets of bands for the signal and masker. When the bands comprising the signal and the masker were mutually exclusive, little energetic masking occurred but large amounts of informational masking were produced. When the masker was a set of noise bands producing minimal informational masking, the noise could either overlap the signal frequency spectrum exactly causing predominantly energetic masking or could be comprised of the same bands as the speech masker (nonsignal bands) to provide a control for the energetic masking produced by the speech masker. Arbogast *et al.* found that the magnitude of the advantage due to spatial separation of sources, for 0 deg vs 90 deg, depended on which type of masker was presented. For the highly energetic (noise) masker, the spatial advantage of about 8 dB was comparable to previous work using speech signals masked by speech-shaped noise in anechoic space. However, for the highly informational masker a much greater advantage was found, averaging about 18 dB across subjects. The interpretation of this large spatial advantage was that the listeners were able to use binaural cues to perceptually segregate the two sources and focus attention on the correct source. Perceptual segregation can provide a significant release from informational masking but would not help in undoing the interactions of signal and masker in the auditory periphery that presumably form the basis for energetic masking.

The contribution of reverberation to masking in multi-source environments, and the role of spatial separation of sources in reducing masking, have been studied in some detail for energetic masking and to a lesser degree for conditions in which significant informational masking likely is present, as for some speech-on-speech masking conditions. In some cases these studies have been conducted in sound fields where the acoustics are manipulated directly and in other cases the reverberation of the acoustic environment is simulated under headphones. From these studies it is quite clear that increasing reverberation decreases the size of the masking level difference due to spatial separation of a pure-tone signal and Gaussian noise masker (e.g. [34]; also see recent work by Zurek *et al.* [35]). And, results from speech discrimination and recognition studies have demonstrated that the advantage of spatial separation of sources for energetic masking of speech also diminishes significantly as reverberation increases (e.g. [36, 37, 38, 39]). However, to date, no one, to our knowledge, has attempted to determine how spatial separation of sources interacts with reverberation in conditions where the proportion of energetic to informational masking is controlled. With respect to studies of speech-on-speech masking in which reverberation was varied, Moncur and Dirks [40] found that binaural listening was superior to near-ear listening, although the differences were modest (roughly 5–14% improvement), for reverberation times ranging from 0 to 2.3 sec for a target speech signal and competing speech masker. Plomp [21] measured speech recognition for connected discourse masked by noise or another talker as a function of both spatial separation of signal and masker and reverberation. He found that a spatial separation of 90 deg improved masked threshold for a speech signal presented with a speech masker by about 5 dB in an anechoic condition and about 2 dB in a highly reverberant (2.3 s reverberation time) environment. The corresponding advantages using a speech-shaped noise masker were also about 5 dB in anechoic space but less than 1 dB in the highly reverberant space. Culling *et al.* [41] and Darwin and Hukin [42] have both demonstrated that adding reverberation can diminish the benefits of interaural differences in speech discrimination tasks. In a recent extension of that earlier work, Culling *et al.* [43] compared speech reception thresholds for sentences in reverberant vs. anechoic conditions at different spatial locations simulated and played through headphones. They found a significant improvement due to spatial separation of sources for both “intonational” (varying $F_0$ contour) and “monotone” (constant $F_0$ contour) stimuli in the anechoic condition with thresholds for the monotone speech some 3–4 dB higher than for intonational speech. In the reverberant case, thresholds increased for both types of speech, especially for the spatially separated presentation, causing the advantage of spatial separation to diminish to less than a decibel.

From the work summarized above it is well established that the release from energetic masking provided by spatial separation of sound sources diminishes as the reverberation in an acoustic space increases. It is much less clear, however, whether the release from informational masking due to spatial separation of sources is likewise affected. In cases where there are multiple simultaneous talkers, the difficulty in determining the effect of reverberation on the informational component of masking is due in large part to the difficulty in controlling for the two types of masking. In the experiments reported in this article, we examined how reverberation affects the release from energetic and informational masking for the task of speech identification. In order to vary the energetic- informational masking distinction, we use the same materials and procedures as Arbogast *et al.* [1].

## 2. Methods

### 2.1. Subjects

The listeners were five young adult college students of ages 19-25 years. Routine audiometric examination indicated that all five listeners had normal hearing. The subjects were paid for their participation in the experiment. All of the listeners had previously participated in other psychoacoustic experiments. Training was minimal. An initial run was conducted in each room condition at 60 dB SPL in order to familiarize the listeners with the task and to assure that their speech identification performance was near 100% correct in unmasked conditions.

### 2.2. Stimuli

The stimuli and signal processing were identical to those used by Arbogast *et al.* [1] and are described in detail in that article. Briefly, both the speech signals and maskers were derived from the CRM test [32] as noted above. The task was forced-choice closed-set speech identification and had the general structure indicated by the phrase "Ready [callsign] go to [color] [number] now." The listener reported the color (1 of a set of 4) and the number (1-8) spoken by the talker who uttered a specific callsign ("Baron"). There were three types of maskers: a talker uttering another phrase from the same corpus, but having a different callsign (e.g., "Charlie" or "Ringo," etc.), color and number from the signal. The other two maskers were multiple bands of noise discussed more fully below.

Both signals and maskers were processed into 15 narrow frequency bands using a modified version of CIS software. The center frequencies of the bands ranged from 215–4891 Hz. The envelopes were extracted from each band and used to modulate pure tone carriers centered in each of the bands. For the noise maskers, the processing was the same except that, after the bands (either 6 or 8 depending on which type of noise masker) were summed, the result was Fourier transformed, multiplied in the frequency domain with white noise and then inverse Fourier transformed. Depending on the type of masker, this yielded noise that had a magnitude spectrum that was nearly identical to the modified CIS speech but was unintelligible.

For the signals, 8 of the 15 available bands were randomly selected on each trial. For the speech masker (called "different-band speech" or DBS), 6 of the 7 remaining bands were selected to comprise the masker. This yielded two sources of speech that were each highly intelligible individually but had very little spectral overlap. For the noise maskers, one type was also 6 bands that were chosen not to overlap the signal bands (termed "different band noise" or DBN) and the other type was chosen to have exactly the same 8 bands as the signal (called "same band noise" or SBN).

### 2.3. Procedures

The task of the listener was to identify the color and number from the sentence having the callsign "Baron." A response was counted correct only if both color and number were reported accurately. Thus, chance performance was 1/32 (4 colors by 8 numbers). Response feedback was given after every trial.

All stimuli were played from two loudspeakers (Acoustic Research model 215) situated 5 feet from the listener approximately at the same height as the head of the seated subject. One loudspeaker was located directly in front of the subject (0 deg azimuth) and the other loudspeaker was located directly to the right of the subject (90 deg azimuth). The signal was always played from the speaker at 0-deg azimuth. The maskers could be played from either speaker although the masker location was held constant throughout any given block of trials. The masker level was set to 60 dB SPL in the BARE room condition (see below). The signal level was varied adaptively using a 1-up 1-down tracking rule that estimates the 50% correct point on the psychometric function. Each experimental run consisted of dual interleaved adaptive tracks that had starting signal levels above and below the level of the masker. Thus, each experimental run (one masker type and location) consisted of a block of 60 trials that yielded 2 threshold estimates. Following training, approximately 8 runs per subject per condition were obtained yielding about 16 threshold estimates per point. Quiet thresholds (50% intelligibility points) for the target sentences alone were measured so that the amount of masking could be estimated in the various masked conditions.

The stimuli were played through 16-bit DACs and an array of programmable analog equipment (Tucker-Davis Technology) at a sampling rate of 50 kHz and were low-pass filtered at 20 kHz. Inverse filters were applied to the stimuli to correct for the measured loudspeaker magnitude responses.

### 2.4. Room characteristics and reverberation estimates

The Sound Field Laboratory (SFL) in which these experiments were conducted consists of a custom-configured single-walled IAC booth having interior dimensions of 12'4" × 13' × 7'6" (length, width, height) located inside a larger room (about 25' × 25' × 10') that houses the computers and supporting experimental equipment. The measured noise floor (bare walls) inside the booth is approximately 27 dBA. The room has been designed so that materials having various sound absorption characteristics can cover all of the surfaces - ceiling, floor, walls and door. When no materials are placed on the surfaces, the interior is like other standard IAC enclosures with a carpeted floor. This condition is the test condition referred to as "BARE" and is the room condition used in the study by Arbogast *et al.* [1]. The other two room conditions were intended to be more and less reverberant than the BARE condition. In one case, the material covering the surfaces was Plexiglas® and is designated "PLEX." In the other case, the material was 8" wedges of polyurethane Silent Source TF-MAX8® foam and that condition is labeled "FOAM."

Three sets of acoustic measurements were obtained at the listener's location with sounds generated from the two loudspeakers in the positions used in the study for each of the 3 room conditions: impulse responses (IRs), modulation transfer functions, and interaural cross-correlation and level difference functions. The IRs and modulation transfer functions were measured using a standard Brüel and Kjær 4192/2669 1/2" microphone suspended from the ceiling pointing downward at the approximate location of the center of the listener's head. The signal from the microphone was then amplified using a Brüel and Kjær 5935L microphone supply, and routed to an analog-to-digital converter (TDT AD1). The interaural cross-correlation and level difference measures were obtained using the Knowles Electronics Mannequin for Acoustics Research (KEMAR) located at the approximate position of the subject's head. The standard Etymotic ER-11 microphones and preamplifiers were used, and their outputs were routed to the TDT AD1 analog-to-digital converter.

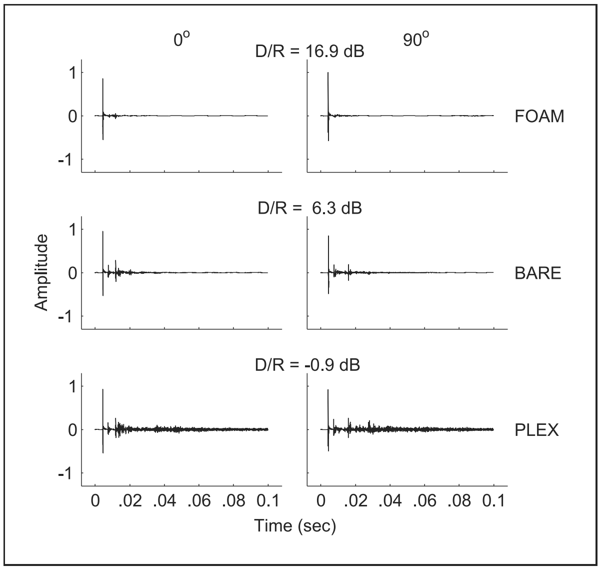


Figure 1. Impulse responses measured at the position of (the center of) the subject's head for FOAM (upper), BARE (middle) and PLEX (lower) room conditions from the two speaker positions (columns). The direct-to-reverberant ratio in dB, averaged across speaker locations, is also shown for each row.

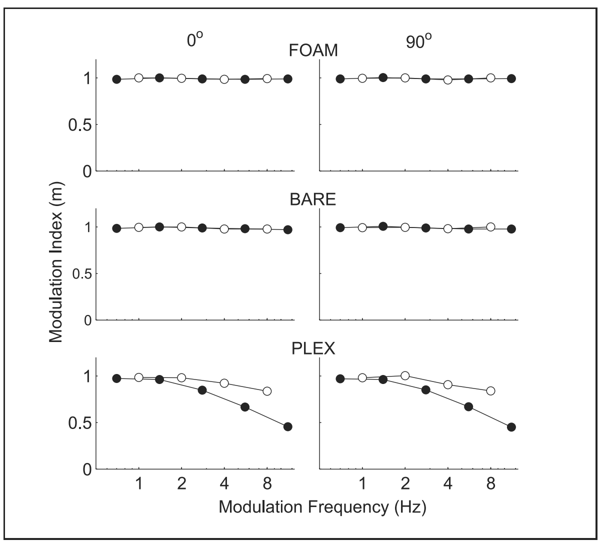


Figure 2. Modulation transfer functions for all three rooms at two noise-band carrier center frequencies: 500 and 2000 Hz (open and filled circles, respectively) and two source azimuths: 0 and 90 deg (left and right columns).

All acoustic measurements were recorded digitally and analyzed by computer using laboratory-written programs calling Matlab functions.

The IRs were based on the average of 500 presentations of a broadband click bandpass filtered from 50 Hz to 20 kHz using a zero-phase digital filter. Each click was presented 80 ms into a recording of 735-ms duration. The modulation transfer functions were obtained using two octave bands of noise as carriers (center frequencies of 500 and 2000 Hz) and the modulation frequencies typically

Table I. Summary of interaural difference measures obtained using KEMAR. The peak values are the peaks in the cross-correlation functions located at the times $T$ ($\mu$sec) indicated. ILD is the Interaural Level Difference (dB) between KEMAR's ears (positive values mean right ear leads in time and/or has higher level).

| | 0° | | | 90° | | |
|---|---|---|---|---|---|---|
| Room | Peak | $T$ | ILD | Peak | $T$ | ILD |
| FOAM | 0.92 | 0 | 0.0 | 0.54 | 680 | 10.2 |
| BARE | 0.87 | 0 | 0.0 | 0.42 | 700 | 8.0 |
| PLEX | 0.56 | 0 | -0.2 | 0.17 | 660 | 3.0 |

specified using RASTI procedures [44]. However, each modulation frequency was tested separately. The cross-correlation functions and interaural level difference measurements were obtained using 1-sec samples of broadband noise.

The IRs are shown in Figure 1 for each room condition and loudspeaker location. To facilitate comparison of the larger early-arriving reflections, only the first 100 ms of the recordings are plotted. The different room conditions yielded IRs that were qualitatively different, with the increasing effect of reflections clearly seen as the room changed from FOAM to BARE to PLEX, and some less obvious effects observed for the different speaker locations. Direct-to-reverberant ratios (D/Rs) were computed on the filtered IRs, and were found to be 16.9, 6.3 and $-0.9$ dB for FOAM, BARE and PLEX conditions, respectively. The D/Rs were measured over the 655 ms duration of the click response, and averaged across the 0-deg and 90-deg speaker locations.

The $m$ values (index of modulation) obtained from the modulated noise measurements are shown in Figure 2 as a function of the modulation frequency. Unlike the IRs and D/Rs, there was very little difference found between FOAM and BARE rooms. This was apparent at both carrier frequencies; however, a decrease in $m$ was observed for the PLEX condition, especially at 2000 Hz for the higher modulation rates.

Cross-correlation and "interaural" level difference measures were obtained using KEMAR for broadband noise played from each loudspeaker in each room condition. Those measurements are summarized in Table I and displayed in Figures 3 and 4. The peak in the cross-correlation function corresponds to the delay between the two ears of KEMAR, so when the source is at 0 deg azimuth, the peaks were at 0 microseconds (1 sample period is 20 microseconds). Further, the magnitude of the peaks of the functions declined from 0.92 in FOAM to 0.87 in BARE and 0.56 in PLEX. When the source was at 90 deg azimuth, the peaks were at about 660–700 microseconds, which is in good agreement with the value one would expect for the human head (e.g. [45]), and the magnitudes of the peaks were 0.54, 0.42, and 0.17, for the three room conditions, respectively. Broadband (0–8 kHz) interaural level differences were also computed. For 0-deg azimuth,

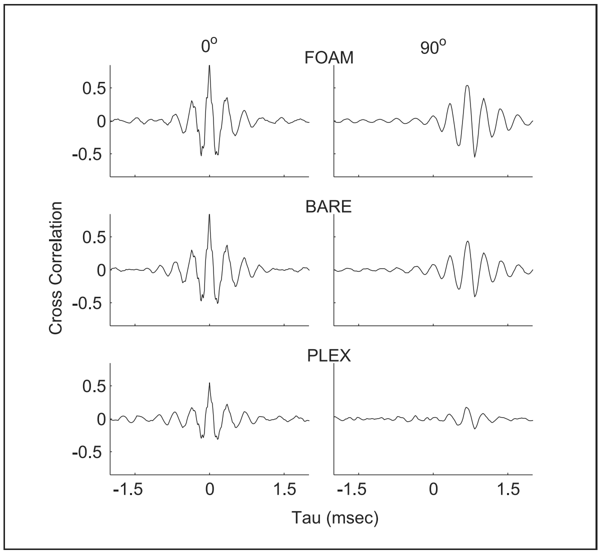


Figure 3. Cross correlation functions measured using KEMAR for both loudspeaker locations (columns) and three room conditions (rows). Relevant values are contained in Table I.

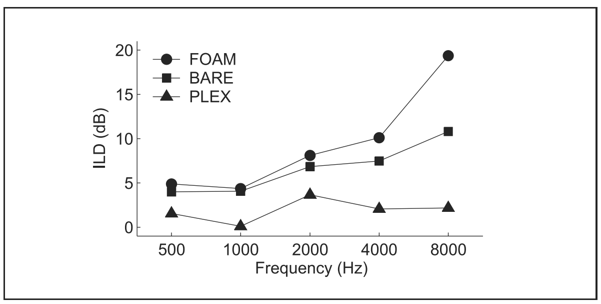


Figure 4. Interaural level differences measured for octave bands of noise for the 90-degree spatial separation condition using KEMAR. The center frequency of each octave band is given on the abscissa while the level difference in dB between ears is given on the ordinate. The three rooms are represented by circles (FOAM), squares (BARE) and triangles (PLEX).

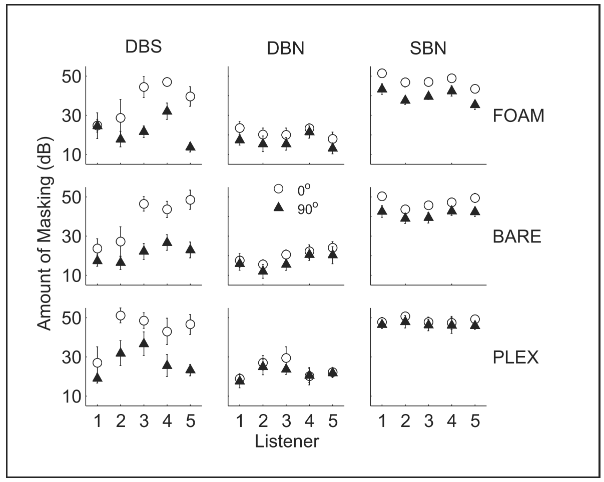


Figure 5. Results from 5 listeners for the three different maskers (columns) and three room conditions (rows) specified as amount of masking (masked threshold minus quiet threshold with threshold defined as 50% correct identification). The open circles are mean thresholds for the 0 deg separation and the filled triangles are for the 90 deg separation conditions. The error bars represent plus and minus one standard deviation of the mean across repetitions.

the measured ILDs were less than 1 dB and are expressed in dB relative to the FOAM condition to eliminate minor asymmetries observed in the calibration channels and pinnae. For the 90 deg azimuth the ILDs were 10.2 dB in FOAM, 8 dB in BARE, and 3 dB PLEX. An octave band analysis was performed on the ILD measurements for 90 deg azimuth, and those results are contained in Figure 4. As above, the values were referenced to the 0-deg measurements in FOAM to eliminate small calibration asymmetries (ranging from $-0.5$ to 2 dB). As expected, the ILDs are greatest at the higher frequencies and decline in all cases with increasing reverberation.

Finally, for a fixed input to the loudspeakers changing the room acoustics changed the overall SPL. It was decided rather arbitrarily that we would specify the levels of the stimuli based on the measurements made in the BARE condition and hold the inputs to the loudspeakers constant in all conditions. The effect of this was that the overall SPL of the stimuli varied somewhat in the different rooms. For a constant-level broadband noise input that produced 70 dB SPL at the head of the listener, the same input produced approximately 69 dB in the FOAM condition and 72 dB in the PLEX condition.

In summary, the acoustic measurements provided a means for quantifying the changes in reverberation observed in the different room conditions. The differences were apparent in all of the measures including reverberation time and direct-to-reverberant ratio, modulation transfer functions, and in the interaural differences of a simulated listener.

## 3. Results

The unmasked adaptive thresholds (50% correct points) did not significantly vary across room conditions, falling between 11.4-12.2 dB SPL using our uncorrected method referenced to the voltage applied to the loudspeakers in the BARE conditions (see above). So, the stimulus levels were within a decibel across room conditions when specified as the level above quiet speech recognition threshold. From this point on the data will be discussed in terms of amount of masking recognizing that the changes in reverberation increased or decreased the SPL of both the signal and masker about the same amount (within a range of about 3 dB).

Figure 5 shows the data from the individual listeners for all conditions. In each panel the mean amount of masking (masked threshold minus quiet threshold) is plotted for 0-deg (open circles) and 90-deg (filled triangles) separation conditions. Listener number is arbitrarily arranged along the abscissa. The three columns are for the three different

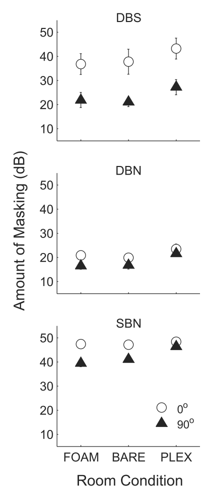


Figure 6. The effect of the different rooms on the amount of masking. Each panel is for a different masker with the rooms plotted along the abscissa. The ordinate is amount of masking computed as group means shown with standard errors of the means. Open circles are for 0-deg spatial separation and filled triangles are for 90-deg spatial separation of signal and masker.

types of maskers and the rows are for the three room conditions. This figure reveals the main effects of interest – how performance is affected by the different maskers and room conditions – and illustrates the differences across subjects. Group mean effects of reverberation and masker type are considered in detail below. With respect to individual differences, it is apparent that there is an interaction between masker type and range of masking amounts across subjects. The ranges of masking amounts for both of the noise maskers in all spatial separation and room conditions were much smaller than the corresponding ranges of masking amounts for the DBS masker. For example, for the DBS masker at the 0-deg separation condition, the range of masking amounts was greater than 20 dB for both FOAM and BARE rooms and about 18 dB for PLEX. In contrast, the range of masking amounts across subjects for the two noise maskers were much smaller falling within about 6 dB for each condition (except for DBN PLEX at 0 deg, where the range was 11 dB). The greater intersubject differences for the DBS masker relative to the two noise maskers is consistent with the expectation that the DBS masker produces primarily informational masking while the DBN and SBN maskers produce primarily energetic masking. Many studies have shown that large intersubject differences in masking are found for informational, as compared to energetic, masking (e.g. [46, 47, 48, 49]) although the intersubject differences for speech tasks have often been found to be smaller than for other tasks (cf. [28, 1]). Large intersubject differences were found in the size of main effects too. For example, consider the release from masking due to spatial separation of signal and masker, which is computed from the difference in amounts of masking for 0 deg and 90 deg separations. For the DBS masker, L1 shows virtually no advantage of spatial separation in FOAM for the DBS masker, due primarily to the small amount of masking found at 0-deg separation, while L3 had a reduction in masking greater than 23 dB. It is apparent from Figure 5 that the range of spatial advantages across subjects was generally much smaller for the DBN and SBN maskers than for the DBS masker.

### 3.1. Effects of Reverberation

The effect of room condition, with the associated change in reverberation, is illustrated for the group-mean data in Figure 6. As in Figure 5, the dependent variable is the amount of masking which is plotted for each room condition (abscissa) for each of the 3 maskers (pamels). Consider first the data from the SBN masker (lower pamels). For 0-deg azimuth, the group means (standard errors) were 47.4 (1.3), 47.2 (1.2) and 48.4 (0.6) dB for the FOAM, BARE and PLEX rooms, respectively. For 0-deg separation, then, increasing reverberation increased the amount of masking by only about 1dB. The group mean amounts of masking for the 90-deg separation were 39.6 (1.5), 41.1 (0.8) and 46.5 (0.4) dB for FOAM, BARE and PLEX, respectively. Thus, as the room became more reverberant, the amount of masking of the signal in the spatially separated masker increased by about 7 dB. Comparison of the results from the two spatial separation conditions for the energetic SBN masker suggests that the effect of reverberation is greater when the signal and masker are spatially separated than when they emanate from the same azimuth.

In contrast, consider the results from the highly informational DBS masker (upper panel). Here, a very different pattern of results was observed than was found for the SBN masker. For the group at 0-deg separation, the mean (standard error) amounts of masking were 36.8 dB (4.4) for FOAM, 37.8 (5.2) dB for BARE and 43.2 (4.3) dB for PLEX. Thus, at 0 deg separation, increasing reverberation increased masked threshold by about 6.4 dB. For the 90-deg separation, the group-mean masking amounts were 21.9 dB (3.1) for FOAM, 21.1 dB (1.9) for BARE and 27.2 dB (3.1) for PLEX. For that spatial separation condition, increasing reverberation increased masking by about 6 dB. Thus, increasing reverberation increased the amount of masking for *both* spatial separation conditions approximately equally for the informational masker.

Finally, for the DBN masker (center panel), which was intended as a control for the (small amount of) energetic masking present in the DBS masker, the group mean amounts of masking (standard errors) at 0 deg separation were 20.9 (1.0), 19.9 (1.5) and 23.5 (2.0) dB SPL for FOAM, BARE and PLEX conditions, respectively, and 16.5 (1.4), 16.8 (1.6) and 21.6 dB (1.3) for the same room conditions at the 90 deg separation. Thus, thresholds varied about 3-5 dB across rooms and were slightly larger for the spatially separated condition.

### 3.2. Spatial Release From Masking

The spatial release from masking (computed as the difference in the amount of masking for 0-deg separation minus the amount of masking for the 90-deg separation) is shown for all conditions in Figure 7. The abscissa is masker type and the ordinate is spatial release in dB.

For the SBN condition, the group mean spatial release from masking decreased from 7.9 dB in the FOAM condition to 2.0 dB in the PLEX condition. For the DBS masker, the spatial release from masking was larger and about the same for the three room conditions: 14.9 dB for FOAM, 16.7 dB for BARE and 16.0 dB for PLEX. And, for the DBN masker, the advantages of spatial separation were 4.4, 3.1 and 1.8 dB, respectively, for the three room conditions.

An analysis of variance was performed on these results with masker type, room condition and spatial separation as main factors. The results indicated that all three main factors were highly significant (masker type: $F(2,8) = 68.1$, $p < 0.0001$; room: $F(2,8) = 24.9$, $p < 0.0001$; spatial: $F(1,4) = 57.7$, $p < 0.005$). Of the possible interactions, only masker type by spatial separation was significant ($F(2,8) = 8.92$, $p < 0.01$) consistent with the trend apparent from Figure 7 that the magnitude of the spatial release from masking depended significantly on masker type.

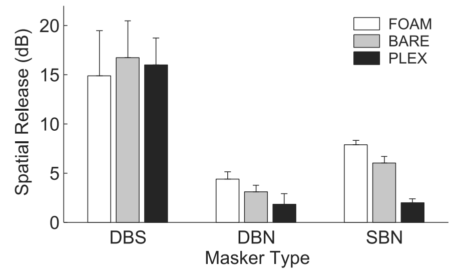


Figure 7. Group mean release from masking due to spatial separation (threshold at 0 deg separation − 90 deg separation) and standard errors of the means for the three maskers in FOAM, BARE and PLEX room conditions.

## 4. Discussion

The three room conditions produced acoustic differences in the sound field that were generally consistent across the different measures. As intended in the design of the rooms, the FOAM condition was the least reverberant, the BARE condition was intermediate and the PLEX condition was the most reverberant. These acoustic differences also significantly affected performance in the speech identification task. First, although unmasked signal thresholds were not affected much by varying reverberation, the amount of masking increased as reverberation increased in all conditions. Averaged over maskers and spatial separations, the PLEX room produced about 8 dB more masking than did the FOAM room. For the energetic SBN masker, the increase was relatively small when signal and masker originated from the same location, but was considerably larger when the signal and masker were spatially separated. The larger increase in threshold with increasing reverberation found for the 90 deg separation condition can be attributed to a loss of interaural timing and level differences which provide a binaural release from masking. Based on the acoustic measurements on KEMAR in section 2.2 above, both interaural time and level differences were greatly reduced for PLEX relative to the other two rooms. Thus, the acoustic advantage due to head shadow that would occur for the FOAM and BARE rooms is much less for PLEX meaning that there was only a small “better ear” advantage. Also, the diminishing of the correlation of the waveforms between ears suggests that binaural analysis would not be very effective in the PLEX room, either. In fact, the peak in the cross-correlation function for the 90 deg separation when measured in the PLEX room condition was only about 0.171. These sound field results are consistent with the finding that the MLD declines as the interaural correlation of the noise masker decreases (e.g. [50, 34, 51]). The current results are also generally consistent with those of Plomp [21] for a speech signal masked by a speech-shaped noise. The spatial advantage we found for the FOAM room was similar to that which he found in an anechoic condition for 90-deg source separation. And, he reported a greater effect of increasing reverberation for the spatially separated conditions than for the case where the two sources had the same azimuth. For example, he reported about a 6 dB increase in thresholds as reverberation was varied for speech masked by noise when both were colocated and about a 10.5 dB increase when the separation was 90-deg. The increase in thresholds he found with increasing reverberation, though, was generally greater than that found here for both conditions, probably due to the much longer reverberation times he imposed on the stimuli.

At the other extreme, for the predominantly informational DBS masker, increasing reverberation by changing room conditions had a relatively large – about 8 dB – and nearly equal effect for both spatial separation conditions. Thus, when masker and target were colocated, increasing reverberation increased the overall amount of masking when the masker was informational but not when it was energetic in nature. This suggests that increasing reverberation made the speech source segregation task more difficult in both spatial conditions. Further, in all three rooms, spatial separation of signal and DBS masker facilitated perceptual segregation of the sources, enhancing the ability of listeners to focus attention on the correct source, leading to a large reduction in the amount of informational masking. What is of particular interest here is that the degradation of the binaural information – interaural time and level differences – noted above as the room conditions changed, apparently had little effect on the benefit of spatial separation of the signal from an informational masker. Although the signal-to-masker ratio needed to achieve criterion identification performance increased by about 8 dB in the PLEX room re. the FOAM room, the advantage of spatial separation was preserved. This finding suggests that the process responsible for improving performance in the spatially separated condition is quite robust with respect to corruption of binaural information. For the energetic masking case, improvements due to spatial separation are thought to be a consequence of binaural interaction due to interaural timing cues and the ability to use the head shadow effect to attend to the ear with the more favorable signal-to-noise ratio. Models of binaural interaction, such as the equalization-cancellation model of Durlach [52, 53], usually assume that the improvement in signal detection – or, by extrapolation, improvements for the task of speech recognition (e.g., [20]) – is a consequence of effectively increasing the signal-to-noise ratio within the frequency channels containing the signal energy. That is, performance improves due to a reduction in energetic masking. However, improving the signal-to-noise ratio in the frequency channels containing the signal energy in the DBS masked condition is not particularly relevant because it is the energy in nonsignal channels that is producing the masking. So the adverse consequence of reverberation on binaural interaction is not relevant for these stimuli. With respect to the headshadow effect, we can only conclude that the loss of the acoustic advantage by focusing attention on the acoustically "better ear" implies that listeners were not depending on that strategy to achieve lower thresholds in the spatially separated condition in the FOAM and BARE rooms. It should be pointed out that the use of the "better ear" as a standard for comparison with binaural performance may be unrealistically conservative in many listening situations as there is some evidence – discussed in the Introduction – that listeners have great difficulty holding the inputs from the two ears separate in many complex listening tasks.

The finding that the large spatial advantage observed for the informational masker appears to be unaffected by a significant degradation of the interaural time and level difference cues suggests that the role of spatial separation of the sources in the perceptual segregation of images is robust and does not depend on the preservation of the steady-state interaural cues. What seems likely is that in reverberant conditions listeners can use the precedence effect to perceive the target as distinct from the masker. The precedence effect derives localization information from signal transients in a way that is very insensitive to increased reverberation [54]. Listeners can then use that location information in the central processing that provides a release from informational masking. By contrast, the precedence effect would not be expected to influence the mechanisms responsible for energetic masking. A similar conclusion was reached by Freyman *et al.* [4] who used lead/lag times to create the perception of a masker originating from a different location than a speech target. When the masker had a high informational masking value, large performance advantages were observed in the perceptually separated masker condition but little, if any, advantage was found for corresponding conditions for an energetic masker. The current work extends Freyman *et al.*'s results to situations where the reflections occur in real rooms rather than being imposed on the stimuli directly.

An important point should be made regarding the comparison of this work to other studies of speech on speech masking. Specifically, the degree to which informational masking is present in the task depends crucially on exactly how the task is structured. For example, one of the inherent problems in designing speech-on-speech masking experiments is how to inform the listener which talker is the signal and which is the masker. One approach, used for example by Hawley *et al.* [16], is to hold the content of the masker constant across trials. The assumption is that the listener can identify the nontarget sentence and direct attention to the other talker. Another approach is that used by Culling *et al.* [43] where the signal and masker are separated by fundamental frequency. The assumption in that case is that the sources can be segregated by $F_0$ and, again, attention directed to the correct source. While such approaches are eminently reasonable and usually successful, they interact with the uncertainty in the listening task, which may affect the expected amount of informational masking present in the experiment. Also, the role of spatial separation in the perceptual segregation process may be difficult to ascertain if it is assumed that other cues are sufficient to accomplish segregation. If the task involves minimal informational masking, then it might not be expected that modes of stimulus presentation that lead to, or strengthen, perceptual segregation of sources would provide a significant performance advantage. At present, there is considerable discussion regarding exactly how to characterize the role of spatial cues in the segregation process (cf. [55, 42, 56]). The findings of this study, while relevant, are not a crucial test of the alternative hypotheses so we will not put forth and evaluate those arguments here. However, the assumptions we make about how the various cues interact with spatial separation are important to state explicitly. In the present experimental paradigm, it is assumed that, in order to solve the task at unfavorable signal-to-masker levels, the listener must follow the speech of the talker who uttered a specific word early in the sentence – the call sign "Baron" – until the test words are presented (otherwise, at favorable signal-to-masker levels the listener might simply report the more salient – i.e., louder – of the target words). This task, combined with the processing we have imposed on the stimuli, is designed to maximize informational masking. There is very little that the subject can use to track the target talker through the sentence. The bands of speech are devoid of harmonic structure and fundamental frequency and hence there is no intonation contour to use to follow the target voice over time. However, even though the sentences all have the same structure and are time-aligned, some prosodic information remains and, throughout a given utterance at least, a timbre that is relatively constant is present for the signal that is clearly different than that of the masker. But given the paucity of information available to the listener it is hardly surprising that presenting the signal and masker from different locations – as long as the locations are perceptually distinct – greatly reduces the listeners' uncertainty about which talker the listener should attend to. A fair question, but one that is difficult to answer, is the extent to which the effects found here occur in real-world listening. If one accepts the premise that masking – either in laboratory conditions or in real-world listening outside of the laboratory – consists of both energetic and informational components that vary in proportion according to circumstance, then studies like this one where we attempt to isolate one type of masking from the other have relevance. The interesting thing here is that the ability to reduce the informational component of masking by spatial separation of sources, however large it may be in any given condition, does not appear to be affected much by reverberation. In contrast, the usefulness of "traditional" binaural mechanisms of attending to the acoustically favored ear and binaural interaction are very much affected by reverberation. Perhaps this implies that in highly reverberant environments we must exert more conscious effort to attend to the correct source and rely more on perceptual factors in communicating than we would in "easy" energetically masked or quiet environments.

One further point to consider has to do with the possibility that subjects are able to learn the acoustic characteristics of specific reverberant rooms and use that information in the segregation process. That is, the experience of the listener in a particular reverberant environment may provide a context in which to interpret even degraded binaural cues. Given the clear role of central factors in informational masking, it is possible that knowledge of the listening environment may be beneficial in reducing uncertainty about the location of sound sources, allowing attention to be directed to the intended source. Even though reverberation greatly disrupts the normal cues used to locate sounds, it may be that we learn the characteristics of reverberant rooms and use that knowledge to disambiguate sources enhancing the ability to perceptually segregate auditory objects.

## Endnote

As pointed out by a reviewer, higher estimates of the magnitude of the cross-correlation between ears may be obtained from different techniques such as computing a running average on a brief time window using the actual stimuli employed in the experiment. Our measurements were higher in most conditions using such a technique although, when we tested the speech samples used in the experiments, the variability increased significantly relative to that which occurred when using samples of noise as the stimulus.

## Acknowledgement

The authors are grateful to Qian-Jie Fu and Robert V. Shannon for providing the software that was modified to generate the cochlear implant simulation speech. They also wish to thank Tanya Arbogast, Christine Carter, Kelly Egan, Erick Gallun, Yosuko Okada and Sally Tressler for their assistance with various tasks related to this project. Work supported by grants DC00100, DC04545 and DC04663 from NIH/NIDCD.

## References

[1] T. L. Arbogast, C. R. Mason, G. Kidd Jr.: The effect of spatial separation on informational and energetic masking of speech. J. Acoust. Soc. Am. **112** (2002) 2086–2098.

[2] I. Pollack: Auditory informational masking. J. Acoust. Soc. Am. **57** (1975) Suppl. 1, S5.

[3] R. L. Freyman, K. S. Helfer, D. D. McCall, R. K. Clifton: The role of perceived spatial separation in the unmasking of speech. J. Acoust. Soc. Am. **106** (1999) 3578–3588.

[4] R. L. Freyman, U. Balakrishnan, K. S. Helfer: Spatial release from informational masking in speech recognition. J. Acoust. Soc. Am. **109** (2001) 2112–2122.

[5] M. Ebata, T. Sone, T. Nimura: Improvement of hearing ability by directional information. J. Acoust. Soc. Am. **43** (1968) 289–297.

[6] K. Saberi, L. Dostal, T. Sadraladabai, V. Bull, D. A. Perrott: Free-field release from masking. J. Acoust.Soc. Am. **90** (1991) 1355–1370.

[7] R. H. Gilkey, M. D. Good: Effects of frequency on free-field masking. Hum. Factors **37** (1995) 835–43.

[8] T. J. Doll, T. E. Hanna: Spatial and spectral release from masking in three-dimensional auditory displays. Hum. Factors **37** (1995) 341–355.

[9] G. Kidd Jr., C. R. Mason, T. L. Rohtla, P. S. Deliwala: Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. J. Acoust. Soc. Am. **104** (1998) 422–31.

[10] T. L. Arbogast, G. Kidd Jr.: Evidence for spatial tuning in informational masking using the probe-signal method. J. Acoust. Soc. Am. **108** (2000) 1803–1810.

[11] I. J. Hirsh: The relation between localization and intelligibility. J. Acoust. Soc. Am. **22** (1950) 196–200.

[12] R. Carhart: Monaural and binaural discrimination against competing sentences. Int. Audiol. **4** (1965) 5–10.

[13] D. D. Dirks, R. H. Wilson: The effect of spatially separated sound sources on speech intelligibility. J. Speech Hear. Res. **12** (1969) 5–38.

[14] A. W. Bronkhorst, R. Plomp: The effect of head-induced interaural time and level differences on speech intelligibility in noise. J. Acoust. Soc. Am. **83** (1988) 1508–1516.

[15] W. A. Yost, J. Dye, R. H., S. Sheft: A simulated "cocktail party" with up to three sound sources. Percept Psyphys. **58** (1996) 1026–36.

[16] M. L. Hawley, R. Y. Litovsky, H. S. Colburn: Speech intelligibility and localization in a multi-source environment. J. Acoust. Soc. Am. **105** (1999) 3436–3448.

[17] W. A. Yost: The cocktail party problem: Forty years later. – In: Binaural and Spatial Hearing in Real and Virtual Environments. R. A. Gilkey, T. R. Anderson (eds.). Erlbaum, NJ, 1997, 329–348.

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

[19] M. Ebata: Spatial unmasking and attention related to the cocktail party problem. Acoust. Sci. & Tech. **24** (2003) 208–219.

[20] P. M. Zurek: Binaural advantages and directional effects in speech intelligibility. – In: Acoustical Factors Affecting Hearing Aid Performance. G. A. Studebaker, I. Hochberg (eds.). Allyn and Bacon, Boston, 1993, 255–276.

[21] R. Plomp: Binaural and monaural speech intelligibility of connected discouse in reverberation as a function of azimuth of a single competing sound source (speech or noise). Acustica **34** (1976) 200–211.

[22] R. Plomp, A. M. Mimpen: Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences. Acustica **48** (1981) 325–328.

[23] E. C. Cherry: Some experiments on the recognition of speech, with one and two ears. J. Acoust. Soc. Am. **25** (1953) 975–979.

[24] I. Pollack, J. M. Pickett: Stereophonic listening and speech intelligibility against voice babble. J. Acoust. Soc. Am. **30** (1958) 131–133.

[25] N. P. Moray: Attention in dichotic listening: Affective cues and the influence of instructions. Q. J. Exp. Psych. **11** (1959) 56–60.

[26] R. Carhart, T. W. Tillman, E. S. Greetis: Perceptual masking in multiple sound backgrounds. J. Acoust. Soc. Am. **45** (1968) 694–703.

[27] D. S. Brungart: Informational and energetic masking effects in the perception of two simultaneous talkers. J. Acoust. Soc. Am. **109** (2001) 1101–1109.

[28] D. S. Brungart, B. D. Simpson: Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. **110** (2001) 2527–2538.

[29] D. S. Brungart, B. D. Simpson: Within-ear and across-ear interference in a cocktail party listening task. J. Acoust. Soc. Am. **112** (2002) 2985–2995.

[30] L. M. Heller, C. Trahiotis: The discrimination of samples of noise in monotic, diotic and dichotic conditions. J. Acoust. Soc. Am. **97** (1995) 3775–3781.

[31] G. Kidd Jr., C. R. Mason, T. L. Arbogast, D. Brungart, B. Simpson: Informational masking caused by contralateral stimulation. J. Acoust. Soc. Am. **113** (2003) 1594–1603.

[32] R. S. Bolia, W. T. Nelson, M. A. Ericson, B. D. Simpson: A speech corpus for multitalker communications research. J. Acoust. Soc. Am. **107** (2000) 1065–1066.

[33] R. V. Shannon, F. G. Zeng et al.: Speech recognition with primarily temporal cues. Science **270** (1995) 303–4.

[34] A. H. Koenig, J. B. Allen, D. A. Berkley, T. H. Curtis: Determination of masking-level differences in a reverberant environment. J. Acoust. Soc. Am. **61** (1977) 1374–1376.

[35] P. M. Zurek, R. L. Freyman, U. Balakrishnan: Auditory target detection in reverberation. J. Acoust. Soc. Am. **115** (2004) 1609–1620.

[36] A. K. Nabelek, J. M. Pickett: Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids. J. Acoust. Soc. Am. **56** (1974) 628–639.

[37] A. K. Nabelek, P. K. Robinson: Monaural and binaural speech perception in reverberation for listeners of various ages. J. Acoust. Soc. Am. **71** (1982) 1242–1248.

[38] J. Koehnke, J. M. Besing: A procedure for testing speech intelligibility in a virtual listening environment. Ear Hear. **17** (1996) 211–217.

[39] S. Devore, B. Shinn-Cunningham: Perceptual consequences of including reverberation in spatial auditory displays. Proc. Int. Conf. Auditory Displays, ICAD, Boston, MA, 2003.

[40] J. P. Moncur, D. D. Dirks: Binaural and monaural speech intelligibility in reverberation. Journal of Speech and Hearing Research **10** (1967) 186–195.

[41] J. F. Culling, Q. Summerfield, D. H. Marshall: Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels. Speech Com. **14** (1994) 71–95.

[42] C. J. Darwin, R. W. Hukin: Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. J. Acoust. Soc. Am. **102** (1997) 2316–2324.

[43] J. F. Culling, K. I. Hodder, C. Y. Toh: Effects of reverberation on perceptual segregation of competing voices. J. Acoust. Soc. Am. **114** (2003) 2871–2876.

[44] T. Houtgast, H. J. M. Steeneken: A review of MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. J. Acoust. Soc. Am. **77** (1985) 1069–1077.

[45] A. W. Mills: Auditory localization. – In: Foundations of Modern Auditory Theory. J. V. Tobias (ed.). Academic Press, New York, 1972.

[46] D. L. Neff, T. M. Dethleffs: Individual differences in simultaneous masking with random-frequency, multicomponent maskers. J. Acoust. Soc. Am. **94** (1995) 125–134.

[47] E. Oh, R. A. Lutfi: Nonmonotonicity of informational masking. J. Acoust. Soc. Am. **104** (1998) 3489–3499.

[48] N. I. Durlach, C. R. Mason, B. G. Shinn-Cunningham, T. L. Arbogast, H. S. Colburn, G. Kidd Jr.: Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. J. Acoust. Soc. Am. **114** (2003) 368–379.

[49] G. Kidd Jr., C. R. Mason, V. M. Richards: Multiple bursts, multiple looks and stream coherence in the release from informational masking. J. Acoust. Soc. Am. **114** (2003) 2835–2845.

[50] D. E. Robinson, L. A. Jeffress: Effect of varying the interaural noise correlation on the detectability of tonal signals. J. Acoust. Soc. Am. **35** (1963) 1947–1952.

[51] M. van der Heijden, C. Trahiotis: Binaural detection as a function of interaural correlation and bandwidth of a masking noise: Implications for estimates of spectral resolution. J. Acoust. Soc. Am. **103** (1998) 1609–1614.

[52] N. I. Durlach: Equalization and cancellation theory of binaural masking level differences. J. Acoust. Soc. Am. **35** (1963) 1206–1218.

[53] N. I. Durlach: Binaural signal detection: Equalization and cancellation theory. – In: Foundations of Modern Auditory Theory, Vol. II. J. V. Tobias (ed.). Academic Press, New York, 1972.

[54] W. M. Hartmann: Localization of sounds in rooms. J. Acoust. Soc. Am. **74** (1983) 1380–1391.

[55] J. F. Culling, Q. Summerfield: Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. J. Acoust. Soc. Am. **98** (1995) 785–797.

[56] W. R. Drennan, S. Gatehouse, C. Level: Perceptual segregation of competing speech sounds: The role of spatial location. J. Acoust. Soc. Am. **114** (2003) 2178–2189.

[57] ISO Standard 3382:1997(E): Measurement of the reverberation time of rooms with reference to other acoustic parameters. 1997.

[58] M. R. Schroeder: Diffuse sound reflection by maximum-length sequences. J. Acoust. Soc. Am. **66** (1975) 497–500.