

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/13737/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Lavandier, Mathieu N. and Culling, John Francis 2007. Speech segregation in rooms: effects of reverberation on both target and interferer. *Journal of the Acoustical Society of America* 122 (3) , pp. 1713-1723. 10.1121/1.2764469 file

Publishers page: <http://dx.doi.org/10.1121/1.2764469>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Speech segregation in rooms: Effects of reverberation on both target and interferer

Mathieu Lavandier<sup>a)</sup> and John F. Culling

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

(Received 3 April 2007; revised 4 June 2007; accepted 27 June 2007)

Speech reception thresholds were measured to investigate the influence of a room on speech segregation between a spatially separated target and interferer. The listening tests were realized under headphones. A room simulation allowed selected positioning of the interferer and target, as well as varying the absorption coefficient of the room internal surfaces. The measurements involved target sentences and speech-shaped noise or 2-voice interferers. Four experiments revealed that speech segregation in rooms was not only dependent on the azimuth separation of sound sources, but also on their direct-to-reverberant energy ratio at the listening position. This parameter was varied for interferer and target independently. Speech intelligibility decreased as the direct-to-reverberant ratio of sources was degraded by sound reflections in the room. The influence of the direct-to-reverberant ratio of the interferer was in agreement with binaural unmasking theories, through its effect on interaural coherence. The effect on the target occurred at higher levels of reverberation and was explained by the intrinsic degradation of speech intelligibility in reverberation. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2764469]

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

Pages: 1713–1723

## I. INTRODUCTION

Studies of the cocktail party problem investigate the fact that we can understand what a person is saying when many others around are talking at the same time. A difference in fundamental frequency (Brox and Nooteboom (1982); Culling and Darwin (1993); Darwin and Culling (1990)) and an azimuth separation of sources (Hawley *et al.* (2004); Plomp (1976)) are both well-established cues helping the segregation of competing voices. The latter effect is termed spatial release from masking or spatial unmasking: an interfering sound masks less efficiently when it is spatially separated from the target speech. Spatial unmasking in anechoic situations results from two cues (Bronkhorst and Plomp (1988)). The first cue is monaural and called head shadow. 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 for the given interferer position. The signal-to-noise ratio between target and interferer can be improved at this ear, and this monaural information helps the listener to better understand the target. The second cue, sometimes called binaural interaction, is caused by interaural time delays (ITDs). The azimuth separation of interferer and target induces a difference in ITDs for these two sources, thus facilitating their segregation.

Reverberation and sound reflections in rooms influence both speech intelligibility in quiet and the segregation of speech from interfering sounds. In quiet, the target intelligibility is decreased by the sound reflections mixing with the direct sound. Some objective measurements are already available to predict this intrinsic degradation of speech by reverberation, such as the Speech Transmission Index (Hout-

gast and Steeneken (1985)) or the useful-to-detrimental ratios (Bradley (1986); Bradley *et al.* (1999)). The room also influences the segregation of speech from interfering sounds, and many experiments have shown that spatial unmasking is reduced by reverberation. Plomp (1976) measured the intelligibility of a speech target against noise or an interfering voice, while varying the interferer's azimuth and the reverberation conditions. The experiment was realized in a room having variable acoustic characteristics, with loudspeakers surrounding the listener. The spatial unmasking observed in the anechoic condition was greatly reduced in reverberation. Bronkhorst and Plomp (1990) measured the Speech Reception Threshold (SRT)—i.e., the level of the target compared to that of the interferer for 50% intelligibility of the target—using a noise interferer and loudspeakers placed at different positions in a reverberant room. They also found that spatial unmasking was reduced in reverberation. Culling *et al.* (2003) measured SRTs under headphones with a speech interferer. They used a room simulation that allowed positioning the interferer and target at chosen positions, as well as varying the absorption coefficient of the room boundaries. The spatial unmasking observed in the anechoic condition was abolished in reverberation. Recently, Beutelmann and Brand (2006) measured SRTs under headphones with a noise interferer, creating their stimuli from binaural impulse responses measured in three different rooms: an anechoic room, a small office and a large cafeteria. Again, the spatial release from masking was reduced in the office and the cafeteria compared to the one obtained in the anechoic room. This effect was also observed in studies concerned with attentional effects and informational masking (Darwin and Hukin (2000); Kidd *et al.* (2005)). Testing the influence of individual echoes rather than full reverberation, Freyman *et al.* (2001, 1999) found that spatial unmasking from a noise interferer was reduced in the presence of an echo. This result

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: lavandiermn@cardiff.ac.uk

was confirmed and extended later using different experimental protocols (Brungart *et al.* (2005); Rakerd *et al.* (2006)).

In Beutelmann and Brand (2006), the cafeteria with the largest reverberation time led to larger spatial unmasking than the office, which had only half the reverberation time. This result shows that a simple reverberation time statistic is not sufficient to predict speech segregation in a room, and that a more sophisticated measurement is needed. This measurement will require binaural information, in order to predict the spatial unmasking effects observed with directional interferers.

The aim of the present study was to investigate the influence of the direct-to-reverberant ratio of sources on the masking of speech in rooms, because reductions in direct-to-reverberant ratio result in reduced interaural coherence, which is known to affect binaural unmasking (Licklider (1948); Robinson and Jeffress (1963)). The interaural coherence of a source evaluates the similarity of the sound waveform it produces at the two ears of the listener. It is calculated through taking the maximum value from the interaural cross-correlation function. The multiple sound reflections reaching the listener in a room degrade the interaural coherence of the source. The coherence is thus related to the direct-to-reverberant energy ratio at the ears, and varies depending on the source and listener positions, and on the reverberation characteristics of the room (Hartmann *et al.* (2005)). Our interest in coherence stems from the observation by Licklider (1948) that, for speech and noise signals having different interaural phase differences, speech intelligibility in noise was reduced when the noise was less correlated at the two ears. A similar effect was observed for tone detection in noise (Robinson and Jeffress (1963)). All binaural unmasking theories predict this effect. In the following, we will consider the equalization-cancellation (E-C) mechanism of Durlach (1972) as an example to interpret our results. It predicts that a less correlated masker will be more difficult to equalize at the two ears, and then more difficult to cancel, resulting in lower speech intelligibility or poorer tone detection.

The four experiments presented in this paper considered interferer and target sources in a virtual room. They all involved SRT measurements under headphones, using target sentences and spatially separated speech-shaped noise or 2-voice interferers. The experiments were designed to investigate how the source direct-to-reverberant ratios might influence target intelligibility. Following Licklider's results, we were expecting this intelligibility to be reduced if the interferer direct-to-reverberant ratio and hence its interaural coherence were reduced by the sound reflections in the room. No binaural effect associated with the direct-to-reverberant ratio of the target was expected, but the direct-to-reverberant ratio also affects the intrinsic intelligibility of the target due to temporal and spectral distortions. The experiments enabled us to separate these two effects, and assess their relative roles in overall intelligibility. The interaural coherence/direct-to-reverberant ratio of sources was controlled independently for interferer and target, by varying the source-listener distance or the absorption coefficient of the room internal surfaces.

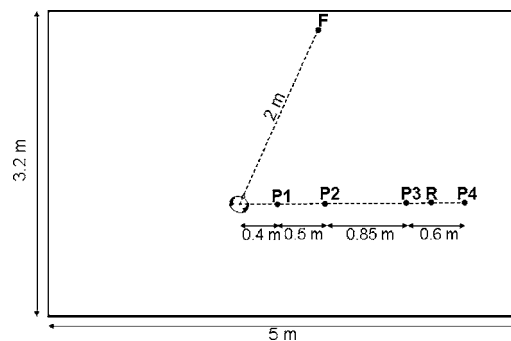


FIG. 1. Spatial configuration and virtual room considered in experiments 1 to 4.

## II. GENERAL METHODS

### A. Stimuli

#### 1. Convolution by the room

The same virtual room was used in the four experiments. It was simulated using the image (ray-tracing) method presented by Allen and Berkley (1979) and Peterson (1986), as it is implemented in the |WAVE signal processing package (Culling (1996)). The room was 5 m long, 3.2 m wide and 2.5 m high (Fig. 1). The absorption coefficients of the room internal surfaces were all set to the same value, and varied all together when the reverberation condition was changed. The listener was modeled as two ears, separated by 20 cm, placed along an axis at  $25^\circ$  to the 5 m wall on either side of a center point located at 1.2 m from the 5 m wall and 2 m from the 3.2 m wall. The ears were modeled as omnidirectional microphones suspended in space, 1.5 m from the floor, with no head between them. The head was not modeled here, in order to avoid head shadow effects which could complicate the interpretation of the effects associated with interaural coherence. The room and listener were identical to those simulated by Culling *et al.* (2003). The interferer and target were placed 1.5 m from the floor, at different positions in the room, keeping their azimuth separation constant ( $65^\circ$ ). Binaural stimuli were produced by calculating the impulse responses between the source positions and each ear, and convolving the speech samples with these impulse responses.

The same source stimuli were involved in the four experiments. A male voice was used for the target. This same voice was used to create 2-voice interferers. As we were interested in segregation cues associated with spatial separation of interferer and target, we chose to use the same voice for both sources in order to minimize segregation cues based on between-voice differences in fundamental frequency (F0) register or voice-specific spectral characteristics (Culling *et al.* (2003); Darwin and Culling (1990)). The original corpus of sentences was from the Harvard Sentence List (IEEE (1969)). The recordings of the male voice DA, made at M.I.T. and digitized at 20 kHz with 16 bit quantization, were used as the basis of all stimuli. The sentences have low predictability, and each has five key words highlighted in capitals. For instance, one sentence was "TAKE the WINDING PATH to REACH the LAKE." Noise interferers were also used. They were speech-spectrum noises, which were obtained by filtering Gaussian noises with an finite impulse

response (FIR) filter designed to match the excitation pattern (Moore and Glasberg (1983)) of all 2-voice interferers concatenated. In each experiment, four spatial configurations or reverberation conditions and two types of interferer (noise and 2 voices) were used. Eighty target sentences were used to test the eight resulting conditions. Although only four interferers of each type were needed, eight were used with the intention of reducing the dependence of the results on the particular sentences involved. Listeners only heard four interferers of each type, different ones for different listeners. The order of the conditions was rotated for successive listeners, while sentence materials remained in the same order. Each of the 80 target sentences was thus presented to every listener in the same order and, across a group eight listeners, contributed equally to each condition. This procedure also ensured that each condition was presented in each serial position within the experimental session, counterbalancing order effects.

## 2. Sound level equalization

The root mean square power of source stimuli was initially equalized, but the convolution by the room impulse responses might change the signal levels differently depending on the positions used within the room and the ear considered (Bradley *et al.* (1999, 2003); Yang and Hodgson (2006)). Using stimuli equalized before the convolution by the room places control of the target-to-interferer level ratio at the emission of sound sources. The influence of the room on sound levels of interferer and target would then be included in the measured SRTs. Our experiments were intended to investigate the effect of the room-induced interaural coherence changes on SRTs. To avoid mixing the effects of sound level and interaural coherence, all stimuli were equalized in level after convolution by the room impulse responses. Left and right channels were equalized independently, removing potential interaural level differences. The target-to-interferer level ratio was then fixed at the ears.

## 3. Computations of direct-to-reverberant ratio and interaural coherence

The direct-to-reverberant energy ratio was calculated at each ear, for all source positions and absorption coefficients involved in our experiments. For each tested configuration, the calculation was similar to the one proposed by Zahorik (2002), using the impulse response between the source position and the considered ear. The direct-path and reverberant portions of the impulse response were separated by a time-windowing procedure. This procedure was trivial as our simulated impulse responses were not contaminated by noise. The energy of the direct-path and reverberant portions were computed, and the direct-to-reverberant energy ratio was expressed in dB.

Interaural coherence of each source at the listening position was also calculated for each experiment. It was computed as the maximum of the absolute value of the cross-correlation of the waveforms reaching the two ears from the source (Hartmann *et al.* (2005)). 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.

## B. SRT measurements

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, each one against the same interferer. For the 2-voice interferers, the sentences corresponding to the interferer were displayed on a computer screen in front of the listeners while they were listening. They were instructed to disregard these sentences, and to listen to the target sentence, not displayed on the screen. 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 could be replayed in this way 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. 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 corresponded to a different tested condition, and used a different interferer. The session began with two practice runs, using monaurally presented and unprocessed speech, in order to familiarize listeners with the task. The following eight runs measured SRTs in each of the eight different conditions.

Signals were digitally mixed, digital/analog converted, and amplified using a 24 bit Edirol UA-20 sound card and a 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.

## C. Participants

Twenty-four listeners took part in each experiment. They were undergraduate students, paid for their participation. None of them was familiar with the sentences used during the test. All listeners reported normal hearing and English as their first language. Each listener participated in a single 50 min session, and in only one of the following experiments.

## III. DISTANCE OF THE INTERFERER (EXP. 1)

SRTs were measured for spatially separated interferers and targets which always had the same azimuth separation, but with the interferer tested at different positions in the

TABLE I. Direct-to-reverberant ratio (D/R) at each ear, for the source at the positions tested in experiments 1 and 3, and an absorption coefficient of 0.5

Position <sup>a</sup>	P1	P2	P3	P4
D/R ratio, left ear (dB)	7.95	1.88	-2.74	-5.04
D/R ratio, right ear (dB)	11.29	3.02	-2.32	-4.79

<sup>a</sup>D/R ratio at position F (dB): left ear=-4.31, right ear=-4.39.

room. These positions were chosen so that they led to different values of direct-to-reverberant ratio and interaural coherence at the listener position.

### A. Design

Only one reverberation condition was considered: an absorption coefficient of 0.5 was used for every room surface. Figure 1 presents the positions considered within the room. The position F was in front of the listener, at 2 m from the center of the ears. In experiment 1, the target was always at position F. The interferer was at one of the positions P1, P2, P3 or P4, all having the same azimuth. These four positions were situated along a line parallel with the 5 m wall passing through the center of the ears, at 0.4, 0.9, 1.75, and 2.35 m from this center. Table I shows that the direct-to-reverberant ratio at the listening position decreased as the source was moved from P1 to P4.

As stated above, positions P1, P2, P3 and P4 were chosen to produce different values of interaural coherence at the listening position. The coherence was calculated for the 80 target stimuli at position F and the eight interferers of each type at the four tested positions. Table II presents the mean coherence for each type of stimulus. Interaural coherence decreased as the interferer was placed further away from the listener, in the reverberant field. These sampled locations reflect the tendency for coherence to decrease with source distance, although the function is not completely monotonic (Hartmann *et al.* (2005)). The interaural coherence of a source seems to depend on its exact position relative to the listener and the room boundaries.

SRTs were measured in eight conditions: two types of interferer (noise and 2 voices) and four interferer positions (P1, P2, P3 and P4).

### B. Results

Figure 2 presents the mean SRTs obtained in experiment 1, for noise or 2-voice interferers at positions P1, P2, P3 or

TABLE II. Mean interaural coherence with standard deviation for each type of interferer in experiment 1.

Interferer position	Interferer mean coherence <sup>a</sup>	
	Noise (Std dev.)	2 voices (Std dev.)
P1	0.95 (0.00)	0.95 (0.01)
P2	0.86 (0.00)	0.86 (0.00)
P3	0.77 (0.01)	0.77 (0.04)
P4	0.71 (0.00)	0.71 (0.05)

<sup>a</sup>Target mean coherence: 0.88 (0.03).

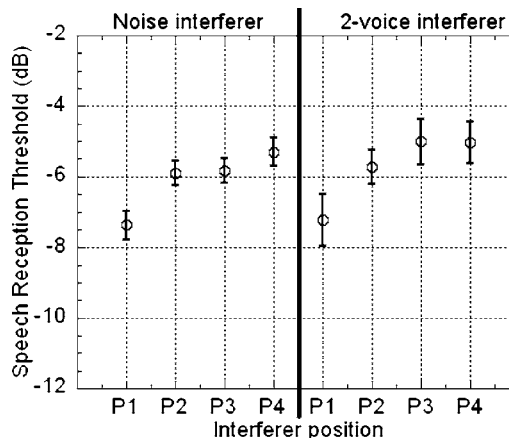


FIG. 2. Mean SRTs with standard errors measured in experiment 1 for noise or 2-voice interferers at positions P1, P2, P3 or P4; the target being always at position F (Fig. 1).

P4; the target being always at position F (Fig. 1). For both types of interferer, SRTs increased as the interferer was moved away from the listener, from P1 to P4, with a 2 dB effect between these two extreme positions. An analysis of variance (ANOVA) showed that only the main effect of interferer position was significant [ $F(3,69)=9.5, p<0.0001$ ]. Tukey pairwise comparisons confirmed that only SRTs involving position P1 were significantly different from SRTs involving the three other positions [ $q>4.5, p<0.01$  in each case].

### C. Discussion

SRTs increased with decreasing interferer coherence (Table II) as the interferer was moved from P1 to P4, in agreement with binaural unmasking theories. For instance, the E-C theory (Durlach (1972)) would predict that, for spatially separated sound sources, the less coherent interferers should be more difficult for the auditory system to cancel, producing less masking release. SRTs measured at positions P2, P3 and P4 were not significantly different. The range of coherence values sampled by our four positions might have been too limited, but these values had to be chosen among those available in the room at a fixed azimuth.

There was no significant difference between noise and 2-voice interferers in this experiment. As our sound sources were always spatially separated (Fig. 1), informational masking produced by the 2-voice interferers was probably minimal (Freyman *et al.* (2001, 1999)).

### IV. ABSORPTION COEFFICIENT FOR THE INTERFERER (EXP. 2)

In order to test a broader range of interaural coherence than in experiment 1, a new experimental design was chosen for experiment 2. SRTs were measured for spatially separated interferer and target at fixed positions. The different interferer coherences were obtained by applying different reverberation conditions to this source, while the reverberation condition of the target was kept constant. This experimental design is not realistic as it implies listening simultaneously to two sources in rooms having different reverberation charac-

TABLE III. Direct-to-reverberant ratio (D/R) at each ear, for the source at position R and the absorption coefficients tested in experiments 2 and 4.

Absorption coeff. <sup>a</sup>	1	0.7	0.5	0.2
D/R ratio, left ear (dB)	$\infty$	0.54	-3.70	-10.93
D/R ratio, right ear (dB)	$\infty$	0.78	-3.47	-10.61

<sup>a</sup>D/R ratio at position F (dB): left ear=-4.31, right ear=-4.39.

teristics, but it offers the advantage of controlling interferer and target coherences individually, keeping distances and azimuth separation constant.

### A. Design

Room dimensions and listener position were the same as in experiment 1, and position F was also unchanged (Fig. 1). The position R was considered. It was situated at the same azimuth as positions P1–P4 in experiment 1, on the parallel of the 5 m wall passing through the center of the ears, but at 2 m from this center, between P3 and P4. In experiment 2, the target was always at F, and the interferer at R. While computing the impulse responses between the sources and each ear, different absorption coefficients were used for the two sources. The absorption coefficient was the same for each room surface. For the target, it was set to 0.5 as in experiment 1. However, four absorption coefficients were tested for the interferer: 1, 0.7, 0.5 and 0.2. Table III shows that the direct-to-reverberant ratio at the listening position decreased with the absorption coefficient used. Interaural coherence of all stimuli was also computed. Table IV presents the mean results for each type of stimulus. The range of coherence values sampled by our four conditions was broader than in experiment 1 (Table II), with mean coherences between 1 and 0.6. As expected, the anechoic room with the absorption coefficient of 1 led to an infinite direct-to-reverberant ratio and a coherence of 1, as there was no sound reflection on the room internal surfaces.

SRTs were measured in eight conditions: two types of interferer (noise and 2 voices) and four reverberation conditions for the interferer (absorption coefficient of 1, 0.7, 0.5 and 0.2).

### B. Results

Figure 3 presents the mean SRTs obtained in experiment 2, for noise or 2-voice interferers for the four levels of absorption applied to the interferer. For both types of interferer,

TABLE IV. Mean interaural coherence with standard deviation for each type of interferer in experiment 2.

Interferer absorption coeff.	Interferer mean coherence <sup>a</sup>	
	Noise (Std dev.)	2 voices (Std dev.)
1	1 (0.00)	1 (0.01)
0.7	0.85 (0.00)	0.85 (0.00)
0.5	0.74 (0.00)	0.74 (0.05)
0.2	0.61 (0.01)	0.61 (0.07)

<sup>a</sup>Target mean coherence: 0.88 (0.03).

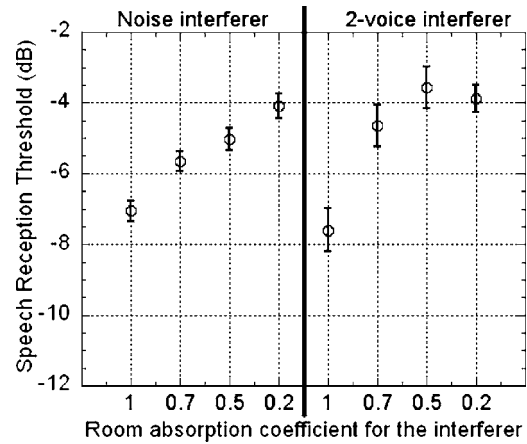


FIG. 3. Mean SRTs with standard errors measured in experiment 2, for noise or 2-voice interferers in the four tested reverberation conditions, with an absorption coefficient of 0.5 used for the target.

SRTs increased as the absorption coefficient for the interferer was decreased, making the room more reverberant for this source. A 3 dB effect was observed between the absorption coefficients 1 and 0.2 with both types of interferer. An ANOVA confirmed that the main effect of the absorption coefficient used for the interferer was significant [ $F(3,69) = 35.8, p < 0.0001$ ]. Tukey pairwise comparisons showed that the absorption coefficient 1 was significantly different from the three other coefficients [ $q > 8, p < 0.001$  in each case]. The absorption coefficient 0.7 was also significantly different from the coefficient 0.2 [ $q = 4.65, p < 0.01$ ]. The other differences were not significant. SRTs measured with noise were significantly lower than those obtained with the 2-voice interferer [ $F(1,23) = 6.0, p < 0.05$ ]. Interferer type and absorption coefficient did not interact significantly.

### C. Discussion

As the absorption coefficient used for the interferer was decreased, this interferer produced more masking, leading to higher thresholds. This result is in agreement with the binaural E-C theory: reverberation decreased the interferer interaural coherence (Table IV), and a less coherent interferer was more difficult to cancel, producing less masking release for spatially separated sources. As the range of coherences involved was larger than in experiment 1, the effect was also larger.

The difference between noise and 2-voice interferers was significant. Some informational masking might have been involved, leading to slightly higher SRTs with the 2-voice interferers. This seems particularly so for the absorption coefficients 0.5 and 0.7, although the interaction with absorption coefficient did not reach significance. When the absorption coefficients 1 and 0.2 were used, the 2-voice interferers did not lead to higher SRTs anymore, indicating a potential reduction of informational masking in these two conditions. This reduction in informational masking may have been caused by an additional cue that helped listeners focus their attention on the target voice rather than on the interferer during the listening task. In these conditions there is a difference in the amount of reverberation between the

TABLE V. Mean interaural coherence with standard deviation for the target in experiment 3.

Target position	P1	P2	P3	P4
Target <sup>a</sup> mean coherence	0.95	0.86	0.78	0.72
(Std dev.)	(0.01)	(0.03)	(0.06)	(0.06)

<sup>a</sup>Interferer coherence: noise=0.88 (0.00), 2 voices=0.89 (0.02).

interferer—anechoic (absorption coefficient of 1) or very reverberant (absorption coefficient of 0.2)—and the target—mildly reverberant (absorption coefficient of 0.5).

After having considered the effect of reverberation on the interferer, we turned to its effect on the target. Experiments 3 and 4 used exactly the same protocol as experiments 1 and 2, respectively, but replacing the interferer by the target and vice versa. In these new experiments, the interferer coherence was fixed across conditions, whereas the target direct-to-reverberant ratio was varied. Here, we did not expect any effect of the direct-to-reverberant ratio to be mediated by interaural coherence and degraded binaural processing. Following for example the E-C theory, binaural processing cancels the interferer and not the target. Reducing the coherence of the target and so the ability to cancel it should not play any role. Instead, the Speech Transmission Index (STI) was computed for the conditions tested, to compare our listening test results with the expected intrinsic degradation of target intelligibility in reverberation.

## V. DISTANCE OF THE TARGET (EXP. 3)

### A. Design

Experiment 3 followed the same protocol as experiment 1. This time the interferer was always at position F, whereas the target was tested at positions P1, P2, P3 and P4 (Fig. 1). The direct-to-reverberant ratios involved are presented in Table I. Interaural coherence of all stimuli was computed. Table V presents the mean results for each type of stimulus.

SRTs were measured in eight conditions: two types of interferer (noise and 2 voices) and four target positions (P1, P2, P3 and P4).

### B. Results

Figure 4 presents the mean SRTs obtained in experiment 3 for the target at positions P1, P2, P3 or P4; noise or 2-voice interferers being always at position F (Fig. 1). The results of STI calculations are also plotted along with the measured SRTs. These calculations are described and discussed in a following section of the paper. An ANOVA showed that, for both types of interferer, SRTs increased slightly as the target was moved away from the listener [ $F(3, 69)=9.4, p < 0.0001$ ]. A 1 dB effect was observed with noise, when the target was moved from P1 to P3. This effect increased to 2 dB with the 2-voice interferers. Tukey pairwise comparisons showed that only position P1 was significantly different from the three other positions [ $q > 5, p < 0.01$  in each case]. The SRT for position P1 was lower with 2-voice interferers

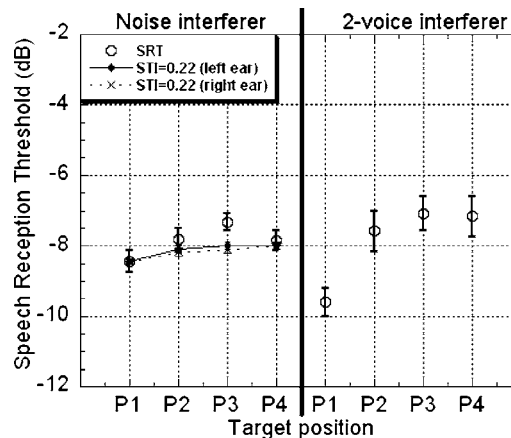


FIG. 4. Mean SRTs with standard errors measured in experiment 3 for the target at positions P1, P2, P3 or P4; the noise or 2-voice interferers being always at position F (Fig. 1). SRTs obtained with the noise interferers are compared to the target-to-interferer ratios required for the STI to be constant across the four conditions (STI=0.22 for the position P1 and a signal-to-noise ratio of  $-8.46$  dB). The STI was computed for each ear separately.

than with noise, but neither the interferer type nor the interaction between interferer type and absorption coefficient were significant.

### C. Discussion

SRTs increased with decreasing target direct-to-reverberant ratio (Table I), as the target was moved away from the listener in the reverberant field. This decrease of target intelligibility might be due to the intrinsic degradation of speech intelligibility in reverberation, occurring even when no interferer is involved, as described, for example, by the STI (Houtgast and Steeneken (1985)) or the useful-to-detrimental ratios (Bradley (1986); Bradley *et al.* (1999)). As in experiment 1, SRTs measured at positions P2, P3 and P4 were not significantly different.

Despite neither the interferer type nor the interaction between interferer type and interferer position being significant, the difference of SRTs for the target at P2 and P1 seemed larger with 2-voice interferers than with noise. This might be due to a release from informational masking when the target was moved very close to the listener. The perceived distance of the very close target might have been an attentional cue for listeners (Brungart and Simpson (2002)). Even if sound levels were equalized at the two ears, the direct-to-reverberant ratio of the target was still varying when changing its position within the room. Perceiving the target very close at position P1 could have helped listeners focus their attention on it. This effect, if present, was weak in experiment 3 and did not reach significance.

With the target in position P1, the SRT was lower with 2-voice interferers than with noise. The 2-voice interferers might produce less energetic masking than noise. Listeners could have taken advantage of some remaining silent periods in the masker to hear the target (Dusquesnoy (1983); Festen and Plomp (1990)), even though “listening into gaps” is smaller with 2-voice interferers than it would have been with 1-voice interferers (Bronkhorst and Plomp (1992)). Listeners

TABLE VI. Mean interaural coherence with standard deviation for the target in experiment 4.

Target absorption coeff.	1	0.7	0.5	0.2
Target <sup>a</sup> mean coherence (Std dev.)	1 (0.00)	0.86 (0.04)	0.74 (0.06)	0.62 (0.07)

<sup>a</sup>Interferer coherence: noise=0.88 (0.00), 2 voices=0.89 (0.02).

could have also used instantaneous F0 differences to cancel one of the interfering voices (Culling *et al.* (2005); de Chevigné *et al.* (1995)).

## VI. ABSORPTION COEFFICIENT FOR THE TARGET (EXP. 4)

### A. Design

Experiment 4 followed the same protocol as experiment 2. This time the interferer was always at F, and the target at R. While calculating the impulse responses between sources and each ear, different absorption coefficients were used for the two sources. The absorption coefficient was the same for each room surface. For the interferer at F, it was set to 0.5, as in Experiment 3. However, four absorption coefficients were tested for the target at R: 1, 0.7, 0.5 and 0.2. The direct-to-reverberant ratios involved are presented in Table III. Interaural coherence of all stimuli was computed. Table VI presents the mean coherence for each type of stimulus. The range of coherence sampled by our four conditions was broader than in experiment 3 (Table V), with mean coherences between 1 and 0.6.

SRTs were measured in eight conditions: two types of interferer (noise and 2-voice) and four reverberation conditions for the target (absorption coefficient of 1, 0.7, 0.5 and 0.2).

### B. Results

Figure 5 presents the mean SRTs obtained in experiment

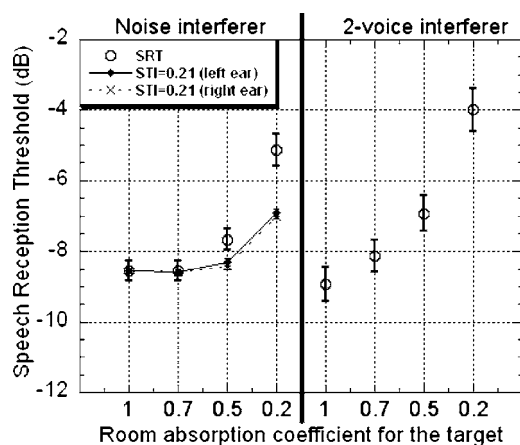


FIG. 5. Mean SRTs with standard errors measured in experiment 4, for the target in the four tested reverberation conditions, with an absorption coefficient of 0.5 used for the noise and 2-voice interferers. SRTs obtained with the noise interferers are compared to the target-to-interferer ratios required for the STI to be constant across the four conditions (STI=0.21 for the absorption coefficient of 1 and a signal-to-noise ratio of  $-8.50$  dB). The STI was computed for each ear separately.

4, for the target in the four tested reverberation conditions (absorption coefficient of 1, 0.7, 0.5 and 0.2), the reverberation condition of the noise or 2-voice interferers being fixed (absorption coefficient of 0.5). The comparison of these SRTs with STI calculations is described in the following section of the paper. For both types of interferer, SRTs increased as the absorption coefficient for the target was decreased, making the room more reverberant for this source. A 3.5 dB effect was observed between the absorption coefficients 1 and 0.2 with noise, this effect being increased to 5 dB with 2-voice interferers. An ANOVA confirmed that the main effect of the absorption coefficient used for the target was significant [ $F(3, 69)=44.8, p < 0.0001$ ]. Tukey pairwise comparisons showed that SRTs were significantly higher for the absorption coefficient of 0.2 than for the three other coefficients [ $q > 9, p < 0.001$  in each case]. The absorption coefficient 1 was also significantly different from the coefficient 0.5 [ $q=5.14, p < 0.01$ ]. The coefficient 0.7 did not differ significantly from the coefficients 0.5 and 1. Neither the interferer type nor the interaction between interferer type and absorption coefficient were significant.

### C. Discussion

SRTs increased with decreasing target direct-to-reverberant ratio (Table III), when the absorption coefficient used for the target was decreased. This decrease of target intelligibility might be due to the intrinsic degradation of speech intelligibility in reverberation. As the range of direct-to-reverberant ratios involved in this experiment was larger than in experiment 3, the effect was also larger.

It should be noted that SRTs from experiments 2 and 4 cannot be compared without caution, because the positions of target and interferer were switched between these experiments. The changes of coherence/direct-to-reverberant ratio introduced by the changes of absorption coefficient were limited for a source at the position F, whereas position R led to a broader range of coherence/direct-to-reverberant ratio. For this reason, for each experiment, the source with varying condition was placed at R, and the source with fixed condition was placed at F. SRTs measured in the two experiments using noise interferers and the absorption coefficient of 0.5 for both sources can be considered as an example. In experiment 2, the SRT was around  $-5$  dB, whereas it was around  $-7.5$  dB in experiment 4. This improvement in intelligibility could be explained by both the increase in interferer coherence—from 0.74 to 0.88 (Tables IV and VI)—and the increase in target direct-to-reverberant ratio—from  $-4.31/-4.39$  to  $-3.70/-3.47$  dB (Tables I and III).

## VII. INTRINSIC DEGRADATION OF TARGET INTELLIGIBILITY IN REVERBERATION

In experiments 3 and 4, the increase of SRTs as the target distance increased or as the absorption coefficient used for the target decreased (Figs. 4 and 5) could result from the intrinsic degradation of target intelligibility in reverberation. To test this hypothesis, the STI was computed in the different conditions involved in the two experiments. The STI evaluates the degradation of speech by reverberation, in the pres-



ence of nondirectional background noise, from the room impulse responses between the target positions and the two ears. As the STI calculation uses only monaural information, the left and right ear were considered independently.

SRTs measured in experiments 3 and 4 corresponded to the target-to-interferer ratios required to get a fixed intelligibility of 50% of the target words. To compare STI calculations and SRTs, we assumed that the STI should be constant at this threshold. We therefore computed the target-to-interferer ratios required for the STI to correspond to the same value across all tested conditions. The reference STI value for experiment 3 was 0.22, calculated from the first SRT measured with noise interferers: impulse responses between each ear and the target position P1 (Fig. 1) were used, and the signal-to-noise ratio was set to the SRT value of  $-8.46$  dB (Fig. 4). The signal-to-noise ratios required to get the same value of STI were computed using the room impulse responses between each ear and the target positions P2, P3 and P4. The reference STI value for experiment 4 was 0.21, again corresponding to the first SRT measured with noise interferers: impulse responses between each ear and the target position R were used, for a room absorption coefficient of 1 and a signal-to-noise ratio set to the SRT value of  $-8.56$  dB (Fig. 5). The signal-to-noise ratios required to get the same value of STI were computed for the absorption coefficients 0.7, 0.5 and 0.2.

Figures 4 and 5 present the results obtained for each ear compared to the SRTs measured with noise interferers. For experiment 3, it seems that STI calculations and intrinsic degradation of target speech in reverberation can explain the SRTs obtained for the target at positions P2 and P4, but the SRT measured with the target at P3 was higher than the STI prediction. For experiment 4, the STI predicted correctly the observed trend: target intelligibility decreased as the absorption coefficient used for the target was decreased. However, SRTs measured with the absorption coefficients 0.5 and 0.2 were higher than the STI predictions. The difference between actual and predicted SRTs also increased as the target coherence decreased with absorption coefficient (Table VI). In both experiments the interferer coherence was kept constant. The binaural masking due to the interferer coherence was constant across conditions and could not account for the SRTs being higher than the STI predictions. As STI calculations do not take into account the interferer coherence, these higher thresholds could result from an interaction between the effects of reverberation on interferer and target. The assumption made that the STI should be constant at threshold in experiments 3 and 4 would then be wrong. A second account for the discrepancies between SRT and STI could come from the STI calculations not being completely reliable in the poor intelligibility conditions considered in our experiments.

## VIII. GENERAL DISCUSSION

### A. Influence of interferer coherence

Experiments 1 and 2 investigated the influence of the interferer interaural coherence by decreasing it in two different ways: the interferer was moved away from the listener

into the reverberant field (experiment 1), or the room absorption coefficient used for the interferer was decreased (experiment 2). In both cases, SRTs increased with decreasing interferer coherence, with effects of 2–4 dB. This influence of the interferer coherence was in agreement with the E-C theory for binaural unmasking (Durlach (1972)): for spatially separated sound sources, a less coherent interferer would be more difficult to equalize and cancel for the auditory system, and so produce less masking release. This unmasking thus depends on the azimuth separation of sources (Beutelmann and Brand (2006); Plomp (1976)), but also on the coherence of the interferer, as shown by our experiments. This means that spatial unmasking will be influenced by the particular position of the interferer and the room characteristics.

Licklider (1948) showed that, for speech and noise having different interaural phase differences, speech intelligibility in noise was reduced when the noise was less correlated at the two ears. He used headphones with diotic target speech and a noise interferer having an interaural correlation of 0, 0.25, 0.5, 0.75 or 1. A difference of 20% word articulation was measured between the extreme correlations 0 and 1, most of the variation in intelligibility occurring between 0.75 and 1. The coherence range we have investigated in virtual rooms corresponds to the range over which he observed the greatest changes in intelligibility (Tables II and IV). Plomp and Mimpfen (1979) found the same effect as Licklider while measuring SRTs, using diotic speech presented against noise which was either identical, uncorrelated or partly correlated (diffuse) at the two ears. SRTs decreased with the noise correlation, from  $-7.3$  dB (noise identical) to  $-8.0$  dB (diffuse noise presentation) and  $-9.6$  dB (noise uncorrelated). Robinson and Jeffress (1963) measured the same trend and an even larger effect than Licklider for tone detection in noise. Zurek *et al.* (2004) investigated the detection of a third-octave band noise source in a virtual room, in the presence of a competing broadband noise source at a different azimuth. After having accounted for the monaural segregation cues, they found a binaural advantage very dependent on the absorption coefficient of the room internal surfaces. The detection threshold increased by up to 10 dB in a very reverberant room compared to an anechoic room. Even if they did not explicitly measure the interferer interaural coherence in their experiments, they showed that this elevation of threshold was well predicted by the E-C theory.

Beutelmann and Brand (2006) measured SRTs for a frontal target with a noise interferer, using binaural impulse responses measured in an anechoic room, a small office and a large cafeteria. As only close target sources and a limited amount of reverberation were involved, the intrinsic degradation of target intelligibility in reverberation was negligible. The thresholds for co-located sources were not significantly different in the three rooms. However, SRTs for spatially separated sources were higher in the office and cafeteria. This might be explained by a decrease of the interferer coherence in these two rooms. Other studies revealing that spatial unmasking observed in anechoic conditions was greatly reduced in reverberation might reflect the same effect (Bronkhorst and Plomp (1990); Culling *et al.* (2003); Plomp (1976)). This effect could also have influenced studies con-

cerned with attentional effects and informational masking (Darwin and Hukin (2000); Kidd *et al.* (2005)), in which spatial unmasking was also reduced in reverberation. The addition of an individual echo to the interferer might reduce its coherence, and then could account for part of the reduction of spatial unmasking observed in Freyman *et al.* (2001; 1999), Brungart *et al.* (2005) and Rakerd *et al.* (2006). Unfortunately, no interaural coherence evaluations were reported in any of these experiments. Even if these results are not directly comparable with those of experiments 1 and 2, the observed trends are the same across studies.

It should be noted that taking into account the influence of the head between the two ears of the listener might lead to larger impairments of intelligibility as reverberation increases, or at least to elevations of threshold appearing earlier. The head will lead to extra interaural decorrelation in asymmetric configurations, particularly for sources to the side, as shown by Lindevald and Benade (1986).

## B. Attentional effects and informational masking

Informational masking seems to have played some role in our experiments. However, the effect of interferer type and its potential interactions with the other variables was not significant (Experiments 1, 3 and 4), or only marginally significant (Experiment 2).

The amount of informational masking depends on the task design and stimuli being used. Informational masking can be seen as an attentional effect, when listeners are presented with several voices at the same time and when there is an ambiguity as to which voice they should attend to (Kidd *et al.* (2005); Shinn-Cunningham *et al.* (2005)). On the one hand, our task design tended to minimize informational masking; for each SRT measurement, interferer sentences were kept constant and displayed on a screen in front of the listener, therefore reducing the chances of confusion and the uncertainty in the listening task (Hawley *et al.* (2004); Kidd *et al.* (2005); Watson (2005)). In addition, interferer and target were always spatially separated (Fig. 1), and a difference in perceived position is an important attentional cue reducing informational masking (Brungart and Simpson (2002); Brungart *et al.* (2005); Darwin and Hukin (2000); Freyman *et al.* (2001, 1999); Kidd *et al.* (2005); Rakerd *et al.* (2006); Shinn-Cunningham *et al.* (2005)). On the other hand, as the same voice was used for the target and the two interfering sources, and because 2-voice interferers produce more informational masking than 1-voice interferers (Freyman *et al.* (2001)), some attentional ambiguity might have remained, even if the interfering sentences were identified on the screen. After our experiments, almost all listeners reported that the task seemed more difficult with 2-voice interferers than with noise.

Even if the effects were weak and often nonsignificant, some informational masking might have occurred in experiments 2, 3 and 4. This masking should be reduced as soon as the similarity between interferer and target was reduced (Watson (2005)), as any difference between the two sources could help listeners focus their attention on the target (Shinn-Cunningham *et al.* (2005)). The difference in the amount of

reverberation between target and interferer might have been used as an attentional cue by listeners in experiments 2 and 4. In experiment 3, with very close targets involved, a difference in perceived distance could have constituted such a cue, as found by Brungart and Simpson (2002).

## C. Influence of target direct-to-reverberant ratio

Experiments 3 and 4 showed that the target direct-to-reverberant ratio influenced SRT measurements in rooms. In experiment 3, SRTs increased as the target was moved away from the listener in the reverberant field. The effect was small, around 1 dB, comparable to the difference of SRTs measured by Bronkhorst and Plomp (1990) while playing the target through loudspeakers that were either close or far from the listener, in a reverberant room. A larger effect was obtained in experiment 4, where SRTs increased as the absorption coefficient used for the target was decreased.

As we were expecting a monaural effect associated with the direct-to-reverberant ratio of the target rather than any binaural effect associated with the coherence of the target, our SRTs were compared to STI calculations involving only monaural information. The decrease of target intelligibility can be partially explained by the intrinsic degradation of speech intelligibility in reverberation, because STI calculations in the different configurations of our experiments were consistent with some elevation of threshold. However, the STI alone could not account for all our results, as differences between SRTs and STI predictions were observed. Experiment 4 showed that these differences increased as the target coherence decreased. As STI calculations do not take into account the interferer coherence, these differences might indicate an interaction between interferer and target coherences. Further experiments will be undertaken to investigate this question. Experiments involving diotic reverberant targets will be conducted to assess the relevance of considering the coherence of the target when describing speech segregation in rooms.

## D. Towards an objective prediction of speech segregation in rooms

Some objective measurements are already available to predict speech intelligibility. The STI (Houtgast and Steeneken (1985)) or the useful-to-detrimental ratios (Bradley (1986); Bradley *et al.* (1999)) evaluate the intrinsic degradation of speech by reverberation, in the presence of non-directional background noise. However, these measurements use only monaural information and do not consider directional interferers. Therefore, the influence of the interferer position and spatial unmasking cannot be taken into account. Zurek (1993) developed a model predicting speech intelligibility in the presence of an interfering source. It simulates head shadow and binaural interaction associated with the azimuth separation of sources, to take into account spatial unmasking. However, it applies only to noise interferers and anechoic situations.

The experiments presented in this paper revealed that, in rooms, speech intelligibility did not depend only on the azimuth separation of sound sources, but also on their inter-

aural coherence. It seems important to take this last parameter into account to predict speech intelligibility in rooms, at least when interferers and speech segregation mechanisms are involved. More experimental data are needed to describe precisely the relation between coherence and intelligibility. Zurek *et al.* (2004) proposed a model predicting the detection of a narrowband noise source in a room when a broadband noise source is also present. The model is based on room statistics and an equalization-cancellation (E-C) mechanism (Durlach (1972)). Binaural detection was quite accurately predicted, even if some discrepancies remained. These discrepancies could be linked to the approximations resulting from the use of room statistics rather than the complete room impulse responses. It is not clear how this model, designed for narrowband noise detection, could be extended to predict speech intelligibility. Beutelmann and Brand (2006) developed a model predicting the intelligibility of a speech target against a noise interferer in rooms. The model is also based on the E-C theory, and so should account for the influence of the interferer interaural coherence. However, the model only holds for near field targets, because the disturbance of the speech itself caused by reverberation is not taken into account. The model is also based on the analysis of the speech sentences spoken in the room, rather than on acoustical measurements more characteristic of the room itself.

It should be noted that we decided here to compute the interaural coherence of our signals as the maximum of the absolute value of the cross-correlation of the signals (filtered between 20 and 1500 Hz) at the two ears. While this calculation is simple and direct, it might need refinement. For example, Hartmann *et al.* (2005) measured the coherence in rooms by third octave bands, as the auditory system would not be able to use the broadband coherence, and there is evidence for the frequency channel independence of binaural unmasking (Akeroyd (2004); Culling and Summerfield (1995)). Weightings could be applied to this within-channel coherence, to take into account the spectrum of speech and the variation of the binaural masking level difference with frequency (Hirsh (1948)).

The influence of interaural coherence on intelligibility has been studied here for a fixed azimuth separation of sources. The influence of azimuth separation has been previously investigated in several studies in anechoic conditions (see Bronkhorst (2000) for a review), that is to say for sources having a fixed interaural coherence of 1. The experimental protocol we have used will allow us to investigate the interaction between coherence and azimuth separation, by varying these two parameters simultaneously.

## IX. CONCLUSION

SRTs measured in a virtual room revealed that speech intelligibility in the presence of noise or speech maskers did not depend only on the azimuth separation of sound sources, but also on their interaural coherence and direct-to-reverberant ratio at the listening position. Speech intelligibility decreased as the coherence/direct-to-reverberant ratio of sources was degraded by reverberation. The influence of the

interferer coherence was in agreement with binaural unmasking theories, and the effect of sound reflections on the target was partially explained by the intrinsic degradation of speech intelligibility in reverberation. It should be noted that, as the reverberation increases, speech intelligibility suffers first from the decorrelation of the interferer at the two ears, before the intrinsic loss of target intelligibility.

## ACKNOWLEDGMENTS

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

- Akeroyd, M. (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., and Berkley, D. (1979). "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.* **65**, 943–950.
- 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**, 131–342.
- Bradley, J. (1986). "Predictors of speech intelligibility in rooms," *J. Acoust. Soc. Am.* **80**, 837–845.
- Bradley, J., Reich, R., and Norcross, S. (1999). "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," *J. Acoust. Soc. Am.* **106**, 1820–1828.
- Bradley, J., Sato, H., and Picard, M. (2003). "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.* **113**, 3233–3244.
- Brokx, J., and Nooteboom, S. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phonetics* **10**, 23–36.
- Bronkhorst, A. (2000). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acust. Acta Acust.* **86**, 117–128.
- Bronkhorst, A., 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., and Plomp, R. (1990). "A clinical test for the assessment of binaural speech perception in noise," *Audiology* **29**, 275–285.
- Bronkhorst, A., 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.
- Brungart, D., and Simpson, B. (2002). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," *J. Acoust. Soc. Am.* **112**, 664–676.
- Brungart, D., Simpson, B., and Freyman, R. (2005). "Precedence-based speech segregation in a virtual auditory environment," *J. Acoust. Soc. Am.* **118**, 3241–3252.
- Culling, J. (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., and Darwin, C. (1993). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by f0," *J. Acoust. Soc. Am.* **93**, 3454–3467.
- Culling, J., Hodder, K., and Toh, C. (2003). "Effects of reverberation on perceptual segregation of competing voices," *J. Acoust. Soc. Am.* **114**, 2871–2876.
- Culling, J., Linsmith, G. and Caller, T. (2005). "Evidence for a cancellation mechanism in perceptual segregation by differences in fundamental frequency," *J. Acoust. Soc. Am.* **117**, 2600 (A).
- Culling, J., and Summerfield, Q. (1995). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," *J. Acoust. Soc. Am.* **98**, 785–797.
- Darwin, C., and Culling, J. (1990). "Speech perception seen through the ear," *Speech Commun.* **9**, 469–475.
- Darwin, C., and Hukin, R. (2000). "Effects of reverberation on spatial, prosodic, and vocaltract size cues to selective attention," *J. Acoust. Soc. Am.* **108**, 335–342.
- de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (1995).

- "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement," *J. Acoust. Soc. Am.* **97**, 3736–3748.
- Durlach, N. (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. (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.
- Festen, J., 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.
- Freyman, R., Balakrishnan, U., and Helfer, K. (2001). "Spatial release from informational masking in speech recognition," *J. Acoust. Soc. Am.* **109**, 2112–2122.
- Freyman, R., Helfer, K., McCall, D., and Clifton, R. (1999). "The role of perceived spatial separation in the unmasking of speech," *J. Acoust. Soc. Am.* **106**, 3578–3588.
- Hartmann, W., Rakerd, B., and Koller, A. (2005). "Binaural coherence in rooms," *Acta. Acust. Acust.* **91**, 451–462.
- Hawley, M., Litovsky, R., and Culling, J. (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. (1948). "The influence of interaural phase on interaural summation and inhibition," *J. Acoust. Soc. Am.* **20**, 536–544.
- Houtgast, T., and Steeneken, H. (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.
- 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.
- Licklider, J. (1948). "The influence of linteraural phase relations upon masking of speech by white noise," *J. Acoust. Soc. Am.* **20**, 150–159.
- Lindevald, I., and Benade, A. (1986). "Two-ear correlation in the statistical sound fields of rooms," *J. Acoust. Soc. Am.* **80**, 661–664.
- Moore, B., and Glasberg, B. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Peterson, P. (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. (1979). "Improving the reliability of testing the speech-reception threshold for sentences," *Audiology* **18**, 43–52.
- Rakerd, B., Aaronson, N., and Hartmann, W. (2006). "Release from speech-on-speech masking by adding a delayed masker at a different location," *J. Acoust. Soc. Am.* **119**, 1597–1605.
- Robinson, D., and Jeffress, L. (1963). "Effect of varying the interaural noise correlation on the detectability of tonal signals," *J. Acoust. Soc. Am.* **35**, 1947–1952.
- 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.
- Watson, C. (2005). "Some comments on informational masking," *Acta. Acust. Acust.* **91**, 502–512.
- Yang, W., and Hodgson, M. (2006). "Auralization study of optimum reverberation times for speech intelligibility for normal and hearing-impaired listeners in classrooms with diffuse sound fields," *J. Acoust. Soc. Am.* **120**, 801–807.
- Zahorik, P. (2002). "Direct-to-reverberant energy ratio sensitivity," *J. Acoust. Soc. Am.* **112**, 2110–2117.
- Zurek, P. (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., Freyman, R., and Balakrishnan, U. (2004). "Auditory target detection in reverberation," *J. Acoust. Soc. Am.* **115**, 1609–1620.