Spatial and temporal effects of room acoustics on the speech privacy of meeting rooms

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

Publisher's version / Version de l'éditeur:
https://doi.org/10.4224/20377126
Research Report (National Research Council of Canada. Institute for Research in Construction); no. RR-265, 2008-09-01

NRC·CNRC CNRC·NRC
Institute for Research in Construction
Institut de recherche en construction

http://irc.nrc-cnrc.gc.ca

Spatial and Temporal Effects of Room Acoustics on the Speech Privacy of Meeting Rooms

RR-265

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

September 2008

National Research Council Canada
Conseil national de recherches Canada
Canada

Spatial and Temporal Effects of Room Acoustics on the Speech Privacy of Meeting Rooms

John S. Bradley, Marina Apfel and Bradford N. Gover

IRC Research Report IRC-RR-265

September 2008

Acknowledgements

Experiments 1, 2 and 3 and initial analyses of these data were carried out by Ms. Marina Apfel for her Diploma Thesis at the Fachhochschule für Technik in Stuttgart, Germany. The data referred to as experiment 4 were obtained at NRC-IRC by Mr. Jarrod Whittington for his Master's thesis at Rensselaer Polytechnic Institute in Troy, New York. The research was jointly funded by the Royal Canadian Mounted Police (RCMP), Public Works and Government Services Canada (PWGSC) and the National Research Council Institute for Research in Construction of (NRC-IRC).

Summary

This report describes the results of a focussed study of the spatial and temporal effects of room acoustics on the intelligibility of speech including speech transmitted from an adjacent room. The purpose of the work was to gain an initial understanding of how these effects influence the speech privacy of enclosed rooms. Earlier speech privacy studies had identified a uniformly weighted (over frequency) signal-to-noise ratio measure (SNR_{UNI32}) as most suitable for rating the audibility and intelligibility of speech from meeting rooms for nearby eavesdroppers. This initial work was carried out in free-field conditions without reflected sound and did not include the effects of room reverberation. In addition, the speech and noise came from two spatially separated single source locations. These conditions were intended to represent a worst-case condition in which it would be most easy to understand speech.

A subsequent two-room validation experiment, in which listeners heard speech from an adjacent room, demonstrated that in those more realistic conditions speech intelligibility scores were significantly reduced relative to the previous results. The validation test results suggested that less sound insulation would be required to meet the same speech privacy criteria. It was thought that the expense of high sound insulation for meeting rooms could be reduced if the spatial and temporal effects of room acoustics were better understood.

To make this possible, a series of speech intelligibility tests were carried out in various simulated conditions to systematically explain the importance of the various spatial and temporal parameters describing the conditions in real rooms. The following details were explained and it is now possible to more accurately estimate the speech privacy of a meeting room.

- The results confirmed how much the intelligibility of speech is increased when the speech and noise sources are separated in free field conditions, i.e. there is a *Spatial Release from Masking*, for horizontally or vertically separated speech and noise sources.
- When a more realistic diffuse noise sound field was produced, the *Spatial Release from Masking* was substantially reduced relative to the case with a single separated noise source.
- Adding early-arriving reflections to the speech sounds, while maintaining a constant overall speech level, had no significant effect on speech intelligibility scores.
- Adding reverberant speech sounds with a reverberation time greater than 0.5 s, while maintaining a constant speech level, decreased speech intelligibility relative to a comparable case without reverberant speech.
- The decrease in intelligibility that resulted from adding reverberation to the speech can be determined from the logarithm of the reverberation time for reverberation times greater than 0.5 s.
- A single noise source from the rear side led to less masking of reverberant speech than a similar single noise source from the front side.

- When semi-diffuse noise was created, the effects on intelligibility were intermediate to those for a single direct noise source and those for completely diffuse noise.
- Although sound transmission through walls attenuates higher frequency sounds much more than lower frequency sounds, this filtering of speech sounds had no significant additional effect on either the influence of noise or of reverberation on the intelligibility of the speech when evaluated in terms of uniformly weighted signal-to-noise ratios (SNR_{UNI32}). This was not true for measures using other frequency weightings.
- The combined effects of the spatial differences in the noise exposures and the reverberation in the rooms adequately explained the difference between the initial laboratory experiment and the two-room validation tests in terms of both speech intelligibility scores and the threshold of speech intelligibility. As a result it is now possible to more accurately estimate the intelligibility of speech from an adjacent meeting room.

Table of Contents

1. Introduction

Previous work determined the uniformly weighted (over frequency) signal-to-noise ratio (SNR_{UNI32}) to be a good predictor of the audibility and intelligibility of speech transmitted from an adjacent room [1]. The SNR_{UNI32} measure was found to be most successful for rating the audibility and intelligibility in very low signal-to-noise ratio conditions of importance to assessing the speech privacy afforded by a meeting room. The initial experiments that led to this conclusion were conducted in approximately free-field conditions intended to represent a worst-case situation in which it would be most easy to hear or understand very low levels of speech in noise. Figure 1 illustrates the experimental conditions of the initial approximately free field, laboratory experiments.

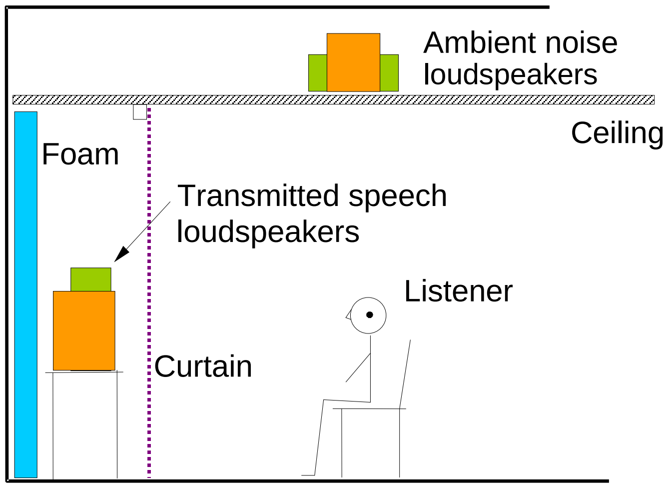


Figure 1. Cross-section of Room Acoustics Test Space showing the arrangement of the listener as well as the speech and noise loudspeakers.

A subsequent validation experiment, in which subjects listened to speech in one room actually transmitted from an adjacent room, confirmed the validity of the initial audibility experiment but found lower intelligibility scores. Figure 2 illustrates the experimental setup of the two-room validation experiment. The difference between the results of the two experiments was attributed to the temporal and spatial room acoustics effects on the sounds that occurred while listening in the real rooms of the two-room validation test. The well-known temporal effect of room reverberation causes one word to mask or interfere with the sounds of subsequent words, making them more difficult to understand. The spatial separation of the speech and noise sounds in the initial experiment is well known to make it easier to understand the speech. In many situations in real rooms the benefit of separated speech and noise sources is compromised by reflected sounds in the room that arrive from many different directions. Although these temporal and spatial effects are known to exist, the many details are not well understood. It was not possible to estimate the magnitude of the effect of each parameter on the intelligibility of speech. In total this is a large complicated set of issues, which cannot all be solved in one relatively small study. This new work was intended to focus on getting an initial understanding of the issues related to the audibility and intelligibility of speech from an adjacent room.

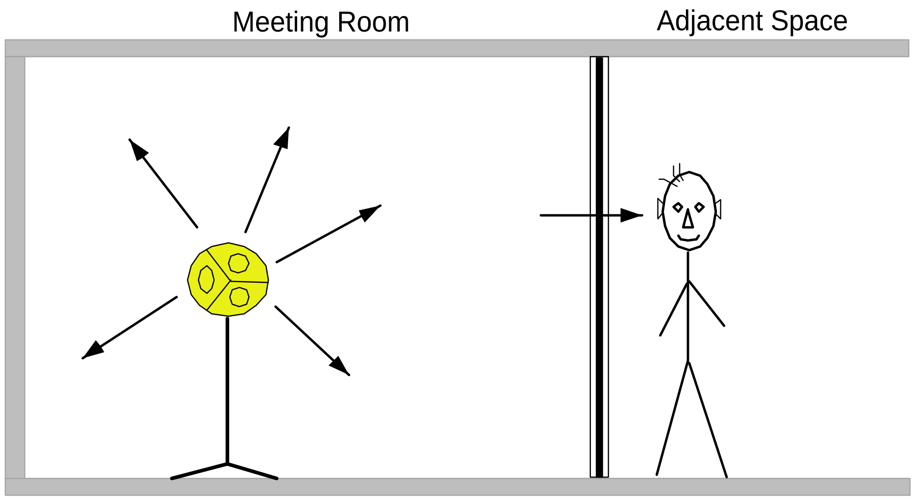


Figure 2. Experimental setup of the two-room validation tests with speech radiated into the source room representing a meeting room with a 0.8 s reverberation time and a subject listening to the speech in an adjacent room.

Review of Previous Research

There have been many studies related to developing an understanding of the temporal and spatial effects of room acoustics on our ability to understand speech heard in combination with competing sounds. Classical room acoustics studies have long identified optimum reverberation times for maximizing the intelligibility of speech [2]. 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* [3]. There have been at least two reviews of the many studies related to the *cocktail party effect* [4, 5].

Interfering sounds can mask the target speech sounds (that we wish to hear) and reduce the intelligibility of the speech. Such masking is influenced by both monaural and binaural effects. For example, 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. The intelligibility of speech is first a signal-to-noise issue. The higher the level of the interfering sounds in frequency bands important for speech communication, relative to the level of the speech sounds in the same frequency bands, the lower will be the intelligibility of the speech. The work of French and Steinberg that led to the Articulation Index can explain the monaural combined signal-to-noise effects over frequencies important for speech [6].

It is usually possible to better understand the target speech mixed with interfering sounds by listening binaurally (i.e. with two ears). 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 inter-aural level, time and phase differences to be important [7]. These interaural differences vary with the direction of the sound source relative to the head of the listener and hence can help us to discriminate among spatially separated sound sources.

Although many of the earlier studies focussed 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 is similar to that of noise with a similar spectrum and level.

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 degrees is detectable and that a 20 degree difference leads to quite significant increases in the intelligibility of the target speech [8]. Systematic studies have reported the resulting increased intelligibility of separating the directions of arrival of the target speech and interfering noise as a function of the angle of the separation [9]. 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*.

Most studies of spatial effects have been conducted in free-field conditions and have not included the effects of reverberation. Where reverberation has been included it has been shown to reduce the magnitude of the spatial release from masking [10]. This indicates, 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. Another study [11] 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 of conditions rather than on intelligibility test results, 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 is used loosely to include all types of reflected sound. Most often the interfering signal has been speech and not typical room noises such as that from ventilation systems [12,13].

Intent of the New Work

The new work was designed to explain the differences between the original laboratory study, that developed the SNR_{UNI32} measure [1], and a subsequent validation study for speech transmission between two moderately reverberant rooms with interfering ambient noise in the listening room [14]. The original laboratory study was in approximately free-field conditions and would have included a significant spatial release from masking because of the well-separated target speech and interfering noise source locations. The effects of reverberation on the two-room validation study exposed listeners to interfering noise in an approximately diffuse sound field as well as degrading intelligibility with the temporal effects of reverberation on the speech sounds. From previous tests with noise sources at multiple locations, we can expect a diffuse noise sound field would lead to a reduced spatial release from masking. We can also expect that the room reverberation would degrade the intelligibility of the speech sounds. However, it was not possible to predict the combined effects of the various differences. In fact, no previous study has considered how the combined spatial and temporal affects of room acoustics would affect speech privacy in rooms and no previous study has tried to separately estimate the relative importance of all of the components of the problem on the resulting speech intelligibility/privacy. This is clearly important because the validation test results would lead to lower and less costly sound insulation requirements in many situations.

The new work was intended to provide a quantitative overview of the spatial and temporal effects of room acoustics on the effective speech privacy of enclosed meeting rooms. Speech intelligibility tests were carried out in which each component that was expected to be significant was varied. Speech sounds were varied from a free-field direct sound only case, to cases with added early-arriving reflections and subsequently added reverberant speech sound with varied reverberation times. The added reflections of the speech sounds were added so that they were also spatially realistic. In a similar manner, the interfering noises were varied from a simple direct sound only (and from varied direction), to semi- and completely diffuse presentations. The noise had a spectrum representative of typical indoor ambient noise [12, 13]. Finally, in some cases the speech sounds were filtered to represent transmission through a wall. The goal was to quantify the most important effects on the intelligibility of the speech transmitted from an adjacent room.

2. Experimental Procedure

General

The importance of each spatial and temporal factor was determined by carrying out speech intelligibility tests in simulated sound fields. Three new experiments were conducted that included a total of 40 different configurations in which the key factors were systematically varied. Experiments #1 and #3 were in an anechoic room with an 8-channel electro-acoustic simulation system and experiment #2 was carried out in the Room Acoustics Test Space illustrated in Figure 1. All tests used the Harvard sentences [15] and responses were in terms of the fraction of the words in each sentence that were correctly identified by the subjects. In some analyses data from a previous experiment in the same anechoic room sound field simulation system were also used [16]. The various analyses in this report compare selected conditions from all 3 experiments. The details of each configuration of the 3 experiments are described in chapter 3 of this report.

Anechoic Room Sound field Simulation System

Figure 3 shows a photograph of the 8-channel sound field simulation system in the anechoic room in building M27 at NRC-IRC. As illustrated in the photograph, the 8 loudspeakers are positioned around the subject at various angles. Five of the loudspeakers are in the horizontal plane of the listener's ears and the other 3 are raised up above this plane. The actual angles of each loudspeaker from the listener are given in Table 1 and the numbering of the loudspeakers is described in Figure 4.

Figure 3. The 8-channel sound field simulation system showing 7 of the 8 loudspeakers and a subject.

The signals to each loudspeaker were processed by four Yamaha DME32 digital signal processing units connected together to form 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 8 loudspeakers were used conditions were perceived as very diffuse.

Loudspeaker		Horizontal angle, degrees	Vertical angle, degrees
1	Centre low	0	0
2	Left low	-32	0
3	Right low	+32	0
4	Centre high	0	25
5	Left high	-37	28
6	Right high	+37	28
7	Left rear	-115	0
8	Right rear	+115	0

Table 1. Horizontal and vertical angles of the loudspeakers relative to the listener. Angle 0,0 is straight ahead of the listener's head.

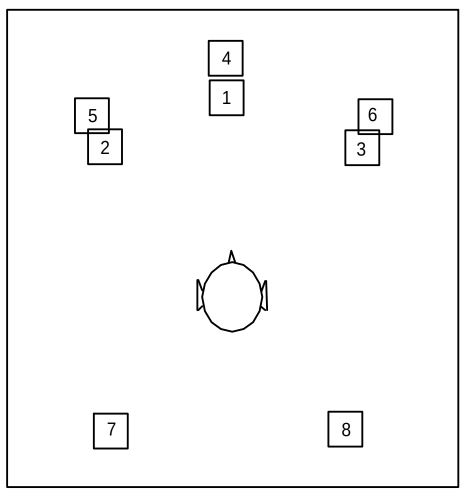


Figure 4. Numbering of the loudspeakers in the anechoic room simulation system. Loudspeakers 1, 2, 3, 7 and 8 were in the horizontal plane of the listener's ears. Loudspeakers 4, 5 and 6 were raised up above that plane as illustrated in the photograph of Figure 3 and the angles of each loudspeaker given in Table 1.

Because early-arriving sounds are perceived differently than later-arriving sounds, early- and later-arriving speech sounds were varied independently. Each loudspeaker could reproduce simulations of 4 early-arriving sounds for a total of 32 early-arriving sounds. The first early-arriving sound arrived from loudspeaker #1 (see Figure 4) 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 represent 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. Figure 5 shows an example of a measured impulse response for the simulation of a room with a 1 s reverberation time.

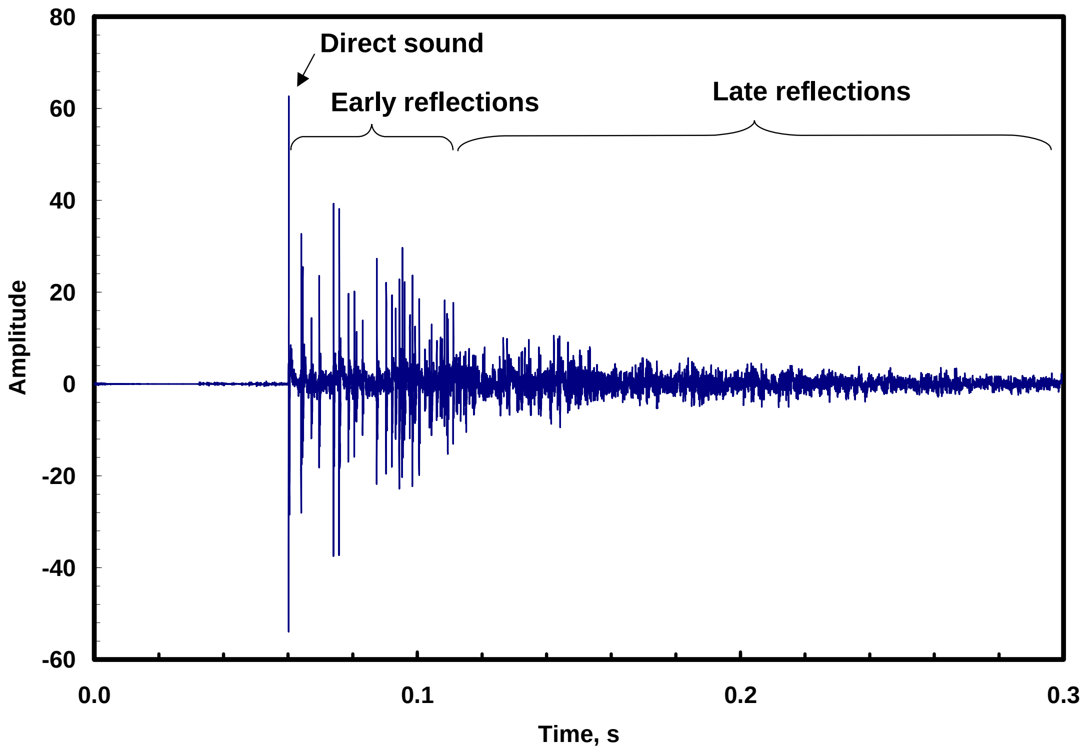


Figure 5. Example impulse response of a simulated sound field with all loudspeakers operating and for the simulation of a room impulse response with a 1 s reverberation time.

Room Acoustics Test Space

Experiment 2 was carried out in the Room Acoustics Test Space located in building M59 at NRC-IRC and illustrated in Figure 1. This is a sound isolated, quiet and acoustically dead space. The room is 9.2 m long, 4.7 m wide and 3.6 m high. The interior walls were covered with 10 cm thick sound absorbing foam behind curtains. There was a 25 mm thick glass fibre suspended ceiling and thin commercial carpet on the floor making the room acoustically very dead. The room is constructed of concrete and mounted on springs to ensure that it is well sound isolated from other parts of the building and with an ambient noise level of only 13 dBA without experimental sounds.

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 were 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.

Yamaha DME32 units were 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 12000 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 light-weight 90 mm steel studs and with glass fibre material in the cavity was simulated. This construction would correspond to a Sound Transmission Class rating of 46 and is typical of many interior office walls.

Simulated Ambient Noise

For all 3 experiments ambient noise with an approximately –5 dB per octave spectrum shape was used. This is often referred to as 'neutral' sounding and is representative of typical indoor noise spectra [12,13]. Although there were small differences in the spectrum shapes, illustrated in Figure 6, they all had an overall level of 45 dBA.

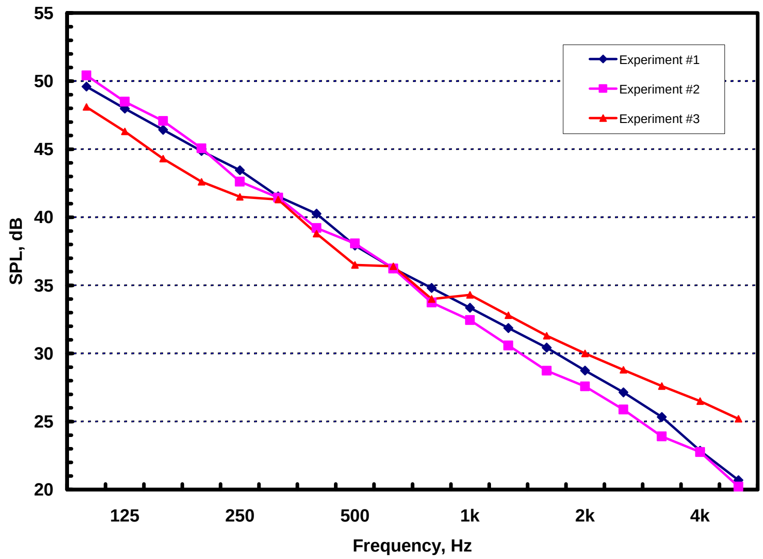


Figure 6. Measured simulated ambient noise spectra from each of the 3 experiments.

Speech Tests and Subjects

Subjects were employees of the National Research Council who volunteered to participate after being contacted by Email. They did not receive any payment or compensation for their help. The tests were approved by the Ethics Review Board of the National Research Council (Protocol #2006-06) and each subject signed a consent form after all of the details of the experiment were explained to them.

In the recruiting process, people were sought having good hearing and English as their first language. Nevertheless, some people responded who were fluent in English, but had a different first language. They were tested, but their results resulted in lower scores than other subjects and their results were not included in the data analyses. Some subjects participated in more than one test. However, each test included different sets of test sentences. The test sentences used were the Harvard sentences [15]. These are phonetically balanced and of low predictability as desired for speech testing conditions of low speech intelligibility.

All subjects were given a hearing threshold test before doing the listening test. Their hearing was judged to be acceptable when their pure tone average for the octave band frequencies from 500 to 4k Hz was less than 20 dB HL and there was no more than a 10 dB HL difference between ears [17]. The symmetric hearing ability was especially important when spatial effects were tested (e.g. when the noise source was moved to one side). Of the 46 subjects considered, the data from only one was rejected because the subject did not meet the hearing test criteria.

After an initial introduction to the experiment, followed by a hearing test, subjects were given detailed instructions on their task in the listening test. Subjects then listened to 10 practice sentences that covered the full range of conditions in the complete test, so that they could get used to the procedure. It was possible to ask questions during and after the practice session. When the experimenter was sure that the person knew what to do, the test would start.

During the actual test the experimenter was outside the test room and monitored the subject using a small video screen and listened to them via a microphone and loudspeaker system. The simulated ambient noise would start before each test sentence and continue a few seconds after the sentence was finished. The subject then had time to repeat back all of the words that they thought they understood, or they could say that they didn't understand any of the words. In all of the conditions, subjects could at least hear the speech sounds even if they did not understand all of the words.

People were asked to concentrate and listen carefully. They were encouraged to guess words, as they had little information from the context of the sentences. The total test time was about 40 minutes. In the middle of each test, people were asked, if they wanted to take a break. It was also possible to stop the test at any time, if necessary. Only one person could be tested at a time.

The words identified correctly were marked by the experimenter on a score sheet and then the percentage of words correctly understood was determined. Words were scored as correct according to the following rules:

- For nouns either the singular or the plural of a word was scored as correct, e.g. chicken and chickens.
- Verbs and adjectives were scored as long as the root was repeated correctly, e.g. rain, rains, raining, rainy.
- Small words like 'the' or 'a' were only scored, if they were repeated at the right position in the sentence.
- Homophones, if detected at all, were scored as correct, e.g. two, too and to.

To create a worst-case condition in terms of speech privacy, good listeners were required and only the 10 best subjects, based on their total percentage of words correctly understood over all conditions of the test, were used in the analyses.

3. Measurement Results

This section includes the measured results of the speech intelligibility tests and acoustical measurements of each configuration of the 3 experiments. Each experiment included the results of 10 subjects. Table 2 includes descriptions of the configurations 1 to 14 included in experiment #1. These included sound fields to demonstrate simple spatial unmasking for direct sounds in free-field conditions as the speech and noise sources were separated. The noise from All (8) loudspeakers made it possible to develop an understanding of the masking effects of a diffuse noise sound field. There were further configurations to illustrate the effects of adding first early reflections, and then reverberant sound to the speech signals. Conditions were repeated for two different signal-to-noise ratios, but no conditions in this first experiment included simulated transmission through a wall. All conditions in experiment 1 were created in the 8-channel sound field simulation system illustrated in Figures 3 and 4.

Config. #	SNRsii22, dB	SNRuni32, dB	Wall	Speech	Noise	S.I.	STDEV
1	-5.52	-6.93	no wall	Direct	Front	0.649	0.105
2	-5.86	-7.13	no wall	Direct	Front side	0.926	0.085
3	-5.22	-7.10	no wall	ER	Front side	0.889	0.067
4	-6.07	-7.77	no wall	T_{60}=1	Front side	0.723	0.070
5	-6.09	-7.39	no wall	Direct	All	0.814	0.098
6	-5.44	-7.36	no wall	ER	All	0.833	0.087
7	-6.30	-8.03	no wall	T_{60}=1	All	0.482	0.127
8	-8.46	-9.90	no wall	Direct	Front	0.451	0.178
9	-8.80	-10.10	no wall	Direct	Front side	0.896	0.115
10	-8.11	-10.01	no wall	ER	Front side	0.819	0.115
11	-8.85	-10.60	no wall	T_{60}=1	Front side	0.542	0.137
12	-9.03	-10.36	no wall	Direct	All	0.492	0.167
13	-8.33	-10.27	no wall	ER	All	0.511	0.133
14	-9.08	-10.86	no wall	T_{60}=1	All	0.178	0.100

Table 2. Description of the 14 configurations included in experiment #1, the measured signal-to-noise ratios (SNR_{SII22} and SNR_{UNI32}), mean speech intelligibility scores (S.I.) and their standard deviation (STDEV). Column 'Wall' indicates whether or not transmission through a wall was simulated. Column 'Speech' describes how the speech sound was processed: (a) 'Direct', only a direct sound, (b) 'ER', direct sound and early reflections, or (c) 'T_{60}=1', direct sound, early reflections and late arriving reflections with a 1 s reverberation time. Column 'Noise' describes from where the simulated noise was radiated: (a) 'Front', from loudspeaker #1 only, (b) 'Front side', from loudspeaker #3 only and (c) 'All', from all 8 loudspeakers.

Experiment #2 was carried out in the Room Acoustics Test space (see Figure 1) and was intended to make it possible to compare the spatial unmasking when speech and noise sources were separated in the horizontal plane as in experiment #1 with a vertical separation of speech and noise sources. This was important because the initial laboratory tests to develop the SNR_{SII22} and SNR_{UNI32} measures [1] included a vertical separation of speech and noise sources. Table 3 includes descriptions of the configurations 15 to 26 that comprised experiment #2. The various configurations to examine spatial unmasking effects were repeated with and without simulated transmission through a wall as well as for different signal-to-noise ratios.

Config. #	SNRsii22, dB	SNRuni32, dB	Wall	Speech	Noise	S.I.	STDEV
15	-6.98	-8.19	no wall	Direct	Ceiling	0.858	0.094
16	-6.96	-8.47	no wall	Direct	Front	0.584	0.132
17	-9.89	-10.70	no wall	Direct	Ceiling	0.651	0.134
18	-10.06	-11.32	no wall	Direct	Front	0.308	0.147
19	-7.03	-4.99	wall	Direct	Ceiling	0.959	0.043
20	-7.03	-4.99	wall	Direct	Front	0.758	0.104
21	-7.07	-5.01	wall	Direct	Ceiling	0.848	0.081
22	-10.02	-8.12	wall	Direct	Front	0.524	0.120
23	-10.14	-7.89	no wall	Direct	Ceiling	0.970	0.031
24	-3.92	-5.02	no wall	Direct	Front	0.856	0.097
25	-7.06	-8.17	no wall	Direct	Ceiling	0.932	0.047
26	-6.92	-7.93	no wall	Direct	Front	0.703	0.131

Table 3. Description of the configurations 15 to 26 included in experiment #2, the measured signal-to-noise ratios (SNR_{SII22} and SNR_{UNI32}), mean speech intelligibility scores (S.I.) and their standard deviation (STDEV). Column 'Wall' indicates whether or not transmission through a wall was simulated. Column 'Speech' shows that in all cases the speech sound included only a direct sound. Column 'Noise' describes from where the simulated noise was radiated: (a) 'Front', from in front of the listener, (b) 'Ceiling', from the loudspeakers in the ceiling above the subject.

Table 4 describes configurations 27 to 40 that made up experiment #3 which was carried out in simulated conditions in the anechoic room. Experiment #3 added configurations to explore the further effects of the transmission of speech through a wall and cases with semi-diffuse noise arriving from several nearby sources but not from All 8 loudspeakers. This semi-diffuse noise was intended to more realistically simulate conditions in some real rooms. Since many of the configurations in experiments #1 and #2 did not include simulation of speech transmission through a wall, it was important to include configurations with a wall to understand issues related to the speech privacy of enclosed meeting rooms.

Config. #	SNRsii22, dB	SNRuni32, dB	Wall	Speech	Noise	S.I.	STDEV
27	-7.15	-4.85	wall	ER	All	0.793	0.104
28	-7.03	-4.75	wall	T_{60}=0.5	All	0.886	0.062
29	-6.93	-5.07	wall	T_{60}=1	All	0.635	0.161
30	-6.99	-4.87	wall	2 s	All	0.381	0.181
31	-11.00	-9.47	wall	ER	All	0.486	0.097
32	-11.06	-9.27	wall	T_{60}=0.5	All	0.528	0.111
33	-11.06	-9.39	wall	T_{60}=1	All	0.147	0.048
34	-11.08	-8.91	wall	T_{60}=2	All	0.1	0.061
35	-8.39	-9.54	no wall	T_{60}=1	Rear side	0.833	0.08
36	-8.51	-9.75	no wall	T_{60}=1	Front side diffuse	0.472	0.171
37	-8.74	-9.91	no wall	T_{60}=1	Rear side diffuse	0.383	0.141
38	-6.87	-4.86	wall	T_{60}=1	Front side diffuse	0.66	0.131
39	-7.13	-5.06	wall	T_{60}=1	Rear side diffuse	0.697	0.102
40	-8.71	-10.06	no wall	ER	All	0.419	0.141

Table 4. Description of the configurations 27 to 40 included in experiment #3, the measured signal-to-noise ratios (SNR_{SII22} and SNR_{UNI32},) mean speech intelligibility scores (S.I.) and their standard deviation (STDEV). Column 'Wall' indicates whether or not transmission through a wall was simulated. Column 'Speech' describes how the speech sound was processed: (a) 'ER', direct sound and early reflections, or (b) 'T_{60}=0.5, direct sound, early reflections and late arriving reflections with a 0.5 s reverberation time, (c) 'T_{60}=1', direct sound, early reflections and late-arriving reflections with a 1 s reverberation time, or (d) 'T_{60}=2' direct sound, early reflections and late-arriving reflections with a 2 s reverberation time. Column 'Noise' describes from where the simulated noise was radiated: (a) 'All', from all 8 loudspeakers, (b) 'Rear side', from loudspeaker #8 only, (c) 'Front side diffuse', from #3, #1 and #6 reduced by 5 dB and all others reduced by 10 dB and (d) 'Rear side diffuse', from #8, #6 and #7 reduced by 5 dB and all others reduced by 10 dB.

Room acoustics measurements were made for configurations where the speech sound included reflected sound and not just the direct sound. The room acoustics measurements included reverberation times (T_{60}), early decay times (EDT) and early-to-late arriving sound ratios (C_{50}) [18]. Measured values of the Speech Transmission Index (STI) were also obtained for configurations that included simulations of reflected speech sounds and all configurations of experiment #2. These results are summarized for the 3 experiments in Tables 5-7.

Config. #	Speech	C50(500), dB	RT(500), s	EDT(500), s	STI
1	Direct	-	-	-	-
2	Direct	-	-	-	-
3	ER	13.76	-	-	0.25
4	T_{60}=1	1.02	1.00	0.92	0.21
5	Direct	-	-	-	-
6	ER	13.76	-	-	0.25
7	T_{60}=1	1.02	1.00	0.92	0.21
8	Direct	-	-	-	-
9	Direct	-	-	-	-
10	ER	13.76	-	-	0.16
11	T_{60}=1	1.02	1.00	0.92	0.13
12	Direct	-	-	-	-
13	ER	13.76	-	-	0.16
14	T_{60}=1	1.02	1.00	0.92	0.13

Table 5. Summary of room acoustics measurements for configurations of experiment #1, showing 500 Hz octave band values of T_{60}, EDT and C_{50} as well as STI values.

Config. #	Speech	C50(500), dB	RT(500), s	EDT(500), s	STI
15	Direct	-	-	-	0.19
16	Direct	-	-	-	0.19
17	Direct	-	-	-	0.19
18	Direct	-	-	-	0.12
19	Direct	-	-	-	0.37
20	Direct	-	-	-	0.34
21	Direct	-	-	-	0.27
22	Direct	-	-	-	0.29
23	Direct	-	-	-	0.30
24	Direct	-	-	-	0.29
25	Direct	-	-	-	0.20
26	Direct	-	-	-	0.24

Table 6. Summary of room acoustics measurements for configurations of experiment #2, showing STI values for each configuration.

Config. #	Speech	C50(500), dB	RT(500), s	EDT(500), s	STI
27	ER	9.72	-	-	0.32
28	T_{60}=0.5	4.73	0.55	0.74	0.31
29	T_{60}=1	0.13	1.07	1.23	0.24
30	T_{60}=2	-2.57	2.01	2.33	0.22
31	ER	9.72	-	-	0.22
32	T_{60}=0.5	4.73	0.55	0.74	0.20
33	T_{60}=1	-0.14	1.07	1.23	0.13
34	T_{60}=2	-2.57	2.01	2.33	0.13
35	T_{60}=1	0.13	1.03	1.09	0.13
36	T_{60}=1	0.13	1.03	1.09	0.13
37	T_{60}=1	0.13	1.03	1.09	0.13
38	T_{60}=1	0.13	1.03	1.09	0.24
39	T_{60}=1	0.13	1.03	1.09	0.24
40	ER	9.72	-	-	0.13

Table 7. Summary of room acoustics measurements for configurations of experiment #3, showing 500 Hz octave band values of T_{60}, EDT and C_{50} as well as STI values.

4. Calculation of Speech Reception Thresholds (SRT)

The Speech Reception Threshold is the value of the signal-to-noise ratio for which the mean speech intelligibility score is 50%. In this report, signal-to-noise ratios are in the form of uniformly weighted signal-to-noise ratio (SNR_{UNI32}) values. The various test results describe the mean speech intelligibility scores for the simulated conditions. Typically each condition was tested for two different signal-to-noise ratio values. That is, mean speech intelligibility scores varied due to changes in signal-to-noise ratio and also acoustical conditions of each test configuration. Converting each mean speech intelligibility score for each condition to a corresponding Speech Reception Threshold (SRT) makes it possible to average results of multiple measurements with varied signal-to-noise ratio but for the same acoustical condition to more reliably and more conveniently compare conditions.

SRT values are estimated by assuming a Boltzmann equation fit to the measured data. The Boltzmann equation is given by the following,

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

where: y is the speech intelligibility score
x is the corresponding signal-to-noise ratio
x_0 is the signal-to-noise ratio for an intelligibility score of 0.5
(i.e. x_0 is the Speech Reception Threshold)
dx is related to the slope of the mid portion of the curve
A_1 is the minimum y value = 0.0
A_2 is the maximum y value = 1.0

Re-arranging this we can solve for x_0, the SRT.

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

$$1 + e^{(x - x_0)/dx} = (A_1 - A_2)/(y - A_2)$$

$$e^{(x - x_0)/dx} = (A_1 - A_2)/(y - A_2) - 1$$

Taking natural logs of both sides,

$$(x - x_0)/dx = \ln\{(A_1 - A_2)/(y - A_2) - 1\}$$

$$x_0 = x - dx \cdot \ln\{(A_1 - A_2)/(y - A_2) - 1\} \qquad (2)$$

The value of dx is related to the type of speech intelligibility test. In previous tests using the same technique, the same speech test material and using the same signal-to-noise ratio measure (SNR_{UNI32}) [1], dx was found to have a value of 2.5259. This previously derived relationship is repeated in Figure 7 (labelled ‘JASA fit’) and dx = 2.5259 was used in all of the current analyses to determine SRT values.

Each mean intelligibility score (SI) was transformed into the corresponding SRT value using equation (2). Figure 7 illustrates the process for the speech intelligibility scores from configurations 27 and 31. These were part of experiment #3 and are described in Table 4. Both corresponded to the same test configuration except that configuration 27 had a higher signal-to-noise ratio than occurred in configuration 31. Figure 7 illustrates a Boltzmann equation fit with dx = 2. 5259 to each mean intelligibility score. The resulting two curves are very close together and having the same dx value, their central portions are parallel to the previously published curve. Figure 7 also shows a curve fitted to the average of the two SRT values that is intermediate to the curves through each data point.

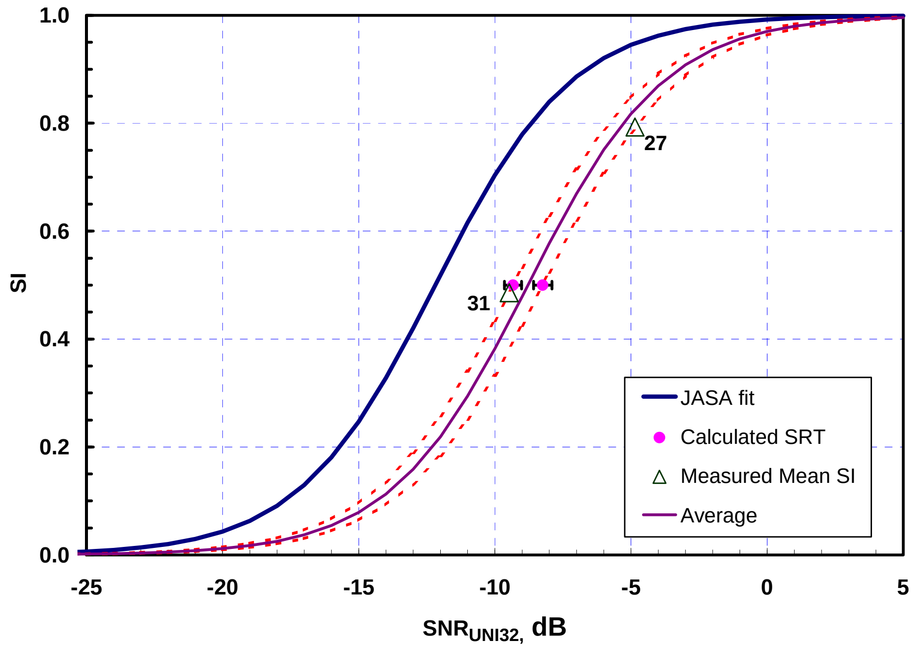


Figure 7. Expected trends of speech intelligibility versus SNR$_{UNI32}$ values based on a previously published relationship [1]. Dashed lines show the curves through each of the two mean intelligibility scores for configurations #27 and #31. The triangles indicate the mean intelligibility scores and the small dots are the calculated corresponding SRT values. The error bars for the SRT values were derived from the standard error of the mean speech intelligibility scores and the slope of the curve at the mid point.

Averaging two or more estimates of the SRT for a particular condition as illustrated in Figure 7 gives a more reliable estimate of the SRT value. Problems arise if one of the signal-to-noise values is too high or too low and the intelligibility scores approach close to 0 or 100%. Close to scores of 0 or 100%, intelligibility changes little for quite large changes in signal-to-noise ratio and hence large errors can occur when estimating the intelligibility score for a particular signal-to-noise value. To avoid these increased uncertainties, cases with intelligibility scores greater than 0.9 or less than 0.1 were excluded from the following analyses.

Table 8 lists the mean SRT values for the various conditions from all 3 experiments of the current work as well as one previous experiment and indicates which configurations were included in each average. The final column indicates the number of the experiment from which the data was obtained. Experiments 1 and 3 were in simulated sound fields in the anechoic room illustrated in Figures 3 and 4. Experiment 2 was carried out in the

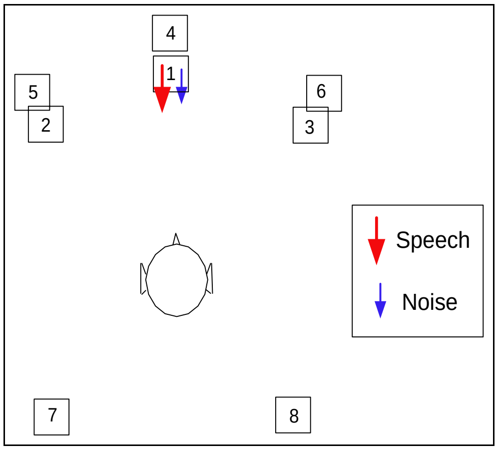
Room Acoustics Test space shown in Figure 1. The data labelled as from experiment 4 were from an earlier experiment [16] using the same procedures and the simulation system in the anechoic room. This table also includes descriptive information about each SRT case to remind the reader of the details of each configuration.

In estimating the significance of the differences between various mean SRT values, the SRT values for each subject's response were first calculated. Statistical tests were then performed on these individual SRT values to determine the significance of differences between the various sound field configurations. The results of these statistical tests are quoted in the following sections where the various SRT cases are compared.

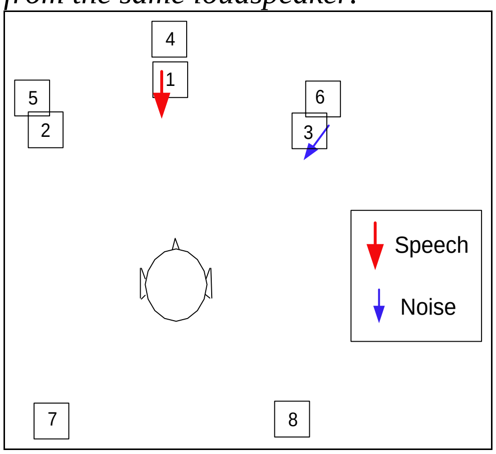
SRT Case	Configurations	Wall	Speech	Noise	SRT, dB	Exp
A	1, 8	no wall	Direct	Front	-9.37	1
B	2, 9	no wall	Direct	Front side	-14.75	1
C	3, 10	no wall	ER	Front side	-13.38	1
D	4, 11	no wall	T_{60}=1	Front side	-10.27	1
E	5, 12	no wall	Direct	All	-10.67	1
F	6, 13	no wall	ER	All	-10.93	1
G	7,14	no wall	T_{60}=1	all	-6.82	1
H	15, 17, 25	no wall	Direct	Ceiling	-13.26	2
I	16, 18, 24, 26	no wall	Direct	Front	-9.55	2
J	21	wall	Direct	Ceiling	-12.46	2
K	20, 22	wall	Direct	Front	-8.01	2
L	27, 31	wall	ER	All	-8.79	3
M	28, 32	wall	T_{60}=0.5	All	-9.74	3
N	29, 33	wall	T_{60}=1	All	-5.71	3
O	30, 34	wall	T_{60}=2	All	-3.50	3
P	64	no wall	ER	All	-9.16	4
Q	51	no wall	T_{60}=0.5	All	-8.83	4
R	55	no wall	T_{60}=1	All	-6.42	4
S	59	no wall	T_{60}=2	All	-3.81	4
T	35	no wall	T_{60}=1	Rear side	-13.60	3
U	36	no wall	T_{60}=1	Rear side diffuse	-9.47	3
V	37	no wall	T_{60}=1	Front side diffuse	-8.71	3
W	38	wall	T_{60}=1	Front side diffuse	-6.54	3
X	39	wall	T_{60}=1	Rear side difuse	-7.16	3

Table 8. Summary of the SRT values indicating which configurations were included in calculating each mean SRT value and giving brief descriptions of the configurations. The final column lists the experiment from which the data was obtained. Conditions 51, 55, 59 and 64 described as Experiment 4 are from a previous experiment [16].

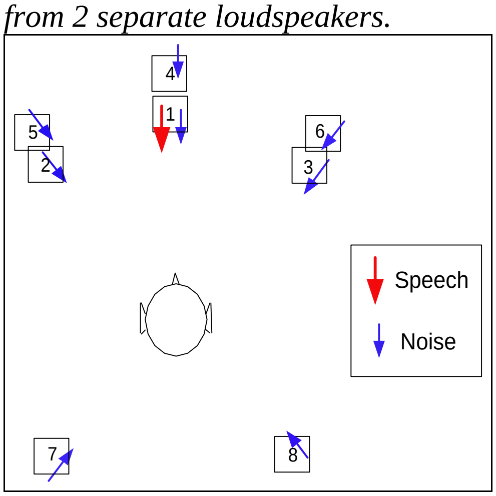
5. Simple Spatial Release from Masking



SRT case A, speech and noise only from the same loudspeaker.



SRT case B, speech and noise only from 2 separate loudspeakers.



SRT case E, speech from ahead only and noise from all loudspeakers.

Figure 8. SRT cases A, B and E.

This section presents the results illustrating the spatial release from masking as the speech and noise sources are spatially separated in free-field conditions (SRT cases A and B) and compares these results with a similar case but with a diffuse noise sound field with incoherent noise from all 8 loudspeakers (SRT case E).

The configurations of each case in the anechoic room sound field simulator are illustrated in the figures at the left. For Case A speech and noise sounds both arrived from only loudspeaker #1 directly in front of the listener.

For SRT case B the speech and noise sounds came from two different loudspeakers separated by an angle of 32 degrees in the horizontal plane as shown in Figure 8.

In SRT case E the speech sounds again came only from loudspeaker #1, which was directly in front of the listener. However, in this case the simulated ambient noise was radiated incoherently by all 8 loudspeakers.

The SRT values for these 3 cases are compared in Figure 9. Of the 3, 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 source 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.

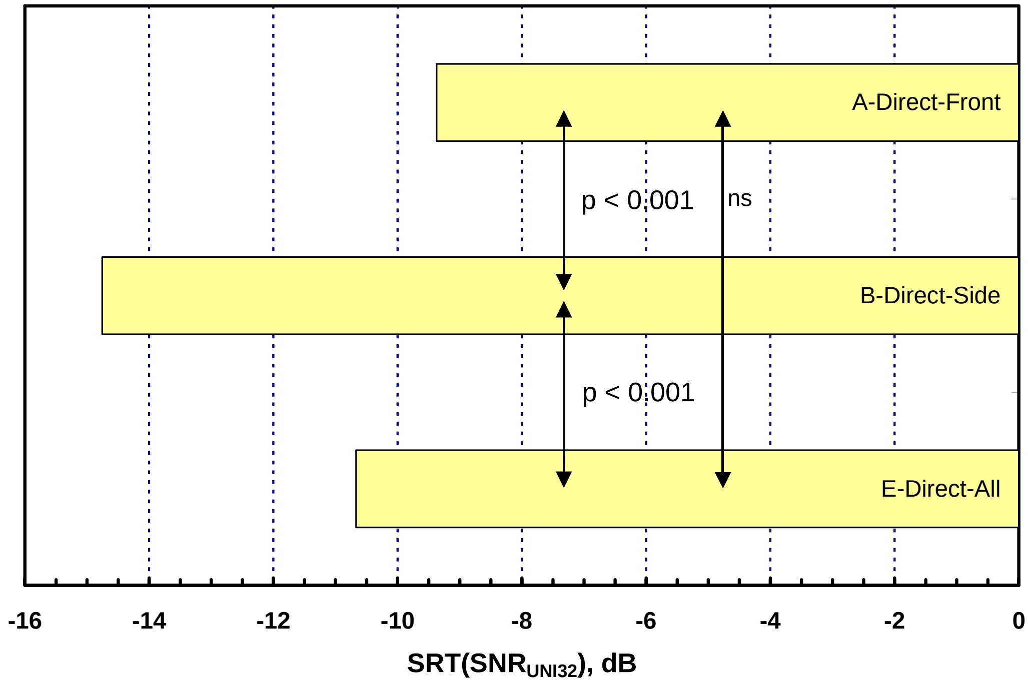
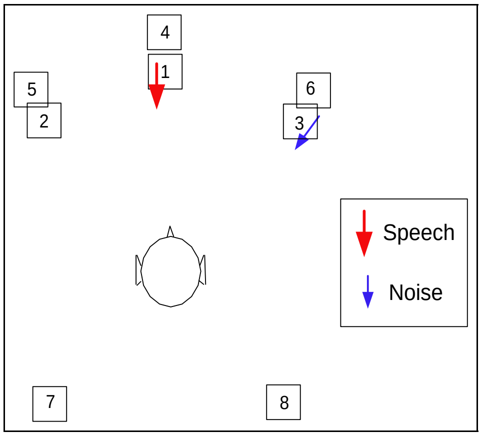


Figure 9. SRT values for Cases A, B and E with configurations illustrated in Figure 8. The arrows on this and the following graphs indicate the two cases for which the significance of the differences in SRT values is given (e.g. "p<0.001", probability of this difference occurring by chance less than 0.001; "ns", not a statistically significant difference).

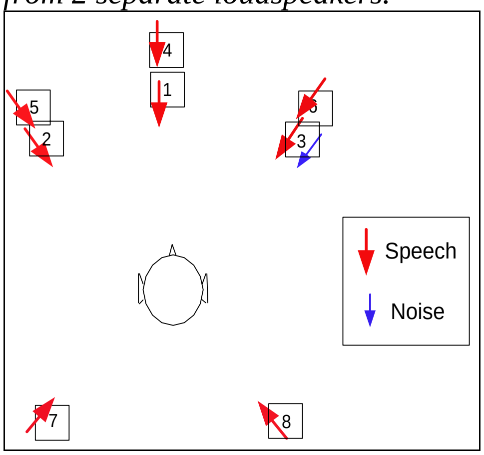
A Oneway analysis of variance for the experiment #1 results, that included these three conditions, indicated significant variations in SRT values ($F=25.15$, $p<0.001$). A post hoc Conferring test of the individual differences indicated that A and B were significantly different ($p<0.001$) as were B and E ($p<0.001$). However, the difference between the SRT values for Cases A and E was not statistically significant.

The comparisons suggests 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. The results also confirm the expected spatial release from masking when speech and noise sources are separated in free-field conditions. When the noise was separated by 32 degrees in the horizontal plane the SRT decreased by 5.4 dB.

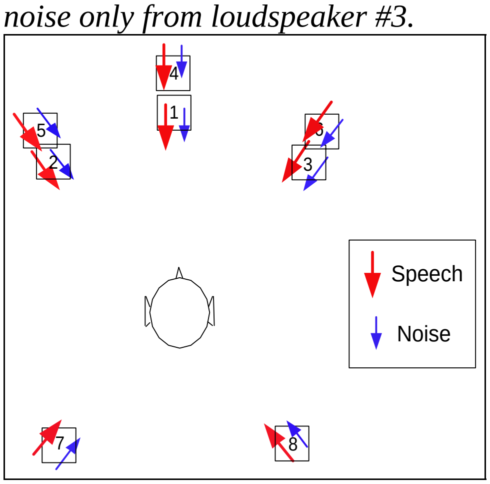
6. Effects of Early-Arriving Reflections of Speech



SRT case B, speech and noise only from 2 separate loudspeakers.



SRT case C, direct speech plus early reflections from all loudspeakers. noise only from loudspeaker #3.



SRT case F, speech direct plus early reflections from all and noise from all 8 loudspeakers.

Figure 10. SRT cases B, C and F.

In the previous section speech sounds arrived at the listener as only a single direct sound from one sound source location. In real rooms the direct sound is followed by many reflections of the speech. The initial reflections arriving within about 50 ms after the direct sound are known to be integrated by our hearing system and give us the impression of a stronger direct sound [19]. On the other hand, later-arriving reflections decrease the intelligibility of the speech by causing one word to be smeared over the following words. The comparisons in this section of the report examine the effect of adding early reflections to the direct speech sound.

The initial reference case considered is SRT case A in which the speech and noise arrive only as direct sounds from loudspeaker #1 as was described in Figure 8.

SRT case B is also included again in these comparisons as a base case for spatial unmasking effects. In SRT case B the direct speech and noise sounds came from sources horizontally separated by 32 degrees.

SRT case C was the same as SRT case B except that simulated early reflections from all 8 loudspeakers were added while maintaining the same overall speech level.

Finally SRT case F was the same as Case C except that in Case F the noise came incoherently from all 8 loudspeakers. Of course the overall levels and spectra of the noises did not change.

The SRT values for each case are compared in Figure 11. The repeated comparison of Cases A and B again shows the much lower SRT when the speech and noise sources are spatially separated. The yellow bars indicate the SRT values for the cases with added early-arriving reflections of the speech sounds. Case C with added early reflections is similar to Case B but Case F, with diffuse, noise has a much higher SRT value and is similar to Case E.

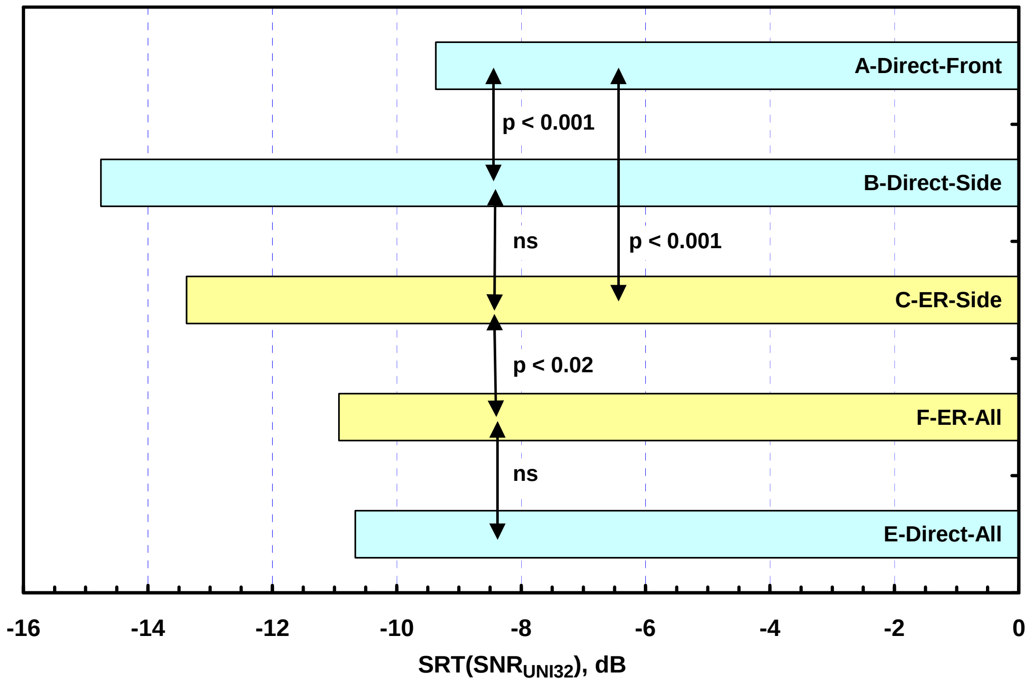


Figure 11. SRT values for Cases A, B C, E and F with configurations illustrated in Figures 8 and 10. (Blue shaded bars are repeated from Figure 9).

As previously mentioned a Oneway analysis of variance on the experiment #1 conditions showed a statistically significant pattern of variations in SRT values ($F=25.25$, $p<0.001$) and a post hoc Bonferroni test showed the SRT values for Cases A and B to be significantly different ($p<0.001$). Although the SRT value for Case C with added early reflections is a little higher than for Case B, the difference was not statistically significant. Case F, that added noise from all loudspeakers to Case C, was significantly different than Case C (Oneway $F=25.25$, $p<0.001$, post hoc Bonferroni $p<0.02$). However, comparing Case F with Case E shows that adding early reflections when there is noise from all directions produced only a very small change in SRT, which was not statistically significant.

Although not statistically significant, adding early reflections 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 early reflections included speech from the same direction as the noise, which might tend to increase the SRT a little for this case. When early reflections were added to configurations in which the noise came from all 8 loudspeakers, the added early speech reflections decreased the SRT slightly but not significantly (Cases E and F). In both cases adding early reflections of the speech energy is largely equivalent to increasing the level of the direct speech sound as was expected. Early reflections did not significantly affect spatial unmasking effects in these simulations.

7. Added Effects of Reverberant Speech

The previous section presented results to show the effect of adding early-arriving reflections to speech sounds. This section goes one step further and presents results showing the effects of adding reverberant speech sounds to the direct and early reflections of the speech. This was done for the cases of noise from only a side source (SRT case D) and noise from all directions (SRT case G). Figure 12 compares the mean SRT values for these cases (yellow bars) with those from Figure 11 in the previous section.

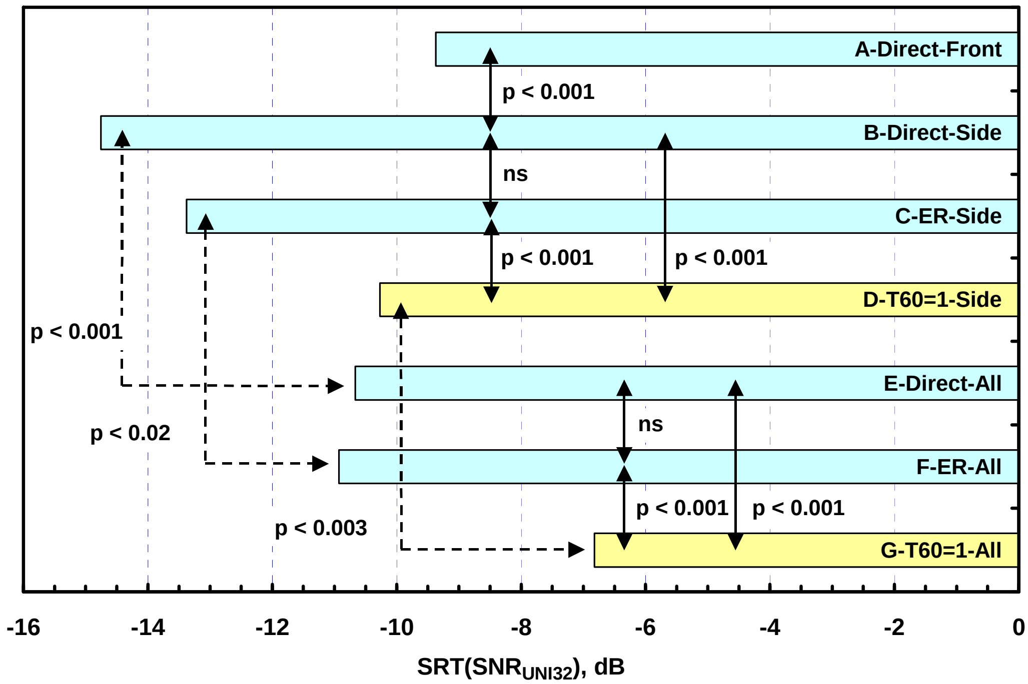


Figure 12. SRT values for Cases A, B C, D, E, F and G with configurations illustrated in Figures 8 and 10. (Blue shaded bars are repeated from previous figures).

Case D is the result of adding reverberant speech to Case C while maintaining the same overall speech level. Both cases included direct speech and early reflections of the speech from all loudspeakers as well as noise from only loudspeaker #3. The early reflections and reverberant speech were reproduced by all 8 loudspeakers. The conditions with added reverberant speech had a mid-frequency reverberation time (T_{60}) of 1 s. When reverberant speech was added to Case C the SRT was changed from -13.78 dB to -10.27 dB and this difference was statistically significant (Oneway F=25.25, p<0.001, post hoc Bonferroni p<0.001).

Case G is the result of adding reverberant speech to Case F while maintaining the same overall speech level. Both cases included direct speech and early reflections of the speech from all loudspeakers as well as noise from all 8 loudspeakers. The early reflections and reverberant speech were reproduced by all 8 loudspeakers. Adding reverberant speech 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 early reflections of the speech did not significantly change the SRT, adding reverberant speech increased the SRT by 3 to 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.

8. Effects of Varied Reverberation Time for Speech

The effect of reverberant speech was further investigated by comparing conditions with reverberation times of 0.5, 1.0 and 2.0 s. These comparisons were repeated with and without simulated speech propagation through a wall.

SRT case P corresponded to no wall and speech with added early reflections. Speech and noise sounds were reproduced by all 8 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 Figure 13.

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 Figure 13. All of these cases included simulated transmission through a wall for the speech sounds.

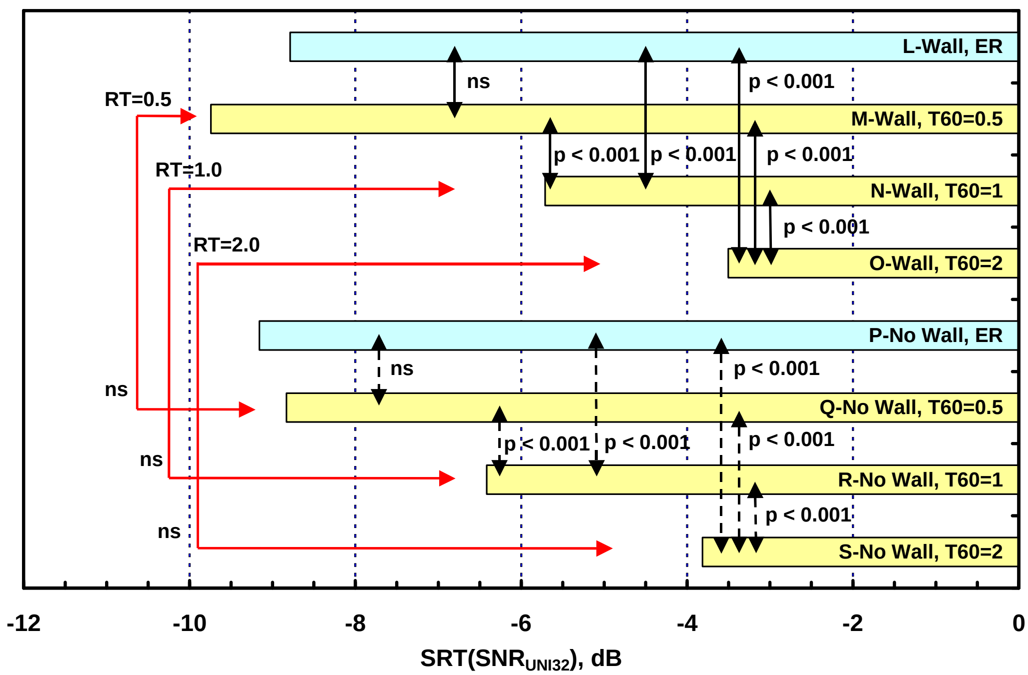


Figure 13. SRT values for Cases L, M, N, O, P, Q, R and S, all with speech and noise sounds from all 8 loudspeakers. (Yellow bars show cases with varied reverberation time).

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 early reflections 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 significant increases in SRT values (Oneway on the experiment #4 data: F=94.15, p<0.001, post hoc Bonferroni, p<0.001).

When a simulated wall was included, the results of Cases L, M, N and O showed a similar progression of changes to 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 #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 red arrows on Figure 13).

In order to better understand the changes to SRT values caused by increasing the reverberant speech sound, the mean SRT values are plotted versus the logarithm of the reverberation time in Figure 14. The data for the cases with only early reflections (ER) are included with a reverberation time of 0.5 s because these 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 reverberation time was increased. The results for the with and without a simulated wall cases lead to very similar results. The two regression equations and an average for all data are:

$$SRT = 8.602\ Log_{10}(T_{60})\ -6.407 \text{ (no wall)} \quad (3)$$

$$SRT = 9.774\ Log_{10}(T_{60}) - 6.201 \text{ (with-wall)} \quad (4)$$

$$SRT = 9.187\ Log_{10}(T_{60}) - 6.304 \text{ (All data)} \quad (5)$$

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

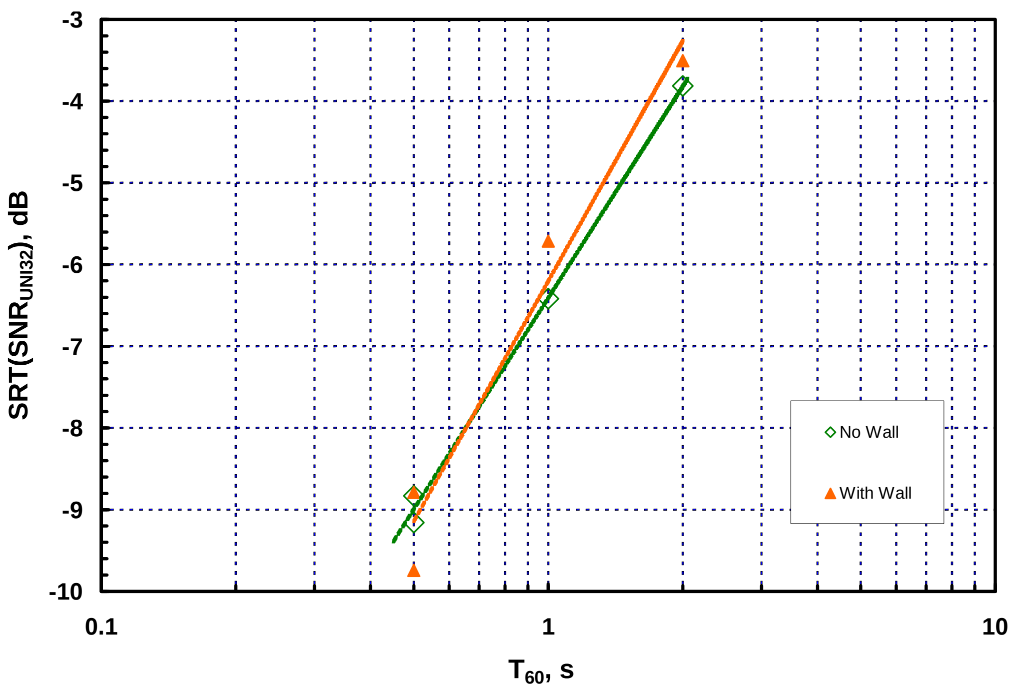


Figure 14. Mean SRT values plotted versus the logarithm of the reverberation time of the simulated speech sounds.

9. Effects of Single, Diffuse and Semi-Diffuse Noise Sources

The change of the masking effects of the ambient noise from a single noise source that was spatially separated from the speech source, to a diffuse noise from all 8 loudspeakers represents two extremes. It is possible that in real rooms intermediate cases could be 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 3 nearby 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 8 loudspeakers.

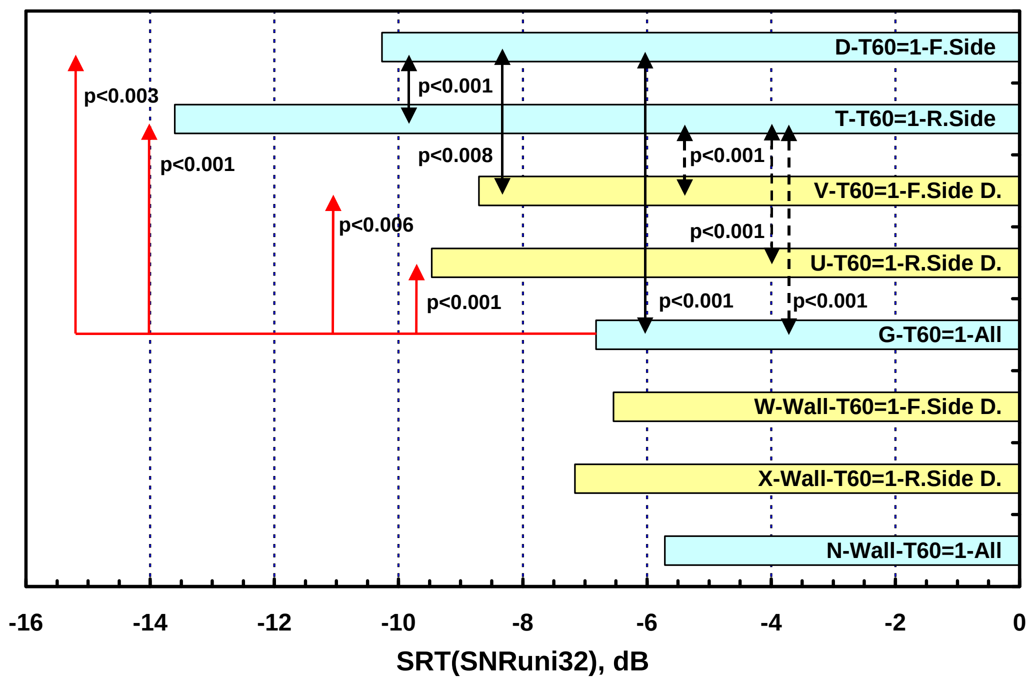


Figure 15. SRT values for Cases D, G, N, T, U, V, W and X with varied masking noise configurations. . (Yellow shaded bars are 'semi-diffuse' cases).

Figure 15 compares SRT values for varied noise masking configurations. The upper 5 configurations (D, T, V, U and G) on the graph contain results for conditions that did not include a simulated wall. Conditions, which included a simulated wall, are in the lower 3 bars of the graph (W, X, N). The speech signal in all conditions included direct sound, early reflections and reverberant sound 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 leads to a larger spatial release from masking (i.e. lower SRT) than for a single noise source from the front side 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 noise from all 8 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 noise source conditions (Cases D and T) and the all 8 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 Figure 15 all of the differences tested were statistically significant. That is, all of the changes in either the direction of the noise, or diffuseness of the noise led to a significant effect.

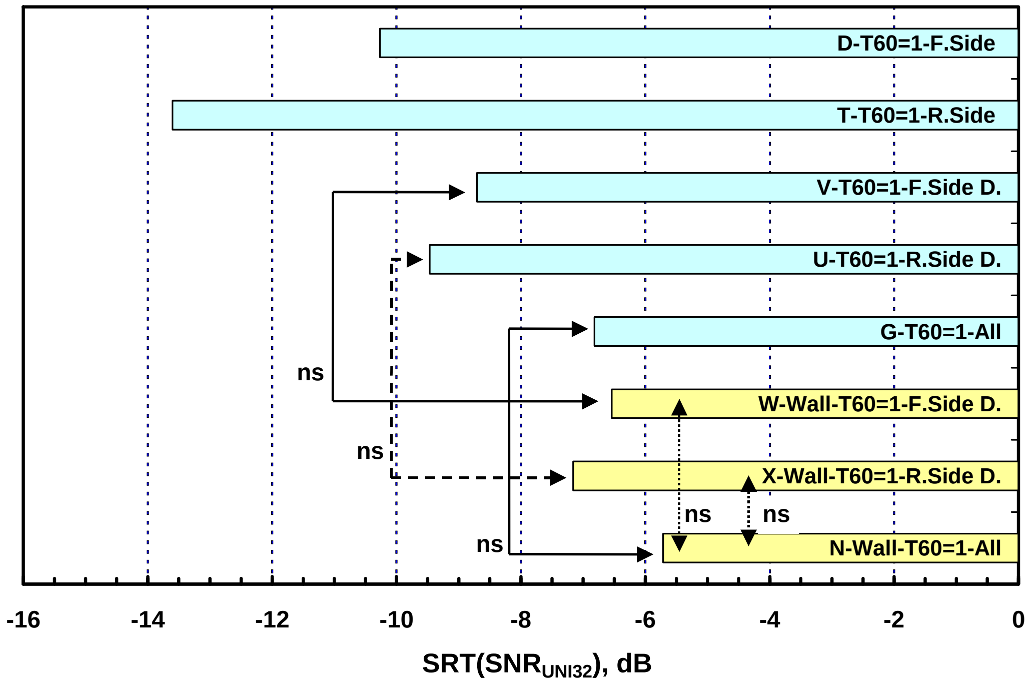


Figure 16. Comparison of effects of masking noise for cases with and without simulated walls. (Yellow shaded bars; data for simulated walls, blue shaded bars: natural speech, no walls).

The lower 3 bars of Figure 15 (configurations W, X, N) include results for similar conditions except that the speech sounds were filtered to simulate transmission through a wall. These show a little higher SRT values than the corresponding cases without walls. For clarity the results of the further statistical tests of these conditions are shown in Figure 16, which repeats the same bar graph of SRT values from Figure 15 but with different statistical results illustrated. 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 8 loudspeaker radiating noise (Cases G and N) were not significantly different. The lack of significance of these differences may be influenced by a lack of data in that several of these SRT cases were based on the measurements of only one condition and were not an average of 2 or more conditions.

From these results one must conclude that there is no proof of an effect of transmission though a wall. For these ‘with-wall’ results the differences between the semi-diffuse conditions (Cases W and X) were not significantly different than the noise from all loudspeakers case (Case N), although the corresponding cases without a simulated wall were significantly different (Cases U and V compared to Case G).

10. Explaining Differences between Initial Laboratory Study and Two-Room Validation Study

The initial tests that developed the SNR_{UNI32} measure [1] were conducted in approximately free-field conditions. These tests included direct speech from only in front of the listener and direct noise from only above the listener. The two-room validation tests included more or less diffuse noise from all directions. The two-room test results were also influenced by the reverberation of the source room (T_{60}=0.8 s, based on averaging over frequency from 160 to 5k Hz), and the reverberation of the receiving space (T_{60}=0.64 s). There were also differences in the presence or absence of early reflections of the speech sounds but these have been shown in the present experiments to not significantly modify the intelligibility of the speech.

We would like to estimate the total expected effect of the differences between these two experiments due to the different directions of arrival of the speech and noise in each case as well as the differences in reverberant speech sound. The results of experiment #2 made it possible to relate the spatial unmasking for a single noise source, separated vertically from the speech source, to the results for coincident speech and noise sources. The results of experiment #1 related the results of coincident speech and noise sources to a case with diffuse noise. Combining these two sets of results with the effects of reverberant speech sound from Figure 14 makes it possible to estimate the total difference in the results between the initial free-field results from reference [1] and the two-room validation test results [14].

Experiment #2 included measurements of SRT values for the case of coincident speech and noise sources directly in front of the listener and vertically separated speech and noise sources. The conditions were repeated for speech with and without the inclusion of simulated transmission through a wall and the resulting SRT values are plotted in Figure 17. As expected, spatially separating the speech and noise sources increased the spatial release from masking for both the with-wall and without-wall cases and spatially separated speech and noise sources had significantly lower SRT values (Oneway, $F=20.30$, $p<0.001$, post hoc Bonferroni $p<0.001$). However, there were no significant differences between the corresponding with-wall and without-wall cases (i.e. the results of Cases K and I are not different and those from Cases J and H are not different). As in the previous section there is no indication that the inclusion of a simulated wall had a significant effect on the resulting SRT values.

As indicated in Figure 17, the shift in SRT for the with-wall results, when the noise source was separated from the speech source was 4.45 dB. Cases I and K from experiment #2 can be assumed to be equivalent to Case A in experiment #1. In all 3 cases speech and noise arrived from the same single source that was directly in front of the listener. We can add onto the 4.45 dB difference, the effect of diffuse noise from the difference in SRT values between Cases A and E in experiment #1 (i.e. a difference of $-10.67-(-9.37)=-1.30$ dB). Thus, the total effect of changing from a single overhead noise source to a diffuse noise would be 4.45 dB reduced by 1.30 dB or a total SRT change of 3.15 dB.

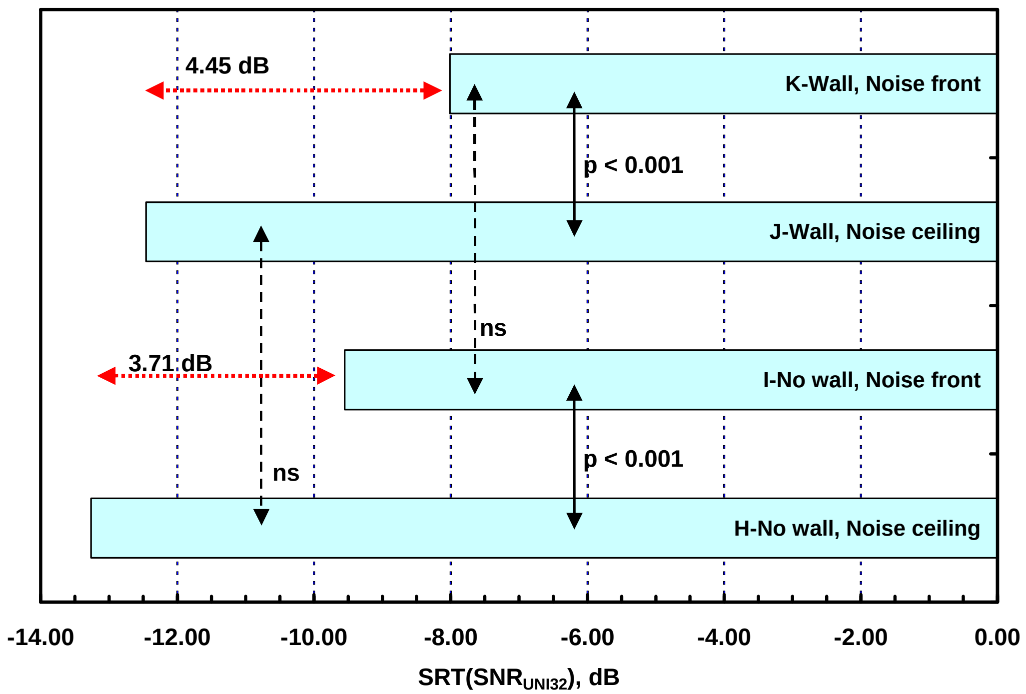


Figure 17. SRT values of the Cases K, J, I and H from experiment #2 showing the effects of vertical separation of the speech and noise sources for cases both with and without a simulated wall. The horizontal red arrows indicate the spatial release from masking for the with-wall (Cases K and J) and without-wall cases (Cases I and H).

The final step is to add on a correction for the effect of reverberation in the rooms during the two-room test. The source room had an average reverberation time of 0.8 s and the receiving room had an average reverberation time of 0.64 s, both averaged over the frequencies from 160 to 5k Hz, These would both reduce the intelligibility of the speech heard by the listeners in the adjacent room. The combined reverberation time of two series resonant systems is similar to two series resonant circuits and can be calculated as the square root of the sum of the squares of the individual reverberation times [20]. This gives a combined effective reverberation time of 1.02 s. The effects of reverberation on SRT can be estimated using a regression line derived from the data in Figure 14. As there was a wall in the previous two-room validation test, regression equation (4) from the with-wall data was used. Because the results indicated that a reverberation time of 0.5 s led to SRT values that were not significantly different than those for a configuration without any reverberant sound, the required correction was determined from the expected change in SRT from a base case of a reverberation time of 0.5 s and the combined effects of the rooms corresponding to a 1.02 s reverberation time. Using equation (5) from Section 8 of this report, the additional effect equivalent to a 1.02 s reverberation time was calculated to be a 3.03 dB increase in SRT.

The total change in SRT between the initial test, with noise from only overhead, and the two-room validation test is therefore expected to be 3.15+3.03 = 6.18 dB. The steps of these calculations are summarised in Table 9.

Explanation	Delta SRT, dB
Difference between coincident noise source and vertically separated noise source.	4.45
Difference between single noise source coincident with speech source and diffuse noise from all 8 loudspeakers.	-1.30
Effect of combined reverberation of both rooms.	3.03
Total change	6.18

Table 9. Summary of conversion calculations from initial laboratory test with vertically separated speech and noise [1] to validation test with reverberation and diffuse noise [14].

Figure 18 compares plots of speech intelligibility scores versus SNR_{UNI32} values from the initial lab tests and the two-room validation tests. In the initial lab study [1] an intelligibility score of 0.5 occurred for an SNR_{UNI32} value of -12.19 dB as indicated on Figure 18. When a parallel best-fit line was fitted to the data from the two-room validation test, an intelligibility score of 0.5 occurred for an SNR_{UNI32} value of -4.65 dB. The difference between the two estimates of the SNR_{UNI32} value for which a mean score of 0.5 occurred (i.e. the SRT values) is (-4.65-(-12.19)) or 7.54 dB.

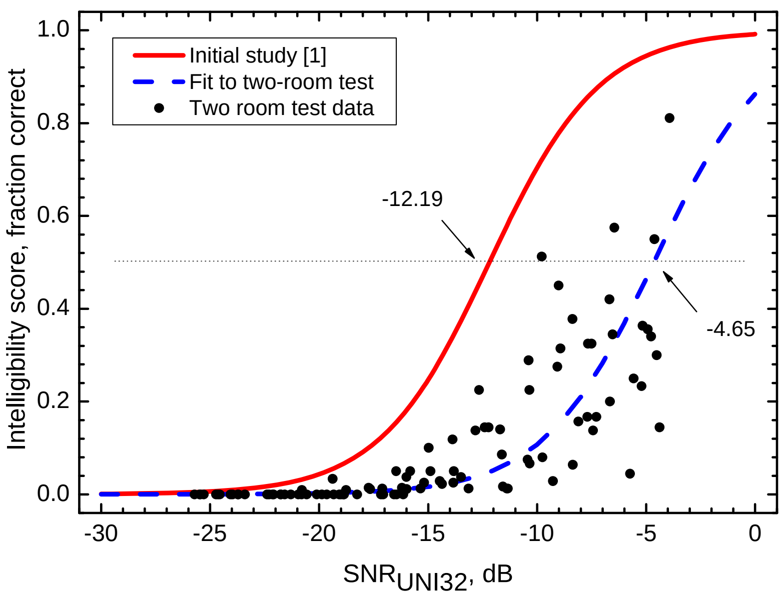


Figure 18. Intelligibility scores versus SNR_{UNI32} values. Solid line is from the initial lab test [1] and the dashed line is a fit to the two-room validation test data points shown.

Given the scatter in the data of Figure 18, the predicted change in speech intelligibility scores in the current work of 6.18 dB is reasonably close to the observed difference of 7.54 dB. It is possible that if cases with higher intelligibility scores had been included in the two-room validation study, the best-fit line would have indicated a little smaller shift relative to the initial study results. Estimates of the observed shift in the threshold of intelligibility are compared in the next section of this report.

As a final consideration of the effects of spatial separation of speech and noise sources, further analyses were included to determine if the spatial release from masking due to the horizontal separation of speech and noise sources between Cases A and B of experiment #1 was different than the spatial release from masking due to a vertical separation of the speech and noise sources in Cases H and J of experiment #2. Because the with-wall and without-wall cases in experiment #2 were not significantly different, their results were averaged and compared to the experiment #1 results. The cases with both speech and noise from directly in front of the subject were Cases A in experiment #1 and cases I and K in experiment #2. Neither the case I nor the case K results were significantly different than the case A results (Independent T-test) and the average SRT difference was -0.59 dB. The SRT values for the two cases with the noise source vertically separated (Case H and J) were both significantly different than those for Case B (Independent T-test, $p < 0.02$) where the noise source was horizontally separated from the speech source. The average SRT difference between the vertically separated and horizontally separated configurations was -1.89 dB. Since the SRT for Case B was 1.89 dB lower than the average SRT for cases H and J, the horizontal separation in this case led to a little larger spatial release from masking. In these cases the vertical separation was 90 degrees and the horizontal separation only 32 degrees. Other angles would be expected to lead to different results.

11. Changes to Speech Intelligibility Threshold Criteria

This report has presented results in terms of speech intelligibility scores and specifically Speech Reception Threshold (SRT) values corresponding to the signal-to-noise ratio for which the mean speech intelligibility score is 50%. However, previously established criteria for acceptable speech privacy have been in terms of the Speech Intelligibility Threshold. The Speech Intelligibility Threshold (SIT) is the signal-to-noise ratio for which 50% of a panel of listeners can just understand at least 1 word of the test sentence. For free-field conditions, the threshold of intelligibility was found to correspond to an SNR_{UNI32} value of -16 dB [1]. However, in the two-room validation tests [14], the threshold of intelligibility was found to be increased by 4.9 dB to an SNR_{UNI32} value of approximately -11 dB.

Figure 19 compares these two previous results. The solid red regression line is the best fit regression line to the data in the initial experiment that developed the SNR_{UNI32} value [1]. This was used to determine the threshold of intelligibility of speech as an SNR_{UNI32} value of -15.64 dB as shown on this figure. The data points and the dashed best-fit line are from the two-room validation study [14]. The dashed best-fit line was obtained by using the same Boltzmann equation as in the line from the initial study [1] and shifting it horizontally to minimize the RMS deviation of the data about the line. This resulted in an estimate of the threshold of intelligibility of an SNR_{UNI32} value of –10.70 dB. The difference between the two estimates of the threshold of intelligibility of speech (-10.70-(-15.64)) is 4.94 dB.

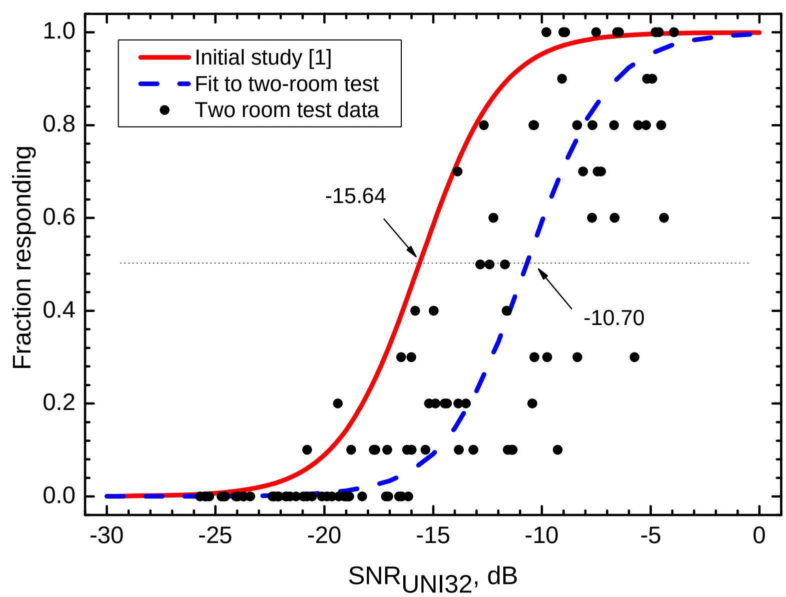


Figure 19. Fraction of listeners who can understand at least one word versus SNR_{UNI32} values. Solid line is from the initial lab test [1] and the dashed line is a fit to the two-room validation test data points shown.

To determine threshold of intelligibility values requires data for conditions with a significant number of responses with low intelligibility scores and values for the fraction of subjects understanding at least one word extending down to close to zero. In most of the new tests discussed in this report such conditions were deliberately avoided to provide data mostly in the range of 10% to 90% intelligibility scores. However, data from a previous study [16], which was included in Table 8 as experiment 4, did include a significant number of low intelligibility cases from which speech intelligibility thresholds can be determined. These data were for natural speech, (without simulated transmission through a wall), with varied reverberation times and signal-to-noise ratios and with ambient noise coming from all 8 loudspeakers. These data were used to determine new estimates of the effects of reverberation and diffuse noise on the speech intelligibility threshold criteria.

Values of the threshold of intelligibility were calculated in a manner similar to the calculation of SRT values described in Section 4 of this report. The fraction of the listeners indicating at least one word was understood for each test configuration were 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 speech intelligibility thresholds [1] corresponding to a value of 1.8739. Speech Intelligibility Threshold values were calculated using equation (2) in Section 4 of this report. In addition to using dx = 1.8739, y was the fraction of subjects understanding at least one word and x_0 is the threshold of intelligibility. As before, x was the SNR_{UNI32} value corresponding to the y value. The resulting Speech Intelligibility Threshold (SIT) values are given in Table 10. The other information in Table 10 is repeated from the description of configurations in Table 8.

SRT Case	Configurations	Wall	Speech	Noise	SIT, dB	Exp
P	64	no wall	ER	All	-12.62	4
Q	51	no wall	T_{60}=0.5	All	-11.89	4
R	55	no wall	T_{60}=1	All	-10.35	4
S	59	no wall	T_{60}=2	All	-7.61	4

Table 10. Speech Intelligibility Thresholds (SIT) in terms of SNR_{UNI32} values in dB.

The calculated Speech Intelligibility Thresholds are plotted versus T_{60} values, on a logarithmic scale, in Figure 20. The Speech Intelligibility Threshold values are seen to be approximately linearly related to the logarithm of the reverberation time similar to the plot of SRT values versus T_{60} in Figure 14. As in Figure 14, the case with only early reflections (ER) was plotted as having a T_{60} value of 0.5 s. From Figure 20 it is seen that the early-reflections-only condition could be interpreted as corresponding to a slightly lower T_{60} value.

Figure 20 suggests that for a typical meeting room, the Speech Intelligibility Threshold would be approximately an SNR_{UNI32} value of -11 dB. On Figure 20 this corresponds to a T_{60} value of 0.75 s. A little lower or higher values are indicated for lower or higher reverberation times. It is therefore proposed that the criterion for the speech intelligibility threshold be –16 dB for free-field conditions with spatially separated speech and noise sources [1] and –11 dB 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 (see Figure 19). Small adjustments for differences in reverberation times could be made if needed but are usually not justified.

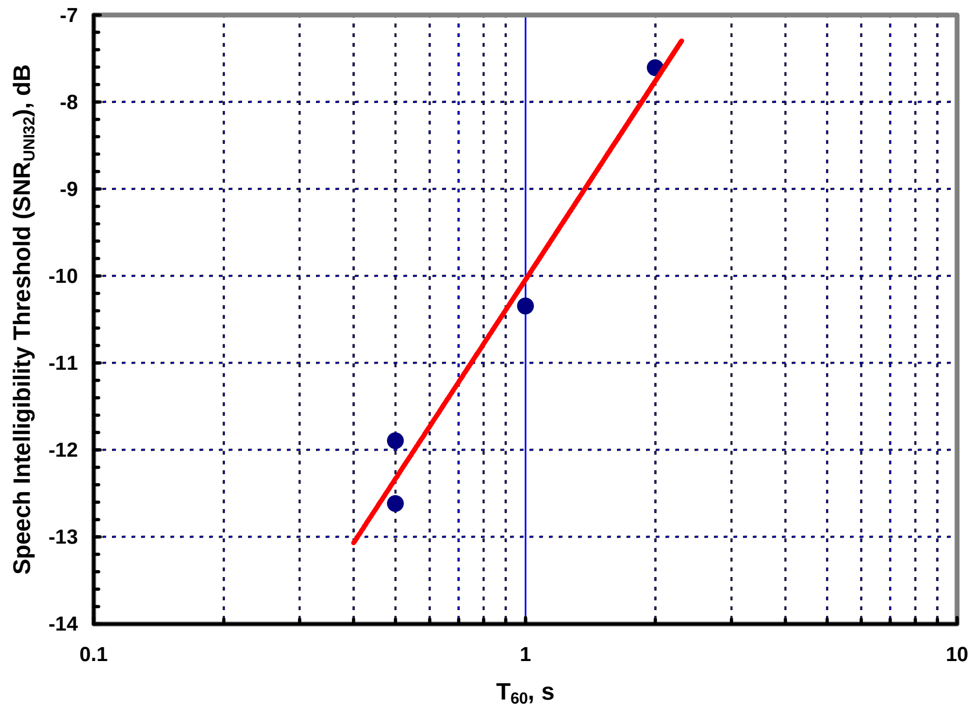


Figure 20. Speech Intelligibility Threshold versus T_{60} for unmodified speech (i.e. no simulated wall) and diffuse ambient noise.

12. Discussion

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 [9, 11]. Kollmeier and Wesselkamp [21] have shown that the results of these two approaches are correlated but lead to differences in the magnitude of the effects and the trends with varying signal-to-noise ratio are also different. A number of studies have used such subjective ratings in an iterative procedure to determine SRT values. In their tests, the subjects heard the same speech material repeatedly and decided when it appeared to be just intelligible. This is quite different than listening to each speech sample only once as in the current work and rating the fraction of words correctly understood.

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 [10]. In previous studies interfering sounds have most frequently been speech and much of the work was focussed 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 but none have used noise representative of typical indoor ambient noises.

The following discussion of the new results includes mention of previous studies where they are at least somewhat comparable.

The new 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 degrees 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 [8, 9, 11] and helps to confirm the validity of the procedures of the new tests in this report. A ninety-degree 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 32-degree 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 8 loudspeakers as in case E of the present work. This resulted in an 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 [9]. Reference 9 quotes work by vom Hövel in a German thesis [22] that apparently found that the spatial release from masking was never more than 3 dB for such diffuse noise conditions but did not provide any details. The 1.3 dB effect in the present work is about half of the reported maximum effect attributed to vom Hövel and confirms the much diminished spatial release from masking for conditions with diffuse noise.

Adding early reflections to the speech sounds, while maintaining constant speech levels, did not significantly change measured SRT values either with a horizontally separated noise or with noise from all loudspeakers. This extends our understanding of the beneficial effects of early-arriving reflections on the intelligibility of speech [19] 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.

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 early reflections and reverberation to the speech was accomplished while maintaining a constant overall speech level at the position of the listener. This would underestimate the benefits of reflected sound in a real room where adding reflected sound would increase the overall level of the speech and lead a to beneficial increase in signal-to-noise ratio.

The spatial release from masking with a single noise source from the rear side (Case T) was larger than for the case of a single noise source from the front side (Case D) where both cases included reverberant speech with a 1 s reverberation time. The spatial release from masking when the noise came from a single noise source separated by 90 degrees vertically from the speech source was similar to that for the front side noise source even though the angular separations were quite different. Previous investigations of the effects of variations with angular separation of speech and noise sources have concentrated on the horizontal plane of the listener’s ears. Plomp [11] found a maximum reduction in the masked threshold for a 135 degree angle from in front of the listener and showed that the variation with the angular separation decreased with increasing reverberation time. Peissig and Kollmeier [9] found the threshold most reduced at approximately a 105 degree angle from in front of the listener. That is, the new results approximately follow the trends of previous work that suggest increased spatial release from masking when noise sources are located to the rear side of the listener. These effects are believed to be due to head-shadow effects at the ear not directly exposed to the test sounds and to the directional properties of human hearing.

When ‘semi-diffuse’ noise was created by radiating incoherent noise predominantly from 3 nearby loudspeakers, SRT values were increased several decibels relative to a single horizontally separated source. Diffuse noise from the rear side (Case U) led to a little lower SRT than did diffuse noise from the front side (Case V). However, for these ‘semi-diffuse' conditions the SRT values were never increased as much as for the case of completely diffuse noise (Case N).

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. As illustrated in Appendix I, this is only true when results are considered in terms of uniformly weighted signal-to-noise ratio (SNR_{UNI32}) values.

The reasons why the with- and without-wall results were not significantly different were not specifically investigated. However, one can speculate that the lack of significant differences is due to the transmission response of the wall being approximately the inverse of the frequency response of human hearing (as described by equal loudness curves). Thus transmission through a wall may essentially "correct" human hearing to have an approximately flat response so that all frequencies in these measurements contribute approximately equally to the perceived loudness of the sounds. It would then be appropriate to weight all frequencies equally as occurs in the SNR_{UNI32} measure.

The main purpose of the new experiments was to better understand the causes of the differences between the initial laboratory study to determine the best signal-to-noise type measure for rating speech privacy of meeting rooms [1] and a subsequent validation test between two adjacent rooms [14]. Two of the various parameters investigated in the current work were found to explain these differences. These were the different spatial relationships between the speech and noise sources in the two studies and the temporal effect of room reverberation in the validation study.

The initial study to evaluate signal-to-noise type measures was intended to consider worst-case conditions for speech privacy in which listeners could most easily eavesdrop and understand speech sounds from an adjacent meeting room. Subjects listened to speech sounds from directly ahead and simulated ventilation noise from overhead in approximately free-field conditions (see Figure 1). The current results confirmed that there was a significant spatial release from masking for this condition and made it possible to estimate the magnitude of the difference between it and a condition with approximately diffuse ambient noise arriving from all around the listener.

It was also possible to determine the magnitude of the effect of various amounts of reverberation on the intelligibility of the speech. The combination of these two effects accurately explained the difference between the two previous experiments. Attempts to accurately predict the intelligibility of speech from an adjacent meeting room must account for the effect of reverberant sound in the meeting room and the spatial relationship between the speech and noise sources at the listening positions. In most cases both the speech and noise sounds will be at least moderately diffuse and the procedure used to explain the difference between the two previous experiments can be used directly. If there are more unusual conditions with very little reflected sound these new results could be used to estimate whether more intelligible speech is likely to occur.

13. Conclusions

The principal findings of this work were as follows:

- The results confirmed how much the intelligibility of speech is increased when the speech and noise sources are separated in free field conditions, i.e. there is a *Spatial Release from Masking*, for horizontally or vertically separated speech and noise sources.
- When a more realistic diffuse noise sound field was produced, the *Spatial Release from Masking* was substantially reduced relative to the case with a single separated noise source.
- Adding early-arriving reflections to the speech sounds, while maintaining a constant overall speech level, had no significant effect on speech intelligibility scores.
- Adding reverberant speech sounds with a reverberation time greater than 0.5 s, while maintaining a constant speech level, decreased speech intelligibility relative to a comparable case without reverberant speech.
- The decrease in intelligibility that resulted from adding reverberation to the speech can be determined from the logarithm of the reverberation time for reverberation times greater than 0.5 s.
- A single noise source from the rear side led to less masking of reverberant speech than a similar single noise source from the front side.
- When semi-diffuse noise was created, the effects on intelligibility were intermediate to those for a single direct noise source and those for completely diffuse noise.
- Although sound transmission through walls attenuates higher frequency sounds much more than lower frequency sounds, this filtering of speech sounds had no significant additional effect on either the influence of noise or of reverberation on the intelligibility of the speech when evaluated in terms of uniformly weighted signal-to-noise ratios (SNR_{UNI32}). This was not true for measures using other frequency weightings (See Appendix for example).
- The combined effects of the spatial differences in the noise exposures and the reverberation in the rooms adequately explained the difference between the initial laboratory experiment and the two-room validation tests in terms of both speech intelligibility scores and the threshold of speech intelligibility. As a result it is now possible to more accurately estimate the intelligibility of speech from an adjacent meeting room.

Appendix I. Some Effects of Frequency Weightings

Some further analyses of the new data from this study illustrate the benefits of the uniform frequency weighting incorporated in the SNR_{UNI32} measure. These results support the use of the SNR_{UNI32} measure in preference to the SII-weighted signal-to-noise ratio (SNR_{SII22}) [1] and point out that other measures such as the more complex Speech Transmission Index (STI) do not have ideal frequency weightings for the assessment of speech privacy when the speech is transmitted through walls.

Figures A-I-1 and A-I-2 compare plots of speech intelligibility scores versus SNR_{UNI32} and SNR_{SII22} values respectively for comparable cases with and without a simulated wall. All data were for the case of noise from all 8 loudspeakers and speech with a 1 s reverberation time.

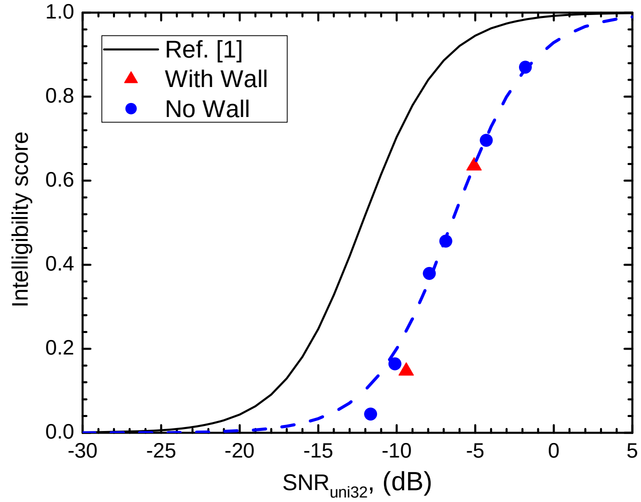


Figure A-I-1. Speech intelligibility scores versus SNR_{UNI32} for cases with noise from all 8 loudspeakers and speech with a 1 s reverberation time with and without a simulated wall and the comparable relationship from [1].

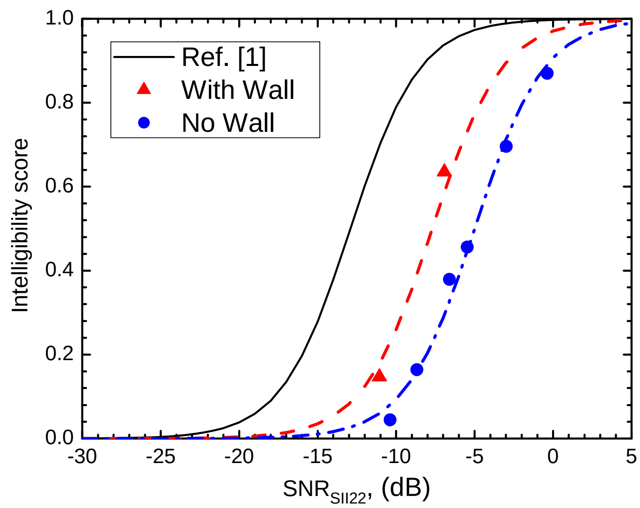


Figure A-I-1. Speech intelligibility scores versus SNR_{SII22} for cases with noise from all 8 loudspeakers and speech with a 1 s reverberation time with and without a simulated wall and the comparable relationship from [1].

When intelligibility scores are plotted versus SNR_{UNI32} values, all data follow the same regression line. However when scores were plotted versus SNR_{SII22} values in Figure A-I-2, the with-wall data and without-wall data clearly follow different trends that are shifted several decibels relative to each other. That is, when using SNR_{SII22} values, different results are obtained due to the process of filtering the speech to represent propagation through a wall. Presumably, different walls with different transmission characteristics would each lead to a slightly different trend. It is clearly more useful to measure the speech privacy in terms of SNR_{UNI32} values that will lead to more universally relevant ratings. SNR_{UNI32} values were recently been shown to successfully predict intelligibility scores for speech transmitted through 20 different walls [23].

A similar problem occurs when using STI values as illustrated in Figure A-I-3. Although the STI measure does seem to accurately account for the effects of a wide range of signal-to-noise and reverberation values, it provides different results depending on the transmission loss versus frequency characteristics when transmission through a wall is included. It might be possible to create an STI rating with a uniform frequency weighting but this was not attempted because it was beyond the scope of the current research.

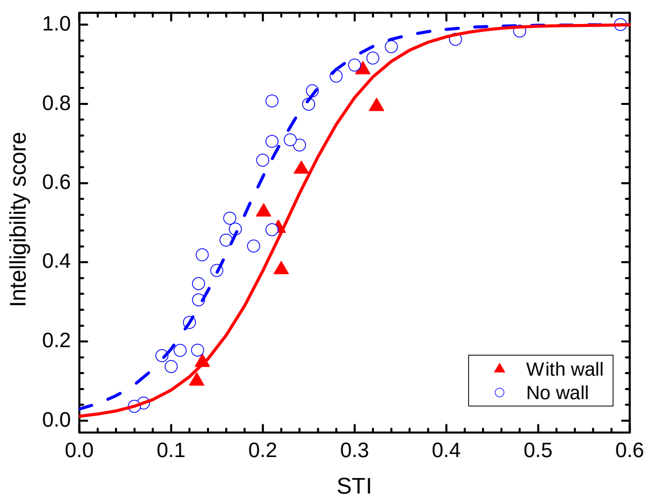


Figure A-I-1. Speech intelligibility scores versus STI values for data including a wide range of combinations of signal-to-noise ratio and reverberation time showing the different trends of results for cases with and without a simulated wall.

References

1. Gover, B.N., and Bradley, J.S., “Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms”, J. Acoust. Soc. Am., 116, (6), 3480-3490 (2004).

2. V.O. Knudsen and C.M. Harris, “Acoustical Designing in Architecture”, John Wiley & Sons, New York, 1965.

3. E.C. Cherry, Some Experiments on the Recognition of Speech with One and Two Ears, J. Acoust. Soc. Am. 25 (5) 975-979 (1953).

4. A. W. Bronkhorst: The Cocktail Party Phenomenon: A review of Research on Speech Intelligibility in Multiple-Talker Conditions. ACUSTICA 86, 117-128 (2000).

5. M. Ebata, Spatial unmasking and attention related to the cocktail party problem, J. Acoust. Sc. And Tech. 24 (5) 208-219 (2003).

6. N.R. French, J.C. Steinberg, Factors Governing the Intelligibility of Speech Sounds, J. Acoust. Soc. Am., 19 (1) 90-119 (1947).

7. A. W. Bronkhorst, 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).

8. D. D. Dirks, R. H. Wilson: The effect of spatially separated sound sources on speech intelligibility. J. Speech Hear. Res. 12, 5-38 (1969).

9. J. Peissig, 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).

10. J. Koehnke, J. M. Besing: A procedure for testing speech intelligibility in a virtual listening environment. Ear and Hearing 17 211-217 (1996).

11. 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), J. Acoust. Soc. Acustica 34, 200-211 (1976).

12. Hoth, D.F., “Room Noise Spectra at Subscribers’ Telephone Locations”, J. Acoust. Soc. Am., vol. 12, 449-504 (1941).

13. Blazier, W.E., “Revised Noise Criteria for Application in the Acoustical Design and Rating of HVAC Systems”, Noise Control Engineering Journal, 16(2), 64-73, (1981).

14. Bradley, J.S. and Gover, B.N., “Validation of Architectural Speech Security Results”, IRC Research Report, RR-221, March 2006.

15. “IEEE recommended practice for speech quality measurements,” IEEE Trans. Audio and Electroacoustics, 17, 227–246 (1969).

16. Whittington, J.E. and Bradley, J.S., “Quantifying the effects of room acoustics on speech intelligibility in very low signal-to-noise ratio environments”, J. Acous. Soc. Am., 120 (5) Part 2, 3359 (2006).

17. ANSI S3.6 (2004): “American National Standard Specifications for Audiometers,” ANSI, New York, NY, U.S.A.

18. ISO3382, Acoustics — Measurement of the reverberation time of rooms with reference to other acoustical parameters, International Organisation for Standardisation, Geneva, Switzerland.

19. Bradley, J.S., Sato, H. and Picard, M., “On the Importance of Early Reflections for Speech in Rooms”, J. Acoust. Soc. Am. 113 (6) 3233-3244 (2003).

20. Valley, G.E. and Wallman H., “Vacuum Tube Amplifiers”, McGraw-Hill (1948).

21. Kollmeier, B, and Wesselkamp, M., “Development and evaluation of a German sentence test for objective abd subjective speech intelligibility assessment”, J. Acoust. Soc. Am., 102 (4) 2412-2421 (1997).

22. Vom Hövel, H., “Zur Bedeutung der Ubertragungseigenshaften des Aussenohres sowie des binauralen Hörsytems bei gestörter Sprachübertsgung”, Ing. Dissertation, RWTH Achen (1984).

23. Park, H. K., Bradley J.S. and Gover, B.N., “Evaluating Airborne Sound Insulation in Terms of Speech Intelligibility”, J. Acoust. Soc. Am., 123 (3) 1458-1471 (2008).