

Prediction of binaural speech intelligibility against noise in rooms

Mathieu Lavandier^{a)} and John F. Culling

School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom

(Received 30 October 2008; revised 3 November 2009; accepted 5 November 2009)

In the presence of competing speech or noise, reverberation degrades speech intelligibility not only by its direct effect on the target but also by affecting the interferer. Two experiments were designed to validate a method for predicting the loss of intelligibility associated with this latter effect. Speech reception thresholds were measured under headphones, using spatially separated target sentences and speech-shaped noise interferers simulated in virtual rooms. To investigate the effect of reverberation on the interferer unambiguously, the target was always anechoic. The interferer was placed in rooms with different sizes and absorptions, and at different distances and azimuths from the listener. The interaural coherence of the interferer did not fully predict the effect of reverberation. The azimuth separation of the sources and the coloration introduced by the room also had to be taken into account. The binaural effects were modeled by computing the binaural masking level differences in the studied configurations, the monaural effects were predicted from the excitation pattern of the noises, and speech intelligibility index weightings were applied to both. These parameters were all calculated from the room impulse responses convolved with noise. A 0.95–0.97 correlation was obtained between the speech reception thresholds and their predicted value. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3268612]

PACS number(s): 43.66.Pn, 43.66.Dc, 43.55.Hy, 43.71.Gv [RLF]

Pages: 387–399

I. INTRODUCTION

Room acoustical measurements such as the useful-to-detrimental ratio (Bradley, 1986; Bradley *et al.*, 1999) or the speech transmission index (Houtgast and Steeneken, 1985) already exist to predict the potential loss of intelligibility in a room, when the original speech signal is being mixed with its delayed versions associated with the sound reflections on the room boundaries. These measurements can take into account a background ambient noise, but they were not designed to predict intelligibility against localized interferers for which signal statistics at the ears depend on the source and listener positions within the room. When listeners need to separate a target speech from a localized interferer, reverberation impairs intelligibility not only by its direct effect on the target but also by affecting the binaural processing of the interferer (Lavandier and Culling, 2007). The corresponding loss of intelligibility appears at lower levels of reverberation and thus occurs more readily than the loss of intelligibility associated with the smearing effect of reverberation on the target (Lavandier and Culling, 2008).

In anechoic situations, the azimuth separation of sources is a well-established cue helping the segregation of competing sounds (Hawley *et al.*, 2004; Plomp, 1976). The spatial unmasking associated with this azimuth separation arises from two cues (Bronkhorst and Plomp, 1988): head shadow and binaural interaction. When the interferer is moved around the listener's head, its sound level is reduced at the ear which is in the acoustic shadow of the head. The target-to-interferer level ratio is improved at this "better" ear, and

this improvement helps the listener to understand the target. Binaural interaction relies on interaural time delays (ITDs). The azimuth separation of interferer and target induces a difference in ITDs for these two sources, which facilitates their segregation. For example, the equalization-cancellation (E-C) theory predicts that the auditory system will be able to "cancel" an interferer if it has a different ITD from the target, thus improving the internal target-to-interferer level ratio (Durlach, 1972).

Levitt and Rabiner (1967b) proposed a model describing the release from masking associated with binaural interaction. Speech intelligibility in noise is predicted from the computation of the articulation index (AI) (Kryter, 1962). Binaural interaction is taken into account by assuming that the effective speech-to-interference ratio in each frequency band is increased by the binaural masking level difference (BMLD) for pure tone detection in noise at the center frequency of the band, using BMLD predictions from Durlach (1963). Predictions based on this model were fairly consistent with previous data collected by the same authors (Levitt and Rabiner, 1967a). From this work, Zurek (1993) proposed a model describing spatial unmasking in anechoic situations. Better-ear listening is simulated by computing the speech-to-interference ratios at the two ears by frequency bands, and taking the better of the left and right ratios in each band. Binaural interaction is then taken into account by increasing the better-ear ratios by the size of the BMLD in each band, this BMLD being estimated for the given set of interaural parameters using a simplified expression proposed by Colburn (1977). The broadband prediction is computed as the AI-weighted sum of the resulting signal-to-noise ratios. The model predictions were compared with measured data from several studies. Predictions of the head-shadow advantage

^{a)}Author to whom correspondence should be addressed. Present address: Université de Lyon, Ecole Nationale des Travaux Publics de l'Etat, Département Génie Civil et Bâtiment (C.N.R.S.), Rue M. Audin, 69518 Vaulx-en-Velin Cedex, France. Electronic mail: mathieu.lavandier@entpe.fr

were generally larger than the measured effects, whereas binaural-interaction advantages were fairly well predicted.

Many experiments have shown that spatial unmasking is reduced by reverberation (Beutelmann and Brand, 2006; Culling *et al.*, 2003; Plomp, 1976). The sound reflections traveling around the listener reduce head shadow (Plomp, 1976), and reverberation impairs binaural interaction by decorrelating the interferer at the listener's ears (Lavandier and Culling, 2008). The models proposed by Levitt and Rabiner (1967b) and Zurek (1993) cannot be applied to reverberant situations, because they do not take into account the interaural coherence of the interferer. The interaural coherence of a source is the correlation of the sound waveform it produces at the two ears of the listener after interaural delay is taken into account. This coherence is reduced by sound reflections in rooms. Lavandier and Culling (2007, 2008) showed that the target intelligibility decreased when they decreased the interferer coherence, either by increasing the listener-interferer distance or making the room more reverberant. 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 two ears, and hence cannot be fully canceled. As a result, there is more masking and lower speech intelligibility, as shown by previous binaural unmasking experiments (Licklider, 1948; Robinson and Jeffress, 1963).

Different approaches have been proposed to predict the effects of reverberation in multiple-source situations. Wijn-gaarden and Drullman (2008) extended the speech transmission index method to predict spatial unmasking effects. This approach offers the advantage of taking into account the influence of reverberation on the target. However, it also makes the initial assumption that the speech target is the only source of modulation in the signals reaching the listener's ears. The aim is to look for modulation to measure the transmission of speech. This approach does not offer any chance to be extended to modulated noise or speech interferers because the modulation is then coming from both target and interferer. Zurek *et al.* (2004) proposed a model predicting the detection of a narrow band noise target against a broadband noise interferer in rooms, which was later extended to predict speech intelligibility (Freymann and Zurek, 2008). 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) and an E-C mechanism. 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 to the use of room statistics rather than room impulse responses, and of a fixed interaural correlation function in the E-C mechanism, independent of the position considered in the room, rather than the measured interaural coherence. Beutelmann and Brand (2006) developed a model predicting the intelligibility of a speech target against a noise interferer in rooms. The stimuli obtained after convolution by the binaural room impulse responses are first processed through an E-C stage which takes into account the deleterious effect of reverberation on binaural interaction and are then resynthesized. The speech intel-

ligibility index (SII) method (ANSI, 1997) is then used to evaluate speech intelligibility in the given configuration. The E-C stage consists of a direct implementation of an E-C process which tests different delays and attenuations for the signals at the ears, and chooses those maximizing the target-to-interferer ratio. The agreement with listening test data was very good, with an overall correlation coefficient of 0.95 between measurement and prediction.

The aim of the present study was to validate another method which quantitatively predicts the loss of intelligibility associated with the impairment of binaural interaction in rooms. Like the model proposed by Beutelmann and Brand (2006), this method is based on the information contained in the room impulse responses. However, whereas Beutelmann and Brand (2006) were primarily concerned with differences between normal-hearing and hearing-impaired listeners in noise and reverberation, we looked for a more computationally efficient prediction method that could be applied in architectural acoustic software. We therefore abandoned the direct implementation of a complex E-C mechanism, in favor of the formula proposed by Culling *et al.* (2005) for the evaluation of BMLDs. This formula has been used to help explain intelligibility in anechoic configurations involving multiple interferers at various azimuths (Culling *et al.*, 2004). Lavandier and Culling (2007, 2008) interpreted the influence of reverberation on binaural interaction as being mediated mainly by the interaural coherence of the interferer, but their experiments were realized with a fixed azimuth separation of target and interferer. The formula used in the present study not only takes into account the dependence of BMLDs on the interaural coherence of the interferer but also on the difference of interaural phase difference between target and interferer. As a result, it seemed to be a good candidate to predict the dependence of binaural unmasking on the level of room reverberation as well as on the azimuth separation of sources. In order to validate the method, predictions were compared to psychoacoustical measurements involving several rooms and source configurations.

Our approach to investigate cocktail-parties in rooms consists of decomposing the various effects taking place simultaneously, in order to predict them individually, and eventually build a global prediction of intelligibility. Because this study was focused on the prediction of binaural interaction, we tried to eliminate any confounding effect associated with better-ear listening. Experiments have attempted to separate the roles of these two components in several ways. Bronkhorst and Plomp (1988) and Culling *et al.* (2004) employed signal-processing techniques in order to create stimuli with only the ITDs that underpin binaural interaction or only the interaural level differences (ILDs) that underpin better-ear listening. In the experiments presented in this paper, the head of the listener was not modeled in order to avoid frequency-dependent head shadow effects which could complicate the interpretation of the effect associated with binaural interaction. Some frequency-dependent ITD effects might have been excluded as well by the removal of the head, but, as discussed at the end of the paper, this limitation should not compromise the validity of the proposed prediction method. The effects of the room and the source-listener distance on

the broadband sound level of sources were also eliminated, by equalizing all stimuli in level independently at the left and right ears, removing broadband interaural level differences.

Speech reception thresholds (SRTs) (i.e., the level of the target compared to that of the interferer for 50% intelligibility of the target) were measured under headphones, using spatially separated targets and interferers simulated in virtual rooms. The simulation allowed variation in the positions of the listener and sources, of the size of the room, and of the absorption coefficient of the room boundaries. To investigate the effect of reverberation on the interferer unambiguously, the target was always anechoic. This experimental design is not realistic, as it implies listening simultaneously to two sources in rooms having different reverberation characteristics, but it offers the advantage of isolating the effect of reverberation on the interferer. The interferer was simulated to be in rooms with different sizes and absorptions, at different distances and azimuths from the listener. Reverberation reduces binaural interaction, but when speech interferers are involved, it impairs intelligibility also by affecting monaural speech segregation mechanisms (Lavandier and Culling, 2008). The sound reflections in the room can disrupt the segregation of competing sentences 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 interaction without mixing it with these additional effects, the experiments presented here used only continuous speech-shaped noise interferers. These continuous noises had no fundamental frequency (F0) and no modulation in their temporal envelope, so that source segregation based on F0 and “dip listening” was eliminated.

The different components of the proposed prediction method were tested in specifically designed experiments. Binaural interaction depends on the interaural phase of the sources and on the interaural coherence of the interferer. Experiment 1 tested whether the proposed method could quantitatively predict the influence of the interaural phase of sources on binaural interaction in rooms. SRTs were measured for four different azimuth separations of target and interferer, at different levels of reverberation. The method based on calculated BMLDs led to good predictions. Experiment 2 tested whether the method could predict the influence of the interaural coherence of the interferer. This coherence was varied in many different ways—changing the absorption and the size of the room, the distance between the listener and the interferer, and the asymmetry of the configuration. The corresponding SRTs were well described by the prediction method. Experiment 2 also showed the importance of taking into account the influence of room coloration on intelligibility.

II. GENERAL METHODS

A. SRT measurements

1. Stimuli

The same voice was used for the target in every experiment. The corpus of sentences was from the Harvard Sen-

tence List (IEEE, 1969). The anechoic recordings of a male voice digitized at 20 kHz with 16-bit quantization were used as the basis of all stimuli. 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 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.

The virtual rooms used in the experiments were simulated using a ray-tracing method (Allen and Berkley, 1979; Peterson, 1986), implemented in the lWAVE signal processing package (Culling, 1996). The absorption coefficients of the room internal surfaces were all set to the same value, and varied together when the reverberation condition was changed. The target was always anechoic, an absorption coefficient of 1 being used for the corresponding room. The listener was modeled as two ears with no head between them, simulated by omnidirectional microphones separated by 18 cm and suspended in space at 1.5 m from the floor. The interferer and target were placed at different positions in the room, also at 1.5 m from the floor. Binaural stimuli were produced by calculating the impulse responses between the source positions and each ear, and convolving the speech and noise samples with these impulse responses. When the interferer was further away from the listener than the target, the arrival time of all interferers was set below the one of the target by deleting part of the silent period at the beginning of their waveforms, to avoid having the target starting before the interferer when stimuli were played to the listeners.

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). The root-mean-square (rms) powers of the left and right channels of the convolved stimuli were equalized independently. The broadband target-to-interferer level ratio was then fixed at the ears. Using stimuli equalized before the convolution would have placed control of the target-to-interferer level ratio at the emission of the sound sources. The influence of the room on the sound levels of interferer and target would then have been included in the measured SRTs. In order to interpret the room-induced binaural interaction changes on SRTs unambiguously, all stimuli were equalized after convolution by the room impulse responses.

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, listeners could either enter a transcript on a computer keyboard or replay the 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.

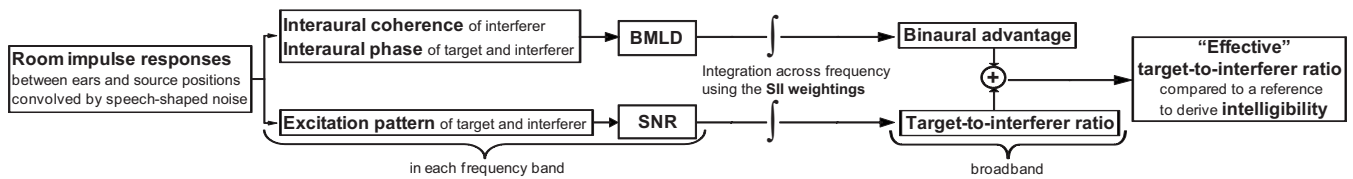


FIG. 1. Proposed method for predicting the intelligibility of a near-field speech target against a localized noise interferer in rooms.

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. Subsequent 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 fixed order randomly chosen ($N=8$ in experiments 1a and 1b, $N=16$ in experiment 2). 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 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 HD650 headphones in a single-walled IAC 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 was familiar with the sentences used during the test. Each listener participated in only a single session of a given experiment.

B. Prediction method

Figure 1 presents the method we used to predict the intelligibility of a near-field speech target against a localized noise interferer in rooms. This method requires calculation of the interaural phase and excitation pattern of target and interferer, and of the interaural coherence of the interferer. These acoustical parameters were derived from the room impulse responses of the tested configuration. First, “target” and “interferer” waveforms were created by convolving speech-shaped noise with the binaural impulse responses be-

tween the listener and source positions. These signals were equivalent to those delivered to the listeners over headphones in the psychophysical experiments, except that the prediction used noise instead of sentences for the target. The signals were used to evaluate separately the binaural and monaural components of masking by the interferer.

The interaural parameters allow the evaluation of the binaural advantage associated with the azimuth separation of sources, by calculating the predicted BMLD of the tested configuration. The signals were passed through a gamma-tone filterbank (Patterson *et al.*, 1987, 1988) with two filters per equivalent rectangular bandwidth (ERB) (Moore and Glasberg, 1983) which covered the frequency range 20 to 10 000 Hz.¹ The left and right channels of the filtered waveforms were cross-correlated within a 100-ms exponentially tapering temporal window, with delays determined between + and -5 ms, as implemented in lWAVE (Culling, 1996). The interaural coherence of the interferer and the interaural phase differences of both target and interferer were obtained in each frequency band from the interaural cross-correlation functions. The coherence was computed as the maximum of the cross-correlation, whereas the phase difference was calculated by multiplying the delay corresponding to this maximum by the center frequency of the band. The BMLD was then computed as a function of frequency following the calculation proposed by Culling *et al.* (2004, 2005), using the formula given in Culling *et al.*, 2005,

$$\text{BMLD} = 10 \log_{10}([k - \cos(\Phi_T - \Phi_I)]/[k - \rho]), \quad (1)$$

with

$$k = (1 + \sigma_e^2) \exp(\omega^2 \sigma_s^2), \quad (2)$$

where ρ is the interaural coherence of the interferer, Φ_T and Φ_I are the interaural phase differences in radians of the target and interferer, ω is the center frequency of the band in rad/s, $\sigma_s = 0.000105$, and $\sigma_e = 0.25$ (Durlach, 1972). Where the formula returns a negative BMLD, the value is set to zero, following the assumption that binaural thresholds are never below their corresponding monaural thresholds (Durlach, 1963). This computation was realized at four epochs (starting at 0.5, 1, 1.5, and 2 s) during the 4.3-s-duration noises, and the four resulting BMLD values were averaged. Finally, the broadband binaural advantage was obtained by integrating the average BMLD function along the frequency scale, using the SII weightings (ANSI, 1997).

The spectrum of a room impulse response depends on the room and the positions of listener and source. Because this spectrum is not flat, the convolution by a room impulse response alters the noise interferer spectrum (in the present study, the target was always anechoic, so its spectrum was not affected by reverberation). The room-induced coloration

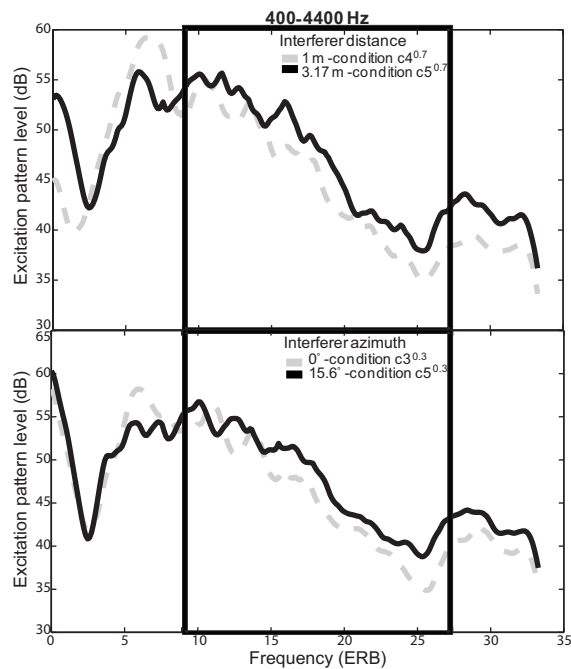


FIG. 2. Minimum excitation pattern levels at the ears for a speech-shaped noise in four interferer conditions of experiment 2, revealing the influence of room coloration. The top panel compares two conditions ($c4^{0.7}$, $c5^{0.7}$) which differed by their listener-interferer distance; the bottom panel compares two conditions ($c3^{0.3}$, $c5^{0.3}$) which differed by their interferer azimuth (see Fig. 7 and Table II for details). Each panel highlights the frequency range most important for intelligibility (400–4400 Hz) according to the SII weightings (ANSI, 1997). In each panel, the dark pattern was almost always above the dashed pattern in this frequency range, indicating an interferer more masking for speech.²

of the interferer can influence target intelligibility, even after rms equalization, because frequency bands are not all equivalent regarding intelligibility. An interferer that has more energy in a frequency range important for intelligibility will cause more masking than an interferer that has more energy in a frequency range not important for intelligibility. Figure 2 presents the excitation pattern of the interferer in four conditions of experiment 2. These conditions are described in detail in Sec. IV (Fig. 7, Table II). The top panel compares two conditions differing in their listener-interferer distance; the bottom panel compares two conditions differing in their interferer azimuth. Even though these conditions were equalized in broadband level, the level of the interferer was not the same in every frequency band in the different conditions. The room did not filter the interferer in the same way for different interferer positions. Figure 2 (box) highlights the frequency range most important for intelligibility according to the SII weightings. In each panel, the solid black line is the excitation pattern of an interferer that had more energy than the other (dashed gray line) in this frequency region. The former was therefore a stronger masker for speech.²

Even if our experiments were specifically designed to test binaural interaction and eliminate head shadow (no head simulated, broadband ILDs equalized), better-ear listening was reduced but still possible due to frequency-dependent ILDs introduced by room coloration. Our prediction method had to take this influence into account by computing a monaural component of masking by the interferer. Better-ear lis-

tening was simulated by frequency bands following the model proposed by Zurek (1993) to describe head shadow in anechoic situations. The target-to-interferer ratio was evaluated using the excitation pattern of the target and interferer waveforms (Fig. 1). The excitation pattern was computed between 0 and 33.25 ERBs (corresponding to the frequency range 0–10 kHz), every 0.13 ERBs, at the left and right ears of the listener. The signal-to-noise ratio (SNR) was evaluated in each frequency band at each ear by computing the difference between the target and interferer patterns.³ To evaluate a better-ear SNR at each frequency, the maximum of the ratio at the left and right ears was considered. The target-to-interferer ratio was computed by integrating this maximum ratio across frequency using the SII weightings.

The prediction method combines the broadband binaural and monaural components by adding them together, in order to estimate the “effective” target-to-interferer ratio associated with the tested configuration. Although some studies decomposing binaural interaction and better-ear listening found that their combined effect was smaller than the sum of the two isolated effects (Bronkhorst and Plomp, 1988; Culling *et al.*, 2004), additive models proved to work well in other cases (Hawley *et al.*, 2004; Zurek, 1993). Because the influence of better-ear listening was limited by our experimental protocol, we used an additive model as a first approximation. The prediction method is relative; it does not give an absolute evaluation of intelligibility. The target-to-interferer ratio needs to be compared to a reference configuration to derive an intelligibility prediction. In the experiments presented here, this ratio was directly compared to the SRT measured for the given room and source configurations. Because SRTs are target-to-interferer ratio by definition, the comparison of measurement and prediction did not require any index calculation (AI, SII) nor any index-to-intelligibility mapping function (Beutelmann and Brand, 2006; Levitt and Rabiner, 1967b). To allow the comparison, the average SRT of the experiment was chosen as a reference. A constant value was added to the predictions, so that the average prediction was equal to the average SRT. This procedure is equivalent to the normalization of all thresholds to a reference condition, as done in spatial unmasking experiments which usually use a reference condition where target and interferer are co-located (Zurek, 1993; Zurek *et al.*, 2004). The prediction of the threshold difference between two conditions is not affected, and it can be compared to measurement without taking into account the influence of speech material on thresholds (ANSI, 1997).

III. VARYING THE AZIMUTH SEPARATION OF SOURCES IN A GIVEN CONFIGURATION (EXP. 1)

Experiment 1 was designed to test whether the proposed prediction method could describe the influence of the relative interaural phase of sources on binaural interaction in rooms. SRTs were measured for an interferer at a fixed position, with an anechoic target tested at different azimuths around the listener. Different levels of reverberation were applied to the interferer, by varying the absorption coefficient of the room. As interaural phase varies with the target azimuth, and as the coherence of the interferer at a fixed position only

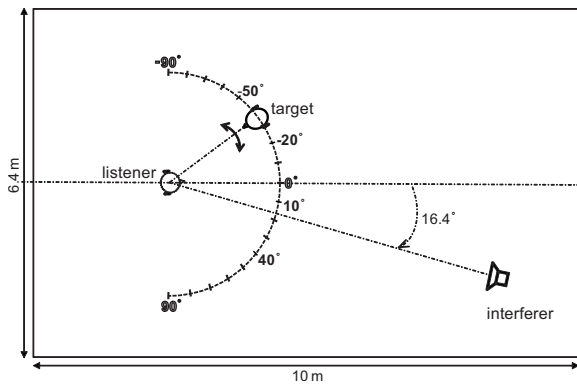


FIG. 3. Virtual room and spatial configuration used in experiment 1, and for the calculations of Fig. 5. The interferer was at 6.16 m from the listener, at 16.4° of azimuth. The target was at 2 m from the listener. In experiment 1, it was tested at -50°, -20°, 10°, and 40° of azimuth. The calculations of Fig. 5 used a target every 10° between -90° and 90° of azimuth, and a target at 16.4° of azimuth.

depends on the level of reverberation, this design produced variation in the BMLDs with target azimuth at several fixed levels of interferer coherence.

A. Design

Experiment 1 was based on the configuration presented in Fig. 3. The target was at 2 m from the listener and was always anechoic. SRTs were measured for four target azimuths: -50°, -20°, 10° and 40°. The room was 10-m long, 6.4-m wide and 2.5-m high. The ears of the listener were placed along an axis parallel to the 6.4-m wall on either side of a center point located at 3.2 m from the 10-m wall and 2.5 m from the 6.4-m wall. The interferer was at 6.16 m from the listener, at 16.4° of azimuth. When computing the room impulse responses used for the interferer, three absorption coefficients were used: 1 (anechoic), 0.7 (mildly reverberant), and 0.2 (very reverberant). These room characteristics and positions were chosen so that they led to different values of interferer coherence at the listener position. Table I shows that the “wide band” coherence of the interferer decreased with increasing reverberation (decreasing absorption coefficient). This form of coherence was considered in two previous studies (Lavandier and Culling, 2007, 2008). It was calculated using the waveforms obtained from the convolution of a speech-shaped noise with the binaural impulse response between the interferer and listener positions. Coherence was computed as the maximum of the cross-correlation of the waveforms.⁴ These waveforms were filtered between 20 and 1500 Hz prior to calculation, in order to consider only the frequency range for which binaural unmasking is most effective in broadband noise. For comparison with

TABLE I. Reverberation time, direct-to-reverberant ratio and interaural coherence, for a source at the interferer position of Fig. 3, as a function of the room absorption coefficient used for the interferer in experiment 1.

Absorption coefficient	RT (s)	D/R ratio (dB)	Interaural coherence
1	0	∞	1
0.7	0.22	-3.1	0.89
0.2	1.30	-15.7	0.75

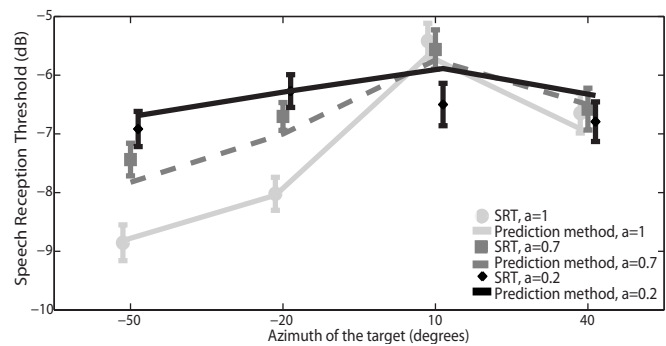


FIG. 4. Mean SRTs with standard errors measured in experiment 1, for four target azimuths and three absorption coefficients (a) used for the interferer. The SRTs are compared to the results of the prediction method.

common room acoustical measurements, Table I also presents the reverberation times (RTs) and direct-to-reverberant ratios involved in experiment 1. They were calculated at both ears, and the results were averaged across ears. The broadband RT was evaluated using a speech-shaped noise convolved with the impulse response between the interferer position and the considered ear, by measuring the time required for the sound level to decrease by 60 dB after the offset of the noise. The direct-to-reverberant energy ratio (D/R ratio) was calculated using the same impulse response. The direct-path and reverberant portions of this response were separated by a time-windowing procedure. This procedure was done by eye and was trivial as our simulated impulse responses were not contaminated by noise. The energy of each portion was computed, and the D/R ratio was expressed in dB. In the anechoic room, the D/R ratio was infinite, the RT was 0, and the coherence was 1, as there was no reverberant field.

In experiment 1a, the interferer was either anechoic (absorption coefficient of 1) or very reverberant (absorption coefficient of 0.2). A session of the experiment consisted of measuring the SRTs for the four target azimuths at two levels of reverberation for the interferer, resulting in eight different conditions. Twenty-four listeners took part in a 40-min session of the experiment. The aim of experiment 1b was to test an intermediate level of reverberation for the interferer. A session of the experiment consisted of measuring the SRTs for the four target azimuths with a mildly reverberant interferer (absorption coefficient of 0.7), twice for each listener, resulting in eight conditions. The experiment used the same target sentences as experiment 1a. Sixteen listeners took part in a 40-min session of the experiment; none of them had participated in experiment 1a.

B. Results

Figure 4 presents the mean SRTs measured in experiment 1, for four target azimuths and three reverberation conditions for the interferer. With the anechoic and mildly reverberant interferers, the SRT was maximal when the target azimuth was 10°, close to that of the interferer (16.4°), and decreased when the target was moved away from the interferer. When reverberation was added to the interferer, SRTs tended to increase and became independent of the target azimuth with the very reverberant interferer. Figure 4 also presents the results of the prediction method. The proposed

method led to good predictions of the measured SRTs, with a 0.95 correlation obtained between the SRTs and their predicted value.

Two within-subject analyses of variance confirmed that, on average, reverberation on the interferer impaired intelligibility [$F(1,23)=18.3$, $p<0.001$, Exp. 1a], and that increasing the azimuth separation of sources could produce a significant release from masking [$F(3,69)=15.5$, $p<0.0001$, Exp. 1a], even with a moderately reverberant interferer [$F(3,45)=6.1$, $p<0.01$, Exp. 1b]. The interaction between the effects of reverberation and target azimuth was significant in experiment 1a [$F(3,69)=8.0$, $p<0.001$]. A simple-main-effect analysis of this interaction showed that the effect of reverberation was significant at all target azimuths except 40° [$F(1,23)>14$, $p<0.0015$ in each case]. The azimuth of the target had a significant effect on SRTs when the interferer was anechoic [$F(3,69)=27.5$, $p<0.0001$], but not when it was very reverberant. Tukey pairwise comparisons showed that, in experiment 1a, the target azimuth of -50° led to lower SRTs than the target azimuths of -20° , 10° , and 40° , and that the target azimuth of -20° also led to lower SRTs than the target azimuth of 10° [$q>5.7$, $p<0.001$ in each case]. The SRTs for the target azimuth of 40° were not significantly different from the SRTs for the target azimuths of -20° and 10° . In experiment 1b, the target azimuth of -50° led to lower SRTs than the target azimuth of 10° [$q>6$, $p<0.001$]. The SRTs for the other target azimuths were not significantly different.

C. Discussion

Experiment 1 validated the prediction method for sources with different azimuth separations in a given room. A high correlation was obtained between measurements and predictions. Lavandier and Culling (2007, 2008) showed that the effect of reverberation on binaural interaction was related to the interaural coherence of the interferer. These new results show that BMLD calculations also incorporating the interaural phase of the target and interferer can account for the effects of both azimuth and reverberation level.

To obtain a more detailed picture of the effect of separation in azimuth, Fig. 5 presents the predicted binaural advantage calculated for the target placed every 10° between -90° and 90° of azimuth around the listener, and at 16.4° of azimuth to test for no azimuth separation between target and interferer. This calculation was done at the three levels of reverberation used for the interferer in experiment 1. It should be noted that the binaural advantage curves were not symmetric relative to the interferer azimuth, because the tested configuration was not symmetric within the room, the interferer being on the right of the listener (Fig. 3). The binaural advantage increased when the target azimuth was moved away from that of the interferer and started to flatten for azimuth separations above 50° . These predictions are in agreement with both the anechoic measurements of Bronkhorst and Plomp (1988) and the results of experiment 1. Bronkhorst and Plomp (1988) measured the azimuth dependency of binaural interaction (“dT” condition) using a frontal target at 0° , and moving the interferer by steps of 30°

between 0° and 180° . They measured a 3-dB effect of the azimuth at 30° , but no effect with increasing azimuth until the source reached the symmetrical position at the back of the listener, at 150° , where the SRTs became sensitive to the source azimuth again. Our measurements were done between -50° and 40° , with the same 30° step size and the competing source at 16.4° . Therefore, the main difference between our anechoic condition and the dT condition of Bronkhorst and Plomp (1988) is that we limited the range of tested positions to those where the influence of source azimuth was maximal, whereas Bronkhorst and Plomp (1988) measured SRTs at equally spaced azimuths over 180° .

For spatially separated sources, reverberation of the interferer reduced both the predicted binaural advantage and its sensitivity to changes in azimuth separation of sources. These reductions resulted from two effects acting in parallel so that an E-C mechanism would gradually become inefficient with increasing reverberation: the decrease in interferer coherence, ρ , which evaluates the proportion of noise energy that is susceptible to cancellation, and the modification of the interferer interaural phase difference, Φ_I , which needs to be different from that of the target for cancellation to result in an improvement in internal signal-to-noise ratio. To evaluate the relative contributions of reverberant coherence and phase to the reduction in the predicted binaural advantage, Fig. 6 presents model calculations applied to stimuli in which phase and coherence were varied independently. For example, the binaural advantage was computed at all target azimuths using the phase of the anechoic interferer and the coherence of the mildly reverberant interferer (dashed line, panel I) or very reverberant interferer (dashed line, panel III). The three levels of reverberation used in Fig. 5 are considered here by pairs. The predictions at the two levels of each pair are compared with predictions involving the phase of one level and the coherence of the other, and reciprocally. Increasing reverberation on the interferer coherence reduced the binaural advantage for all target azimuths (e.g., light gray line above dashed line in panel III), in agreement with binaural unmasking theories, with a floor effect observed for azimuths close to that of the target. Increasing reverberation on the interferer phase reduced the binaural advantage sensitivity to changes in azimuth separation of sources (e.g., light gray line vs dotted line in panel III). For small azimuth separations, where BMLDs are small in anechoic situations, the room reflections distributed part of the interferer energy at different phases, corresponding to large “apparent” azimuth separations for the associated virtual images of the interferer, so that this energy could be canceled and the binaural advantage increased. For large azimuth separations, where anechoic BMLDs are large, reflections brought part of the interferer energy at phases close to that of the target. This energy could not be canceled and the binaural advantage decreased. Beutelmann and Brand (2006) interpreted the influence of early reflections as responsible for the creation of mirror sources acting as secondary interferers with considerably different azimuths, thus disturbing the binaural system (cafeteria condition). These mirror sources would indeed greatly reduce head shadow if they were on the other side of the head compared to the interferer (Culling *et al.*, 2004; Hawley *et al.*,

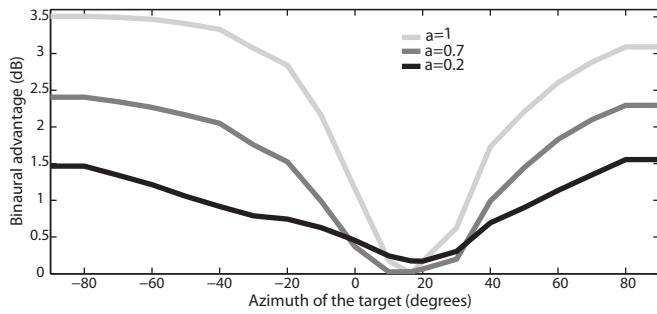


FIG. 5. Predicted binaural advantage as a function of target azimuth in the configuration of Fig. 3, for the three absorption coefficients (a) used for the interferer in experiment 1.

2004), and our results show that they would also influence binaural interaction if they affect the interferer coherence and phase at the ears. It should be noted, however, that mirror sources might not systematically reduce binaural interaction. When considering only interaural phase, if the interferer azimuth is close to that of the target, then a mirror source at a widely different angle could be far from the target azimuth, distributing interfering energy where it can be canceled, so that the binaural advantage might increase depending on the associated effect of reflections on coherence.

When sources have the same interaural phase, reducing the coherence of the interferer can produce a small release

from masking (Licklider, 1948). A corresponding effect occurred in experiment 1a (Fig. 4). For the target azimuth of 10° , close to that of the interferer (16.4°), the SRT significantly decreased when reverberation was added to the interferer. This small effect was somewhat larger than the prediction based on the BMLD calculations, which indicated no effect for the target at 10° and a very limited effect for the target at 16.4° (black line above the light gray line in Fig. 5). No such effect was predicted for the mildly reverberant interferer (dark gray line below the light gray line).

In experiment 1, the three levels of reverberation tested for the interferer led to very limited differences in room-induced coloration. The SII-weighted target-to-interferer ratio in the anechoic room was 0.19 dB lower than in the very reverberant room, and 0.21 dB lower than in the mildly reverberant room. Even without taking into account these small monaural differences, a 0.94 correlation was obtained between the SRTs and their predicted value based only on the binaural component of the method. The coloration did not play an important role here, because only one interferer-listener configuration was tested. More configurations and rooms were involved in experiment 2, so that among other things the monaural component of the prediction method was more thoroughly tested.

IV. VARYING THE CONFIGURATION FOR A GIVEN AZIMUTH SEPARATION OF SOURCES (EXP. 2)

Experiment 2 was designed to further validate the prediction method in a wide variety of configurations. To consider other aspects than those already measured and predicted in experiment 1, the azimuth separation of sources was kept constant. The aim of experiment 2 was to test whether the method could quantitatively predict the influence of the interferer coherence. This coherence was varied in many different ways. SRTs were measured while changing the absorption and the size of the room, the distance between the listener and the interferer, and the asymmetry of the configuration. Because all these design parameters could not be tested systematically here, the aim of the experiment was not to give a general conclusion on the link between each individual parameter and intelligibility. For example, we did not want to link distance or absorption to intelligibility. We were rather interested in the link between interferer coherence and intelligibility. Because the coherence of a source can be modified in many different ways in a room, we did not want to limit ourselves to one way or another. So, we varied the various design parameters that could influence coherence, and then we selected conditions leading to a large range of interferer coherence using these different parameters.

A. Design

In all conditions of experiment 2, the target was anechoic and placed at 2 m from the listener, at 60° on the left of the interferer.⁵ The interferer was tested at different positions in different rooms.

Figure 7 and Table II present the room, absorption coefficient, and position used for the interferer in each condition of experiment 2. Room m (medium) was the room used in

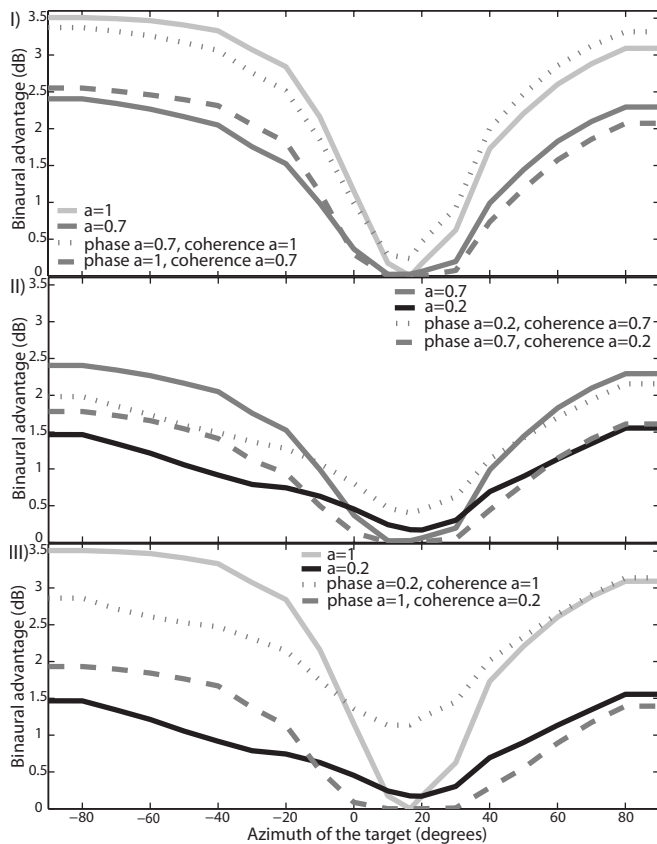


FIG. 6. Relative contributions of reverberant coherence and phase to the reduction of predicted binaural advantage. The predictions of Fig. 5 obtained with three absorption coefficients (a) used for the interferer are presented by pairs ($a=1-0.7$ in I, $a=0.7-0.2$ in II, and $a=1-0.2$ in III). Each pair is associated with the corresponding predictions involving the interferer phase obtained with the first coefficient and the interferer coherence obtained with the second, and reciprocally.

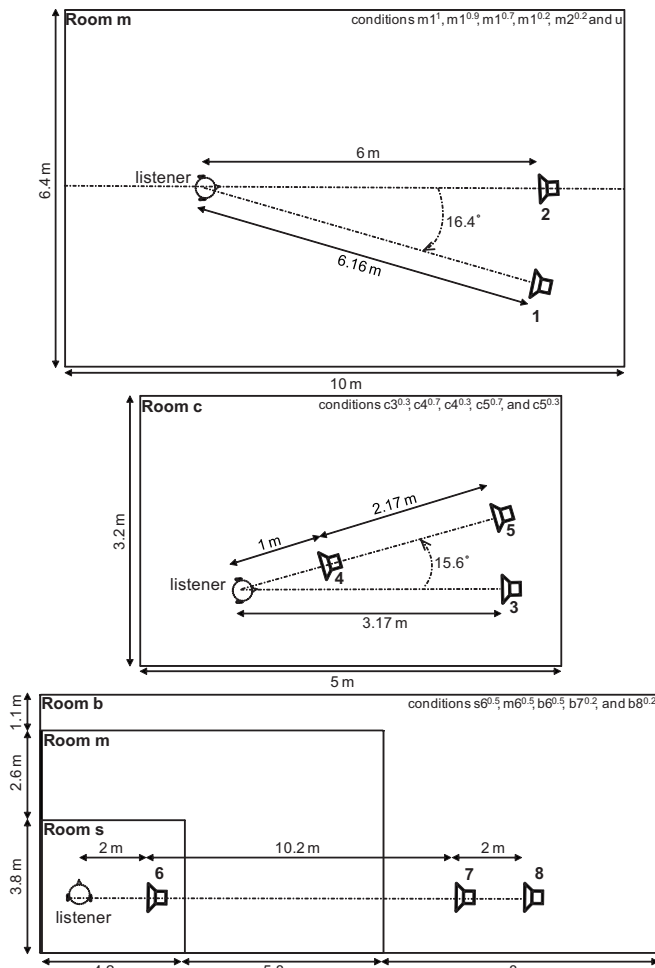


FIG. 7. Virtual rooms and positions used for the interferer in experiment 2. The anechoic target was always at 2 m from the listener, at 60° on the left of the interferer.⁵

experiment 1, with the same listener position used in conditions $m1^1$, $m1^{0.9}$, $m1^{0.7}$, $m1^{0.2}$, $m2^{0.2}$, and u . Position 1 was

identical to the interferer position of experiment 1. Position 2 was on the median axis of the room, directly in front of the listener. In room c [common with Culling *et al.* (2003); Lavandier and Culling (2007)], the ears of the listener were placed along an axis parallel to the 3.2-m wall on either side of a center point located at 0.9 m from the 5-m wall and 1.2 m from the 3.2-m wall. Position 3 was directly in front of the listener, whereas positions 4 and 5 were at 15.6° azimuth, at two different distances. Conditions $s6^{0.5}$, $m6^{0.5}$, $b6^{0.5}$, $b7^{0.2}$, and $b8^{0.2}$ used rooms s (small), m , or b (big), with the ears of the listener placed along an axis parallel to the longer wall on either side of a center point located at 1.6 m from the longer wall and 1.2 m from the shorter wall. Positions 6, 7, and 8 were at different distances from the listener, at 90° azimuth. The sizes of the rooms and the listener-interferer distances are detailed in Fig. 7.

Table II presents the reverberation time, direct-to-reverberant ratio, and interaural coherence for a source at the interferer position, as well as the predicted binaural advantage, for all tested conditions. Because interaural coherence was the parameter of interest here, the conditions were ranked by decreasing interferer coherence. Condition $m1^1$ used an anechoic interferer, with a coherence of 1. The configuration of condition $m2^{0.2}$ was symmetric regarding the room, the interferer, and the ears of the listener. In this special case, listener and interferer were both on the median axis of the room, so that the reflections getting to the listener from the interferer were exactly the same at the two ears. The room was very reverberant (D/R ratio of -16.7 dB), but because the reverberation was identical at the two ears, it did not reduce the coherence of the interferer. This coherence was 1 as in the anechoic case, and the predicted binaural advantage was similar in the two conditions. Conditions $m1^{0.9}$ to $c5^{0.3}$ used reverberant interferers with progressively lower interaural coherence. Condition u was introduced to include the reference of an interferer completely uncorrelated

TABLE II. Room, absorption coefficient and position used for the interferer in each condition of experiment 2 (see Fig. 7 for details of rooms and positions). Reverberation time, direct-to-reverberant ratio, and interaural coherence for a source at the interferer position, and predicted binaural advantage in each condition. The conditions are ranked by decreasing coherence.

Condition	Room	Absorption coefficient	Position	RT (s)	D/R ratio (dB)	Interaural coherence	Binaural advantage (dB)
$m1^1$	m	1	1	0	∞	1	3.32
$m2^{0.2}$	m	0.2	2	1.23	-16.7	1	3.08
$m1^{0.9}$	m	0.9	1	0.09	3.7	0.96	2.87
$c4^{0.7}$	c	0.7	4	0.12	4.9	0.92	2.40
$m1^{0.7}$	m	0.7	1	0.22	-3.1	0.89	2.12
$c3^{0.3}$	c	0.3	3	0.44	-12	0.83	1.92
$c4^{0.3}$	c	0.3	4	0.47	-3	0.79	1.82
$m1^{0.2}$	m	0.2	1	1.30	-15.7	0.75	1.01
$c5^{0.7}$	c	0.7	5	0.16	-2.7	0.72	1.61
$m6^{0.5}$	m	0.5	6	0.39	-2	0.68	0.34
$b6^{0.5}$	b	0.5	6	0.64	-1.6	0.67	0.43
$b7^{0.2}$	b	0.2	7	2.34	-19.2	0.65	0.18
$s6^{0.5}$	s	0.5	6	0.20	-4.4	0.61	0.42
$b8^{0.2}$	b	0.2	8	2.45	-20.4	0.54	0.51
$c5_{0.3}$	c	0.3	5	0.47	-12.2	0.48	0.97
u	m	1	1 ^a	0	∞	0.05	0.52

^aUsing independent noises for each ear.

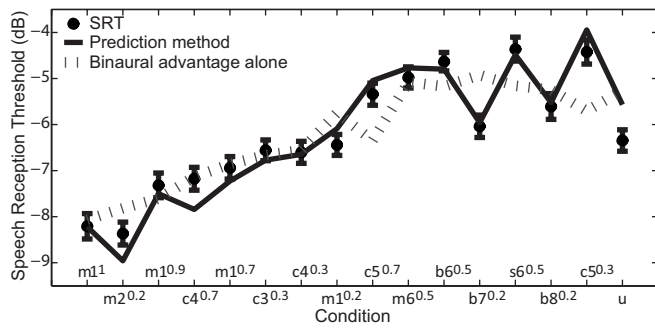


FIG. 8. Mean SRTs with standard errors measured in the sixteen conditions of experiment 2. The SRTs are compared to the results of the prediction method. The predictions based on the binaural advantage alone are also plotted.

at the two ears, with a coherence very close to 0. It used exactly the same configuration as the anechoic condition $m1^1$, but instead of convolving one speech-shaped noise with the binaural impulse response, two independent speech-shaped noises were used. One noise was convolved with the impulse response corresponding to the left ear; the other noise was convolved with the impulse response corresponding to the right ear.

A session of experiment 2 consisted of measuring the SRTs for the sixteen conditions. The experiment used different target sentences from experiment 1. Forty-eight listeners took part in a 70-min session of the experiment.

B. Results

Figure 8 presents the mean SRTs measured in the 16 conditions of experiment 2, the conditions being ranked by decreasing coherence. The SRTs initially increased with decreasing coherence (from conditions $m1^1$ to $b6^{0.5}$), but the relation was not monotonic at low coherence values (from conditions $b6^{0.5}$ to u). The results of the prediction method in the different conditions are also presented. The proposed method led to good predictions of the measured SRTs, with a 0.97 correlation obtained between the SRTs and their predicted value. The predictions based on the binaural advantage

alone are also plotted. They predicted the general trend of increasing SRTs with decreasing coherence, but the correlation between the SRTs and their predicted value was only 0.84; they could not predict the non-monotonic behavior of the SRTs from condition $b6^{0.5}$ to condition u . Here, room coloration had substantial effects on SRTs, according to our model. Some conditions caused more masking than others because the corresponding interferers had more energy in frequency channels important for intelligibility.

Table III compares the SRTs of experiment 2 using a paired t-test analysis. The significant differences are indicated by crosses. The individual significance level was fixed to 0.0004, following the Bonferroni correction (Keppel and Wickens, 2004), in order to keep the family-wise significance level below 0.05 for the 120 comparisons realized. The changes in coherence resulting from the variation in the different design parameters often led to significant differences in measured SRTs. For example, conditions $m1^1$, $m1^{0.9}$, $m1^{0.7}$, and $m1^{0.2}$ differed only by the absorption used for the interferer, this absorption decreasing from $m1^1$ to $m1^{0.2}$. The anechoic condition $m1^1$ led to a significantly lower SRT than the two most reverberant conditions $m1^{0.7}$ and $m1^{0.2}$. Conditions $m2^{0.2}$ and $m1^{0.2}$ differed in the azimuth of the interferer. They used the same absorption and roughly the same level of reverberation (Table II), but condition $m2^{0.2}$ used a symmetric interferer with a coherence unaffected by reverberation. Condition $m1^{0.2}$ with the reduced coherence led to a significantly higher SRT than condition $m2^{0.2}$. The reverberation in condition $m2^{0.2}$ did not impair intelligibility, as shown by its SRT not being significantly different from that of the anechoic condition $m1^1$. The three pairs of conditions $c4^{0.7}$ – $c5^{0.7}$, $c4^{0.3}$ – $c5^{0.3}$, and $b7^{0.2}$ – $b8^{0.2}$ correspond to three different rooms, in which the distance between interferer and listener was varied. Increasing this distance led to significantly higher SRTs in the case $c4^{0.7}$ – $c5^{0.7}$ and $c4^{0.3}$ – $c5^{0.3}$, but not in the case $b7^{0.2}$ – $b8^{0.2}$. Conditions $m6^{0.5}$, $b6^{0.5}$, and $s6^{0.5}$ used the same configuration, but the size of the room was changed around the listener and sources. There was no significant effect of the room size on the SRTs.

TABLE III. Comparison of the SRTs measured in the sixteen conditions of experiment 2 using a paired t-test analysis. The significant differences are indicated by crosses.

	$m2^{0.2}$	$m1^{0.9}$	$c4^{0.7}$	$m1^{0.7}$	$c3^{0.3}$	$c4^{0.3}$	$m1^{0.2}$	$c5^{0.7}$	$m6^{0.5}$	$b6^{0.5}$	$b7^{0.2}$	$s6^{0.5}$	$b8^{0.2}$	$c5^{0.3}$	u
$m1^1$	0	0	0	×	×	×	×	×	×	×	×	×	×	×	×
$m2^{0.2}$.	0	×	×	×	×	×	×	×	×	×	×	×	×	×
$m1^{0.9}$.	.	0	0	0	0	0	×	×	×	0	×	×	×	0
$c4^{0.7}$.	.	.	0	0	0	0	×	×	×	0	×	×	×	0
$m1^{0.7}$	0	0	0	×	×	×	0	×	0	×	0
$c3^{0.3}$	0	0	×	×	×	0	×	0	×	0
$c4^{0.3}$	0	×	×	×	0	×	0	×	0
$m1^{0.2}$	0	×	×	0	×	0	×	0
$c5^{0.7}$	0	0	0	×	0	0	0
$m6^{0.5}$	0	0	0	0	0	×
$b6^{0.5}$	×	0	0	0	×
$b7^{0.2}$	×	0	×	0
$s6^{0.5}$	0	0	×
$b8^{0.2}$	0	0
$c5^{0.3}$	×

C. Discussion

While experiment 1 showed that the prediction method correctly took into account the influence of the interaural phases of target and interferer, experiment 2 showed that it accurately described the influence of reverberation on the coherence and spectrum of the interferer, in a wide variety of configurations. A high correlation was obtained between measurements and predictions. The predictions based on the binaural advantage alone described the general trend of increasing SRTs with decreasing coherence, but they could not predict the non-monotonic behavior of the SRTs measured at high levels of reverberation. At low levels of reverberation, SRTs seemed to be predominantly influenced by interferer coherence and binaural interaction. At high levels of reverberation, the binaural advantage was very limited and it became important to take into account room coloration.

If the room reflections coming from the interferer were different at the listener's ears, then reverberation reduced the interferer coherence, resulting in lower binaural advantages and higher SRTs. No matter how late or energetic they were, these reflections did not impair intelligibility if they were identical at the two ears⁶: intelligibility in the reverberant but symmetric configuration of condition $m2^{0.2}$ was as good as in the anechoic condition $m1^1$. The reverberation on the interferer impaired intelligibility only if it reduced its coherence, in agreement with Lavandier and Culling (2008). Measuring target relevant characteristics such as the direct-to-reverberant ratio or the reverberation time could not predict this result, whereas a prediction based on the E-C theory can: similar values of binaural advantage led to similar SRTs.

In experiment 2, the binaural advantage and interferer coherence were reduced in various ways, which led to significant increases in SRT. Decreasing the room absorption gave more energy to the sound reflections, which made the interferer less coherent if these reflections were different at the two ears (conditions $m1^1$ vs $m1^{0.7}$ and $m1^{0.2}$). When the distance between interferer and listener increased, the reverberant sound, which was incoherent, became more important relative to the direct sound, which was coherent, so that the interferer coherence decreased (conditions $c4^{0.7}$ vs $c5^{0.7}$ and $c4^{0.3}$ vs $c5^{0.3}$). These effects of absorption and distance were consistent with previous measurements (Lavandier and Culling, 2007). Varying the azimuth of the interferer could also decrease its coherence and increase the SRT (condition $m2^{0.2}$ vs $m1^{0.2}$ and $c3^{0.3}$ vs $c5^{0.3}$). There was no significant effect of the size of the room used for the interferer in conditions $m6^{0.5}$, $b6^{0.5}$, and $s6^{0.5}$; but it should be noted that the predicted binaural advantage was very similar in these conditions.

Table II shows that the main influence of interferer coherence on the predicted binaural advantage took place at low levels of reverberation, for coherence between 1 and 0.75 (conditions $m1^1$ to $m1^{0.2}$). Further decreases in coherence had less influence. This is in agreement with the results of Licklider (1948), who showed that most of the variation in intelligibility for speech in noise occurs for a noise coherence between 1 and 0.75. This reduced sensitivity at low coherence might explain why the SRTs in conditions $c5^{0.7}$ vs

$c5^{0.3}$ and $b7^{0.2}$ vs $b8^{0.2}$ did not differ significantly. Conditions $m1^{0.9}$ vs $m1^{0.7}$ and $c4^{0.7}$ vs $c4^{0.3}$ did not lead to significant differences either, probably because their coherences and resulting binaural advantages were not sufficiently different. Conditions $m1^{0.9}$ and $m1^{0.2}$ were expected to result in different SRTs, due to the magnitude of their difference in interferer coherence (Lavandier and Culling, 2007). However, the 1.86-dB predicted difference of binaural advantage was slightly reduced by room coloration, so that the overall difference predicted from the model was only 1.41 dB, and the results of experiment 2 showed a non-significant difference of 0.9 dB (Fig. 8).

V. GENERAL DISCUSSION

The method proposed here to predict binaural interaction in rooms is an extension of the models of Levitt and Rabiner (1967b) and Zurek (1993). Binaural interaction is modeled by increasing the speech-to-interference ratio by the size of the BMLD for pure tone detection in noise in each frequency band. BMLDs are estimated for the given set of interaural parameters using the formula proposed by Culling *et al.* (2005). Unlike the BMLD estimations used by Levitt and Rabiner (1967b) and Zurek (1993), this formula depends on the interaural coherence of the interferer, so it can describe the decorrelation of the interferer at the listener's ears by reverberation and the corresponding impairment of binaural interaction, allowing us to extend the predictions to reverberant situations. Like the model proposed by Beutelmann and Brand (2006), the method is based on the information contained in the room impulse responses rather than room statistics (Zurek *et al.*, 2004), but the direct implementation of a complex E-C mechanism is replaced by the more computationally efficient BMLD estimation. It should be noted that the threshold difference between two compared configurations is predicted without any model parameter being fitted to the data.

The method correctly took into account the influence of the azimuth separation of sources and the importance of their interaural phases in experiment 1, with a 0.95 correlation obtained between the SRTs and their predicted value. It also accurately described the influence of reverberation on the coherence and spectrum of the interferer in experiment 2, leading to a 0.97 correlation between measurement and prediction. These results are comparable with those of Beutelmann and Brand (2006), who obtained an overall correlation of 0.95 between measured and predicted SRTs (0.97 in the anechoic room, 0.94 in the office and cafeteria). In terms of prediction errors, the maximum and mean prediction errors were 0.6 and 0.2 dB in experiment 1; they were 0.8 and 0.3 dB in experiment 2. The maximum and mean prediction errors obtained by Beutelmann and Brand (2006) were 3 and 1.6 dB in the anechoic room, about 1 and 0.5 dB in the office, and about 1 and 0.3 dB in the cafeteria. Zurek *et al.* (2004) obtained maximum prediction errors ranging between 2 and 5 dB for the binaural detection of narrow band noises in rooms. These discrepancies could be due to the approximations inherent to the use of room statistics and of a fixed correlation function in the E-C model.

Reverberation impairs binaural interaction for noise and speech interferers (Lavandier and Culling, 2008), and room coloration also affects both types of interferer. The proposed prediction method is not limited to noise interferers and can be used to evaluate part of the effect of reverberation on speech interferers. However, the prediction would not be complete, because additional speech segregation mechanisms are involved with speech interferers (Lavandier and Culling, 2008). Room reflections can disrupt the segregation of competing sentences based on fundamental frequency differences (Culling *et al.*, 2003, 1994), and they can also fill the potential silent periods in the speech interferers (Bronkhorst and Plomp, 1990; George *et al.*, 2008). These additional effects need to be described to refine the intelligibility prediction. The direct effect of reverberation on the target needs to be considered as well (Bradley, 1986; Houtgast and Steeneken, 1985).

In the experiments presented in this paper, the head of the listener was not modeled. The aim was to avoid head shadow effects which could complicate the interpretation of the effects associated with binaural interaction. The head would lead to extra interaural decorrelation in asymmetric configurations, particularly for sources to the side (Lindevald and Benade, 1986). Because of the floor effect associated with coherence, the maximum intelligibility loss with increasing reverberation should be little affected by the presence or absence of a head, but the additional decorrelation associated with the head would make the interferer coherence drop at lower levels of reverberation. There is no reason to believe that the binaural component of our prediction method would not be able to take this additional decorrelation into account. Adding the head between the ears would also introduce a frequency-dependent head shadow. This head shadow can be seen as a coloration component which is different at the two ears. The better-ear target-to-interferer ratio of our prediction correctly described the coloration introduced by the room, so it should be able to predict this head coloration as well. Zurek (1993) used a similar SNR calculation to predict the head shadow in anechoic situations. Though it seems plausible that our prediction method will describe the influence of the head, it remains to be tested. Moreover, by simply adding the binaural and monaural components of masking, the method neglects any potential interaction between better-ear listening and binaural interaction. It led to very good results in this study where the influence of better-ear listening was limited by our experimental protocol. With the extra ILDs introduced by the head, the assumption of additivity might not hold anymore. The interaction would then need to be modeled. The applicability of the method to real-life situations also requires it to be tested in real rooms. It could be done using binaural room impulse responses similar to those measured by Watkins (2005) instead of our simulated room responses.

Several studies have shown that binaural interaction can still be efficient to partly unmask a target among multiple interferers placed at different positions in an anechoic environment (Bronkhorst and Plomp, 1992; Carhart *et al.*, 1969; Culling *et al.*, 2004). Culling *et al.* (2004) showed that BMLD calculations could be used to describe SRTs mea-

sured in such configurations. The prediction method proposed in this paper might generalize to a setting with more than one interfering source in rooms. The number of these sources should, in principle, be immaterial to the model, since it is already handling multiple virtual images of one source. The parameters—interaural phase and coherence, and excitation pattern—of the single interferer of Fig. 1 would need to be replaced by the same parameters calculated on the overall interfering signals, obtained by adding the signals produced at the ears by each interferer. The prediction method does not need to “know” if the interaural phase corresponds to the position of a single interferer, to multiple interferers at different positions, or to room reflections mixed with the direct sound at the ears. In the same way, it does not matter if the source spectra at the ears result from head shadow, room coloration, or multiple sources. A prediction of intelligibility should be possible in each case because the auditory mechanisms involved remain the same.

VI. CONCLUSION

Using SRT measurements and acoustical analyses done in several virtual rooms, a method was proposed to quantitatively predict the influence of reverberation on the binaural interaction mechanism used to improve speech intelligibility in the presence of noise interferers. This method, based on predicted BMLDs, took into account the decorrelation of the interferer at the listener’s ears by room reflections, and the dependence of binaural interaction on the azimuth separation of sources. To fully predict the measured losses of intelligibility caused by reverberation, the coloration introduced by the room also had to be considered. A 0.95–0.97 correlation was obtained between the SRTs and their predicted value. The method requires calculation of the excitation pattern and interaural phase of target and interferer, and of the interaural coherence of the interferer. These parameters were all calculated from the room impulse responses convolved with noise. At low levels of reverberation, SRTs seemed predominantly influenced by binaural interaction, so that the azimuth separation of sources and the interferer coherence were critical. At high levels of reverberation, the binaural advantage was very limited and it became important to take into account room coloration.

ACKNOWLEDGMENTS

This work was supported by the UK EPSRC. The authors are grateful to the associate editor Richard Freyman and three anonymous reviewers for their helpful comments on the first version of this paper, and to all listeners who took part in the experiments.

¹The upper limit of the filterbank was unnecessarily high, given that BMLDs occur mostly below 1500 Hz (Culling *et al.*, 2004). We choose to do the calculation up to 10 000 Hz to use the same frequency range as the one used for the computation of the excitation patterns.

²The peak of each excitation pattern observed at very low frequencies (below 2.5 ERBs) was due to an artifact inherent to the room simulation program. This dc component increasing with the level of reverberation was not audible, and this irrelevant energy was not taken into account while equalizing the rms level of the stimuli.

³In this study, the target was always anechoic and equalized in rms level at

the two ears, so that the excitation pattern of the target waveform was constant across ears and conditions. Therefore, the target-to-interferer ratio only depended on the excitation pattern of the interferer.

⁴Coherence was not computed as the maximum of the absolute value of the cross-correlation, as indicated by mistake in Lavandier and Culling, 2007, 2008.

⁵The azimuth separation of sources was 60° in all conditions of experiment 2 except in the configuration used for conditions $m1^1$, $m1^{0.9}$, $m1^{0.7}$, $m1^{0.2}$, and u (which correspond to the configuration of Fig. 3 with a target azimuth of -50°), where it was 66.4°, due to a calculation error. It resulted in an azimuth separation of sources which was not perfectly constant across conditions, as had been planned, but this does not affect the validity of the results obtained and does not affect the conclusions of the study. With such a separation, following the calculations of Fig. 5, the binaural advantages in the tested configurations were mainly influenced by the changes in the interferer interaural coherence.

⁶Of course, if the configuration is such that the reflections coming from the interferer are different at the two ears, the more energetic they are, the more deleterious effect they should have on interferer coherence and target intelligibility.

- Allen, J. B., and Berkley, D. A. (1979). "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.* **65**, 943–950.
- ANSI (1997). "Methods for calculation of the speech intelligibility index," ANSI S3.5-1997, 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.
- Bradley, J. S. (1986). "Predictors of speech intelligibility in rooms," *J. Acoust. Soc. Am.* **80**, 837–845.
- 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.
- 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.
- Colburn, H. S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," *J. Acoust. Soc. Am.* **61**, 525–533.
- 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., 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.
- 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.
- Freyman, R. L., and Zurek, P. M. (2008). "Effects of reverberation on spatial release from masking (A)," *J. Acoust. Soc. Am.* **123**, 2977.
- 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.
- 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.
- 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.
- Keppel, G., and Wickens, T. D. (2004). *Design and Analysis. A Researcher's Handbook*, 4th ed. (Pearson Prentice-Hall, Englewood Cliffs, NJ).
- 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.
- Levitt, H., and Rabiner, L. R. (1967a). "Binaural release from masking for speech and gain in intelligibility," *J. Acoust. Soc. Am.* **42**, 601–608.
- Levitt, H., and Rabiner, L. R. (1967b). "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.
- Lindevald, I. M., and Benade, A. H. (1986). "Two-ear correlation in the statistical sound fields of rooms," *J. Acoust. Soc. Am.* **80**, 661–664.
- 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.
- Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1988). "Spiral VOS final report, Part A: The auditory filterbank," Cambridge Electronic Design, Contract Report No. APU 2341.
- 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.
- 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.
- Watkins, A. J. (2005). "Perceptual compensation for effects of reverberation in speech identification," *J. Acoust. Soc. Am.* **118**, 249–262.
- Wijngaarden, S. J., and Drullman, R. (2008). "Binaural intelligibility prediction based on the speech transmission index," *J. Acoust. Soc. Am.* **123**, 4514–4523.
- 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.