The influence of informational masking in reverberant, multi-talker environments[a)]

Adam Westermann[b),c)] and Jörg M. Buchholz[b)]
National Acoustic Laboratories, Australian Hearing, 16 University Avenue, Macquarie University, New South Wales 2109, Australia

(Received 18 September 2014; revised 11 June 2015; accepted 22 June 2015; published online 5 August 2015)

The relevance of informational masking (IM) in real-world listening is not well understood. In literature, IM effects of up to 10 dB in measured speech reception thresholds (SRTs) are reported. However, these experiments typically employed simplified spatial configurations and speech corpora that magnified confusions. In this study, SRTs were measured with normal hearing subjects in a simulated cafeteria environment. The environment was reproduced by a 41-channel 3D-loudspeaker array. The target talker was 2 m in front of the listener and masking talkers were either spread throughout the room or colocated with the target. Three types of maskers were realized: one with the same talker as the target (maximum IM), one with talkers different from the target, and one with unintelligible, noise-vocoded talkers (minimal IM). Overall, SRTs improved for the spatially distributed conditions compared to the colocated conditions. Within the spatially distributed conditions, there was no significant difference between thresholds with the different- and vocoded-talker maskers. Conditions with the same-talker masker were the only conditions with substantially higher thresholds, especially in the colocated conditions. These results suggest that IM related to target-masker confusions, at least for normal-hearing listeners, is of low relevance in real-life listening. © *2015 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4923449]

[FJG] Pages: 584–593

I. INTRODUCTION

Many studies have attempted to uncover the factors involved when listening in reverberant multi-talker environments, often labeled as the "cocktail party effect" (Cherry, 1953; Bronkhorst, 2000). The cocktail party effect is a highly complex problem covering acoustical phenomena, auditory masking, attention, binaural processing, and spatial processing. Thereby, masking is often divided between *energetic masking* (EM) and *informational masking* (IM) (Freyman *et al.*, 1999; Brungart *et al.*, 2001; Watson, 2005). Historically, the need for such a differentiation arose from studies like Carhart *et al.* (1969), where higher speech reception thresholds (SRTs) were reported in presence of a speech-masker when compared to a noise-masker. In most studies, EM is defined as masking, which degrades the peripheral representation of the target signal (Cherry, 1953; Kidd *et al.*, 2007), and various auditory models have been successfully applied to account for the effect of EM (Fletcher, 1940; Dau *et al.*, 1996). On the other hand, there are no comprehensive frameworks explaining how the peripheral representation of a mixture of sound sources is formed to a perception of discrete auditory objects, and how attention is steered to select only one specific source out of the mixture of sources. This has lead to the concept of IM, which describes failures of the auditory system to perform these higher-level tasks, being loosely defined as everything that cannot be accounted for by EM. To abrogate such vague explanations, several authors have proposed different definitions of IM (Durlach *et al.*, 2003; Watson, 2005; Shinn-Cunningham, 2008). Watson (2005) specifically divides IM between effects attributed to uncertainty and similarity. Uncertainty is caused by the listener not knowing where to listen (e.g., to which part of the stimulus to attend) and is often linked to experiments with tone-complexes (e.g., Watson *et al.*, 1976). IM due to similarities between target and masker is caused by failures to segregate the target and masker, and is often associated with speech-on-speech masking (e.g., Carhart *et al.*, 1969). Shinn-Cunningham (2008) adopted the concept of auditory scene analysis (Bregman, 1994) to better define IM, in which the auditory system segments and integrates auditory elements into basic objects from which streams are segregated and selected. Mainly, Shinn-Cunningham (2008) introduces a conceptual model that distinguishes between bottom-up salience and top-down attention, and argues that IM is due to failures in either auditory object formation or object selection. Failures in object formation are caused by target-masker similarities hindering basic bottom-up grouping and streaming processes. Failures in object selection are caused by both target-masker similarity and target uncertainty, where similarities can interfere with the correct selection of properly segregated streams and uncertainty can either inhibit direction of top-down attention or draw exogenous attention (e.g., a person saying your name). The present study follows the main principles described by Shinn-Cunningham (2008), assuming that IM either occurs due to (1) confusions caused by the presence of

[a)]Aspects of this work were presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2013.
[b)]Also at: Department of Linguistics, Macquarie University, Building C5A, Balaclava Road, North Ryde, NSW 2109, Australia.
[c)]Electronic mail: adam.westermann@nal.gov.au

maskers and target similarities, or (2) distractions attributable to the presence of maskers that compete with and capture the exogenous attention of the listener.

In a reverberant multi-talker environment, sound sources will often be spread out in a room. Many studies have shown that speech intelligibility increases when a target and masking talker are spatially separated rather than colocated (i.e., on top of each other), resulting in a spatial advantage. This advantage, or spatial release from masking (SRM), can result in differences in SRTs of up to 20 dB and is observed for changes in direction (Freyman *et al.*, 1999; Bronkhorst, 2000), distance (Shinn-Cunningham *et al.*, 2001; Westermann and Buchholz, 2015), and head orientation of the masking talkers (Strelcyk *et al.*, 2014). The degree of masking release is influenced by factors such as the spatial configuration, the number of interfering maskers, and the applied speech corpus (Bronkhorst, 2000; Brungart *et al.*, 2001). To explain SRM, the concepts of EM and IM are often applied (e.g., Freyman *et al.*, 1999; Glyde *et al.*, 2013). Whereas the release from EM may be linked to binaural auditory mechanisms that provide an improvement in "effective" signal-to-noise ratio (SNR), such as better-ear glimpsing or equalization-cancellation processes (Durlach, 1963), release from IM is linked to a perceptual segregation of target and masker signals (e.g., Freyman *et al.*, 1999), which does not necessarily involve a SNR advantage. Following the conceptual framework proposed by Shinn-Cunningham (2008), perceptual segregation can facilitate both the auditory object formation and selection stage. However, when attempting to quantify SRM, and thus the involved binaural mechanisms, most studies rely on the rather unnatural colocated spatial configuration as a reference condition. In particular, in the case that speech maskers are used, such colocated condition provides very few segregation cues and thereby overly increases target-masker confusions, which magnifies IM.

In addition to the issues surrounding the colocated reference condition, a majority of SRM studies use speech corpora, which further increase the effect of IM, such as the coordinate response measure (CRM, e.g., Bolia *et al.*, 2000; Brungart *et al.*, 2001; Best *et al.*, 2013b) or corpora with the same target and masking speaker (e.g., same-voice condition in the listening in spatialized noise-sentences test; Cameron and Dillon, 2007). Generally, for these speech corpora, they involve an unrealistic amount of confusion, from either similarity between talkers or context, which is expected to inhibit auditory object formation and selection. Together, the similarities, due to spatial configuration as well as speech corpora (both talker and contextual), provide not only a substantially increased amount of IM but also in EM, which are both easily resolved by spatial separation and thereby cause the large SRM effect reported in many studies. This leaves the obvious question about how to translate the corresponding SRM results into real-life listening. While the influence and consequence of reduced EM from spatial separation might be clear, the effect of reduced IM is hard to interpret. It is not clear how much IM is actually left when masking talkers have different voices, are located in different places, and room acoustics provide additional cues for speaker segregation, such as distance (direct-to-reverberation ratio), coloration, and changes in speech modulations (Westermann and Buchholz, 2015).

Many studies have tried to separate EM and IM by measuring the effect of each individually. This requires reference maskers that result in only one type of masking. Arbogast *et al.* (2002) and, later, Ihlefeld and Shinn-Cunningham (2008) quantified IM in the CRM corpus by eliminating spectral overlap between target and masker thereby removing EM. The study showed that spatial separation improved thresholds for an IM-only masker by 18 dB, whereas the EM-only masker improved by 7 dB. In other studies, EM has been isolated and IM removed or, at least, reduced by reversing the masking speech (Freyman *et al.*, 2001), mixing genders (Brungart *et al.*, 2001) or constructing unintelligible noise with similar peripheral masking characteristics (Brungart *et al.*, 2001; Best *et al.*, 2013b). Similarly to the goals of the present study, Culling (2013) investigated the influence of IM in a simulated reverberant environment using headphone presentation. He measured SRTs for a varying number of either same-talker, different-talker, or speech-shaped noise maskers and found only small differences between the same- and different-talker masker. However, the speech-shaped noise maskers were stationary and therefore did not fully capture the EM contribution of the different-talker masker. As a consequence, the authors could not draw any conclusions on the overall IM content of this more realistic condition.

With respect to most of the above studies, the effectiveness of the different types of maskers that were chosen as an EM reference can be discussed. The EM, or peripheral representation, is often different between the EM-reference masker and mixed EM and IM masker under investigation. Reversed speech, for example, has very different temporal envelopes, vowel transitions, and pitch contours from normal speech; or broad-band modulated maskers miss the spectro-temporal modulations inherent in speech. Moreover, IM caused by failures in the bottom-up formation process of low-level auditory objects is still likely to be present in all of the EM-reference masker solutions mentioned above. Hence, the EM-reference masker will only partially replicate the amount of EM involved in a given masker signal and additionally might not completely remove all types of IM. Naturally, this needs to be considered when drawing conclusions from such studies.

To better understand the involvement and relevance of IM in reverberant multi-talker environments, the current study implemented a speech intelligibility test in a simulated cafeteria where ongoing background conversations masked the target talker. The cafeteria was simulated using a room-acoustical model and auralized using a three-dimensional loudspeaker array. A number of conditions were designed to investigate (1) the strength of IM in a realistic cafeteria setting by comparing intelligibility scores between a speech masker and an unintelligible vocoded masker, (2) the effect of spatially distributing sound sources throughout the room as opposed to presenting them in the same location, i.e., taking into account the distance, direction, and head-orientation of the masking talkers, (3) the effect of talker similarity, i.e., considering masking talkers with the same and different voices to the target talker, and (4) the effect of different number of masking talkers, which, on the one hand, is expected to change the amount of involved EM and thus the considered SNR range, and, on the other hand, may affect the strength of the observed IM (Freyman *et al.*, 2004).

As IM is the central topic of this study, and there are so many different aspects to IM, it is pertinent to clarify the nature of the IM that is considered here. Since the isolated EM condition is realized by unintelligible vocoded speech, this study mainly examines confusion-based IM. In other words, it considers IM related to target-masker similarities, which causes failures in streaming and object selection, thereby following the lines of literature on SRM (Best *et al.*, 2013b). In turn, this also means that IM, which interferes with basic auditory object formation, is still present in the applied EM-reference condition and IM caused by failures in exogenous attention is not considered.

II. METHODS

A. Subjects

In this study, 17 (13 female, 4 male) subjects with Australian English as their first language participated, all with normal hearing (thresholds $\leq$ 20 dB hearing loss at audiometric frequencies from 250 Hz to 8000 kHz). Mean age was 30 yrs (ages from 18 to 42 yrs), and all subjects reported normal cognitive function. Subjects were either employed at the National Acoustic Laboratories, or they were students at Macquarie University. Those that were not employed at the National Acoustic Laboratories were given a gratuity for their participation. All subjects gave written consent before participating in the study.

B. Stimuli

A sentence test was implemented using the Bamford–Kowal–Bench (BKB) sentence material (Bench *et al.*, 1979). The corpus contains 336 sentences, organized in 21 lists, spoken by a native Australian-English male speaker and sampled at 44.1 kHz. The sentences have a simple syntactical structure (e.g., "The angry man shouted") and an average length of about 1.5 s. The original BKB material is filtered so that its long-term spectra matches the "universal" long-term average speech spectrum (LTASS) defined by Byrne *et al.* (1994). However, this filtering made the sentences sound unnatural when presented inside the simulated cafeteria environment (Sec. II C). Therefore, the unfiltered long-term spectrum of the monologues used for the same-talker speech masker (described below) recorded with the original BKB talker were used to construct a 512-tap finite impulse response (FIR) filter to reverse the LTASS equalization filtering applied to the BKB sentences. This filter was then applied to all of the BKB sentences. After filtering, the sentences sounded more natural and cohesive with the cafeteria background.

The BKB target sentences were used to measure speech intelligibility in 12 different masker conditions. These masker conditions were realized by three versions of the four spatial configurations shown in Fig. 2. The four spatial configurations were all realized in a simulated cafeteria environment using either two or seven two-talker dialogues in either a colocated or a spatially separated configuration. Each of the masking stimuli were about 5 min long and looped during the measurement of each condition. Therefore, the alignment of the target sentence and masking speech was random. The three masker versions differed in the way the individual talker signals were generated:

(i) Different-talker speech masker: The different two-talker dialogues were realized using anechoic recordings of seven scripted dialogues taken from published examinations of the International English Language Testing System (IELTS). The recordings were made in the anechoic chamber of the National Acoustic Laboratories, were about 5 min long, and were spoken by eight female and six male talkers. In the conditions with two dialogues, both dialogues were between a male and a female talker. After recording, they were post-processed so that root-mean-square levels were equal during speech segments, following the procedure outlined in IEC 60268–16 (2011).

(ii) Vocoded-talker masker: In order to create an EM-reference masker, a noise vocoder was implemented and applied to the different anechoic recordings used in the different-talker speech masker described above. The aim of the vocoder was to make speech unintelligible while maintaining the EM components of the different-talker dialogues over time and frequency. In addition, the vocoding process was designed so that it would not destroy localization cues present in the different-talker masker, namely, interaural time and level differences and, therefore, the spatial percept of discrete sources in a room would be maintained. In order to accomplish this, the short-time Fourier transform (20 ms windows and 75% overlap) was used to convert each of the anechoic speech maskers described above to the time-frequency domain. The individual time-frequency representations were then spectrally smoothed across rectangular windows with a width of either one octave for the seven-dialogue condition or two octaves for the two-dialogue condition. The additional smoothing in the two-dialogue condition was applied to ensure that the combined masker signal was unintelligible even with fewer concurrent talkers. By using very short temporal-windows and broad frequency smearing, the transients and inherent level fluctuations were preserved as well as possible. The smoothed magnitude spectrum of the individual talkers was combined with the phase-spectrum from white noise, and the vocoded signal was reconstructed using the inverse short-time Fourier transform. In order to ensure that the long-term spectrum of the vocoded speech masker was identical to that of the corresponding different-talker masker, a spectral matching filter was applied. The filter was implemented as a 512-tap FIR filter using the critical band smoothed spectrum of the different- and vocoded-talker maskers measured with a Brüel and

Kjær (B&K) (Nærum, Denmark) 4134 condenser microphone in the center of the loudspeaker array.

(iii) Same-talker speech masker: To maximize the effect of IM (or target-masker confusions), a same-talker condition was implemented. In this case, all the different two-talker dialogues shown in Fig. 2 were realized by monologues recorded with the same talker as used for the BKB target sentences. The monologues were based on scripts taken from published IELTS examinations. To create two-talker dialogues, each of the monologues was segmented to form dialogues with the same approximate temporal pattern as the different-talker dialogues. The segmentation was done by hand to ensure that the breaks did not occur in the middle of words. Compared to the different-talker dialogues, the created same-talker dialogues did not contain a clear semantical stream, i.e., the location switching within the individual dialogues did not make sense. However, since participants were not continuously following the background dialogues, the lack of semantical validity was not expected to affect target intelligibility.

Figure 1 shows the long-term spectra in critical bands of the three different types of seven-dialogue maskers together with the BKB-sentence material measured with a B&K 4134 condenser microphone in the center of the loudspeaker array. Note that the different- and vocoded-talker maskers have very similar spectra because of the spectral matching described above. The spectrum of the same-talker masker seems to be marginally more similar to the target sentences than the other maskers and, thus, may provide slightly more EM.

C. Spatialization of sounds

A cafeteria scene with multiple masking talkers and a fairly long reverberation time was chosen as it represents a complex scene often encountered in real-life listening. However, environmental non-speech sounds like foot steps, moving of chairs, and noise generated by cutlery and plates were not considered. The acoustic scene was created using the room simulation software ODEON (Rindel, 2000) and subsequently processed with the loudspeaker-based room

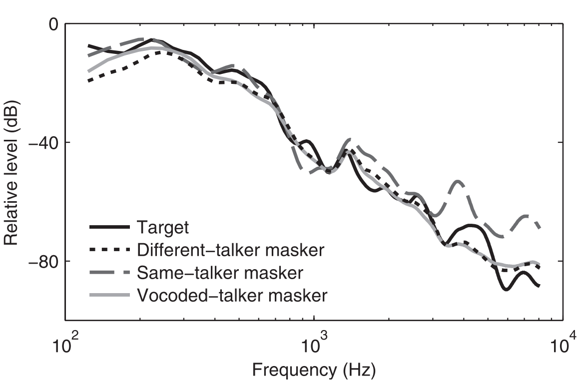

FIG. 1. Long-term spectra in critical bands for each of the three applied speech maskers and the target sentences.

auralization (LoRA) toolbox (Favrot and Buchholz, 2010). The resulting stimuli were presented to the test subjects using the 41-channel loudspeaker system available in the anechoic chamber of the National Acoustic Laboratories.

The simulated cafeteria [shown in Figs. 2(a)–2(d)] was 15 m long by 8.5 m wide by 2.8 m high and had a reverberation time of $T_{30} \approx 0.6$ s. The model included windows, tables, and chairs, and all talkers were simulated by sound sources with a directivity measured on a real talker (i.e., applying ODEON's directivity file Tlknorm_natural.so8). The target speaker was always placed 2 m in front of the listener [corresponding to *T* in Figs. 2(a)–2(d)].

For each of the 15 source-receiver pairs (Fig. 2), reflectograms and decay curves were computed with ODEON. These were converted to room impulse responses (RIRs) and auralized using the LoRA toolbox to create filters for each of the 41 loudspeaker-channels. The toolbox divides the RIRs into three parts accounting for the direct sound, specular early reflections (here, up to third order), and late reverberation. The direct sound and early reflections were mapped to the nearest loudspeaker in the array. The late reverberation was realized by applying uncorrelated noise to the calculated frequency- and direction-dependent decay curves. Details can be found in Favrot and Buchholz (2010). The final target and masker signals were then spatialized by convolving the derived 41-channel filters of the corresponding talker position with the anechoic source signals described in Sec. II B. The combined multi-talker masker scenarios shown in Fig. 2 were then derived by simply adding the individual 41-channel masker signals. Note that each two-talker dialogue is denoted from M_1 through to M_7, i.e., in the bottom left panel all seven two-talker dialogues are colocated with the target.

D. Procedures

Subjects were seated at the center of a spherical loudspeaker array inside an anechoic chamber. The loudspeaker array consisted of 41 Tannoy (Coatbridge, United Kingdom) V8 loudspeakers arranged in multiple rings covering a sphere with a radius of 1.85 m. The loudspeaker responses were individually equalized from the critical band smoothed response measured with a B&K 4134 condenser microphone placed in the center of the array by applying a 1024-tap FIR equalization filter. A height-adjustable chair ensured that the head of the subject was in the exact center of the array. The signal path originated outside the chamber with a personal computer running MATLAB fitted with an RME MADI (Haimhausen, Germany) sound card. This was connected to two RME M-32 digital-to-analog converters which feed 11 four-channel Yamaha XM4180 (Hamamatsu, Japan) amplifiers connected to the loudspeakers through an acoustically dampened passage. In order to communicate with the operator, the subjects wore a lavalier microphone connected to the RME MADI sound card via an RME M-16 analog-to-digital converter. Additionally, a video camera was used to monitor the subjects.

The experiment was conducted in one visit lasting ~1 h. Following an audiometric screening, the listeners were

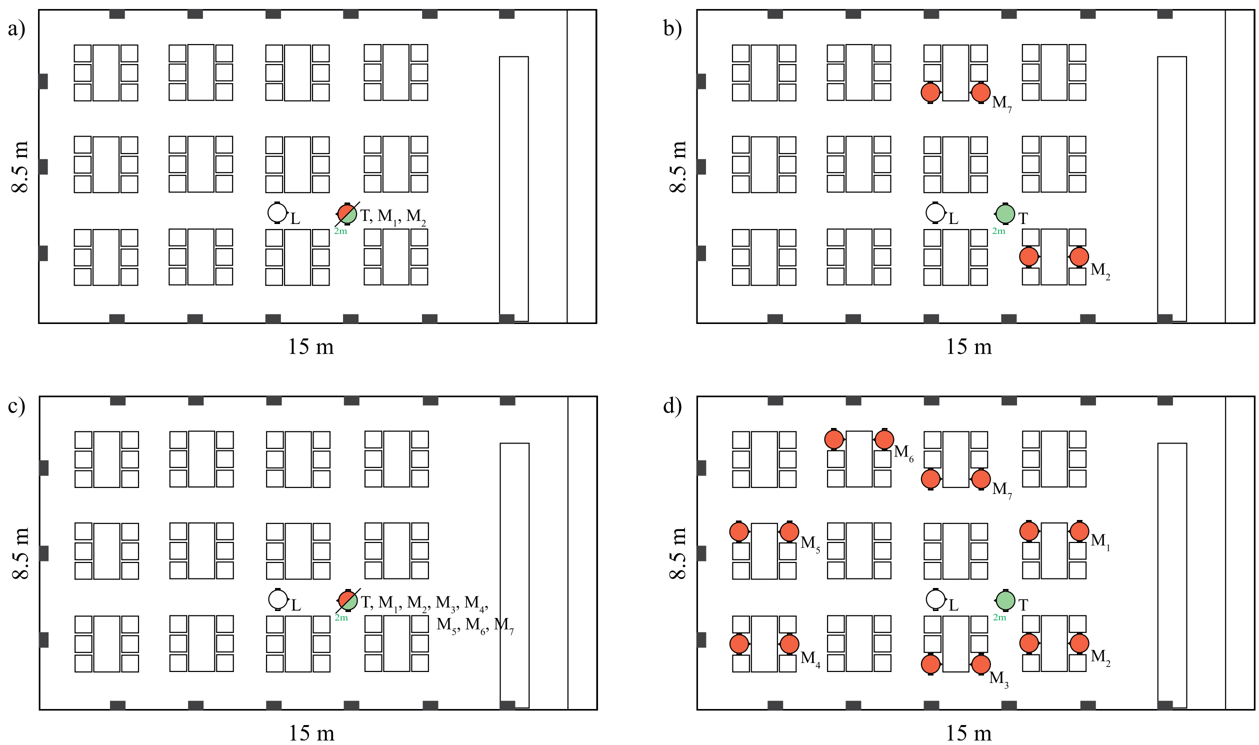

FIG. 2. (Color online) Top-down view of cafeteria simulated in ODEON. The listener position (*L*) faces the target (*T*) at 2 m distance. The two-talker maskers (with each dialogue denoted as $M_1 - M_7$) were either distributed in the room or colocated with the target (*T*).

seated in the loudspeaker array, and the height of the chair was adjusted. The listeners were told to imagine being in a cafeteria environment and instructed to repeat the BKB sentence following a beep. Head movements were allowed, but subjects were instructed to sit still to ensure that they stayed in the "sweet-spot" of the reconstructed sound fields. The SNR resulting in 50% correct performance was adaptively measured using an one-up one-down staircase method (Keidser *et al.*, 2013) varying the level of the target and keeping the masker level fixed at 65 dB(A) for all masker conditions. Sound pressure levels (SPLs) were measured and calibrated *in situ* with a B&K 2250 sound level meter with a 1/4-inch microphone placed in the center of the loudspeaker array, using an integration time of 30 s. The level of the target sentences was initially set to 70 dB(A), which was calculated by first concatenating the entire BKB speech material and then applying the speech level calculation described in IEC 60268–16 (2011), which excludes speech pauses.

Performance was scored morphemically, i.e., for each morpheme in a given sentence. Thus, the SNR decreased or increased when responses were above or below 50% correct morphemes, respectively. If the number of correct morphemes was exactly 50%, the SNR was not changed. The SRTs were calculated using the algorithm presented in Keidser *et al.* (2013). This requires a minimum of 16 presentations with decreasing step sizes of 5, 2, and 1 dB. A run completed when the unbiased standard error, estimated by 2 times the standard deviation of the SNRs over the root of the number of presented sentences, fell below 0.8 dB or the maximum number of 32 sentences was reached (for further details, see Keidser *et al.*, 2013). The BKB-sentence material contains 21 lists of 16 sentences. One list was used for training purposes and results were discarded. The masker in the training was always the different-talker speech masker (Sec. II B) using seven simultaneous dialogues. For each SRT, two randomly chosen lists were combined to 32 sentences. Since the experiment measured 12 SRTs and there were only enough lists for 10 conditions, 2 of the SRTs for each subject reused sentences that the subjects had already been exposed to. To minimize learning effects on overall mean results, the SRTs where sentences had to be reused were balanced over all conditions (i.e., each condition had the same number of SRTs where sentences were heard before). Throughout the experiment, the order of presentation and list/masker combination was randomized. No feedback was provided during testing.

III. RESULTS

The mean SRTs and corresponding 95% confidence intervals measured when applying the two- and seven-dialogue maskers are shown in the top panels of Figs. 3 and 4, respectively. The results are grouped based on the spatial configuration, i.e., colocated or spatially separated. The three masker types containing either different-, vocoded-, or same-talker maskers are marked by the circles, squares, and diamonds, respectively. The difference between the SRTs in the colocated and spatially separated condition, defining the spatial advantage, was calculated individually for each subject. The mean and 95% confidence intervals of the spatial advantage are shown in the lower panels of Figs. 3 and 4. A three-way repeated measures analysis of variance (ANOVA) was applied to all measured results. It showed significance for

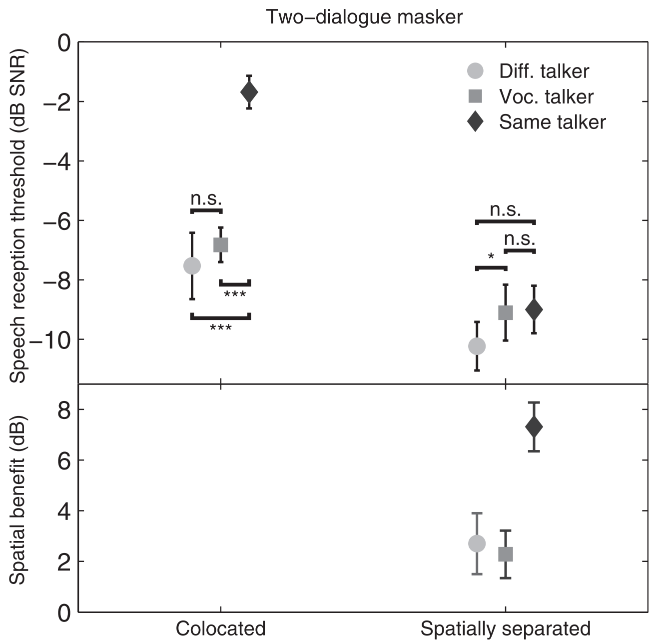

FIG. 3. (Top) Mean and across-subject 95% confidence interval of SRTs (divided between colocated and spatially separated) for different-, vocoded-, and same-talker maskers (circles, squares, and diamonds, respectively). (Bottom) Mean and across-subject 95% confidence intervals of the spatial advantage, i.e., the difference between the spatially separated SRT and the colocated SRT calculated individually for each subject. Stars indicate level of significance between conditions (i.e., $*$, $**$, and $***$ correspond to $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively).

type of masker applied [$F(2,16) = 61.7$, $p < 0.001$], spatial configuration [$F(1,16) = 316.5$, $p < 0.001$], and the number of dialogues in the cafeteria [$F(1,16) = 337.9$, $p < 0.001$]. The ANOVA also showed significance for all types of interaction effects, namely, between type of masker and spatial configuration [$F(2,32) = 29.0$, $p < 0.001$], type of masker and number of dialogues [$F(2,32) = 11.7$, $p < 0.001$], spatial configuration and number of dialogues [$F(1,16) = 10.8$, $p < 0.005$], and,

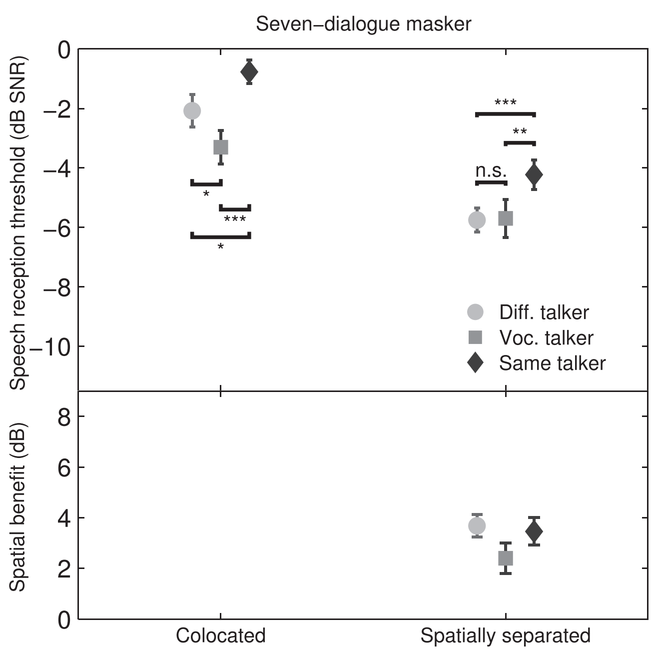

FIG. 4. Same as for Fig. 3, but for the seven-dialogue masker.

finally, for interaction between all three dependent variables [$F(2,32) = 15.2$, $p < 0.001$].

A. Two-dialogue cafeteria

For the two-dialogue masker results shown in Fig. 3, a *t*-test with Bonferroni correction did not reveal a significant difference between the different- and vocoded-talker maskers in the colocated condition ($p = 0.24$), but found weak significance in the spatially distributed condition ($p < 0.05$). However, the same-talker masker was significantly different from both the different- and vocoded-masker in the colocated ($p < 0.001$), but not in the spatially separated condition ($p = 0.11$ and $p = 0.88$ for each masker, respectively).

The lack of significant difference between the SRTs measured with the different- and vocoded-talker masker in the colocated condition suggests that IM has little relevance in the different-talker masker. A similar conclusion can be drawn for the spatially separated condition, where the SRT for the different-talker masker was even lower (by 1 dB) than for the vocoded-talker masker. The small difference may be linked to an increase in EM due to the temporal and spectral smearing applied in the vocoding process, which may reduce dip-listening cues, as well as pitch cues that may have helped speech segregation in the different-talker condition. Reconsidering the colocated results, if such dip-listening cues are indeed reduced with the vocoded-talker masker, they might counteract a small amount of IM in the different-talker masker and consequently produce the non-significant difference in the colocated condition.

In the colocated condition, the SRT for the same-talker masker was 6 dB larger than for both the vocoded- and different-talker maskers. This significant difference indicates that the same-talker masker produced a substantial amount of IM as well as a bit of EM due to more spectral overlap (see Fig. 1 and Sec. II C). When the spatial separation was introduced, the SRTs for the same-talker masker improved by almost 8 dB to a level that was similar to that of the spatially separated different- and vocoded-talker maskers. Here, the similarity in SRTs measured with the same- and vocoded-talker masker suggests that the contribution of EM in these two maskers is roughly equivalent, as the potential IM in the same-talker masker is resolved by the spatial separation between target and maskers. Furthermore, in this condition, the slightly higher SRTs measured with the same-talker masker (albeit not significantly) could indicate that this masker provides slightly more EM than the different-talker masker, as mentioned in Sec. II C.

In general, the SRTs decreased in all conditions when shifting from colocated to spatially separated masker presentation. For the different- and vocoded-talker maskers, this spatial benefit, or SRM, was ~2.5 dB and for the same-talker masker, it was ~8 dB.

B. Seven-dialogue cafeteria

For the results of the seven-dialogue masker shown in Fig. 4, a paired *t*-test with Bonferroni correction revealed that the different- and vocoded-talker SRTs were significantly different in the colocated condition ($p < 0.05$), but not in the spatially separated condition ($p = 0.29$). The same-talker condition was significantly different from both the different- and vocoded-talker maskers in the colocated ($p < 0.05$ and $p < 0.001$, respectively) as well as the spatially distributed condition ($p < 0.01$ and $p < 0.001$, respectively).

In the colocated condition, the SRT for the different-talker masker was ∼1.2 dB higher than for the vocoded-talker masker. This difference was not observed in the two-dialogue masker condition (Sec. III A) and might indicate a minor involvement of IM. However, the difference in SRT between the vocoded and different-talker masker was removed by spatially separating the masking talkers from the target talker.

Comparing the individual SRTs in the two- and seven-dialogue conditions, SRTs in the seven-dialogue condition with the different- and vocoded-talker maskers are substantially higher both in the colocated and spatially separated conditions. This may be explained by the fact that there are fewer temporal and spectral fluctuations when more masking talkers are present, which results in a decreased number of available spectro-temporal gaps with high SNR. Hence, increasing the number of maskers decreases the possibility of dip-listening. However, the spatial benefit was very similar (of ∼ 2.5 dB) for the two- and seven-dialogue background for both the different- and vocoded-talker maskers.

The same-talker masker as in the two-dialogue conditions always resulted in the highest SRTs. However, here the difference between the same-talker and the other two maskers in the colocated condition was reduced to ∼2 dB as opposed to the 6 dB observed in the two-dialogue condition. This substantial reduction in SRM was mainly due to an increase in SRTs in the spatially separated condition, whereas the SRTs in the colocated condition were very similar between the two- and seven-dialogue maskers.

IV. DISCUSSION

Throughout literature it has been shown that differences in talker characteristics, as well as spatial location, resolve target-masker confusions and thereby reduce or even completely remove IM (Bronkhorst, 2000; Brungart *et al.*, 2001; Glyde *et al.*, 2013). In these studies, spatial cues mainly due to angular differences between target and masking talkers were considered, but Westermann and Buchholz (2015) have additionally shown that room-reverberation cues provided by a separation in distance can similarly reduce IM. In Sec. III, both spatial separation in angle and distance, as naturally occurring in the real world, was applied to spatially separate the maskers from the target. This resulted in SRTs, which for the different-talker condition were equal or even lower than for the vocoded-talker condition. This was true for the two-dialogue as well as seven-dialogue masker condition and indicates that the spatially separated different-talker conditions are dominated by EM. The interaction, or relative importance of differences either in talker or location, is exemplified with the two-dialogue masker (Fig. 3) where the different-talker masker shows no or, at least, very little IM in the colocated condition and spatial separation removes IM from the same-talker masker. However, for the seven-dialogue masker (Sec. III) it was found that talker cues alone could significantly reduce but not fully remove IM. The SRT for the different-talker masker in the colocated condition was ∼1.3 dB lower than for the same-talker masker, but still 1.2 dB higher than for the vocoded-talker masker. This difference was removed by providing spatial cues. Hence, in realistic scenarios where both spatial and talker cues are available, listeners can rely on both cues to severely reduce or even completely remove IM. Throughout the experiment, while some IM effects might have been present and could have been teased apart by substantially increasing the amount of test subjects, the current study indicates that such effects would be minor with questionable practical relevance. In addition, listeners in real-life communication would also often have access to visual cues, which have been shown to be even more effective in reducing IM with speech maskers (Helfer and Freyman, 2005). Thus, confusion-based IM, as it is discussed here and in numerous other studies (Sec. I), seems to have a negligible effect on speech intelligibility in reverberant multi-talker scenarios when one or more of these cues are available.

However, these findings rely on two assumptions: (1) that the vocoded-talker masker does not result in IM and (2) that the EM contributions of the different- and vocoded-talker maskers are the same. According to the definition of IM set fourth in Sec. I (and several other studies: Kidd *et al.*, 2007; Ihlefeld and Shinn-Cunningham, 2008) to successfully remove confusion-based IM from the vocoder-talker masker, the vocoder processing needs to ensure that no target speech segments are confused with the masker. An informal listening test found that the masker was unintelligible and the vocoding method made the target clearly stand out, but no formal testing was done. Other studies have applied similar processing to construct EM-only maskers, but only applied the broadband envelope to the noise with the masker long-term spectra (Brungart *et al.*, 2001; Best *et al.*, 2013b). However, this type of processing does not capture any of the fluctuations of speech within each frequency band, thereby likely violating the second assumption. The vocoder used in this study applied spectral smoothing (one or two octaves) that was wider than the auditory critical bands in order to ensure that the masking speech was unintelligible while maintaining most within frequency band fluctuations. However, there is no clear solution to how to best solve this trade-off when designing a vocoder for creating EM-only stimuli that does not violate the two assumptions and thus the parameters were chosen heuristically. Furthermore, as noted in Sec. I, this type of EM-only reference may include some IM occurring in the basic auditory object formation stage.

For the two-dialogue cafeteria conditions, the different-talker SRT was significantly lower than the vocoded-talker SRT ($p < 0.05$) in the spatially separated condition. This further confirms this different-talker condition is not affected by IM, but also suggests that the vocoded-talker masker provides more EM than the different-talker masker. If this is the case, then this effect might have been counteracted in the colocated condition by a small amount of IM resulting in the insignificant difference between SRTs measured with the different- and vocoded-talker maskers. Several studies have shown that SRTs measured with modulated speech-shaped noise (similar to single-channel vocoders) maskers are higher than SRTs measured with "irrelevant" different-talker maskers (Festen and Plomp, 1990; Qin and Oxenham, 2003; Bernstein and Grant, 2009). The difference in SRTs found in these studies are between 1 and 5 dB, which seems to be on the larger side when compared to the differences observed in this study. In the present study, the discrepancy between vocoded- and different-talker SRTs might be explained by the vocoding process smearing the signal over time and frequency and thereby reducing dip-listening, as well as spatial cues, in particular cross-ear glimpsing cues. Glyde *et al.* (2013) showed how reduced spectral resolution, following a (moderate) hearing loss, significantly reduces cross-ear glimpsing cues. This resulted in increases in SRTs of ∼1–2 dB in their spatially separated condition (±90 degrees). The reduction in spectral resolution following a hearing loss is similar to the effect of spectral smoothing applied during the vocoding process and, thus, a similar increase in SRT may be expected here.

In the colocated condition, the SRTs for the same-talker masker were substantially higher than for the vocoded-talker, clearly indicating involvement of IM. This is expected, since the target-masker similarities, both in terms of voice and spatial location, cause failures in streaming and auditory object selection. For the two-dialogue cafeteria condition, the measured spatial benefit was ∼8 dB with the same-talker masker. This benefit is comparable or slightly lower than benefits reported in other SRM studies with two masking talkers (e.g., Freyman *et al.*, 1999; Bronkhorst, 2000; Glyde *et al.*, 2013). However, these others studies compared colocated masker conditions with two maskers that were spatially separated to +/−90 degrees in an anechoic environment. In the current study, the two-dialogue maskers were not spaced symmetrically (see Fig. 2) as one masker (M_2) was positioned considerably closer to the target than the other (M_7) masker. In addition, room reverberation was included here, which has been shown to reduce the effect of SRM (Kidd and Mason, 2005). Finally, the maskers employed here highly supported listening in dips, which is significantly reduced in corpora such as the CRM because of time-aligned target and maskers.

The measured SRM for the different- and vocoded-talker maskers was ∼3 dB, both in the seven- and two-dialogue condition. Other studies that applied maskers that mainly resulted in EM found spatial benefits of up to 8 dB (Kidd and Mason, 2005; Best *et al.*, 2013b). However, Kidd and Mason (2005) measured the SRM with several different levels of reverberation and, in their most reverberant condition, the spatial benefit was only 3 dB. Plomp (1976) conducted similar experiments and found spatial benefits of 3.2 dB in a room with a reverberation time $T_{30} = 0.4$ s. Generally, they explained the smaller spatial benefit by a reduction in interaural fluctuations caused by reverberation. The results found in this study are in good agreement with those found in the studies that considered a significant amount of room reverberation.

Generally, the SRTs for the two-dialogue masker were lower than for the corresponding seven-dialogue masker conditions. Other studies that measured the effect of number of masking talkers have found similar behavior (Brungart *et al.*, 2001; Freyman *et al.*, 2004). This effect is commonly linked to the increasing advantage of dip listening as the masker fluctuations increase with decreasing number of masking talkers. However, this is only the case when EM is dominant, but when IM is additionally involved, this increase in SRT with increasing number of talkers is counteracted by at least two additional mechanisms. As the number of masking talkers increases, the masker becomes more noise-like. In consequence, stream formation of individual maskers that can be confused with the target speech is less likely to occur and the maskers tend to form a single, fused background stream instead. In addition, if the number of masking talkers increases and the total masker SPL is kept constant, the level of each talker decreases in relation to the target and, as a result, the difference in loudness between target and individual masking speakers increases. This difference in loudness provides a strong segregation cue that very much limits the occurrence of IM at high SNRs or, more accurately, at high target-to-masker energy ratios (TMRs) (Agus *et al.*, 2009).

The observation that the applied SNR has an effect on the occurrence or strength of IM has already been discussed by other studies. Best *et al.* (2013b), for instance, argued that in the case of spatially separated target and masker, the auditory system is able to fully segregate the target from the masker signals and, as a consequence, SRTs are dominated by EM effects. Hence, when sufficient speech segregation cues are available, SRTs seem to be limited by EM and no IM can be observed. Similarly, Brungart *et al.* (2001) and Agus *et al.* (2009) showed that when the level of the target in reference to each individual masker, as defined by the TMR, exceeds 0 dB TMR, the effect of IM dramatically drops. Naturally, above 0 dB TMR, the target is louder than each individual masker and loudness cues can be used for talker segregation. Hence, IM seems to have a limited dynamic range, which is limited at low SNRs by EM and above 0 dB TMR by loudness cues. The exact details are complicated and will depend on a large number of scene-related factors (e.g., the number, sex, and spatial configuration of masking talkers), as well as subject-related factors (i.e., hearing ability). This complexity might be reflected in the highly significant two- and three-way interaction effects found in the ANOVA (Sec. III) for each of the dependent variables (number of talkers, spatial configuration, and masker type).

Since in real-life the scene-related factors are dictated by the encountered acoustic scene, this leaves the obvious question: which TMRs are encountered in real-life listening? TMRs applied in typical anechoic SRM experiments are easily estimated, but this estimation becomes harder when including room acoustics and masking sources at varying distances. For the scenario applied in this study (Fig. 2), ODEON supplies the predicted levels of each individual source. Assuming normal vocal effort for the target and masking talkers, the predicted TMR in the spatially distributed cafeteria scene with seven-dialogues [Fig. 2(d)] ranges between 1.8 dB (masker M_3 facing toward the listener) and 6.5 dB (masker M_1 facing away from the listener). When combining each of the masking dialogues by averaging their predicted levels, the predicted SNR is -4.5 dB. Hence, even at negative SNRs, the TMRs are still positive.

Smeds *et al.* (2014) investigated the SNRs that listeners typically experience in their daily life and found most of the relevant SNRs observed in multi-talker, cafeteria-like environments to be positive (~3 dB SNR). Hence, TMRs in such conditions would be very positive and loudness cues are abundant for target segregation. Hence, in most challenging, multi-talker environments encountered in real-life, the involvement of confusion-based IM is even more unlikely than in the experiments considered here.

A. Perspectives

It can be argued that morphemic sentence tests are a poor representation of real-life communication. Not only do they lack conversational dynamics and listener involvement, they also disregard any form of comprehension. In addition, they neglect attention switching between different target sources and do not take into account listening effort. Several studies have looked into increasing the amount of realism in speech tests. In addition to applying more true-to-life SNRs, speech material and comprehension tasks that mimic real-life conversations are desirable. Best *et al.* (2013a) compared sentence recall and comprehension in the same complex cafeteria environment as applied in this study [seven-dialogue cafeteria; Fig. 2(d)]. They measured comprehension in 18 normal-hearing and 28 hearing-impaired listeners by conducting an on-going questionnaire assessing the listeners understanding of the monologues presented in the cafeteria. However, they found a strong correlation ($r = 0.77$, $p < 0.001$) between the comprehension scores and the SRT measured with a sentence test (same as applied here).

Cognition and IM have often been linked (Kidd *et al.*, 2007), but as far as the authors are aware, no studies have shown a significant correlation between the individual susceptibility to IM and cognitive measures (i.e., Glyde *et al.*, 2012). However, some studies have applied a dual-task paradigm and found a relation between working memory capacity and SRM (Helfer *et al.*, 2010). It could be of interest to apply a similar methodology to the study presented here, especially if it was possible to create scenarios with distraction-based IM, thus, investigating other types of IM as defined in Sec. I. This type of IM could be driven by the salience or novelty of a stimulus (e.g., Knudsen, 2007), adding cues in the masker familiar to the subject (such as their name, e.g., Wood and Cowan, 1995) or by including maskers closer to the listener than the target (e.g., Westermann and Buchholz, 2015). Attention- or distraction-related IM may well be observed in real-world environments, but this needs to be further investigated.

Overall, further investigations will be required to generalize the contribution of IM to other real-life listening environments. In particular, it would be interesting to increase the possibility of talker confusions by using a less reverberant room, maskers more similar to the target (e.g., all male talkers), or other spatial configurations. The reproduction method limits the minimum distance of simulated sound sources to the distance of the loudspeakers in the array (here, 1.85 m), although real-world listening often occurs at nearer distances, which could significantly change the applied TMRs. In order to generalize the results, more conditions with both close target and maskers are needed. In addition, the study could be expanded to include subjects with a hearing impairment. Since hearing loss is often accompanied by reduced temporal, spectral, and spatial resolution, these subjects may be more susceptible to IM or show an increased dynamic range of IM. If this group exhibits significant susceptibility to IM, it might be possible to design algorithms for hearing devices that reduce IM.

V. SUMMARY AND CONCLUSIONS

This study investigated the role of confusion-based IM in a simulated, reverberant cafeteria environment by systematically varying the similarity between target and masking talkers as well as their spatial configuration. The results demonstrated the following:

(1) Significant IM was only observed in the colocated condition with the same-talker masker (i.e., when the target and masking talkers were all the same person).
(2) Differences in either location or talker resolved target-masker confusions and effectively removed IM. It was further argued that the involvement of IM is even less likely when visual cues are additionally included and increased SNRs are considered, as in many real-world environments, the SNR is slightly higher than the SNRs considered in this study.
(3) The SRM observed in the simulated cafeteria environment is considerably smaller than the SRM typically reported in literature, where the effect of room reverberation is often excluded and symmetrical two-talker masker conditions are considered. The SRM with the two-dialogue same-talker masker was about 8 dB, but only ~3 dB for all other maskers. The diminished spatial benefit was explained by reduced interaural differences and lack of IM.

Overall, this study suggests that, at least in the considered cafeteria environments, IM due to target-masker similarity plays no or only a minor role. Hence, these results question the real-world relevance of IM encountered in psychoacoustic experiments when the target and masker are colocated and corpora with excessive confusions are used. However, this does not question the relevance of these studies for clinical purposes or in general terms. Moreover, the present study focused mainly on confusion-based IM, and did not explicitly consider aspects such as attention switching, distraction-related IM, or IM encountered at a basic bottom-up grouping stage. It may well be that these aspects of IM play a significant role in real-world listening, in particular, when listeners with a hearing impairment or cochlear implant are considered.

ACKNOWLEDGMENTS

This work was supported by Widex A/S (Lynge, Denmark), the Australian Government Department of Health, and an International Macquarie University Research Excellence Scholarship (iMQRES).

Agus, T. R., Akeroyd, M. A., Gatehouse, S., and Warden, D. (**2009**). "Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise," J. Acoust. Soc. Am. **126**, 1926–1940.

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

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**, 108–112.

Bernstein, J. G. W., and Grant, K. W. (**2009**). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **125**, 3358–3372.

Best, V., Keidser, G., Buchholz, J. M., and Freeston, K. (**2013a**). "Psychometric effects of adding realism to a speech-in-noise test," in *Proceedings of Meetings on Acoustics*, Vol. 19, p. 050067.

Best, V., Thompson, E. R., Mason, C. R., and Kidd, G. (**2013b**). "An energetic limit on spatial release from masking," J. Assoc. Res. Otolaryngol. **14**, 603–610.

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

Bregman, A. (**1994**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (The MIT Press, Cambridge, MA), pp. 47–184.

Bronkhorst, A. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acta Acust. Acust. **86**, 117–128.

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," J. Acoust. Soc. Am. **110**, 2527–2538.

Byrne, D., Dillon, H., and Tran, K. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Cameron, S., and Dillon, H. (**2007**). "Development of the listening in spatialized noise-sentences test (LISN-S)," Ear Hear. **28**, 196–211.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

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

Culling, J. F. (**2013**). "Energetic and informational masking in a simulated restaurant environment," in *Basic Aspects of Hearing*, edited by B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, and H. E. Gockel (Springer, New York), Vol. 787, pp. 511–518.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Durlach, N. (**1963**). "Equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **426**, 416–426.

Durlach, N., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (L)," J. Acoust. Soc. Am. **113**, 2984–2987.

Favrot, S., and Buchholz, J. (**2010**). "LoRA: A loudspeaker-based room auralization system," Acta Acust. Acust. **96**, 364–375.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

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

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

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," J. Acoust. Soc. Am. **106**, 3578–3588.

Glyde, H., Buchholz, J., Dillon, H., Best, V., Hickson, L., and Cameron, S. (**2013**). "The effect of better-ear glimpsing on spatial release from masking," J. Acoust. Soc. Am. **134**, 2937–2945.

Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (**2012**). "The effects of hearing impairment and aging on spatial processing," Ear Hear. **34**, 15–28.

Helfer, K., and Freyman, R. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.

Helfer, K. S., Chevalier, J., and Freyman, R. L. (**2010**). "Aging, spatial cues, and single- versus dual-task performance in competing speech perception," J. Acoust. Soc. Am. **128**, 3625–3633.

IEC 60268-16 (**2011**). *Sound System Equipment—Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index* (International Electrotechnical Commission, Geneva, Switzerland).

Ihlefeld, A., and Shinn-Cunningham, B. (**2008**). "Spatial release from energetic and informational masking in a selective speech identification task," J. Acoust. Soc. Am. **123**, 4369–4379.

Keidser, G., Dillon, H., Mejia, J., and Nguyen, C.-V. (**2013**). "An algorithm that administers adaptive speech-in-noise testing to a specified reliability at selectable points on the psychometric function," Int. J. Audiol. **52**, 795–800.

Kidd, G., and Mason, C. (**2005**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acta Acust. Acust. **91**, 526–536.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2007**). "Informational masking," in *Auditory Perception of Sound Sources* (Springer, New York), pp. 143–189.

Knudsen, E. I. (**2007**). "Fundamental components of attention," Annu. Rev. Neurosci. **30**, 57–78.

Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acta Acust. Acust. **34**, 200–211.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Rindel, J. (**2000**). "The use of computer modeling in room acoustics," J. Vibroengin. **3**, 219–224.

Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cognit. Sci. **12**, 182–186.

Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. (**2001**). "Spatial unmasking of nearby speech sources in a simulated anechoic environment," J. Acoust. Soc. Am. **110**, 1118–1129.

Smeds, K., Wolters, F., and Rung, M. (**2014**). "Estimation of signal-noise ratios in realistic sound scenarios," J. Am. Acad. Audiol. **26**, 183–196.

Strelcyk, O., Pentony, S., Kalluri, S., and Edwards, B. (**2014**). "Effects of interferer facing orientation on speech perception by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **135**, 1419–1432.

Watson, C. S. (**2005**). "Some comments on informational masking," Acta Acust. Acust. **91**, 502–512.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.

Westermann, A., and Buchholz, J. (**2015**). "The effect of spatial separation in distance on the intelligibility of speech in rooms," J. Acoust. Soc. Am. **137**, 757–767.

Wood, N., and Cowan, N. (**1995**). "The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel?," J. Exp. Psychol.: Learn., Mem. Cognit. **21**, 255–260.

Acoustical Society of America. Journal

Table of Contents:	Click here to view
ISSN:	0001-4966
Title:	Acoustical Society of America. Journal — Additional Title Information
Publishing Body:	Acoustical Society of America
Country:	United States
Status:	Active
Start Year:	1929
Frequency:	Monthly
Volume Ends:	Jun - Dec
Document Type:	Journal; Academic/Scholarly
Refereed:	Yes
Abstracted/Indexed:	Yes
Media:	Print
Alternate Edition ISSN:	1520-8524, 1520-9024, 0163-0962, 0163-0970
Size:	Standard
Language:	Text in English
Price:	USD 1,865 combined subscription per year domestic to institutions (Print & Online Eds.) USD 1,990 combined subscription per year in the Americas to institutions (Print & Online Eds.) USD 2,010 combined subscription per year elsewhere to institutions (Print & Online Eds.) Free to members (effective 2009)
Subject:	PHYSICS - SOUND
Dewey #:	534
LC#:	QC221
CODEN:	JASMAN
Circulation:	9000 unspecified
Special Features:	Includes Advertising, Abstracts, Illustrations, Patents
Article Index:	Index, cum.index: vols.1-94, 1929-1993
Pages per Issue:	700
Cols per Page:	2
Composition:	Offset
Editor(s):	Allan D Pierce (Editor-in-Chief)
URL:	http://asa.aip.org/jasa.html
Description:	Covers all phases of research and engineering of interest to acoustical scientists and engineers like linear and nonlinear acoustics; aeroacoustics, underwater sound and acoustical oceanography; ultrasonics and quantum acoustics; architectural and structural acoustics and vibration; speech, music and noise; psychology and physiology of hearing; engineering acoustics, sound transducers and measurements; bioacoustics, animal bioacoustics and bioresponse to vibration.

ADDITIONAL TITLE INFORMATION

Alternate Title:	Medline Abbreviated title: J Acoust Soc Am; Abbreviated title: J A S A