Some spatial and temporal effects on the speech privacy of meeting rooms

Bradley, J. S.; Apfel, M.; Gover, B. N.

Publisher’s version / Version de l'éditeur:
https://doi.org/10.1121/1.3097771
Journal of the Acoustical Society of America, 125, 5, pp. 3038-3051, 2009-01-01

Some spatial and temporal effects on the speech privacy of meeting rooms

NRCC-53559

Bradley, J.S.; Apfel, M.; Gover, B.N.

August 2010

A version of this document is published in / Une version de ce document se trouve dans:
Journal of the Acoustical Society of America, 125, (5), pp. 3038-3051, January 01, 2009, DOI: 10.1121/1.3097771

Some spatial and temporal effects on the speech privacy of meeting rooms

John S. Bradley,[a] Marina Apfel, and Bradford N. Gover
National Research Council, 1200 Montreal Road, Ottawa K1A 0R6, Canada

(Received 16 December 2008; revised 19 February 2009; accepted 19 February 2009)

This paper reports on initial experiments concerning how key spatial and temporal effects in rooms influence the speech privacy provided by enclosed rooms. The first part of the work demonstrates that for the same signal-to-noise ratio, the intelligibility of speech and the threshold of intelligibility are significantly different for transmission between real rooms than in the previous results in approximately free-field conditions [B. N. Gover and J. S. Bradley, J. Acoust. Soc. Am. **116**, 3480–3490 (2004)]. The second part investigates the influence of aspects of the spatial and temporal components of sound fields in typical rooms, to explain these differences for transmission between real rooms. These components included the separate effects of early-arriving and later-arriving reflected speech sounds. They also included the effects of spatially separated speech and noise sources as well as more diffuse noise representative of typical meeting rooms. In realistic combinations these effects are of practical importance and can change privacy criteria by 5 dB or more. Ignoring them could lead to costly over-design of the sound insulation required to achieve adequate speech privacy. [DOI: 10.1121/1.3097771]

PACS number(s): 43.55.Hy [NX] Pages: 3038–3051

I. INTRODUCTION

Most previous work has assumed that speech privacy is only influenced by the signal-to-noise ratio of the speech and the concurrent ambient noise. The primary importance of signal-to-noise ratios (SNRs), as determinants of speech privacy, was established by the pioneering work of Cavanaugh *et al.*[1] on speech privacy of enclosed offices, which related privacy ratings to articulation index values. A more recent study[2] found a uniformly weighted signal-to-noise ratio measure (SNR_{uni32}) to be a good predictor of ratings of both the audibility and intelligibility of speech sounds from an adjacent room. Although it is well known that the intelligibility of speech can be reduced by reverberation[3] and increased by the spatial separation of speech and noise sources,[4] these effects have not previously been considered in studies of the speech privacy of enclosed rooms.

This new work is an initial investigation of the key spatial and temporal effects in rooms on the audibility and intelligibility of speech in noise for conditions representative of typical rooms where speech privacy is required. The new work includes two studies. The first more exploratory study (Sec. III) demonstrates that the intelligibility of speech and the threshold of intelligibility are significantly different for propagation between real rooms than in the previous results for simulated conditions in approximately free-field conditions.[2] The second part of the work (Sec. IV) investigates the magnitude of the effects of each aspect of the spatial and temporal components of sound fields in typical rooms. These included the separate effects of early-arriving and later-arriving reflected sounds. They also included the effects of spatially separated speech and noise sources as well as more diffuse noise representative of typical meeting rooms.

There have been many previous studies related to understanding the temporal and spatial effects of room acoustics on our ability to understand speech heard in combination with competing sounds. It is often considered that reduced speech intelligibility corresponds to increased speech privacy, but in some situations higher degrees of privacy might require speech to be inaudible. Classical room acoustics studies have long identified optimum reverberation times for maximizing the intelligibility of speech.[3] Earlier work to explain spatial effects was concerned with explaining our ability to understand speech in the presence of other interfering speech sounds, the so-called *cocktail party effect*.[5] There have been at least two reviews of the many studies related to the cocktail party effect.[6,7]

Interfering sounds can mask the target speech sounds (that we wish to hear) and reduce the intelligibility of the speech. The intelligibility of speech is first a SNR issue and the work of French and Steinberg,[8] which led to the articulation index, can explain the monaural signal-to-noise effects on speech intelligibility. Masking is influenced by both monaural and binaural effects. Even for monaural listening, head shadow effects can influence the intelligibility of speech as a function of the relative directions of the target speech and interfering sounds.

It is usually possible to better understand the target speech mixed with interfering sounds by listening binaurally. The benefit of listening binaurally rather than monaurally is referred to as a *binaural advantage*. Many studies have tried to explain the cause of binaural advantages and have shown interaural level, time, and phase differences to be important.[9] These interaural differences vary with the direction of the sound source relative to the head of the listener and hence

[a]Author to whom correspondence should be addressed. Electronic mail: john.bradley@nrc-cnrc.gc.ca

can help us to discriminate among spatially separated sound sources.

Although many of the earlier studies focused on the effect of interfering speech sounds on the target speech, interfering noises can lead to larger reductions in intelligibility. For example, with a single interfering talker, it is possible for listeners to hear the target speech in the gaps of the interfering speech. This is not possible with more or less constant noises such as ventilation type noise common in buildings. When the interfering speech is made up of a number of talkers, the masking effect of the speech tends to be similar to that of noise with a similar spectrum and level.[10]

The masking effect of an interfering sound is greatly influenced by the direction of arrival of the masking sound relative to that of the target speech. Experiments in free-field conditions have shown that separating the target speech and masking sound by as little as 10° is detectable, and that a 20° difference leads to quite significant increases in the intelligibility of the target speech.[4] Systematic studies have reported the resulting increased intelligibility as a function of the angle of separating the speech and noise sources.[6,11] This reduction in the masking effect of the interfering noise by spatially separating the noise and the speech sources is referred to as a *spatial release from masking*.

Studies of spatial effects have mostly been conducted in free-field conditions and have not often included the effects of reverberation. Where reverberation has been included, it has been shown to reduce the magnitude of the spatial release from masking,[12] indicating that in typical rooms with reverberant sound, listeners are less able to benefit from a spatial release from masking when the target speech source and the interfering sound source are spatially separated. Plomp[13] systematically investigated the combined effects of varied reverberation time with varied separation of the speech and interfering noise sources. Although based on subjective ratings of the intelligibility rather than on intelligibility test scores, the study showed a gradual decrease in the spatial release from masking as reverberation time was increased.

Most of the studies to date have focused on understanding individual parts of the overall issue of spatial and temporal effects of room acoustics and have most often been carried out in free-field conditions. Only a few studies have included the effects of room reverberation and usually the term reverberation has been used loosely to include all types of reflected sound. Most often the interfering signal has been speech[14,15] and not typical room noises such as that from ventilation systems.

II. COMMON EXPERIMENTAL PROCEDURES FOR EXPERIMENTS

This paper describes two different studies to investigate the effects of room acoustics on the speech privacy of meeting rooms. The first investigation was intended to explore how closely the results of previous tests in approximately free-field conditions[2] could be replicated in more realistic acoustical conditions. In these validation tests subjects listened to speech transmitted through real walls from an adjacent room. The second series of tests was carried out in simulated sound fields to determine the effects of various details of the spatial and temporal characteristics of speech and noise sounds on the intelligibility of speech for speech privacy situations. Although the experimental setups and goals of each investigation were different, many aspects of the experimental procedure were the same.

Experimental conditions were characterized in terms of uniformly weighted (over speech frequencies) SNRs as defined by the following:

$$\mathrm{SNR}_{\mathrm{uni32}} = \frac{1}{16}\sum_{f=160}^{5000} \{L_s(f) - L_n(f)\}_{-32} \quad (\mathrm{dB}), \qquad (1)$$

where $L_s(f)$ and $L_n(f)$ are the 1/3-octave band speech and noise levels at the listener's position. The −32 indicates that differences in the brackets are clipped so that they can never be less than −32 dB. $\mathrm{SNR}_{\mathrm{uni32}}$ values were previously shown[2] to be the most successful compromise for predicting subjective judgments of both the audibility and the intelligibility of speech when rating the speech privacy of meeting rooms.

Conditions were subjectively evaluated using speech tests to determine the audibility and intelligibility of speech for each test condition following procedures similar to those in the initial study.[2] Recordings of the Harvard sentences[16] were used as the speech test material. They were high quality digital recordings spoken by a male talker in an anechoic room. The Harvard sentences are phonetically balanced and of low predictability, which is necessary for conditions of low intelligibility for which guessing could distort the scores in some other types of speech intelligibility tests.

The test protocol consisted of first turning on the noise signal and then a few seconds later playing one sentence. After the noise had stopped, the subject could tell the experimenter what they had heard, using a microphone to communicate with the researcher who was outside the test room. The subjects first did a practice test consisting of ten sentences that included $\mathrm{SNR}_{\mathrm{uni32}}$ values distributed over the complete range included in the full test.

Because in some cases no screening of subjects for hearing sensitivity was possible, some may have had some hearing loss. In addition, others may not have had English as their first language. Therefore, the data analyses are based on the results of the listeners with the ten highest intelligibility scores over all conditions of the test. Using only the best ten listeners was expected to exclude listeners with less sensitive hearing or others who may have not been sufficiently fluent in English.

Subjects were all adults and were employees of the National Research Council who volunteered to participate after being contacted by electronic mail. They did not receive any compensation for their participation. The tests were approved by the Ethics Review Board of the National Research Council (Protocol No. 2006-06) and each subject signed a consent form after all of the details of the experiment were explained to them.

TABLE I. Construction details of the three walls used in the two-room validation tests, (RC, resilient channels; STC, sound transmission class).

Wall No.	Face No. 1 gypsum board (mm)	Stud type	Cavity insulation	RC	Face No. 2 gypsum board (mm)	STC
1	2×16	92 mm steel	90 mm mineral fiber	No	2×16	56
2	13	90 mm wood	90 mm glass fiber	Yes	2×13	53
3	13	90 mm wood	90 mm glass fiber	Yes	13	46

III. TWO-ROOM VALIDATION TEST

A. Procedure

The initial tests[2] had included a wide range of speech and noise levels but were carried out in approximately free-field conditions with spatially separated speech and noise sources. The test subjects listened to speech sounds, modified to represent transmission through one of several walls, from a loudspeaker system in front of them. At the same time they heard ambient noises from a second loudspeaker system located above them. These conditions were intended to represent a worst-case condition (i.e., minimum speech privacy) in which it would be most easy to understand speech. The new two-room validation tests[17] were intended to explore how the results might differ for more typical room acoustics conditions.

In the new validation tests, speech was radiated into one room and listeners heard it while located close to the common wall in an adjacent room. The recorded test sentences were reproduced using a dodecahedron loudspeaker located approximately 2 m from the center of the test wall in a large ($250\ \mathrm{m}^3$) reverberation chamber. The speech sounds were naturally transmitted through the test wall into the adjacent room ($140\ \mathrm{m}^3$) where the listener was located. The constructions of the three walls are described in Table I and they had sound transmission class (STC) values of 46, 53, and 56.

Sound absorbing material was added to both rooms to reduce the reverberation times (averaged over frequencies from 160 to 5000 Hz) in the receiving room to 0.64 s and in the source room to 0.80 s. With the added sound absorbing material present, the listeners heard speech sounds in realistic conditions representative of typical meeting rooms and heard speech sounds that had been transmitted through real walls.

The integrated 1/3-octave band levels of each test sentence were obtained as room average levels in the source room. The attenuation of speech sounds was determined using a broadband white noise test signal, as the difference between the average level in the source room and the level at a point 0.25 m from the test wall in the receiving room (similar to the procedure of ASTM E2638-08).[18] The average levels of each test sentence in the source room and the measured attenuations were used to determine the speech levels in the receiving room at the location of the listener 0.25 m from the test wall. This was a more accurate estimate of the speech levels at the receiver position than would be possible by directly measuring the, often low level, transmitted speech sounds.

Ambient noise was added to the receiving room from a single loudspeaker located across the room from the listener. Random noise, which was equalized to have an approximately −5 dB/octave spectrum shape representative of typical ventilation noise,[14,15] was used. The ambient noise levels were measured using a single microphone at the location of the listener's head. Three different ambient noise levels were used: 24, 29, and 34 dBA. These were selected so that at the listener's position, speech sounds varied from barely audible to completely intelligible.

From the transmitted speech levels and the ambient noise levels at the location of the listener's head, SNR_{uni32} values were calculated. For each of three test walls, speech and noise levels were adjusted to give the most useful range of SNR_{uni32} values. The intention was that the SNR_{uni32} values should be evenly distributed over conditions from, not audible to most listeners, to quite intelligible to most listeners. An even distribution of SNR_{uni32} values from just below −25 dB to a maximum of about −3 dB was used.

The test protocol consisted of first turning on the noise signal and then a few seconds later playing one sentence. A few seconds after the end of the sentence, the noise stopped. In this experiment the start and stop of the noise was marked by a chime sound so that subjects did not miss the start of the often very low levels of noise. In a few cases there was no sentence present to minimize the probability of subjects always guessing that they did hear some speech. After the noise had stopped, the subject told the experimenter what they had heard using a microphone to communicate with the researcher who was outside the test rooms. The subject first said whether any speech sounds were audible, and then if some speech was audible said whether they could hear the cadence or rhythm of the speech. Finally if they had heard the cadence, they repeated the words that they had understood back to the experimenter.

After all subjects had listened to all sentences, it was possible to calculate the fraction of the subjects, for each test sentence, who were able to hear any speech sounds, and determine an estimate of the audibility threshold as the SNR_{uni32} value for which 50% could just hear some speech sound. Similarly, from the number of subjects who were able to hear the cadence of the speech, an estimate of the thresh-

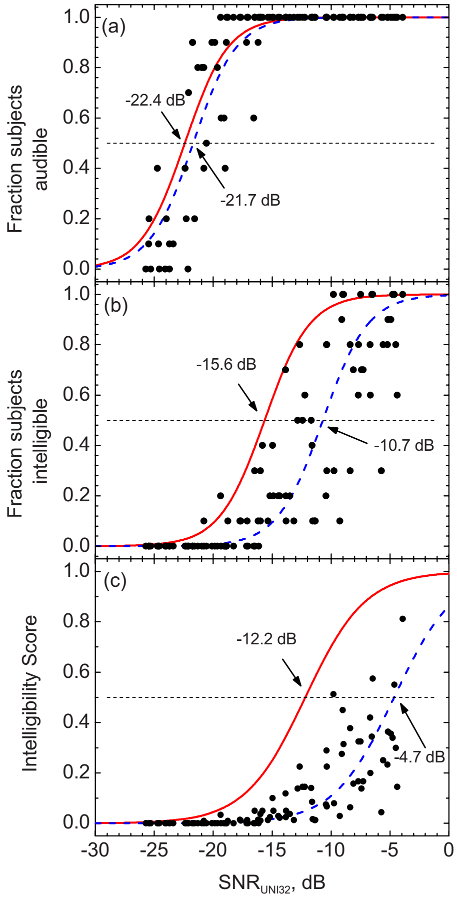

FIG. 1. (Color online) Comparison of results of initial study in approximately free-field conditions (solid line) (Ref. 2) with the results of a two-room validation study (data points and dashed line): (a) fraction of subjects finding some speech audible, (b) fraction of subjects able to understand at least one word, and (c) speech intelligibility scores, all plotted versus values of the uniformly weighted signal-to-noise ratio (SNR$_{uni32}$). The decibel values and arrows on each plot indicate the various threshold values [(a) and (b)] and speech reception threshold (c).

old of the cadence was determined. The intelligibility threshold was determined as the point at which 50% of the subjects could just understand at least one word of each test sentence and the speech intelligibility score was determined as the fraction of the words correctly understood.

B. Results

The scores were first plotted versus SNR$_{uni32}$ values separately for each wall tested. However, it was not possible to detect systematic differences among the results for the different walls and the data for all three wall tests were combined.

The upper part of Fig. 1 plots the fraction of listeners indicating that they heard some speech sounds versus SNR$_{uni32}$ values for the combined data from all three walls. Each point indicates the fraction of the ten best listeners who heard speech sound for that particular condition. The solid line in this plot is the best-fit regression line obtained in the previous work.[2] The dashed line is the same form of curve but shifted to the right to better fit the new measured data. The point at which 50% of the subjects found the speech to be just audible was described as the threshold of audibility of the speech and corresponded to a SNR$_{uni32}$ of −22.4 dB in the initial study. The best-fit lines were Boltzmann equations given by

$$y = \frac{(A_1 - A_2)}{1 + e^{(x-x_0)/dx}} + A_2, \qquad (2)$$

where y is the fraction responding or intelligibility score, x is the corresponding SNR, x_0 is the SNR for a y value of 0.5, dx is related to the slope of the midportion of the curve, A_1 is the minimum y value=0.0, and A_2 is the maximum y value =1.0.

TABLE II. Regression coefficients for the Boltzmann best-fit equations in Fig. 1. "Initial" identifies results from the previously published (Ref. 2) initial study and "validation" results form the new two-room validation study (Ref. 17).

Response	Expt.	dx	x_0
Threshold of audibility	Initial	1.8053	−22.41
	Validation	1.8053	−21.70
Threshold of cadence	Initial	1.4037	−20.05
	Validation	1.4037	−17.50
Threshold of intelligibility	Initial	1.8379	−15.64
	Validation	1.8379	−10.70
Speech intelligibility score	Initial	2.5259	−12.19
	Validation	2.5259	−4.65

The regression fits to the new data were obtained by shifting the corresponding regression line from the initial study[2] to the right an amount that minimized the rms deviation about the regression line for the new data. Thus the dx parameter was not changed and only the x_0 value was changed corresponding to shifting the regression line horizontally.

Table II lists the x_0 and dx parameters for each of the regression lines in Fig. 1. For the audibility scores, the shift between the new results and the previous results for the threshold of audibility was only 0.7 dB (i.e., −22.4 to −21.7 dB).

To examine the significance of the differences between the initial study results and the new results, the standard deviations of the measured values about the new best-fit line in terms of SNR$_{uni32}$ values were calculated. This cannot be done for scores of either 0 or 1 for which SNR$_{uni32}$ values are not defined for this type of equation. However, the standard deviation of the data points about the new best-fit line was calculated for all other points to give an indication of the relevance of the shift between the two best-fit lines. For the results in Fig. 1, the standard deviation of the remaining data points (excluding points with scores of 0 or 1) about the new best-fit line was ±2.2 dB in terms of SNR$_{uni32}$ values. This is much larger than the 0.7 dB shift, and indicates that within the limits of the data, the new audibility data replicate and hence validate the previous results with respect to the audibility of transmitted speech sounds.

Figure 1(b) plots the fraction of listeners who understood at least one word versus SNR$_{uni32}$ values. For these data the new best-fit line that is shifted 4.9 dB to the right minimizes the vertical deviations of the data about the line. This is a quite large shift and almost all of the data points are to the right of the previous best-fit line (the solid line in this figure). The standard deviation of the data about the new best-fit line in terms of SNR_{uni32} values was ± 3.0 dB. The shift in the new best-fit line relative to the initial best-fit line is much larger than the scatter of the data about the new best-fit line. For these data, listeners were much less likely to understand at least one word for a given SNR_{uni32} value than in the previous work. These results do not replicate the previous work and indicate a greater degree of speech privacy for a given SNR_{uni32} value than did the previous results. According to the new results, the threshold of intelligibility (the point where 50% of the subjects could just understand at least one word) corresponds to a SNR_{uni32} value of -11 dB rather than the -16 dB value from the initial work.

The fractions of the best ten subjects indicating that they heard the cadence of the speech sounds versus SNR_{uni32} values were also determined.[17] For these data the new regression line was shifted 2.5 dB to the right to best fit the new data and the threshold of cadence was shifted from -20.0 to -17.5 dB. This shift was intermediate to that for the audibility threshold and that for the intelligibility threshold.

The speech intelligibility scores are plotted versus SNR_{uni32} values in Fig. 1(c). The results are quite similar to those for the speech intelligibility threshold (SIT) in Fig. 1(b) in that a quite large shift is required to align the new best-fit regression line with the measured speech intelligibility scores. In this case the new best-fit line was shifted 7.5 dB to the right relative to the previous best-fit line shown in Fig. 1(c). The standard deviation about the new best-fit line in Fig. 1(c) in terms of SNR_{uni32} values was ± 2.9 dB. Again the new results cannot be said to replicate the old results. The new data indicate much lower speech intelligibility scores at a given SNR_{uni32} value.

C. Discussion of results of validation tests

The main purpose of these listening tests was to compare the results of the initial study with the new results obtained in a more realistically valid acoustical environment. The new results presented here do partially agree with the previous results, but also indicate some significant differences.

There were not significant differences between the audibility threshold results from the previous work and the new results presented here. These new results validate the previous threshold of audibility of speech sounds and confirmed it to be a SNR_{uni32} value of -22 dB.

On the other hand, the mean trends for the threshold of intelligibility and for the speech intelligibility scores indicate differences of 5–7 dB relative to the previous studies. These are large and important differences that could lead to requiring walls with STC values 5–7 points larger than necessary. Such a large difference could lead to significant additional construction costs and it is essential to understand the cause of these differences.

There are many possible sources of error that could influence the results. For example, there could be errors in the measurement of speech and noise levels. However, this is unlikely to be the cause of the differences, since the same SNR values were used for both the audibility threshold and for the intelligibility threshold results. Since there is good agreement for the threshold of audibility results, it is reasonable to assume that the speech and noise level measurements were correct in both studies and that the subjects were capable of hearing the speech sounds.

It is likely that the differences in intelligibility results are due to factors that would affect the understanding of speech and not the simple perception of the presence of speech sound as in the audibility test. Perhaps the most obvious factor is reverberation. It is well known that the intelligibility of speech is influenced by both the SNR and the reverberation time of the listening space. However, the effects of reverberation on speech intelligibility at the very low SNRs that are of concern in speech privacy issues are not well defined. In our new two-room validation experiment, the subjects listened to just audible speech modified by the reverberation of both the source and the receiving rooms, but no measure of the effects of reverberation was included in determining the SNR_{uni32} values.

One other possible contributing factor was the differences in the spatial characteristics of the speech and noise sounds to which the listeners were exposed, which are well known to influence the intelligibility of speech.[6,7] In the initial work[2] subjects listened in approximately free-field conditions. The speech sounds arrived from a loudspeaker system directly in front of the seated subject and the noise sounds arrived from another loudspeaker system directly overhead. Such conditions would maximize a listener's ability to understand speech.

In the new tests, speech sounds were radiated into one room with an average reverberation time of 0.80 s (averaged over the frequencies 160–5000 Hz); they traveled through a real wall and from the wall a further 0.25 m to the listener's ear. Other speech sounds would reflect about the rooms and arrive a little later at the listener's ears. The noise source was located at the opposite side of the room to the listener and hence the direct speech and noise sounds were spatially separated. However, the noise source was in a room with an average reverberation time over speech frequencies of 0.64 s and the listener would hear reflected noise sounds from many directions.

Although the conditions in the new experiment were more realistic and more representative of listening in real rooms, it is not known how the differences in added reflected speech and noise sounds would each affect speech intelligibility scores. Small amounts of reverberant sound do not usually have large effects on the intelligibility of speech. However, it is possible that we are more sensitive to the negative effects of reverberation for the very low signal-to-noise conditions in these experiments. Our knowledge of the effects of the spatial separation of speech and noise sources is almost entirely based on the perception of only the direct sounds in free-field conditions.[6,7] It is quite possible that the addition of reflected sound to the speech and noise signals considerably modifies these effects.

TABLE III. Horizontal and vertical angles of the loudspeakers relative to the listener's head. Angle 0,0 is straight ahead of the listener's head.

Loudspeaker		Horizontal angle (deg)	Vertical angle (deg)
1	Center	0	0
2	Left	−32	0
3	Right	+32	0
4	Center high	0	25
5	Left high	−37	28
6	Right high	+37	28
7	Left rear	−115	0
8	Right rear	+115	0

IV. INVESTIGATIONS OF SPATIAL AND TEMPORAL EFFECTS

A. Procedure

A second series of four experiments[19] was carried out to better understand the magnitude of particular details of the spatial and temporal differences in the sound fields of the initial tests and the new two-room validations tests. The experiments were mostly carried out in simulated sound fields in an anechoic room using an eight-channel simulation system. The eight loudspeakers of the system were positioned around the subject at angles listed in Table III and illustrated in Fig. 2. Five of the loudspeakers (1, 2, 3, 7, and 8) were in the horizontal plane of the listener's ears and the other three were raised up above this plane. The listener sat facing loudspeaker No. 1, which reproduced the direct sound.

The signals to each loudspeaker were processed by four Yamaha DME32 digital signal-processing units connected together to function as one large unit. Speech and noise signals were separately processed and mixed together for each loudspeaker. The loudspeakers were Tannoy model 800A units with concentric drivers so that all frequencies were radiated from the same location. In some cases all loudspeakers radiated speech and noise signals and in other cases only selected loudspeakers were used depending on the purpose of each test condition. For the noise signals, large delays were introduced between the signals to the different loudspeakers so that they arrived at the listener incoherently and when all eight loudspeakers were used conditions were perceived as very diffuse.

Because early-arriving sounds are perceived differently than later-arriving sounds,[20] early- and later-arriving speech sounds were varied independently. Each loudspeaker could reproduce simulations of four early-arriving sounds for a total of 32 early-arriving sounds. The first early-arriving sound arrived from loudspeaker No. 1 (see Table III) to simulate the direct sound. The other early-arriving sounds arrived within a 50 ms interval after the arrival of the simulated direct sound and were intended to simulate various early-arriving reflections. Digital reverberator components in the DME32 units were used to simulate the many later-arriving reflections that would occur in a room.

One of the new experiments was carried out in the same conditions as the initial study[2] so that the same geometry of speech and noise sources could be included. This is a sound

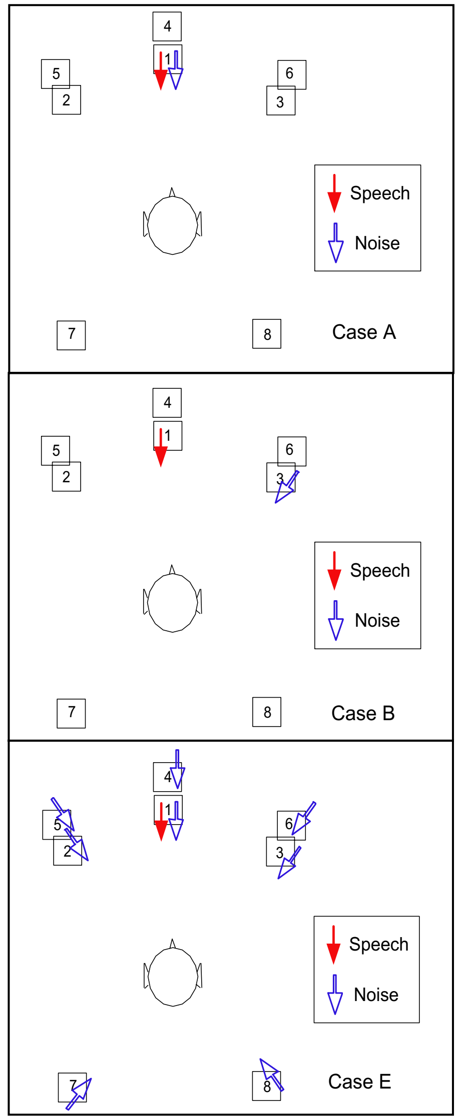

FIG. 2. (Color online) Plan of loudspeaker locations and descriptions of simulated sound fields: Case A: speech and noise only from the same loudspeaker (No. 1) directly in front of the listener, case B: speech and noise only from two separate loudspeakers (Nos. 1 and 3), and case E: speech from ahead only and noise from all loudspeakers.

isolated, quiet (background level of 13 dBA), and acoustically dead space. In this space, speech sounds were reproduced by loudspeakers approximately 2 m in front of the listener and located behind a curtain. A second set of loudspeakers in the ceiling void above the subject was used to produce simulated ventilation noise. In some cases the loudspeakers in front of the subject were used to reproduce both speech and noise sounds. A Yamaha DME32 unit was used to control the sounds to each of the loudspeaker systems. Each set of loudspeakers consisted of two Paradigm Compact Monitors and a Paradigm PW sub-woofer with a response corrected to be flat ±1 dB from 60 to 12 000 Hz at the listener's position.

In some cases speech sounds were filtered by the DME32 unit to represent transmission through a wall. A wall consisting of 16 mm gypsum board on both sides of lightweight 90 mm steel studs and with glass fiber material in the

TABLE IV. Descriptions of cases A, B, and E to demonstrate simple spatial release from masking and cases C, D, and F to demonstrate the effects of added early-arriving reflections of speech sounds.

Case	Speech	Noise	Wall
A	Direct from No. 1	Direct from No. 1	None
B	Direct from No. 1	Direct from No. 3	None
C	Direct+early reflections	Direct from No. 3	None
D	Direct+early+T_{60}=1	Direct from No. 3	None
E	Direct from No. 1	Diffuse from all	None
F	Direct+early reflections	Diffuse from all	None
G	Direct+early+T_{60}=1	Diffuse from all	None

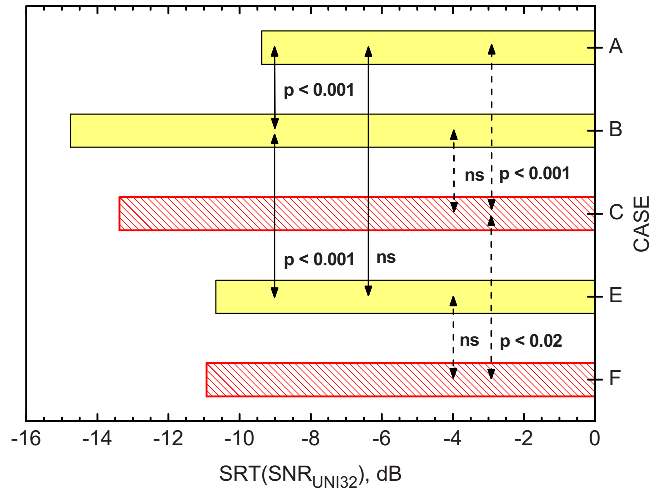

FIG. 3. (Color online) SRT values for cases A, B, C, E, and F with configurations illustrated in Fig. 2 and described in Table IV. Arrows indicate the two cases for which significance of the difference is given. Solid filled bars: simple spatial release from masking cases. Hatched bars: cases with added early-arriving reflections.

cavity was simulated. This construction would correspond to a STC rating of 47 and is typical of many interior office walls.

For these experiments ambient noise with an approximately −5 dB per octave spectrum shape was again used with an overall level of 45 dBA. This is often referred to as "neutral" sounding and is representative of typical indoor noise spectra.[14,15]

To focus on the main differences between the two previous studies, in these tests only speech intelligibility scores were obtained. A number of spatially different combinations of speech and noise sources were compared as well as differences in added reflected sounds. These are later described with the results of each comparison. For each configuration, tests were carried out at two different SNRs. To make it possible to combine the results for two different SNRs all speech intelligibility scores were converted to speech reception threshold (SRT) values. The SRT is the SNR for which the mean intelligibility score is 50%. This provides more accurate descriptors of the intelligibility scores for each configuration by basing the results on a larger number of responses because the results of two SNRs could be combined.

SRT values were calculated using the same Boltzmann equations described previously. When plotting speech intelligibility scores versus SNR_{uni32} values, the SRT value is the x_0 value in Eq. (2). Using the same dx values as in the previous studies (Table II), the SRT can be calculated using the Boltzmann equation and the mean intelligibility scores with the corresponding SNR_{uni32} values from each test. The results were obtained in four separate experiments, which are described in Table VII in the Appendix along with the SRT values for each configuration tested.

B. Simple spatial release from masking

The first tests demonstrated simple spatial release from masking effects to confirm the validity of the new results as well as to serve as a reference case with which subsequent results can be compared. Configurations A, B, and E are described in Fig. 2 and Table IV. Figure 2 shows a top down view of the subject and loudspeakers with solid arrows indicating sources of speech sounds and open arrows for sources of simulated ventilation noise. As illustrated in Fig. 2, speech and noise were produced by only loudspeaker No. 1 directly in front of the subject in case A. In case B the speech and noise sources were separated in the horizontal plane so that the speech came from only loudspeaker No. 1 (directly in front of the subject) and the noise came from only loudspeaker No. 3 (32° to the right of straight ahead). In case E, the speech again came from only loudspeaker No. 1 but the noise was radiated from all eight loudspeakers. Because of the large delays between the noise signals to each loudspeaker, the subject experienced a very diffuse noise sound.

The SRT values for these three cases are compared in the filled bars of Fig. 3. (SRT values and details of all cases are summarized in Table in the Appendix). Of the three, case A has the highest SRT value (−9.37 dB) indicating the lower intelligibility scores that result when the speech and noise come from exactly the same source location. Case B has a much lower SRT value (−14.75 dB), indicating a spatial release from masking when the sources of the speech and the noise are spatially separated. However, when the noise comes from all directions, as in case E, the SRT (−10.67 dB) is similar to that for case A.

A oneway analysis of variance for the experiment No. 1 results, which included these three conditions, indicated significant variations in SRT values (F=25.15, $p<0.001$). A post hoc Bonferroni test of the individual differences indicated that cases A and B were significantly different ($p<0.001$) as were cases B and E ($p<0.001$). However, the difference between the SRT values for cases A and E was not statistically significant.

When the noise source was separated from the speech source by 32° in the horizontal plane the SRT decreased by 5.4 dB, which agrees with previously published results.[6,11] The comparisons with case E results suggest that a diffuse masking noise arriving from all directions leads to results that are quite similar to the case of coincident speech and noise sources (case A) and in this experiment cases A and E were not significantly different. That is, in real rooms with at least somewhat diffuse noise, and where the listener is in the reverberant field of the talker, we would expect less spatial release from masking for separated speech and noise sources.

C. Effects of early-arriving reflections of speech

Early-arriving reflections were radiated from all eight loudspeakers, so that they arrived within about 50 ms after the direct sound and decreased realistically in amplitude with increasing time. Such early-arriving reflections have been shown to increase speech intelligibility equivalent to increasing the level of the direct sound by the increase in energy added by the early-arriving reflections.[20] In these experiments the speech levels were adjusted to be the same for both with and without early-arriving reflections cases. Case C included direct speech and early reflections (ERs) of speech from all loudspeakers combined with simulated ventilation noise from only loudspeaker No. 3. Case F was similar to case C but included simulated ventilation noise arriving diffusely from all eight loudspeakers (as in case E).

Figure 3 compares the SRT values for cases C and F (hatched bars) with those configurations previously described (solid filled bars). Case C is essentially the result of adding ERs to the conditions of case B without any change in the overall speech level. Figure 3 indicates a small increase in SRT value for case C relative to case B. Case F is essentially the result of adding early-arriving reflections of speech sounds to case E. The results in Fig. 3 indicate a small decrease in the SRT value for case F relative to case E.

A oneway analysis of variance on the experiment No. 1 conditions showed a statistically significant pattern of variations in SRT values ($F=25.25$, $p<0.001$) but a post hoc Bonferroni test showed that the SRT values for cases B and C (with added ERs) were not significantly different. Similarly the SRT values for cases E and F were not significantly different.

Although not statistically significant, adding ERs, while maintaining the same overall speech level, did increase the SRT value a little for case C compared to case B. This may have been because the speech energy in the ERs included some speech from the same direction as the noise, which might tend to increase the SRT a little for this case. When ERs were added to configurations in which the noise came from all eight loudspeakers, the added early-arriving speech reflections decreased the SRT slightly but not significantly (cases E and F). In both cases adding ERs of the speech energy was largely equivalent to increasing the level of the direct speech sound as was expected. ERs did not significantly affect spatial unmasking effects in these simulations.

D. Added effects of reverberant speech

The digital reverberators in the DME32 units were used to add a reverberant decay of the speech sounds after the early-arriving reflections, but with the overall speech level not varied. In the configurations presented here, the reverberant sound had a mid-frequency reverberation time of 1.0 s. Figure 4 compares the resulting SRT values with those for the previously described cases.

Case D was constructed by adding a reverberant decay to the case C configuration, which had included direct speech, early-arriving reflections of the speech sounds, and noise from only loudspeaker No. 3. When reverberant speech was added to case C, the SRT was increased from

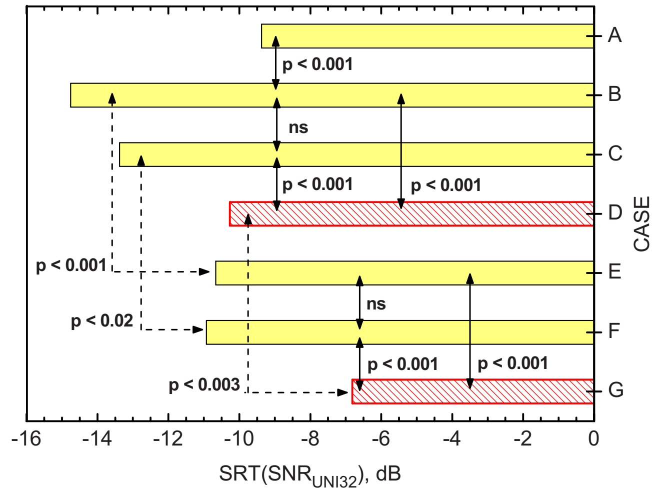

FIG. 4. (Color online) SRT values for cases A–G with configurations illustrated in Fig. 2. Solid filled bars are repeated from previous figure. See Table IV for descriptions of cases.

−13.78 to −10.27 dB and this difference was statistically significant (oneway $F=25.25$, $p<0.001$, post hoc Bonferroni $p<0.001$).

Similarly, for the cases with diffuse noise from all loudspeakers, adding reverberant sound increased the SRT from −10.93 dB for case F to −6.82 dB for case G. This difference was statistically significant (oneway $F=25.25$, $p<0.001$, post hoc Bonferroni $p<0.001$).

While adding ERs of the speech did not significantly change the SRT (for constant speech level), adding reverberant speech increased the SRT by 3–4 dB independent of the spatial differences in the simulated ambient noise. The addition of reverberant speech adds to the masking of the speech independently from that of adding diffuse noise.

E. Effects of varied reverberation time

The effect of reverberant speech was further investigated by comparing conditions with reverberation times of 0.5, 1.0, and 2.0 s. All cases included direct speech and early-arriving reflections with added reverberant speech with one of three different reverberation times. These comparisons were also repeated with and without modifications to the spectrum of the speech sound to simulate the effect of propagation through the wall.

Case P corresponded to no wall, and speech with added ERs. Speech and noise sounds were reproduced by all eight loudspeakers. This base case without reverberant sound was compared with cases Q, R, and S, which had added reverberant speech with reverberation times of 0.5, 1.0, and 2.0 s, respectively. The resulting mean SRT values for each case are given in Fig. 5 (see Table V).

For the cases without a simulated wall, adding reverberant speech with a 0.5 s reverberation time (case Q) to case P (which had only ERs of the speech) only increased the SRT by a small amount and the difference was not statistically significant. However, adding more reverberant speech corresponding to a 1.0 s reverberation time (case R) and a 2.0 s reverberation time (case S) produced larger and statistically

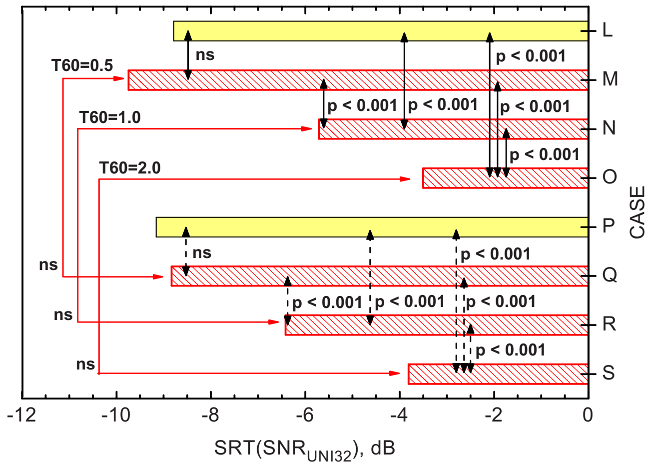

FIG. 5. (Color online) SRT values for cases L–S, all with speech and noise sounds from all eight loudspeakers. Bars with hatched lines show cases with varied reverberation time. See Table V for descriptions of cases.

significant increases in SRT values (oneway on the experiment No. 4 data: $F=94.15$, $p<0.001$, post hoc Bonferroni, $p<0.001$).

Case L was similar to case P but with the inclusion of filtering to simulate transmission through a wall for the speech sounds. The SRT value from case L is compared with those for cases M, N, and O with reverberation times of 0.5, 1.0, and 2.0 s, respectively, in Fig. 5. All of these cases included simulated transmission through a wall for the speech sounds.

When a simulated wall was included, the results of cases L, M, N, and O showed a similar progression of changes in the SRT values. Adding only reverberant speech with a 0.5 s reverberation time (case M) led to a non-significant change. However, adding reverberant speech with a 1.0 or a 2.0 s reverberation time each led to large and statistically significant increases in SRT values (oneway on the experiment No. 3 data: $F=31.51$, $p<0.001$, post hoc Bonferroni $p<0.001$).

For these cases with reverberation times of 0.5, 1.0, and 2.0 s, the corresponding with-wall and without-wall cases were not significantly different (independent t-test) (indicated by the solid U-shaped arrows in Fig. 5).

In order to better understand the changes in SRT values caused by increasing the reverberant speech sound, the mean SRT values are plotted versus the logarithm of the reverberation time in Fig. 6. The data for the cases with only ERs are included with a reverberation time of 0.5 s because these

TABLE V. Descriptions of the effects of added reverberation: cases P, Q, R, and S without a wall and cases L, M, N, and O with a wall.

Case	Speech	Noise	Wall
L	Direct+early reflections	Diffuse from all	Wall
M	Direct+early+T_{60}=0.5	Diffuse from all	Wall
N	Direct+early+T_{60}=1	Diffuse from all	Wall
O	Direct+early+T_{60}=2	Diffuse from all	Wall
P	Direct+early reflections	Diffuse from all	None
Q	Direct+early+T_{60}=0.5	Diffuse from all	None
R	Direct+early+T_{60}=1	Diffuse from all	None
S	Direct+early+T_{60}=2	Diffuse from all	None

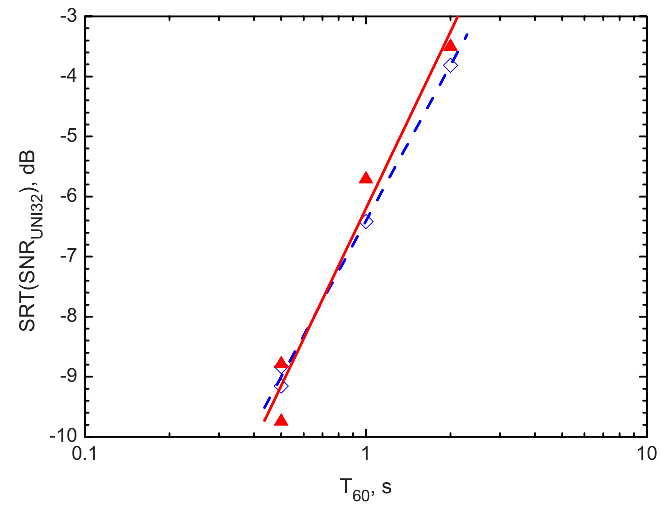

FIG. 6. (Color online) Mean SRT values plotted versus the logarithm of the reverberation time (T_{60}) of the simulated speech sounds. Dashed line and open symbols: no wall; solid line and filled symbols: speech transmission through a wall.

results were not statistically different than the results for a 0.5 s reverberation time. On this plot the SRT values are seen to increase linearly as the logarithm of the reverberation time increased. The results for the with and without a simulated wall cases led to very similar results. The two regression equations and an average for all data are

$$\mathrm{SRT} = 8.602 \log_{10}(T_{60}) - 6.407 \quad (\text{no wall}), \tag{3}$$

$$\mathrm{SRT} = 9.774 \log_{10}(T_{60}) - 6.201 \quad (\text{with wall}), \tag{4}$$

$$\mathrm{SRT} = 9.187 \log_{10}(T_{60}) - 6.304 \quad (\text{all data}). \tag{5}$$

These can be used to predict the effects of reverberation in meeting rooms on the SRT of the transmitted speech.

F. Effects of horizontal and vertical separations of speech and noise sources

Figure 3 showed the results of separating single speech and noise source by 32° in the horizontal plane. Because the initial study[2] included vertically separated speech and noise sources, such cases were repeated here both with and without a simulated wall. The SRT values for the horizontally and the vertically separated conditions are compared in Fig. 7. In all cases separating speech and noise sources led to a highly significant decrease in SRT values ($p<0.001$). A comparison of the SRT values for cases A and B indicated a 5.4 dB shift in SRT values when the speech and noise sources were separated by 32° horizontally.

Case I was essentially a repeat of case A except in a completely different experiment and test facility. In both cases speech and noise sounds came from only the loudspeaker directly in front of the listener. The difference between the SRT values of cases A and I was very small and was not statistically significant. This makes it possible to compare the effects of horizontal separation of speech and noise sources (cases A and B) with vertical separation (cases H and I). In case H the speech and noise sources were separated by 90° in the vertical plane but the spatial release from

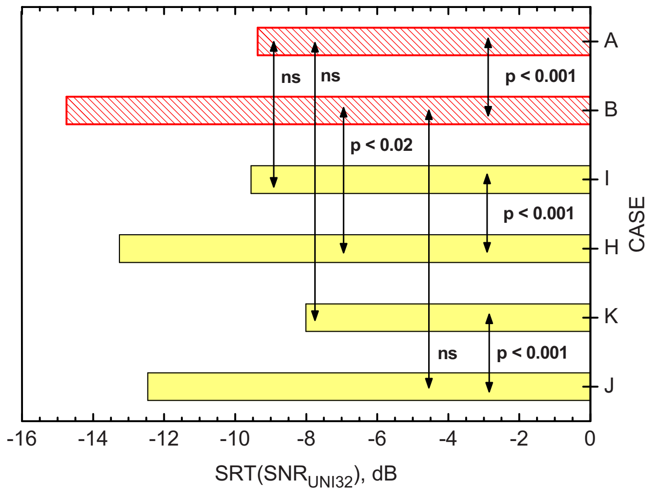

FIG. 7. (Color online) SRT values for cases A, B, I, H, K, and J with separations of speech and noise in the horizontal plane (hatched bars) and in the vertical plane (solid filled bars).

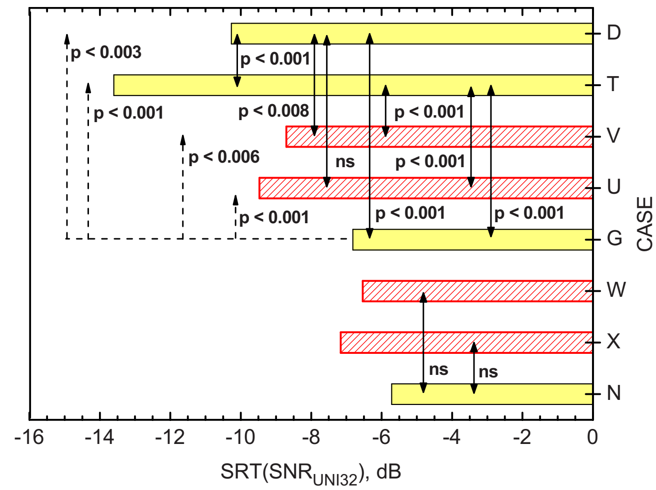

FIG. 8. (Color online) SRT values for cases D, G, N, T, U, V, W, and X with varied masking noise configurations. Cases D, T, V, U, and G do not include a simulated wall; cases W, X, and N include a simulated wall. (Bars with hatched lines are semi-diffuse cases.)

masking (decrease in SRT value) was a little less than when speech and noise sources were separated by 32° in the horizontal plane (cases A and B).

Comparing SRT values for cases K and J indicates a similar spatial release from masking when the speech and noise were separated by 90° in the vertical plane and when speech sounds were in both cases modified to represent transmission through a wall. The SRT values tended to be slightly higher than for the corresponding cases without a simulated wall. However, the differences were not statistically significant. (It is interesting to note that the SRT value for case J was −12.46 dB, which was very similar to the value of −12.19 obtained previously.[2])

G. Effects of single, diffuse, and semi-diffuse noise sources

The difference between the masking effects of ambient noise from a single noise source, which was spatially separated from the speech source, and speech with a diffuse noise from all eight loudspeakers represents two extremes. It is likely that in real rooms intermediate cases are found for which the noise might be described as "semi-diffuse." Such semi-diffuse conditions were produced by radiating the simulated ambient noise predominantly from three adjacent loudspeakers. Because one case included a cluster of noise sources from the rear side of the listener, a single rear-side noise source was also tested as a reference case. These new noise source conditions were compared with the previously described conditions that included either a single noise source or diffuse noise from all eight loudspeakers.

Figure 8 compares SRT values for varied noise masking configurations. The upper five configurations (D, T, V, U, and G) on the graph are for conditions that did not include a simulated wall. Conditions, which included a simulated wall, are in the lower three bars of the graph (W, X, and N). The speech signal in all conditions included direct sound, ERs, and reverberant speech with a 1 s reverberation time to represent conditions in a real room.

Comparing SRT cases D and T shows that the single noise source from the rear side (case T with noise from only loudspeaker No. 8) led to a larger spatial release from masking (i.e., lower SRT) than for a single noise source from the front side (case D) and the difference was highly significant (independent t-test, $p<0.001$). SRT case V for a semi-diffuse noise from the front side has a higher SRT than the single noise source from the front side (i.e., case D) (independent t-test, $p<0.008$). Similarly, a semi-diffuse noise source from the rear side (SRT case U) had a higher SRT than the single noise source from the rear side (SRT case T) (independent t-test, $p<0.001$). However, SRT case G, with diffuse noise from all eight loudspeakers, had the highest SRT of all the cases without a simulated wall.

The semi-diffuse conditions (cases U and V) were intermediate to the single separated noise source conditions (cases D and T) and the all eight loudspeaker noise source condition (case G). Noise sources to the rear side (cases T and U) led to lower SRT values than the corresponding noise sources from the front side (cases D and V). As shown in Fig. 8 all of the differences due to changes in either the direction or diffuseness of the noise were statistically significant.

The lower three bars of Fig. 8 (configurations W, X, and N) include results for similar conditions except that the speech sounds were filtered to simulate transmission through the wall. These results show a little higher SRT values than the corresponding cases without walls. Although the corresponding conditions seemed to have higher SRT values when a simulated wall was included, none of the differences between the with- and without-wall cases were statistically significant. That is, cases with semi-diffuse noise from the front side (cases V and W) were not significantly different and neither were the two semi-diffuse noise from the rear-side conditions (cases U and X). Similarly, the two cases with all eight loudspeaker radiating noise (cases G and N) were not significantly different. From these results one must conclude

TABLE VI. SITs in terms of SNR_{uni32} values in decibels. "ER," direct sound and early-arriving reflections; "T_{60}=0.5," direct sound early-arriving reflections and reverberant sound with a 0.5 s reverberation time.

SRT case	Wall	Speech	Noise	SIT (dB)	Expt.
P	No wall	ER	All	−12.62	4
Q	No wall	T_{60}=0.5	All	−11.89	4
R	No wall	T_{60}=1	All	−10.35	4
S	No wall	T_{60}=2	All	−7.61	4

that there is no proof of an effect of transmission though a wall when conditions are described in terms of SNR_{uni32} values.

H. Changes in speech intelligibility thresholds

The main focus of this work was on speech intelligibility scores expressed in terms of SRT values. However, previously established criteria for acceptable speech privacy were in terms of the SIT values.[2] The SIT is the SNR for which 50% of a panel of listeners can just understand at least one word of a test sentence. For approximately free-field conditions, the threshold of intelligibility was found to correspond to a SNR_{uni32} value of −16 dB.[2] However, in the two-room validation tests [see Fig. 1(b)], the threshold of intelligibility was increased by 4.9 dB to a SNR_{uni32} value of approximately −11 dB.

To determine threshold of intelligibility values requires data for conditions with a significant number of responses with low intelligibility scores extending down to zero. In most of the new tests such conditions were deliberately avoided to provide data mostly in the range 10%–90% intelligibility scores. However, one series of conditions with varied reverberation times and with ambient noise coming from all eight loudspeakers did include a significant number of low intelligibility cases from which SITs could be determined. These data were for natural speech without simulated transmission through a wall and were used to determine new estimates of the effects of reverberation and diffuse noise on the SIT criteria.

Values of the threshold of intelligibility were calculated in a manner similar to the calculation of SRT values described in Sec. IV A of this paper. The fraction of the listeners indicating at least one word was understood for each test configuration was considered in terms of plots of these values versus SNR_{uni32} values. A Boltzmann equation was fitted to the data using the same dx value as previously obtained for SITs (Ref. 2) corresponding to a value of 1.8739 (see Table II). SIT values were calculated using Eq. (2) and using dx=1.8739. The fraction of subjects understanding at least one word was y and x_0 was the threshold of intelligibility. As before, x was the SNR_{uni32} value corresponding to the y value. The resulting SIT values are given in Table VI. The other information in Table VI is repeated from the description of configurations in Table VIIVII in the Appendix.

The calculated SITs are plotted versus the logarithm of T_{60} values in Fig. 9. The SIT values are approximately linearly related to the logarithm of the reverberation time similar to the plot of SRT values versus T_{60} in Fig. 6. As in Fig. 6, the case with only ERs was plotted as having a T_{60} value of 0.5 s.

Figure 9 suggests that for a typical meeting room, the SIT would be approximately a SNR_{uni32} value of −11 dB, corresponding to a T_{60} value of 0.75 s. A little lower or higher values are indicated for lower or higher reverberation times. The criterion for the SIT should be raised from −16 dB for free-field conditions with spatially separated speech and noise sources,[2] to −11 dB for conditions where there is reverberant speech and diffuse ambient noise as found in most meeting rooms. This agrees well with the result from the two-room validation study as shown in Fig. 1(b). Small adjustments for differences in reverberation times could be made if needed but are usually not justified.

V. DISCUSSION

The new results clearly demonstrate that temporal and spatial effects influence speech privacy in that they can significantly change the intelligibility of speech and the threshold of intelligibility. Both the results in Sec. IV H and the results of the two-room validation study indicate that shifts in the threshold of intelligibility of 5 dB or more are possible. Ignoring these effects could lead to a 5 dB over-design of the sound insulation of the meeting room.

The SRT (expressed in terms of SNR_{uni32} values) for the various conditions similarly indicate that speech intelligibility scores can be significantly affected by spatial and temporal effects of room acoustics. These results can be used to explain the difference between the initial results in approximately free-field conditions[2] and the new two-room validation study results. In the initial study, subjects heard speech

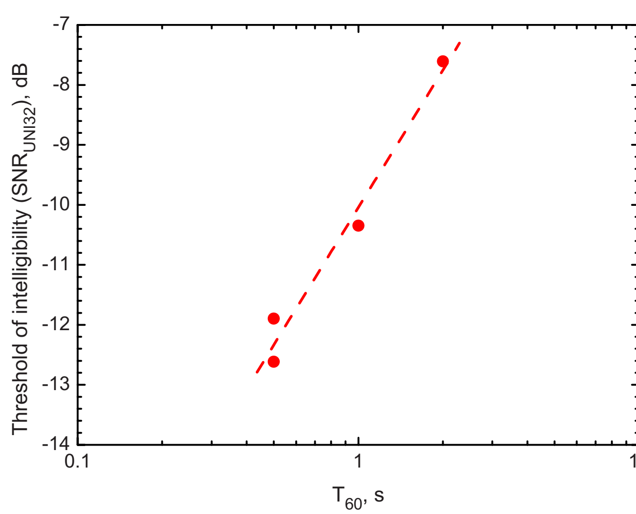

FIG. 9. (Color online) SIT versus T_{60} for unmodified speech (i.e., no simulated wall) and diffuse noise.

sounds from only directly in front of them and simulated ventilation noise from only above them. Going from this (case J) to the co-located speech and noise configuration (case K) decreased the SRT by 4.45 dB. Going from a co-located speech and noise configuration to a condition with diffuse noise would increase SRT by a further 1.30 dB. Combining these two differences would give a total SRT change of 3.15 dB. Finally adding on the effects of room reverberation from Fig. 9 leads to a further increase in SRT of 3.03 dB to give a total expected increase in SRT of 6.18 dB between the conditions of the initial study[2] and those of the two-room validation tests. This is reasonably close to the 7.5 dB difference in SRT values seen in Fig. 1(c).

Clearly these results can be used to estimate the effects for other conditions including varied sound diffusion of the rooms, varied reverberation times, and the effects of various spatial separations of speech and noise sources in spaces with high sound absorption. In spaces that are highly absorptive, listeners will benefit from spatial separation of speech and noise sources and speech privacy will be reduced. Strong early-arriving reflections of the speech sound will increase the effective speech level because the early-arriving reflections are not perceived as spatially or temporally separated from the direct sound and are equivalent to increasing the level of the direct sound. However, reverberant sound will minimize any spatial release from masking and lead to decreased intelligibility and hence increased speech privacy.

It is difficult to precisely compare the new results with those from the many previous studies in the literature because of the methodological differences among the various investigations. For example, subjective ratings of conditions have frequently been used rather than speech intelligibility scores.[11,13] Kollmeier and Wesselkamp[21] showed that the results of these two approaches are correlated but there are differences in the magnitude of the effects and the form of the trends with varying SNR can also be different. A number of studies have used such subjective ratings in an iterative procedure to determine SRT values. In their tests, the subject heard the same speech material repeatedly and decided when it appeared to be just intelligible. This is quite different than listening to each test sentence only once and counting the fraction of words correctly understood, as in the current work.

No previous work has examined the separate effects of early-arriving reflections of speech sounds on the various spatial effects. Descriptions of room acoustics conditions and reverberation are often not very detailed and conditions with as little as a 0.4 s reverberation time have been tested as a reverberant extreme.[12] In previous studies interfering sounds have most frequently been speech and much of the work was focused on explaining the cocktail party effect. Where the interfering sound has been noise, it has most often been noise with a speech spectrum shape. At least one study used white noise[22] but none have used noise representative of typical indoor ambient noises.

The initial test results comparing SRT cases A and B confirmed the expected spatial release from masking when the speech and noise sources were horizontally separated by 32° in free-field conditions. The 5.4 dB difference in SRT values for these two cases is of similar magnitude to results in several previous studies.[4,11,13] In Bronkhorst's review[6] he indicated that values between about 4 and 6 dB occur for this angular separation. The decreased SRT for the rear-side noise source (case T) is also supported by the results in Bronkhorst's summary[6] that indicate maximum spatial release from masking when the interfering source is at an angle of 110°–120° from straight ahead. A 90° vertical separation of the speech and noise sources (speech from in front of the listener and noise from overhead, as in cases H and I) had a 4.7 dB spatial release from masking, a little smaller than the 5.4 dB difference for the 32° horizontal separation. No previous measurements of the effect of a vertical separation of speech and noise sources were found.

There are very few previous results that can be compared with the diffuse interfering noise in the present study where the noise came incoherently from all eight loudspeakers as in case E of the present work. This resulted in a SRT only 1.3 dB lower than the case of coincident speech and noise sources (case A). That is, with diffuse noise, there is very little spatial release from masking. Some previous work has showed that the magnitude of the spatial release from masking decreases as the number of spatially separated noise sources increases.[11] In a similar manner, the semi-diffuse cases in the present work show that the spatial release from masking is significantly reduced when there were three spatially close masking sources compared to a single noise source.

Adding ERs to the speech sounds did not significantly change measured SRT values either with a horizontally separated speech and noise or with noise from all loudspeakers. This extends our understanding of the beneficial effects of early-arriving reflections on the intelligibility of speech[20] and it can be said that early-arriving reflections of speech sounds do not reduce our ability to benefit from spatially separated speech and noise sources. In rooms, the addition of early-arriving reflections will increase the effective SNR and enhance the intelligibility of speech.

Adding reverberant speech sound does degrade the intelligibility of speech in noise. The addition of reverberant speech with a 1 s reverberation time had about the same magnitude of increase in SRT as did adding diffuse noise to the case B results with neither reverberation nor diffuse noise. That is, although they are completely independent effects, adding diffuse noise or adding reverberant speech in these cases led to about the same 4 dB increase in SRT values. When both diffuse noise and reverberant speech were included (case G), then the SRT was increased by about 8 dB or approximately the sum of the individual effects.

The effect of adding reverberant speech increased linearly with the logarithm of the reverberation time above a reverberation time of about 0.5 s. The addition of reverberant sound with a 0.5 s reverberation time did not significantly change the measured SRT relative to the case with only early-arriving reflections added to the speech sound. It is only for more reverberant conditions that the negative effects of reverberation became significant. In these experiments adding ERs and reverberation to the speech was accomplished while maintaining a constant overall speech level at

TABLE VII. SRT values and descriptions of conditions for 24 test configurations. Column "Expt." indicates to which of the four experiments each case belonged (each experiment used different subjects). Column "Wall" indicates whether a simulated wall was included or not. Column "Speech" indicates the composition of the speech signal: "Direct," direct sound only; "ER," direct sound and early-arriving reflections, "$T_{60}=0.5$," direct sound early-arriving reflections and reverberant sound with $T_{60}=0.5$ s; "$T_{60}=1$," direct sound early-arriving reflections and reverberant sound with $T_{60}=1.0$ s; "$T_{60}=2$," direct sound early-arriving reflections and reverberant sound with $T_{60}=2.0$ s. Column "Noise" indicates the composition of the noise signal: "Front," from only loudspeaker No. 1 directly in front of subject; "Front side," from only loudspeaker No. 3; "All," uncorrelated noise from all eight loudspeakers; "Ceiling," from only immediately overhead; "Rear side," from only loudspeaker No. 8; "Front-side diffuse," predominantly from loudspeaker Nos. 1, 3, and 6; and "Rear-side diffuse," predominantly from loudspeaker Nos. 6, 8.

SRT case	Expt.	Wall	Speech	Noise	SRT (dB)
A	1	No wall	Direct	Front	−9.37
B	1	No wall	Direct	Front side	−14.75
C	1	No wall	ER	Front side	−13.38
D	1	No wall	$T_{60}=1$	Front side	−10.27
E	1	No wall	Direct	All	−10.67
F	1	No wall	ER	All	−10.93
G	1	No wall	$T_{60}=1$	All	−6.82
H	2	No wall	Direct	Ceiling	−13.26
I	2	No wall	Direct	Front	−9.55
J	2	Wall	Direct	Ceiling	−12.46
K	2	Wall	Direct	Front	−8.01
L	3	Wall	ER	All	−8.79
M	3	Wall	$T_{60}=0.5$	All	−9.74
N	3	Wall	$T_{60}=1$	All	−5.71
O	3	Wall	$T_{60}=2$	All	−3.50
P	4	No wall	ER	All	−9.16
Q	4	No wall	$T_{60}=0.5$	All	−8.83
R	4	No wall	$T_{60}=1$	All	−6.42
S	4	No wall	$T_{60}=2$	All	−3.81
T	4	No wall	$T_{60}=1$	Rear side	−13.60
U	3	No wall	$T_{60}=1$	Rear-side diffuse	−9.47
V	3	No wall	$T_{60}=1$	Front-side diffuse	−8.71
W	3	Wall	$T_{60}=1$	Front-side diffuse	−6.54
X	3	Wall	$T_{60}=1$	Rear-side diffuse	−7.16

the position of the listener. In a real room, the addition of reflected sound would increase the overall level of the speech, which could further modify the intelligibility of speech depending on the relative amounts of early-arriving and later-arriving speech sounds.

In some cases the speech sounds were modified to simulate propagation through a wall and cases were compared both with and without the effect of a simulated wall. When this was done for cases with varied reverberation time and also for varied noise diffusion, there were no significant additional effects of adding a simulated wall. That is, the results apply equally well to natural speech as they do to speech filtered by propagation through a wall. Similarly in the validation study, there were no differences for the results using three different walls. However, this was only true when results were considered in terms of uniformly weighted signal-to-noise ratio (SNR_{uni32}) values and not for SNR measures with other frequency weightings.[19] The uniform frequency weighting of the SNR_{uni32} measure seems to make it an ideal measure for assessing speech privacy conditions.

VI. CONCLUSIONS

Although speech privacy may be primarily a signal-to-noise issue, it is also significantly influenced by the spatial and temporal characteristics of sound in rooms. The two-room validation study results and the subsequent investigations of the various combinations of speech and noise showed that SRTs can be changed by 5 dB or more by the differences in the spatial relationships between speech and noise sources as well as the effects of reflected sound in rooms. As speech privacy is most often considered to correspond to reduced speech intelligibility, these effects are practically important and ignoring them could lead to expensive over-design of the sound insulation of meeting rooms, to provide the required speech privacy against eavesdroppers.

These new results have made it possible to revise previous estimates of the threshold of intelligibility from a SNR_{uni32} value of −16 to a value of −11 dB to be more representative of conditions in typical meetings rooms. The new results also make it possible to estimate further modifications to speech privacy criteria to better suit particular combinations of room reverberation and spatial separations of speech and noise sources.

These results do not suggest that the audibility of speech in noise is influenced by the same spatial and temporal room acoustics effects. That is, the threshold of audibility of speech transmitted from an adjacent room is confirmed to correspond to a SNR_{uni32} value of −22 dB as previously determined.[2]

ACKNOWLEDGMENTS

Most of the original experimental work and initial analyses was carried out by M.A. for her diploma thesis at the Fachhochschule für Technik in Stuttgart, Germany. Funding for this work was provided by Public Works and Government Services Canada (PWGSC) and the Royal Canadian Mounted Police (RCMP).

APPENDIX: SUMMARY OF THE TEST CONFIGURATIONS

The measured SRT values and details of the 24 measurement conditions are summarized in Table VII. The data were obtained from four different experiments that each used different subjects. The configurations included in each experiment are indicated in col. 2 of the table.

[1]W. J. Cavanaugh, W. R. Farrell, P. W. Hirtle, and B. G. Watters, "Speech privacy in buildings," J. Acoust. Soc. Am. **34**, 475–492 (1962).

[2]B. N. Gover and J. S. Bradley, "Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms," J. Acoust. Soc. Am. **116**, 3480–3490 (2004).

[3]V. O. Knudsen and C. M. Harris, *Acoustical Designing in Architecture* (Acoustical Society of America, 1980).

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

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

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

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

[8]N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119 (1947).

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

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

[11]J. Peissig and B. Kollmeier, "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670 (1997).

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

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

[14]D. F. Hoth, "Room noise spectra at subscribers' telephone locations," J. Acoust. Soc. Am. **12**, 449–504 (1941).

[15]W. E. Blazier, "Revised noise criteria for application in the acoustical design and rating of HVAC systems," Noise Control Eng. **16**, 64–73 (1981).

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

[17]J. S. Bradley and B. N. Gover, "Validation of architectural speech security results," IRC/NRC Research Report No. RR-221, National Research Council, Ottawa, Canada, March 2006.

[18]ASTM Standard E2638-08, "Standard test method for objective measurement of the speech privacy of closed rooms," ASTM International, West Conshohocken, PA, 2008.

[19]J. S. Bradley, M. Apfel, and B. N. Gover, "Spatial and temporal effects of room acoustics on the speech privacy of meeting rooms," IRC/NRC Research Report No. RR-265, National Research Council, Ottawa, Canada, October 2008.

[20]J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**, 3233–3244 (2003).

[21]B. Kollmeier and M. Wesselkamp, "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," J. Acoust. Soc. Am. **102**, 2412–2421 (1997).

[22]H. Levitt and L. R. Rabiner, "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**, 601–608 (1967).