Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources

Mathieu Lavandier[a)]
Université de Lyon, Ecole Nationale des Travaux Publics de l'Etat, Département Génie Civil et Bâtiment (CNRS), Rue M. Audin, 69518 Vaulx-en-Velin Cedex, France

Sam Jelfs
Welsh School of Architecture, Cardiff University, Bute Building, King Edward VII Avenue, Cardiff CF10 3NB, United Kingdom

John F. Culling
School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, CF10 3AT, United Kingdom

Anthony J. Watkins, Andrew P. Raimond, and Simon J. Makin
Department of Psychology, Reading University, Reading, RG6 6AL, United Kingdom

(Received 19 July 2010; revised 19 October 2011; accepted 20 October 2011)

When speech is in competition with interfering sources in rooms, monaural indicators of intelligibility fail to take account of the listener's abilities to separate target speech from interfering sounds using the binaural system. In order to incorporate these segregation abilities and their susceptibility to reverberation, Lavandier and Culling [J. Acoust. Soc. Am. **127**, 387–399 (2010)] proposed a model which combines effects of better-ear listening and binaural unmasking. A computationally efficient version of this model is evaluated here under more realistic conditions that include head shadow, multiple stationary noise sources, and real-room acoustics. Three experiments are presented in which speech reception thresholds were measured in the presence of one to three interferers using real-room listening over headphones, simulated by convolving anechoic stimuli with binaural room impulse-responses measured with dummy-head transducers in five rooms. Without fitting any parameter of the model, there was close correspondence between measured and predicted differences in threshold across all tested conditions. The model's components of better-ear listening and binaural unmasking were validated both in isolation and in combination. The computational efficiency of this prediction method allows the generation of complex "intelligibility maps" from room designs. © *2012 Acoustical Society of America*. [DOI: 10.1121/1.3662075]

PACS number(s): 43.55.Hy, 43.66.Pn, 43.71.Gv, 43.66.Dc [LMW] Pages: 218–231

I. INTRODUCTION

Human listeners show remarkable abilities to segregate speech from noisy backgrounds, so-called "cocktail-party listening" (Cherry, 1953), compared with even the most sophisticated of automatic speech recognition systems (Lippmann, 1997). Nevertheless, segregation is often severely impaired by sound reflections in rooms (Bronkhorst, 2000). Purely acoustical measures of temporal smearing of speech are useful in determining overall intelligibility in many reverberant spaces, especially where reverberation levels are sufficiently high for smearing to be the overriding factor (Bradley *et al.*, 1999; Houtgast and Steeneken, 1985). These essentially monaural measures account for the effect of diffuse ambient noise, but they neglect the listener's abilities to separate target speech from interfering sounds using the binaural system, as well as the deleterious effect of reverberation on these abilities (Beutelmann and Brand, 2006; Culling *et al.*, 2003; Lavandier and Culling, 2007; Plomp, 1976). In the presence of discrete interfering sources, when source segregation becomes the overriding factor, intelligibility can be reduced at relatively low levels of reverberation, and thus more readily than would be predicted from the temporal smearing of speech (Lavandier and Culling, 2008). This paper presents a binaural model which can efficiently predict speech intelligibility in rooms, in the presence of several discrete noise sources.

Cherry (1953) used the term "cocktail-party" to illustrate a general class of situations where a listener attempts to understand target speech among competing-sound interferers. Other examples include open-plan offices and open-plan classrooms, where competing sources can be other people talking, or any other sound source that might mask the target (e.g., an air conditioner or road noise from an open window). Possessing two ears is useful for understanding speech in these situations. Comprehension is improved by better-ear listening and binaural unmasking (Bronkhorst and Plomp, 1988), both of which rely on differences in the intensity and timing of the sound at the two ears—interaural level and time differences (ILDs and ITDs, respectively). For sources located to one side of a listener, the sound level is reduced at the far ear—the ear for which the head throws an acoustic shadow, creating an ILD. In addition, because the sound must travel farther from the source to the far ear, it arrives later, generating an ITD. Target and interferers at

[a)]Author to whom correspondence should be addressed. Electronic mail: mathieu.lavandier@entpe.fr different locations often produce different ILDs so one ear will usually offer a better target-to-interferer level ratio than the other, and listeners can simply attend to whichever ear offers the better ratio. Differences in the ITD and ILD generated by the target and an interferer also provide for binaural unmasking, in which the central auditory nervous system is able to "cancel" to some extent sounds generated by this interferer [equalization-cancellation (E-C) theory; Durlach, 1972], thus improving the internal target-to-interferer level ratio. Better-ear listening and binaural unmasking are both frequency-dependent. The binaural advantage produced by the combination of these two components of binaural hearing to unmask the target speech from a spatially-separated interferer is called spatial unmasking.

In rooms, sound reflections reduce the magnitude of acoustic shadowing on which better-ear listening depends (Plomp, 1976). The modification of source spectra by room "coloration" (resulting from the constructive/destructive interference of sound reflections and the frequency-dependent absorption characteristics of room materials) directly influences speech intelligibility (for a review see Ratnam *et al.*, 2003), but it also influences better-ear listening by creating frequency-dependent ILDs varying with position (Lavandier and Culling, 2010). Moreover, reflections impair binaural unmasking by decorrelating the interfering sound at the two ears. The interaural coherence of a sound source evaluates the similarity of the waveforms it produces at the two ears. The source coherence in a room is degraded by the multiple sound reflections reaching the listener (Hartmann *et al.*, 2005), because these reflections are not identical at the two ears (as long as the configuration is not perfectly symmetrical). An E-C mechanism would be less effective against an interferer that is not perfectly correlated, because a less correlated interferer cannot be fully equalized at the ears, and hence cannot be fully canceled. As a result, there is more masking and lower speech intelligibility (Licklider, 1948; Robinson and Jeffress, 1963). Lavandier and Culling (2007, 2008) showed that binaural unmasking and target intelligibility decreased when interferer coherence was decreased, either by increasing the listener-interferer distance or making the room more reverberant. Room reflections also modify the signal phases at the ears, further affecting binaural unmasking which depends on the interaural phase differences of target and interferer (Lavandier and Culling, 2010).

Different approaches have been proposed to predict the deleterious effects of reverberation on intelligibility in cocktail-party situations. Van Wijngaarden and Drullman (2008) extended the speech transmission index method to take into account binaural hearing. This approach offers the advantage of predicting the smearing effect of reverberation on the speech target. However, it also makes the initial assumption that the target is the only source of modulation in the signals reaching the listener's ears. This approach does not offer any opportunity for extension to more realistic cases where interferers are modulated noise or speech, because in these cases, the modulation is now coming from both the target and the interferer and this attribute no longer distinguishes them. Zurek *et al.* (2004) proposed a model predicting the detection of a narrow band noise target against a broadband noise interferer in rooms, which could be extended to predict speech intelligibility. The model is based on room statistics (surface area and average absorption coefficient of the room, assuming a perfectly diffuse reverberant sound-field, independent of the direct sound). Binaural detection of the narrow band noises was quite accurately predicted, even if some discrepancies remained. These discrepancies could be linked to the initial approximations inherent in the use of room statistics rather than room impulse responses, and of a fixed interaural correlation function, independent of the position considered in the room, rather than the measured interaural coherence.

Three binaural models based on the E-C theory have been proposed recently to predict intelligibility against a discrete noise source (Beutelmann and Brand, 2006; Lavandier and Culling, 2010; Wan *et al.*, 2010). Following Durlach (1972) or vom Hövel (1984), the models of Beutelmann and Brand (2006) and Wan *et al.* (2010) use a direct implementation of an E-C process. The stimuli simulated at the ears are first processed through an E-C stage which tests different delays and attenuations for these signals, and chooses those maximizing the effective target-to-interferer ratio. The speech intelligibility index (SII) method (ANSI S3.5, 1997) is then used to evaluate intelligibility. The model of Wan *et al.* (2010) gave accurate predictions for speech intelligibility against up to three noise interferers, but it was only tested in anechoic conditions. Beutelmann and Brand (2006) obtained very good agreement with listening test data involving single noise interferers in three different rooms, with an overall correlation coefficient of 0.95 between measurement and prediction. The agreement was even slightly better with the revised version of this model (Beutelmann *et al.*, 2010), which was further extended to deal with non-stationary noise. Lavandier and Culling (2010) obtained similar agreement following a different approach proposed by Levitt and Rabiner (1967) and Zurek (1993). Better-ear listening and binaural unmasking are modeled as two separate components. The direct implementation of cancellation is replaced by a predictive equation similar to those developed by Durlach (1972), and the resulting prediction of binaural unmasking is added to a better-ear target-to-interferer ratio. Like the models of Beutelmann and Brand and Wan *et al.*, this method is based on the signals produced by sources in rooms, requiring averaging across signals (i.e., across time) to predict reliably the effect of interfering sources. Jelfs *et al.* (2011) further improved the computational efficiency of Lavandier and Culling's method, by applying their model directly to binaural impulse responses, thus producing fast and accurate non-stochastic predictions.

None of these models have been tested using multiple interferers in reverberation. Anechoic studies have shown binaural hearing to be efficient against multiple interferers (Bronkhorst and Plomp, 1992; Carhart *et al.*, 1969; Culling *et al.*, 2004; Hawley *et al.*, 2004) and both Wan *et al.* (2010) and Jelfs *et al.* (2011) have successfully modeled such data. Lavandier and Culling (2010) showed that their model accurately predicts the effect of binaural unmasking in reverberation, as well as the effect of room coloration on better-ear listening, but their experiments involved simplified virtual rooms. Moreover, the effect of head shadow was not involved because the listener was modeled without a head. Broadband ILDs were also removed by equalizing the stimuli independently at each ear. Beutelmann and Brand (2006) showed that their model was accurate using real-room reverberation, but only for single interferers. The present study investigates situations involving multiple interferers in a variety of spatial configurations in the reverberation from real-room acoustical measurements, and asks whether the revised model of Jelfs *et al.* (2011) can predict the effects of both binaural unmasking and better-ear listening in these conditions. Moreover, the individual effects of ITDs and ILDs were modeled for these cases.

The prediction method was tested against measured differences in speech reception threshold (SRT) (the level of the target compared to that of the interferer for 50% intelligibility). For SRT measurements, real-room listening over headphones was simulated by convolving anechoic stimuli with binaural room impulse responses (BRIRs) (Watkins, 2005; Zahorik, 2002). These BRIRs were measured in different rooms with dummy-head transducers that had the directional characteristics of a human talker and listener. In some conditions, spectral-envelope impulse responses (SEIRs) were used. These SEIRs were obtained by removing the temporal characteristics of the BRIRs whilst preserving their spectral envelopes. This manipulation removed the ITDs necessary for binaural unmasking while preserving the frequency-dependent ILDs necessary for better-ear listening, thereby allowing the two prediction components to be tested separately.

Reverberation affects binaural speech segregation mechanisms, but when speech interferers are involved, it also impairs intelligibility by affecting monaural segregation mechanisms (Lavandier and Culling, 2008). Room reflections can disrupt the segregation of competing voices based on fundamental frequency differences (Culling *et al.*, 2003, 1994). They can also fill the potential silent periods in the speech interferers which otherwise allow one to hear the target better (Bronkhorst and Plomp, 1990; George *et al.*, 2008). To study the influence of reverberation on binaural hearing without mixing it with these additional effects, the experiments presented here used only continuous speech-shaped noise interferers which had no modulation in their temporal envelope. Like the other models presented above [except the one of van Wijngaarden and Drullman (2008)], the prediction method tested here does not consider the potential smearing of target speech in very reverberant environments, so the prediction only holds for targets not too far from the listener in these environments, at positions where the direct-to-reverberant ratio is not too low and segregation from interferers is the overriding factor for intelligibility. Thus, the experiments presented below involved only near-field targets.

Experiments 1 and 2 assessed the prediction method for the case of single stationary noise interferers affected by various levels of reverberation. In experiment 3, the method was confronted with multiple interferer situations, involving one to three stationary noise interferers in reverberation. In each case, the model's components of better-ear listening and binaural unmasking were tested both in isolation and in combination. The last two sections of the paper present intelligibility maps of virtual rooms to illustrate the efficiency and modularity of the prediction method, and then its limitations and the improvements required before periodic and modulated speech interferers could be handled are discussed.

II. GENERAL METHODS

A. Prediction method

The prediction method is based on the model of Lavandier and Culling (2010) revised by Jelfs *et al.* (2011), an extension of the anechoic models of Levitt and Rabiner (1967) and Zurek (1993). The better-ear listening and binaural unmasking components are predicted independently, from the BRIRs measured between the sources and listener positions. Better-ear listening is estimated from the target-to-interferer ratios (TIRs) computed as a function of frequency at each ear, selecting band-by-band the ear for which the ratio is higher. Ratios are weighted according to their relevance for speech (ANSI S3.5, 1997), and integrated across frequency to provide a broadband "better-ear target-to-interferer ratio" in dB. Binaural unmasking is estimated from the interaural phase differences of target and interferer (Φ_T and Φ_I) and the interaural coherence of the interferer (ρ_I). The binaural masking level difference (BMLD) is obtained in each frequency band using Eq. (1) proposed by Culling *et al.* (2004, 2005) following a development of the E-C theory (Durlach, 1972),

$$BMLD = 10\log_{10}([k - \cos(\Phi_T - \Phi_I)]/[k - \rho_I]) \tag{1}$$

with

$$k = (1 + \sigma_\varepsilon^2)\exp(\omega^2\sigma_\delta^2) \tag{2}$$

and $\omega =$ center frequency of the band in rad/s, $\sigma_\delta = 105$ μs and $\sigma_\varepsilon = 0.25$ (standard deviations of the time and amplitude jitters, respectively, characterizing the internal noise in the E-C model; Durlach, 1972). It should be noted that, following the "revised" model of Durlach (1972), the model assumes that the sound source is sufficiently distant so that interaural level differences are negligible at low frequencies where binaural unmasking is effective. The ILDs of target and interferer are thus not included in Eq. (1). However, in order to broaden the model's application, it would be desirable to take into account the detrimental effect that masker ILDs can have on binaural unmasking (Egan, 1965). The accuracy of the present version of the model relies on the fact that, in the cases considered, the magnitude of low-frequency ILDs is quite small. Where Eq. (1) returns a negative value, the BMLD is set to zero, following the assumption that binaural thresholds are never above either of the corresponding monaural thresholds (Durlach, 1963). The BMLD values are then weighted (ANSI S3.5, 1997) and integrated across frequency to provide a broadband binaural unmasking advantage. To predict the overall effect of binaural hearing, the "effective" target-to-interferer ratio is obtained by adding the binaural unmasking advantage to the better-ear ratio [Eq. (3)],

$$\text{Effective TIR} = \sum_i w_i \times TMR_i + \sum_i w_i \times BMLD_i, \quad (3)$$

where i is the frequency-band index, w_i is the SII weighting of the band (ANSI S3.5, 1997), TMR_i is the better-ear TMR selected independently for each frequency band, and $BMLD_i$ is computed with Eq. (1).

The BRIRs are decomposed into simulated peripheral frequency channels using a gammatone filterbank (Patterson *et al.*, 1987) with two filters per equivalent rectangular bandwidth (ERB) (Moore and Glasberg, 1983). The target-to-interferer ratio for each channel is calculated as the energy ratio between the filtered BRIRs for target and interferer. In the single-interferer case, the filtered BRIRs for each source are cross-correlated to derive the interaural parameters for the application of the EC model. The coherence is taken as the maximum of the cross-correlation function, and the phase difference is obtained by multiplying the corresponding delay by the center frequency of the band. In the case of multiple interfering sound sources, the interferer BRIRs have to be combined into a single binaural pair. The BRIRs are concatenated rather than added to avoid constructive/destructive interference. Concatenation has the effect of summing the frequency-dependent energy of each contributing impulse response, and generating an averaged cross-correlation function. It may seem intuitively reasonable to add together the BRIRs, just as one would add together different interfering sounds. However, summing directly the BRIRs would result in spectral distortion due to interference, which does not occur when summing statistically independent interfering signals that have been convolved with those BRIRs. The BRIRs are not themselves independent, because they were produced by the same source in the room: the impulse used to measure them. Concatenation is the appropriate approach when the interfering sources are independent. Only in the particular case of different interfering sources driven by the same signal (e.g., different loudspeakers driven by the same input) should the BRIRs be summed, to take into account the interference of the signals produced by the sources at the ears.

The method was used to predict measured differences in SRT, without any model parameter being fitted to the data. To be compared to SRTs, which are by definition speech-to-noise ratios, effective target-to-interferer ratios are simply inverted, so that high ratios correspond to low thresholds. This comparison assumes that a reduction by 10 dB of the interferer level (at a fixed target level) induces a 10-dB improvement in SRT. This assumption might not hold for very high or very low source levels. Predicted differences in inverted effective ratio can be directly compared to SRT differences across experimental conditions. To compare absolute thresholds and ratios rather than relative differences, a reference needs to be chosen. For each experiment presented here, the reference was the average SRT across conditions and participants. Before the comparison, inverted ratios were centered to this average SRT (by subtracting their mean and adding the average SRT), or, in other words, the average inverted ratio was aligned to the average SRT of the experiment. It should be noted that, unless one needs to model listeners with different receptive capacities or speech varying in intelligibility (differing in word frequency or the presence of syntactic and/or semantic constraints), there is no requirement to calculate speech indices such as articulation index (Kryter, 1962) or SII (ANSI S3.5, 1997), or to conduct index-to-intelligibility mapping (Beutelmann and Brand, 2006; Levitt and Rabiner, 1967).

B. SRT measurements

SRTs for 50% intelligibility were measured with headphones using an adaptive threshold task in which listeners transcribed semantically unpredictable English sentences heard against spatially separated noise interferers. Real-room listening over headphones was obtained by convolving anechoic stimuli with BRIRs.

1. Stimuli

The anechoic recordings of the same male voice digitized at 20 kHz with 16-bit quantization were used as the basis of all target speech sentences in the three experiments. The corpus of sentences was from the Harvard Sentence List (IEEE, 1969). The sentences have low predictability, and each sentence contains five key words. For instance, one sentence was "TAKE the WINDING PATH to REACH the LAKE." The speech-shaped noise interferers were obtained by filtering continuous Gaussian noises with a finite impulse response filter designed to match the speech long-term excitation pattern (Moore and Glasberg, 1983). These interferers all lasted longer than the longest target sentence.

Binaural stimuli were produced by convolving the speech sentences and noise samples with the room impulse responses measured between the source positions and each ear (see BRIR measurements section below). Within a given room, the relative amount of reverberation imposed on a source was increased by moving the source further away from the listener (Lavandier and Culling, 2007; Watkins, 2005). Convolution by a room impulse response can change the sound level of a stimulus differently depending on the source position in the room and the ear considered (Bradley *et al.*, 1999). Because the impulse response measurements did not preserve the broadband sound level differences between positions and rooms,[1] the broadband target-to-interferer level ratio was fixed at the ears rather than at the emission of the sources. The left-right average of the root-mean-square (RMS) powers of all convolved stimuli was equalized before the experiments. As a consequence, every source produced the same average sound level at the ears. This level was independent of the room considered and of the distance of the source from the head. The equalization preserved the potential influence of ILDs in better-ear listening. Multiple interferers were obtained by summing equalized single interferers corresponding to independent noise samples, and by re-equalizing the resulting signal so that they had the same mean level across the ears as single interferers.

2. Procedure

SRTs were measured using a 1-up/1-down adaptive threshold method (Plomp and Mimpen, 1979). For each SRT measurement, ten target sentences were presented one after another against the same noise interferer. The target-to-interferer level ratio was initially very low (−32 dB). On the first trial (first target sentence), listeners could either enter a transcript on a computer keyboard, or replay the same stimuli. If stimuli were replayed, the target level was increased by 4 dB. Stimuli had to be replayed until the target was loud enough to be judged partially intelligible. Listeners were instructed to attempt a transcript of this first target sentence when they believed that they could hear more than half the words of the sentence. Once the first transcript was entered, the correct transcript was displayed on the computer terminal, with the five key words in capitals. The listener self-marked the number of correct key words. The subsequent nine target sentences were presented only once, and self-marked in a similar manner. The target level was decreased by 2 dB if the listener correctly identified three or more of the five key words in the previous sentence, and otherwise increased by 2 dB. The SRT for a given condition was taken as the mean target-to-interferer level ratio on the last eight trials.

Each SRT measurement used a different set of ten target sentences and a different noise interferer. The session began with two practice runs using unprocessed stimuli, in order to familiarize listeners with the task. The following runs measured SRTs in each of the *N* tested conditions in a randomly chosen order ($N = 12$ in experiments 1 and 2, $N = 16$ in experiment 3). The order of the conditions was then rotated for successive listeners, while sentence materials remained in the same order. Each target sentence was thus presented once to every listener in the same order and, across a group of N listeners, a complete rotation of conditions was achieved. Each experiment therefore used a multiple of N listeners. This procedure also ensured that each condition was presented in each serial position within the experimental session.

Signals were digitally mixed, D/A converted, and amplified using a 24-bit Edirol UA-20 sound card and an MTR HPA-2 Headphone Amplifier. They were presented to listeners over Sennheiser HD480 headphones in a single-walled sound-attenuating booth within a sound-treated room. A computer terminal screen was visible outside the booth window. A keyboard was inside the booth to gather the transcripts of listeners.

3. Listeners

Listeners all reported normal hearing and English as their first language. They were undergraduate students, paid for their participation. None of them were familiar with the sentences used during the test. Each listener participated in only a single session of a given experiment. Experiments 1 and 2 each involved 24 listeners, and 32 listeners took part in experiment 3. For each experiment, mean SRTs are presented with standard errors.

C. BRIR measurements

Real-room reverberation was introduced into the anechoic stimuli by convolution with BRIRs measured in different rooms with dummy-head transducers (a speaker in a Bruel and Kjaer 4128 head and torso simulator, and Bruel and Kjaer 4134 microphones in the ears of a KEMAR mannequin), which incorporate the directional characteristics of a human talker and a human listener. The BRIRs were measured using doubled maximum-length sequences in a corridor and an L-shaped room (Watkins, 2005), and using log sine sweeps (Farina, 2000) in two meeting rooms and a lecture hall. To obtain signals at the listener's eardrum that match the signal at KEMAR's ear, the frequency-response characteristics of the dummy-head loudspeaker and of the listener's headphones were removed using appropriate inverse filters. All measurements were done at 48 kHz, and BRIRs were resampled at 20 kHz before convolution with the anechoic stimuli.

BRIRs were obtained with the transducer mannequins facing each other, both on stands to fix their height at 1.53 m. The talker's position was varied to give different distances from the listener (0.65, 1.25, 2.5, 5, or 10 m), at a selection of bearings (−25°, −5°, 0°, 5°, or 25°). These bearings and distances are relative to the listener's fixed location. In the L-shaped room and meeting rooms, the listener was located near a corner facing diagonally across the room. In the corridor, the listener was central and faced along the room. In the lecture hall, the listener was where the lecturer would normally stand, i.e., near one wall half way along it, facing the opposite wall. The amount of reverberation at these locations is indicated by the ratio of early-to-late impulse response energy, C50 when "early" is defined as the first 50 ms of the impulse response (ISO 3382, 1997). The present measurements do not comply with the ISO standard's recommendations for omni-directional transducers and spatial averaging, because the purpose here was to capture features present for listeners. A-weighted C50 values measured in the five rooms[2] are shown in Fig. 1, which also indicates the shape and size of each room.

As expected, C50 systematically decreased with increasing source distance, indicating that the relative amount of reverberation increased when the source was moved away from the listener in the five rooms. The C50 value at a given distance was of course dependent on the room considered. C50 was very similar at the two ears for frontal sources at 0°; but for lateral sources, it was higher at the ear which was on the side of the source (left ear for a source at −25° and right ear for a source at 25°) compared to its level at the contra-lateral ear. This difference indicates that head shadow reduced the level of the direct sound at the contra-lateral ear. The difference of C50 across the ears decreased with increasing source distance, suggesting that the influence of head shadow was limited when the sound at the ears was dominated by reverberation. Figure 1 finally demonstrates that the rooms and positions considered in this study gave access to a broad range of reverberation levels.

To be sure of the good quality of the BRIR recordings, their measurement-noise level was assessed. The presence of noise is indicated by non-linear (dB vs time) energy decay (Zahorik, 2002). For each BRIR, the energy decay curve was obtained by reverse integration of the impulse response (Schroeder, 1965). The BRIR's amplitude resolution ("bit

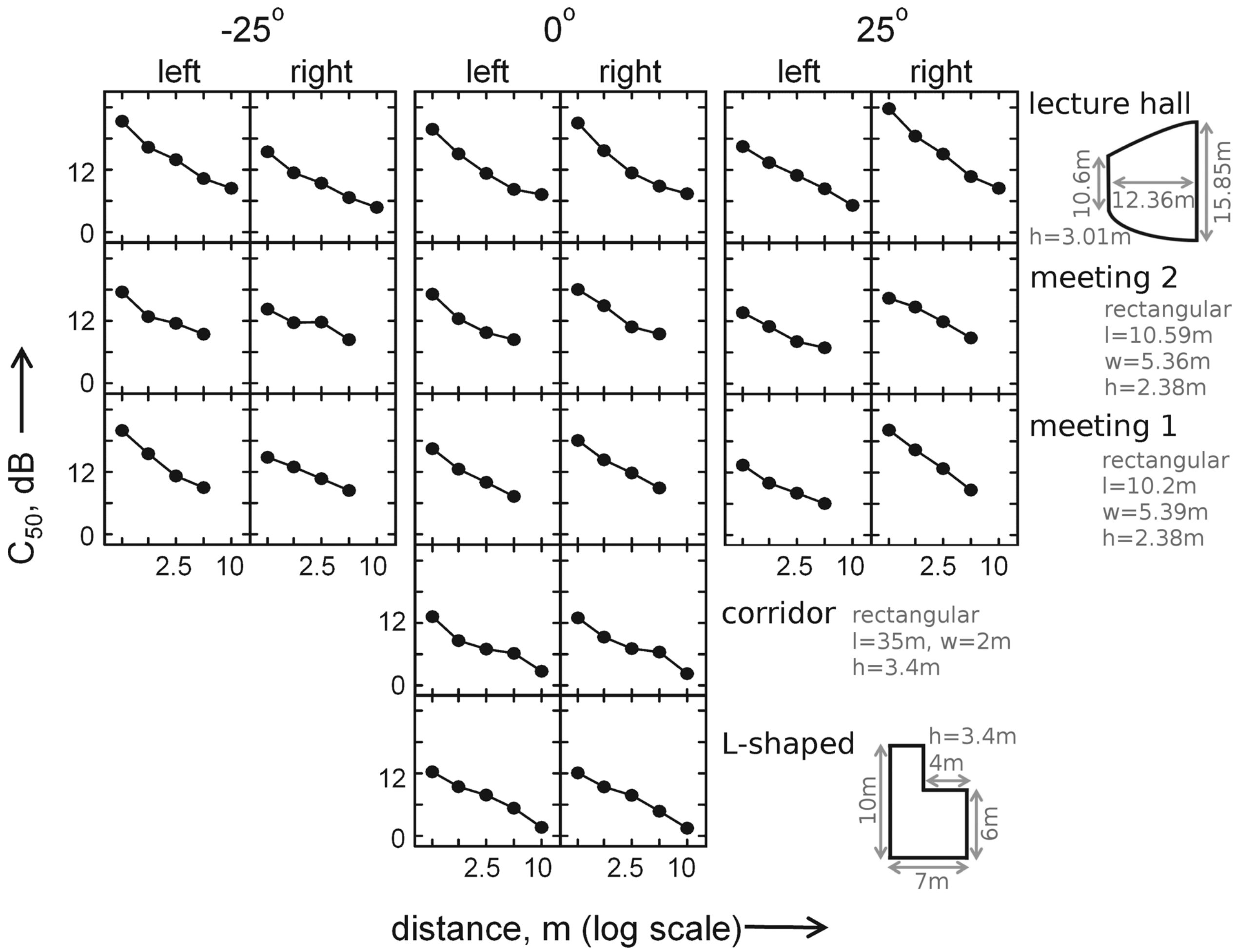

FIG. 1. A-weighted C50 values (ratios of early-to-late impulse response energy for a 50-ms early/late limit) at the left and right ears of the listener mannequin for the talker mannequin at different locations in the five rooms in which BRIRs were measured.[2] Bearings and distances are relative to listener's location, which was fixed in each room. A broad range of reverberation levels was considered in this study, with C50 systematically decreasing with increasing distance. For lateral sources, C50 was higher at the ear which was on the side of the source (left for a source at −25° and right for a source at 25°), indicating that the contra-lateral ear suffered from head shadow. The difference of C50 across the ears decreased with increasing source distance.

depth") was progressively reduced from 16 bits by integer division, until the energy decay was linear. The bit depth required for linear decay indicates the signal-to-noise ratio of the measured BRIR, which was better than 45 dB in all cases. The original 16-bit depth was then restored by multiplication.

In some conditions, spectral-envelope impulse responses (SEIRs) were used. They were obtained by removing the temporal characteristics of the BRIRs whilst preserving their spectral envelopes and associated ILDs. This manipulation removed the ITDs necessary for binaural unmasking while preserving the ILDs necessary for better-ear listening, thereby allowing the corresponding prediction components to be tested separately. The SEIRs were obtained using the fast Fourier transform of the BRIRs, whose frequency components were rotated to cosine phase (independently for the left and right channels), before taking the inverse transform and applying a short, 42.6-ms Hann window to the resulting symmetrical time-function. Consequently, the long decaying "tails" of the original BRIRs were no longer present. Left and right channel SEIRs were aligned in time, thereby removing any ITD at the onsets and elsewhere in the original BRIRs.

The resulting SEIRs were short binaural impulse responses, with very different waveforms from the BRIRs (with no long tails nor ITDs), but they had the same spectral envelope as their corresponding BRIR, with the same frequency-dependent ILDs responsible for better-ear listening. These ILDs corresponded to differences in the left and right temporal waveforms of a SEIR, but these waveforms did not have any ITD because they were created independently. As a result, there was no binaural unmasking possible with the SEIRs (as verified in the model predictions), but better-ear listening was similar to the one obtained with their corresponding BRIR.

III. INTELLIGIBILITY AGAINST SINGLE INTERFERERS IN ROOMS (EXPERIMENTS 1 AND 2)

The prediction method was first assessed for the case of single sources of interference. The difference in bearing between the target and interferer impulse responses was varied with the aim of highlighting the spatial unmasking associated with ILDs and ITDs. The relative amount of reverberation was varied by varying the interferer's distance from the listener in the five rooms tested (Fig. 1).

A. Design

In experiment 1, SRTs were measured using BRIRs and SEIRs from meeting room 1. The interferer was tested at two distances, 0.65 m (near) and 5 m (far), and three bearings, −25° (left), 0° (front), and 25° (right); whereas the target was always at near-right (0.65 m, 25°).[3] The two distances and three bearings for the interferer and the two BRIR processings resulted in twelve tested conditions.

Experiment 2 aimed to generalize the results of experiment 1, considering the four other rooms, five interferer distances and the same three source bearings.[3] Twelve configurations were tested. These conditions were chosen to maximize the differences between the corresponding predicted SRTs. In the corridor, both sources were in front, with the target at 0.65 m and the interferer at 1.25 or 5 m. In the L-shaped room, both sources were in front, with the target at 0.65 m and the interferer at 2.5 or 10 m. In meeting room 2, the target was at 0.65 m on the left (−25°) and the interferer was on the right (25°) at 0.65, 1.25, or 5 m. In the lecture hall, the target was always at 0.65 m and 25° on the opposite side of the interferer (for example, when the interferer was on the left at −25°, the target was on the right at 25°). The interferer was on the right (25°) at 0.65, 2.5, or 10 m, or on the left (−25°) at 0.65 or 5 m. Only BRIRs were used in experiment 2.

B. Results

Figure 2 presents the mean SRTs measured in experiment 1. The difference between BRIRs (black) and SEIRs (gray) indicates the contribution of binaural unmasking. The model predictions are also plotted, showing a close correspondence between measured and predicted thresholds (Bravais–Pearson correlation $r = 0.98$, $p < 0.0001$, $n = 12$). Spatial unmasking is obtained by comparing the SRT measured in each condition to the SRT of the co-located condition (near-right). For nearby interfering sources, the contribution of better-ear listening (SEIR data, gray, 4 dB at near-left) was larger than that of binaural unmasking (black minus gray, about 1.5 dB at near-left). Increasing the interferer's distance from the listener increased the relative amount of reverberation, which had the effect of reducing the influence of a bearing separation between target and interferer (the difference between the near-left and near-right conditions is substantial while the unmasking in the far-left and far-right conditions is similar). This reduced influence indicates that head shadow was very limited in the far conditions; but better-ear listening benefited from room coloration (which provided about 3 dB of unmasking in these conditions). Note that coloration is dependent on the positions of both the sound source and the listener within a room. When important frequencies for speech are attenuated in the masking noise, then speech intelligibility can improve (as seen in experiment 1). Equally a worsening of intelligibility may occur if these frequencies are amplified. Given the two ears of the listener, coloration might provide an advantage at one ear or the other. Binaural unmasking was still apparent in the far conditions (just below 1 dB at far-left and far-front).

An analysis of variance (ANOVA) confirmed that the main effects of BRIR processing, interferer's distance and interferer's bearing were significant (Table I). Tukey pairwise comparisons showed that, on average, the three tested bearings led to significantly different SRTs ($q > 4.6$, $p < 0.01$ in each case). This effect was driven by the conditions at 0.65 m (near). The interaction between the effects of interferer's distance and bearing was significant. Tukey pairwise comparisons confirmed that, on average, the three bearings led to significantly different SRTs at 0.65 m ($q > 7.0$, $p < 0.001$ in each case), but none of these differences were significant at 5 m. The effect of distance was significant at

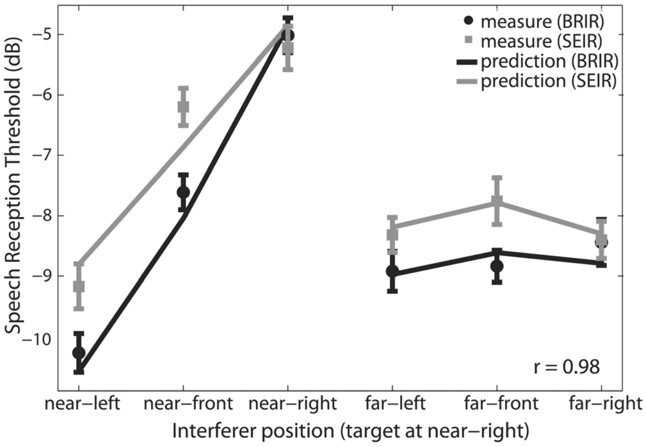

FIG. 2. Mean SRTs with standard error measured in experiment 1. The difference between BRIRs (black) and SEIRs (gray) indicates the contribution of binaural unmasking. Measurements were well predicted by the proposed method (Bravais–Pearson correlation $r = 0.98$, $p < 0.0001$, $n = 12$). For nearby interferers, the contribution of better-ear listening (SEIR data) was larger than that of binaural unmasking. Increasing reverberation reduced the influence of a bearing separation between sources, indicating that head shadow was very limited in the far conditions; but better-ear listening then benefited from room coloration. Binaural unmasking was still apparent in the far conditions.

TABLE I. Repeated measure analysis of variance (ANOVA) for Experiment 1 (Fig. 2) and Experiment 3 (Fig. 6). The factors involved in Experiment 1 were BRIR processing (BRIR), interferer's distance (Dist.) and interferer's bearing (Bear.). The factors involved in Experiment 3 were BRIR processing, interferer's distance and interferer configuration (Config.).

Factor	Sum of squares	df	Mean square	F	p
Experiment 1					
BRIR	32.7	1	32.7	19.6	<0.001
Dist.	103.9	1	103.9	34.6	<0.0001
Bear.	286.7	2	143.3	45.7	<0.0001
BRIR × Dist.	0.7	1	0.7	0.3	n.s.
BRIR × Bear.	22.5	2	11.3	4.7	<0.05
Dist. × Bear.	235.4	2	117.7	48.1	<0.0001
BRIR × Dist. × Bear.	1.9	2	0.9	0.4	n.s.
Experiment 3					
BRIR	59.1	1	59.1	33.2	<0.0001
Dist.	12.2	1	12.2	4.7	<0.05
Config.	493.5	3	164.5	75.9	<0.0001
BRIR × Dist.	11.6	1	11.6	4.9	<0.05
BRIR × Config.	12.1	3	4.0	2.0	n.s.
Dist. × Config.	65.6	3	21.9	12.9	<0.0001
BRIR × Dist. × Config.	1.7	3	0.6	0.2	n.s.

all bearings ($q > 4.4$, $p < 0.01$ in each case). The interaction between the effects of BRIR processing and interferer's bearing was also significant. Tukey pairwise comparisons confirmed that binaural unmasking (difference between BRIR and SEIR) was significant for sources at different bearings [interferer on the left and in front ($q > 4.5$, $p < 0.01$ in each case)], but not for sources at the same bearing (interferer on the right). On average, all bearings led to significantly different SRTs with the BRIRs ($q > 5.3$, $p < 0.01$ in each case). The SRTs measured with the interferer in front and on the right were not significantly different using the SEIRs. These SRTs were both significantly different from the SRT measured with the interferer on the left using the SEIR ($q > 6.9$, $p < 0.001$ in each case).

Figure 3 presents the results of experiment 2, comparing the model predictions to the measured SRTs. Because all experimental parameters (room, distance, bearing) were not varied systematically, results are presented as a scattergram rather than plotted as a function of these parameters. The aim here was not to investigate a systematic effect of distance or room for example, but to validate a prediction method which will allow these future investigations. Experiment 2 confirmed the good performance of the model observed in experiment 1, with again a close correspondence between measured and predicted thresholds (Bravais–Pearson correlation $r = 0.98$, $p < 0.0001$, $n = 12$). An ANOVA confirmed that the effect of the tested condition was significant [$F(11,253) = 29.7$, $p < 0.0001$], with Tukey pairwise comparisons indicating that forty pairs of conditions (out of sixty-six) led to significantly different SRTs ($q > 4.7$, $p < 0.05$ in each case).

C. Discussion

Experiments 1 and 2 showed that the proposed model accurately predicts the effects of binaural unmasking and better-ear listening, both in combination (BRIRs) and isolation [results with SEIRs indicate that effects of better-ear listening alone are well predicted, while results from Lavandier and Culling (2010) indicate good predictions of effects of binaural unmasking alone], in the presence of single interferers in reverberation. The correspondence between measured and predicted thresholds was as good as for other models [overall correlation of 0.95 for Beutelmann and Brand (2006), 0.95–0.97 correlation for Lavandier and Culling (2010)] and for previous validations of this model in anechoic situations [correlations between 0.86 and 0.99 for Jelfs *et al.* (2011)]. These results show that the model's utility in artificial situations (Lavandier and Culling, 2010) extends to the real-room conditions used in the current validation, which involved five very different rooms, five interferer's distances ranging from 0.65 to 10 m, and three source bearings with target and interferer both tested in front and on both sides of the listener.

It should be noted that all model parameters are fixed and come from the literature. The frequency selectivity of the auditory system is taken from Moore and Glasberg (1983); the two jitter parameters of the E-C model are taken from Durlach (1972); the SII weightings are taken from the ANSI standard (ANSI S3.5, 1997). Because the proposed method does not require any parameter to be fitted to the measured data to predict differences in SRT, it could be used to predict the SRTs measured by Beutelmann and Brand (2006) (using the average SRT across conditions as a reference for the model in each experiment). These SRTs were obtained in a different laboratory, in different rooms and at different bearings and distances, using a different measurement procedure and a different language. Figure 4 shows that a close correspondence between measured and predicted thresholds was obtained (Bravais–Pearson correlation $r = 0.99$, $p < 0.0001$, $n = 16$).

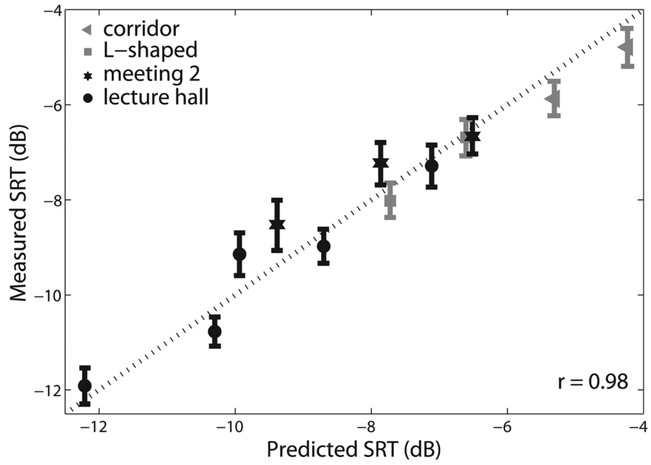

FIG. 3. Comparison of the mean SRTs with standard errors measured in experiment 2 with the model predictions. The dashed reference line is a line of unit slope passing though the origin and represents a 1:1 relationship between the predicted and measured SRTs. Measurements were well predicted by the model (Bravais–Pearson correlation $r = 0.98$, $p < 0.0001$, $n = 12$), generalizing the good performance observed in experiment 1 while considering other rooms and distances.

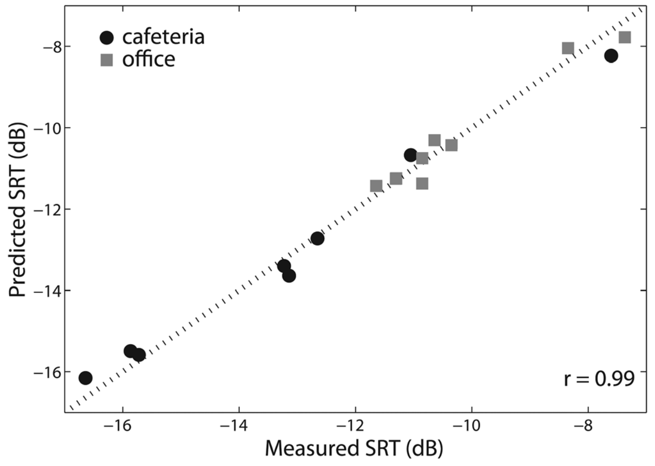

FIG. 4. Comparison of the model predictions with the mean SRTs measured by Beutelmann and Brand (2006) with normal-hearing listeners in a cafeteria (black circles) and an office (gray squares). The dashed reference line is a line of unit slope passing though the origin and represents a 1:1 relationship between the predicted and measured SRTs. Measurements were well predicted by the model (Bravais–Pearson correlation $r = 0.99$, $p < 0.0001$, $n = 16$), confirming the good performance observed in experiments 1 and 2 while considering measurements done in a different laboratory, at different bearings and distances, using a different procedure and language.

Beutelmann and Brand (2006) measured SRTs for speech against a single noise interferer, using German sentences and an adaptative procedure (Brand and Kollmeier, 2002), with ten normal-hearing listeners. Measurements were carried out in two separate experiments involving BRIRs from two different rooms, an office and a cafeteria. In the office, BRIRs were measured with a loudspeaker at 1.45 m from a listener mannequin placed in the middle of the room. No indication was found concerning the distance between the source and the mannequin in the cafeteria, but it should have been less than 3 m, which was the distance between the mannequin and an adjacent window. In both experiments, the speech target was always in front (0°), whereas the noise interferer was tested at eight bearings: −140°, −100°, −45°, 0°, 45°, 80°, 125°, and 180° in the office; −135°, −90°, −45°, 0°, 45°, 90°, 135°, and 180° in the cafeteria. The resulting SRTs, scanned from their Fig. 2, are plotted against the predictions obtained with our model directly applied to the corresponding BRIRs (Fig. 4). The correlation coefficient between measured and predicted SRTs is 0.99 when comparing between all conditions. It is 0.98 when considering only the office data, and 0.99 for the cafeteria data. These values are comparable with the 0.94 correlation obtained in each room by Beutelmann and Brand (2006).

IV. INTELLIGIBILITY AGAINST MULTIPLE INTERFERERS IN ROOMS (EXPERIMENT 3)

The prediction method was then assessed for the case of multiple sources of interference. One to three interferers were tested using impulse responses on one or both sides of the listener and target, as in previous measurements in anechoic experiments (Culling *et al.*, 2004; Hawley *et al.*, 2004). The relative amount of reverberation was varied by varying the interferers' distance from the listener (Fig. 1).

A. Design

In experiment 3, SRTs were measured using BRIRs and SEIRs from meeting room 1, in the configurations illustrated with the sketches of Fig. 5. For all configurations, interferers were tested at the same distance, either 0.65 m (near) or 5 m (far). In configurations 1, 2, and 3, the target was always at 0.65 m and 25° (right), whereas a single interferer was at −25° (left) in configuration 1, a second interferer was added at −5° (left) in configuration 2, and a third interferer was added at 5° (right) in configuration 3. In configuration bilateral, the target was at 0.65 m and 0° (front), with one interferer on each side, at −25° and 25°. These four configurations, two interferer distances and two BRIR processings resulted in 16 tested conditions.

B. Results

Figure 6 presents the mean SRTs measured in experiment 3. The difference between BRIRs (black) and SEIRs (gray) indicates the contribution of binaural unmasking. The model predictions are also plotted, showing again a close correspondence between measured and predicted thresholds (Bravais–Pearson correlation $r = 0.95$, $p < 0.0001$, $n = 16$). In the presence of between one and three interferers placed on one or both sides of the listener's head, the effects of binaural unmasking remained apparent with multiple interferers (about 1 dB improvement in intelligibility across the different configurations of nearby interferers). The main loss of intelligibility with increasing number of interferers appeared to arise through the loss of better-ear listening when interferers were on both sides of the listener (near-3 vs near-2), and this loss of intelligibility was even greater when the interferers were on both sides of the target (near-bilateral vs near-2). As for single interferers, increased reverberation in the far conditions reduced the effects of both head shadow (reduced unmasking between the bilateral and 2-conditions) and

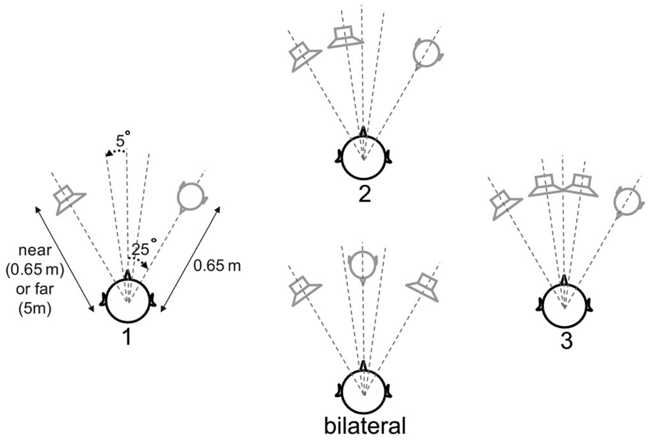

FIG. 5. Spatial configurations used for experiment 3 in meeting room 1 (Fig. 1). The target was always at 0.65 m, at 25° (1, 2, and 3) or 0° (bilateral). A single interferer was at −25° (1), two interferers were at −25° and −5° (2), three interferers were at −25°, −5°, and 5° (3), or two interferers were at −25° and 25° (bilateral). All interferers were at the same distance, either 0.65 m (near) or 5 m (far).

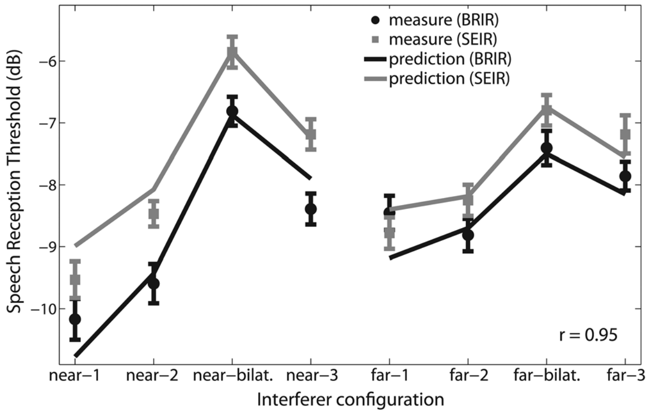

FIG. 6. Mean SRTs with standard error measured in experiment 3 with multiple interferers (Fig. 5). The difference between BRIRs (black) and SEIRs (gray) indicates the contribution of binaural unmasking. Measurements were well predicted by the proposed method (Bravais–Pearson correlation $r = 0.95$, $p < 0.0001$, $n = 16$). Binaural unmasking remained apparent with increasing number of interferers, whereas the main loss of intelligibility was associated with the loss of better-ear listening when interferers were on both sides of the listener (near-3 vs near-2) and on both sides of the target (near-bilateral vs near-2). As for single interferers, increased reverberation reduced binaural unmasking and head shadow (reduced unmasking between the bilateral and 2-conditions), but on average better-ear listening was not reduced because the loss of head shadow might have been compensated by the beneficial effect of room coloration (see experiment 1).

binaural unmasking (about 0.5 dB across conditions). On average, increasing distance did not reduce better-ear listening, because the loss of head shadow with increased reverberation might have been compensated by the beneficial effect of room coloration (see experiment 1 in the same room).

An ANOVA confirmed that the main effects of BRIR processing, interferer distance and interferer configuration were significant (Table I). Tukey pairwise comparisons showed that, on average, all configurations except 1 and 2 led to significantly different SRTs [$q > 7.1$, $p < 0.001$ in each case]. The interaction between the effects of BRIR processing and interferer distance was significant. Tukey pairwise comparisons showed that binaural unmasking (difference between BRIR and SEIR) was significant at 0.65 and 5 m ($q > 3.2$, $p < 0.05$ in each case). The effect of distance was significant with the BRIRs ($q > 4.2$, $p < 0.01$), but not with the SEIRs. The interaction between the effects of interferer distance and configuration was also significant. On average, all configurations led to significantly different SRTs at 0.65 m ($q > 4.4$, $p < 0.05$ in each case). It was also the case at 5 m ($q > 5.4$, $p < 0.01$ in each case), except for configurations with interferers on only one side of the listener (1 vs 2), or on both sides of the listener (bilateral vs 3). The effect of distance was significant for configurations 1 and bilateral ($q > 3.7$, $p < 0.05$ in each case), but not for configurations 2 and 3.

C. Discussion

Experiment 3 showed that, as in anechoic situations (Bronkhorst and Plomp, 1992; Carhart *et al.*, 1969; Culling *et al.*, 2004; Hawley *et al.*, 2004), binaural hearing is still effective against multiple interferers in rooms, and that the proposed model accurately predicts the corresponding effects of binaural unmasking and better-ear listening (both in combination and isolation). The correspondence between measured and predicted thresholds with multiple interferers was as good as with single interferers (see Sec. III C of experiments 1 and 2). We are not aware of any other prediction model tested using multiple interferers in reverberation. Jelfs *et al.* (2011) obtained similar agreement while testing the model for multiple noise interferers in anechoic conditions, with correlations of 0.98 and 0.99 between measured and predicted thresholds. The model of Wan *et al.* (2010) also accurately predicted the SRTs measured by Hawley *et al.* (2004) using one to three noise interferers in different anechoic configurations. The direct comparison with the present results is difficult though, because their SII criterion was changed each time the number of interferers varied. This criterion is equivalent to the reference SRT used in each experiment presented here to compare inverted effective target-to-interferer ratios and SRTs (no reference is needed to compare directly differences in ratios and in SRTs).

In experiment 3, configurations 1 and 2 did not lead to significant differences in SRT. The two interferers of configuration 2 were in the same hemifield and opposite to the target side, so head shadow was not greatly affected by the second interferer. It was greatly reduced when the interferers were spatially distributed in both hemifields, as previously measured in anechoic studies (Culling *et al.*, 2004; Hawley *et al.*, 2004). Also in agreement with these studies, binaural unmasking was robust in all spatial configurations, whether there were one or multiple interferers distributed across locations in the same hemifield or in both (no significant interaction between the effects of BRIR processing and interferer configuration in experiment 3). Based on the interaural phase differences associated with ITD [Eq. (1)], binaural unmasking can still be effective against multiple interferers at different positions with different ITDs, because this mechanism then acts on the composite interferer at the ear. The interaural phase differences of this composite interferer do not correspond to any real interferer position anymore.

V. MAPPING INTELLIGIBILITY IN NOISY ROOMS

Unlike previous intelligibility models based on source signals in rooms (Beutelmann and Brand, 2006; Lavandier and Culling, 2010), the proposed method is applied directly to BRIRs, producing fast and accurate non-stochastic predictions (Jelfs *et al.*, 2011). Thanks to its resulting computational efficiency, the method can be used to generate intelligibility maps of rooms containing multiple interfering sources, as long as these sources are stationary noises. These spatial representations offer visualization of the space accessible to a listener who would wish to maintain a given level of intelligibility whilst moving within the room. This section of the paper presents examples of such maps obtained in simple simulated rooms. The aim here was not to demonstrate systematic effects of room parameters, but to illustrate the potential applications of the prediction method to support the design of social interaction spaces.

A. Room simulations

Virtual rooms were simulated using a ray-tracing method (Allen and Berkley, 1979; Peterson, 1986), implemented in the WAVE signal processing package (Culling, 1996). They were $10 \times 6.4 \times 2.5$ m^3, each surface having a uniform frequency-independent absorption coefficient. Figure 7 shows the effect of the level of reverberation controlled by setting the absorption coefficient to a single value for all surfaces (0.9 for dry, 0.5 for mildly reverberant, and 0.1 for very reverberant). Figure 8 shows a decomposition of the effects of binaural unmasking and better ear listening using a more realistic allocation of absorption coefficients (0.4 on walls, 0.9 on ceiling, and 0.2 on floor). In all computations reported here, an adapted version of the program was used, so that the listener's head was modeled by filtering each ray by the appropriate head-related transfer function of a KEMAR mannequin (Gardner and Martin, 1995) in accordance with its angle of incidence. All sources (of equal power level) and receivers were at 1.5-m height, and the positions considered were (in m): target (5.5; 2), interferer 1 (2; 2.5), interferer 2 (4; 5), interferer 3 (6.5; 5.5) or (8.5; 3.5), listener positions centered on a grid 0.3×0.3.

B. Efficiency and modularity of the method

Maps illustrating situations for between one and three stationary noise interferers (increasing number from left to

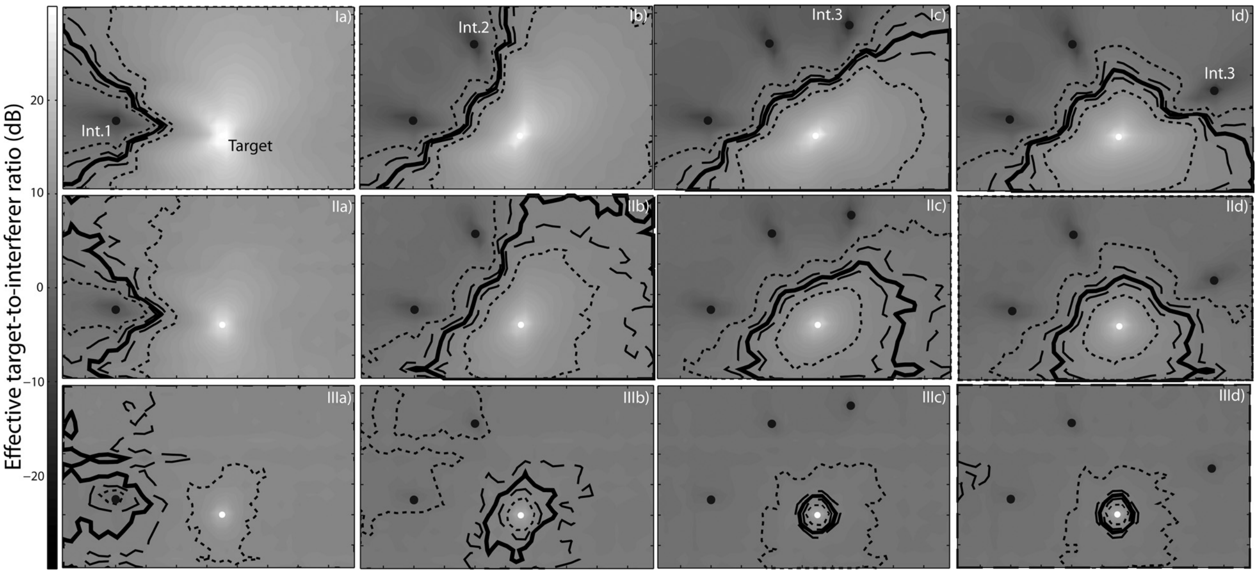

FIG. 7. Intelligibility maps of rooms. The effective target-to-interferer ratio was predicted as a function of listener's position (facing the target) in a virtual room modeled as dry (I), mildly reverberant (II), or very reverberant (III), in the presence of 1 (a), 2 (b), or 3 (c and d) stationary noise interferers. Also shown are the 0 dB ratio contour (solid line), the 1 dB contours (dashed lines), and the 3 dB contours (dotted lines). Increasing reverberation and surrounding interferers limited the space available to listeners.

right in panels a–d of Fig. 7), in a room where the absorption was varied to change the level of reverberation (which increases from top to bottom in rows I to III), show the most desirable locations for understanding speech (equal-intelligibility zones corresponding to high effective target-to-interferer ratios are shaded lighter). These zones were greatly narrowed for increasingly reverberant conditions (Fig. 7, top to bottom), as they were when multiple interfering sources enter the room (left to right). This latter reduction in intelligible listening space was associated with the loss of positions that offer substantial head shadow against interferers. The maps also demonstrate that reverberation tended to spread the target and interferer energy throughout the room, making target-to-interferer ratios more uniform as they tended towards 0 dB (one interferer), −3 dB (two interferers), or −4.8 dB (three interferers). Even though adding absorbent material in a room does not directly eliminate the interfering sources, it might enable the auditory system to work more efficiently and to effectively "cancel" (at least) part of the interfering sound, resulting in more freedom to stand in different parts of the room for the listener.

For natural listening conditions, it should be borne in mind that the level of intelligibility corresponding to a given effective ratio is dependent on hearing abilities. To ensure the same level of understanding, hearing-impaired listeners (Beutelmann and Brand, 2006) and cochlear implantees (Qin and Oxenham, 2003), for example, will require a better ratio than normally hearing listeners. The prediction method could

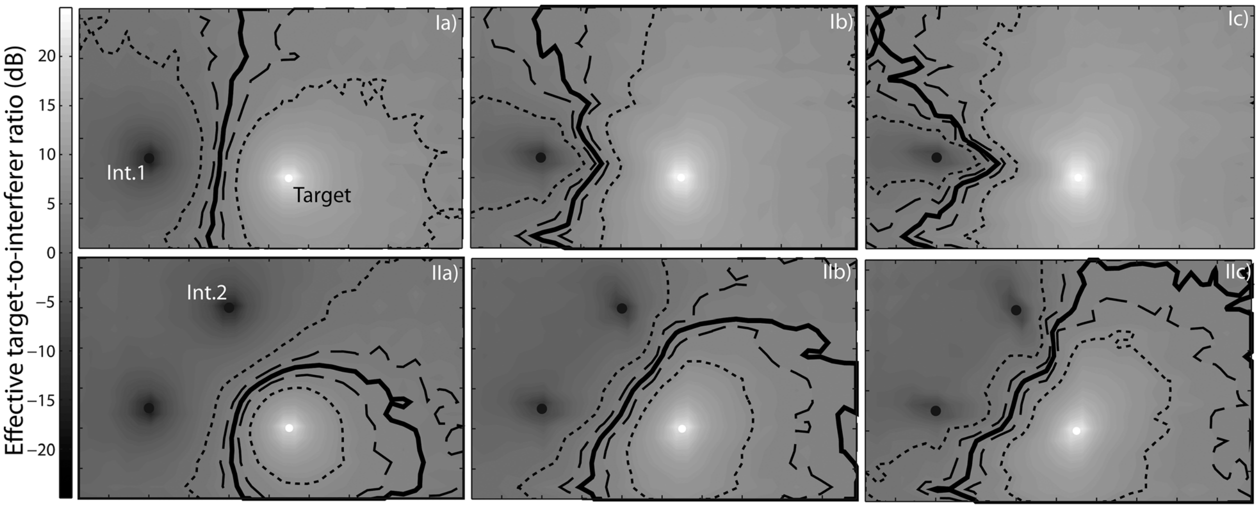

FIG. 8. Decomposition of speech segregation mechanisms. The effective target-to-interferer ratio was predicted as a function of position in a virtual mildly reverberant room, in the presence of 1 (I) or 2 (II) stationary noise interferers (solid line for 0 dB contour, dashed lines for 1 dB contours and dotted lines for 3 dB contours). The listener facing the target was modeled with binaural unmasking ability (c), without this ability (b, only better-ear listening), or simply as an omnidirectional microphone (a, no head shadow/better-ear listening, no binaural unmasking). This decomposition showed that better-ear listening and binaural unmasking both enabled the listener to stand in more places within the room.

take into account specific forms of hearing impairment in order to guide technical applications directed towards the listener (e.g., directional microphones on hearing aids) or environmental policies concerning room design. As we show, this can be achieved in realistic environments including those with multiple sources and reverberation. As mechanisms of speech segregation are modeled separately, their influence can be predicted independently. This is particularly relevant for cochlear implantees, who benefit from better-ear listening (if bilaterally implanted) but not binaural unmasking, because current implants encode the temporal envelope of incoming sounds but not the temporal-fine structure (Majdak *et al.*, 2006). Prediction maps obtained with binaural unmasking [Fig. 8, panel (c)] and without [panel (b)], as well as maps where the two-eared head was replaced by an omnidirectional microphone (panels a), in the presence of one or two stationary noise interferers (panels I,II), indicate that better-ear listening (a vs b) and binaural unmasking (b vs c) resulted in a listener being able to stand farther away from the target without losing understanding.

VI. LIMITATIONS OF THE METHOD

The prediction method relies on several assumptions. The short-term variations of interaural phases and levels are ignored in the model which implicitly considers only mean statistics over signals (speech sentence or noise sample) when processing overall BRIRs. It should be noted that the measured data were also averaged over listeners and sentence lists. The model assumes that each frequency channel operates independently. There is evidence for within-channel processing of binaural unmasking (Akeroyd, 2004; Edmonds and Culling, 2005), but it is currently unclear whether better-ear listening operates independently in each frequency channel, or whether the same ear is selected for all frequencies. It might be the case that, in real listening situations, it does not make a big difference. For example, for sources which are not too reverberant, head shadow dominates better-ear listening and tends to favor the same ear at all frequencies. The model assumes additive contributions of binaural unmasking and better-ear listening, neglecting their potential interaction. In particular, the effect of ILDs on binaural unmasking is not taken into account, even if it is known that binaural unmasking of tonal targets is reduced when target or masker has a large ILD (Egan, 1965). The accurate predictions of the model might indicate that this effect is very limited when realistic ILDs are involved. The additivity assumption might also hold because binaural unmasking and better-ear listening tend to operate in different frequency regions (low frequencies for binaural unmasking and high frequencies for better-ear listening), such that when they are summed, one of them is always negligible. As in the original E-C theory, the model does not predict any BMLD at high frequency [Durlach (1972), pp. 435–436], whereas a BMLD of up to 3 dB can be observed for tonal signals up to at least 4 kHz in broadband noise (Hirsh and Burgeat, 1958). Despite these limitations and assumptions, the fit between predictions and data was good, in the experiments presented here and also in the validation presented by Jelfs *et al.* (2011), who successfully modeled a range of anechoic data sets from the literature.

In the experiments used to validate the prediction method, all sources had the same sound level and long-term spectrum. The application of the method is not limited to these situations. Sources at different sound levels can be modeled by scaling their respective BRIR to the appropriate level. Note that only level differences between sources are relevant. Sources with different spectra can also be modeled by appropriate filtering of their BRIRs. Again, differences in spectrum are the relevant parameter. If sources have all the same spectrum, no filtering is required. In the case of multiple interferers, concatenation of the scaled or filtered BRIRs would have the effect of summing the frequency-dependent energy of each contributing impulse response, and generating an averaged cross-correlation function weighted according to the energy in each impulse response. This generalization of the method has not been directly tested; but, because the model successfully predicted differences in source spectra introduced by room coloration and head shadow, there is no reason to believe that the proposed processing of the BRIRs should not also result in accurate prediction.

The model does not consider the potential smearing of target speech in very reverberant environments, so that prediction only holds for targets not too far from the listener in these environments, at positions where the direct-to-reverberant ratio is not too low and segregation from interferers is the overriding factor for intelligibility. The model needs to be extended to take into account this direct effect of reverberation on target speech. It could be combined with existing models predicting temporal smearing (Bradley *et al.*, 1999; Houtgast and Steeneken, 1985). Such a combined approach was used by van Wijngaarden and Drullman (2008) when they introduced binaural-hearing inspired modifications to the speech transmission index method.

A model that can completely describe cocktail-party situations in rooms needs to handle competing speech sources. Interferer periodicity and modulation need to be incorporated in the model to refine the predictions. Fundamental frequency (F0) differences facilitate segregation of competing voices (Brokx and Nooteboom, 1982; Culling and Darwin, 1993), and Culling *et al.* (2003, 1994) showed that reverberation was detrimental to segregation by F0 differences where the F0 was non-stationary. Modulations in the temporal envelope of the interferer allow one to hear the target better (Dusquesnoy, 1983; Festen and Plomp, 1990), so-called “listening in the gaps,” and this ability is impaired by reverberation which reduces modulations (Bronkhorst and Plomp, 1990; George *et al.*, 2008), filling the “gaps” in the interferer. Beutelmann *et al.* (2010) extended their model to take this effect into account, following an approach proposed by Rhebergen and Versfeld (2005), which consists in applying a stationary model to short time frames of the target and interferer signals, and then averaging the predictions over time. This signal-based approach would need to be adapted to be applied to our model based on BRIRs. If it cannot be assumed that the listener knows who/where to listen to, then additional attentional effects also have to be modeled (Kidd *et al.*, 2005; Shinn-Cunningham *et al.*, 2005).

VII. CONCLUSION

A binaural intelligibility model combining better-ear listening and binaural unmasking was validated in real rooms, in the presence of multiple stationary noise interferers. Correlation coefficients ranging from 0.95 to 0.99 were obtained between measured and predicted differences in threshold, without any model parameter being fitted to the data. The prediction method is based on BRIR measurements and can accurately predict speech intelligibility against any number of noise interferers, in any spatial distribution within a room and for any orientation of the listener. The method is sufficiently computationally efficient to generate intelligibility maps from room designs. These visualizations of the space accessible to listeners could form the basis of powerful architectural tools, and provide guides to treatment strategies for the hearing impaired. The method still needs to be refined to be able to predict the temporal smearing of target speech in very reverberant spaces and the segregation mechanisms associated with the temporal envelope modulations and the periodicity of speech interferers.

ACKNOWLEDGMENTS

This work was supported by a grant from the UK EPSRC. The authors are grateful to Thomas Brand for providing J.F.C. with the BRIRs used by Beutelmann and Brand (2006), to David McAlpine for comments on an earlier version of the manuscript, to Steve Colburn and one anonymous reviewer for their helpful comments, and to all listeners who took part in the experiments.

[1] In order to maximize signal-to-noise ratio, the sound level of the loudspeaker was varied during the BRIR measurements. This procedure did not affect the BRIR ILDs.

[2] Only a subset of all BRIRs were used in the experiments. BRIRs could not be measured at 10 m in the meeting rooms. Measurements done at −5° and 5° in these rooms and used in experiment 3 are not displayed on Fig. 1.

[3] To give an idea of how decorrelated the signals were in the rooms and configurations tested, the interaural coherence of the interfering sources below 1500 Hz (the frequency range for which binaural unmasking is most effective in broadband noise) was calculated as described by Lavandier and Culling (2010). For experiment 1, the coherence was around 0.63 at near-left, 0.64 at near-front and 0.68 at near-right. It was around 0.20 at far left, 0.35 at far-front and 0.65 at far-right. The variations in coherence observed in the far conditions confirm the asymmetry of the configurations, with the listener mannequin placed in a corner of the room. For experiment 2, the coherence varied between about 0.34 and 0.89.

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–1148.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

ANSI S3.5 (**1997**). *Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Beutelmann, R., and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 331–342.

Beutelmann, R., Brand, T., and Kollmeier, B. (**2010**). "Revision, extension, and evaluation of a binaural speech intelligibility model," J. Acoust. Soc. Am. **127**, 2479–2497.

Bradley, J. S., Reich, R. D., and Norcross, S. G. (**1999**). "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. **106**, 1820–1828.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–281.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

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

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

Bronkhorst, A. W., and Plomp, R. (**1990**). "A clinical test for the assessment of binaural speech perception in noise," Audiology **29**, 275–285.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Release from multiple maskers: Effects of interaural time disparities," J. Acoust. Soc. Am. **45**, 411–418.

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

Culling, J. F. (**1996**). "Signal processing software for teaching and research in psychoacoustics under UNIX and X-windows," Behav. Res. Methods Instrum. Comput. **28**, 376–382.

Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0," J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2005**). "Erratum: The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **118**, 552.

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

Culling, J. F., Summerfield, Q., and Marshall, D. H. (**1994**). "Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels," Speech Commun. **14**, 71–96 (**2003**).

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York), Vol. II, pp. 371–462.

Dusquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Edmonds, B. A., and Culling, J. F. (**2005**). "The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.

Egan, J. P. (**1965**). "Masking-level differences as a function of interaural disparities in intensity of signal and noise," J. Acoust. Soc. Am. **38**, 1043–1049.

Farina, A. (**2000**). "Simultaneous measurement of impulse response and distorsion with swept-sine technique," in *AES 108th Convention*, Preprint 5093 (D-4).

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.

Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.

George, E. L. J., Festen, J. M., and Houtgast, T. (**2008**). "The combined effects of reverberation and nonstationary noise on sentence intelligibility," J. Acoust. Soc. Am. **124**, 1269–1277.

Hartmann, W. M., Rakerd, B., and Koller, A. (**2005**). "Binaural coherence in rooms," Acta Acust. Acust. **91**, 451–462.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Hirsh, I. J., and Burgeat, M. (**1958**). "Binaural effects in remote masking," J. Acoust. Soc. Am. **30**, 827–832.

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

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

ISO 3382 (**1997**). "Acoustics—measurement of the reverberation time of rooms with reference to other acoustical parameters" (International Organization for Standardization, Geneva).

Jelfs, S., Culling, J. F., and Lavandier, M. (**2011**). "Revision and validation of a binaural model for speech intelligibility in noise," Hear. Res. **275**, 96–104.

Kidd, G., Mason, C., Brughera, A., and Hartmann, W. (**2005**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acta Acust. Acust. **91**, 526–535.

Kryter, K. D. (**1962**). "Methods for the calculation and use of the Articulation Index," J. Acoust. Soc. Am. **34**, 1689–1697.

Lavandier, M., and Culling, J. F. (**2007**). "Speech segregation in rooms: Effects of reverberation on both target and interferer," J. Acoust. Soc. Am. **122**, 1713–1723.

Lavandier, M., and Culling, J. F. (**2008**). "Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer," J. Acoust. Soc. Am. **123**, 2237–2248.

Lavandier, M., and Culling, J. F. (**2010**). "Prediction of binaural speech intelligibility against noise in rooms," J. Acoust. Soc. Am. **127**, 387–399.

Levitt, H., and Rabiner, L. R. (**1967**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**, 820–829.

Licklider, J. C. R. (**1948**). "The influence of interaural phase relations upon masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.

Lippmann, R. P. (**1997**). "Speech recognition by machines and humans," Speech Commun. **22**, 1–15.

Majdak, P., Laback, B., and Baumgartner, W.-D. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," presented to the Institute of Acoustics speech group on auditory modelling at the Royal Signal Research Establishment.

Peterson, P. M. (**1986**). "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," J. Acoust. Soc. Am. **80**, 1527–1529.

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)," Acustica **34**, 200–211.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech-reception threshold for sentences," Audiology **18**, 43–52.

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.

Ratnam, R., Jones, D. L., Wheeler, B. C., O'Brien, Jr., W. D., Lansing, C. R., and Feng, A. S. (**2003**). "Blind estimation of reverberation time," J. Acoust. Soc. Am. **114**, 2877–2892.

Rhebergen, K. S. and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

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

Schroeder, M. (**1965**). "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412.

Shinn-Cunningham, B., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acta Acust. Acust. **91**, 967–979.

van Wijngaarden, S. J., and Drullman, R. (**2008**). "Binaural intelligibility prediction based on the speech transmission index," J. Acoust. Soc. Am. **123**, 4514–4523.

vom Hövel, H. (**1984**). "Zur bedeutung der übertragungseigenschaften des außenohrs sowie des binauralen hörsystems bei gestörter sprachübertragung (On the importance of the transmission properties of the outer ear and the binaural auditory system in disturbed speech transmission)," Ph.D. thesis, RTWH, Aachen, as cited by Beutelmann *et al*. (2010).

Wan, R., Durlach, N. I., and Colburn, H. S. (**2010**). "Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers," J. Acoust. Soc. Am. **128**, 3678–3690.

Watkins, A. J. (**2005**). "Perceptual compensation for effects of reverberation in speech identification," J. Acoust. Soc. Am. **118**, 249–262.

Zahorik, P. (**2002**). "Assessing auditory distance perception using virtual acoustics," J. Acoust. Soc. Am. **111**, 1832–1846.

Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. Studebaker and I. Hochberg (Allyn and Bacon, Needham Heights, MA), pp. 255–276.

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