Predicting Binaural Speech Intelligibility in Architectural Acoustics

J. F. Culling, M. Lavandier and S. Jelfs

1 Introduction

1.1 Measures of Acoustic Quality

Speech intelligibility can be impaired by poor room acoustics. This may happen as a result of distortion of the speech signal itself, because many delayed versions of the speech are summed at the ear, causing both spectral coloration and temporal smearing. Reverberation may also exacerbate the effects of background masking noise by impeding the processes by which the auditory system can overcome such masking. The relative importance of these effects depends on the type of listening situation. However, when listening to speech and noise from equidistant sources, it has been shown that the effects of reverberation on noise masking occur at lower levels of reverberation, and thus occur more readily, than the distorting effects on the speech [1].

In the planning and regulation of buildings, the acoustic quality of a room is generally summarized using statistics such as the reverberation time, T_{60}, and noise level. For instance, in the U.K., *Building Bulletin 93*, BB93, specifies upper limits for unoccupied ambient noise levels and for T_{60} in different types of classrooms. Ambient noise may vary across the space, in which case an average measure is needed, but the T_{60} should, at least in principle, be independent of measurement position. While single-value indices are convenient, they may not always accurately

J. F. Culling (✉)
School of Psychology, Cardiff University, Cardiff, UK
e-mail: cullingj@cf.ac.uk

M. Lavandier
Laboratoire Génie Civil et Bâtiment, Université de Lyon, Lyons, France

S. Jelfs
Philips Research Europe, Eindhoven, The Netherlands

J. Blauert (ed.), *The Technology of Binaural Listening*, Modern Acoustics and Signal Processing, DOI: 10.1007/978-3-642-37762-4_16,

reflect the speech intelligibility that will result. As will be demonstrated below, the T_{60} in particular, can be misleading.

In some circumstances, the *speech transmission index* [2], STI, or the useful-to-detrimental ratio [3] may be considered. The STI evaluates the degree to which amplitude modulation of speech survives the temporally smearing effect of reverberation. It is dependent upon the positions of the speech source and the receiver within the room. This makes it appropriate for lecture theatres and public address systems, for instance, where one individual communicates from a fixed location to an audience. The STI can be evaluated for each location in the listening space in order to ensure that adequate intelligibility is achieved in all listening locations. It can also produce predictions for intelligibility in noise, provided that noise is continuous and totally diffuse. However, these methods fail to produce accurate results where noise sources are nearby, such as in a busy social environment, where noise sources, such as background voices, may not be diffuse.

1.2 Binaural Speech Intelligibility

When speech and noise sources are spatially separated, speech intelligibility always improves compared to a situation in which they are co-located. This effect is known as *spatial release from masking*, SRM, and is likely related to a combination of at least two binaural processes, binaural unmasking and better-ear listening [4, 5]. Since speech and noise generally come from different sources, some SRM occurs in almost all natural listening situations. However, SRM is adversely affected by reverberation [6] and by the presence of multiple noise sources [4]. In order to accurately predict intelligibility in noisy rooms, it is therefore essential to take into account SRM and the influence that reverberation has upon it. This task is complicated by the dependence of these effects on the exact spatial layout of the speech and noise sources—it is not possible to characterize a room as facilitating a given level of SRM. However, it has now become possible to predict SRM for any given situation with considerable speed and accuracy.

Two very successful models of SRM have been developed by research groups in Oldenburg [7, 8] and Cardiff [9–11]. The current version of the Oldenburg model is the more comprehensive, because it can accommodate modulated masking noises and also hearing-impaired listeners. However, this chapter will employ the Cardiff model, which is well adapted to the rapid computation needed for many of the analyses below. This model explicitly evaluates the benefit to intelligibility expected from binaural unmasking and better-ear listening and regards their effects on the *speech reception threshold*, SRT, in noise as additive in decibels. The model has been applied to a wide range of data sets from the literature in both anechoic conditions with multiple noise sources [10] and in reverberant situations [9, 11] and generally provides a very high correlation with the empirical data—see Table 1. At present this model is only strictly applicable to continuous random noise sources. In order to apply them to more structured masking noises, such as voices, additional perceptual

Table 1 Summary of correlations between empirically measured SRTs from different experiments and corresponding model predictions

Experiment	Room	Number of noise sources	Correlation
Bronkhorst and Plomp [5]	Anechoic	1	0.86
Bronkhorst and Plomp [12]	Anechoic	1–6	0.93
Peissig and Kollmeier [13]	Anechoic	1–3	0.98
Hawley et al. [4]	Anechoic	1–3	0.99
Culling et al. [14]	Anechoic	3	0.94
Lavandier and Culling [9]	Simulated room #1	1	0.91
	Simulated room #2	1	0.98
Beutelmann and Brand [7]	Two real rooms	1	0.99
Lavandier et al. [11]	One real room	1	0.98
	Four real rooms	1	0.98
	One real room	3	0.95

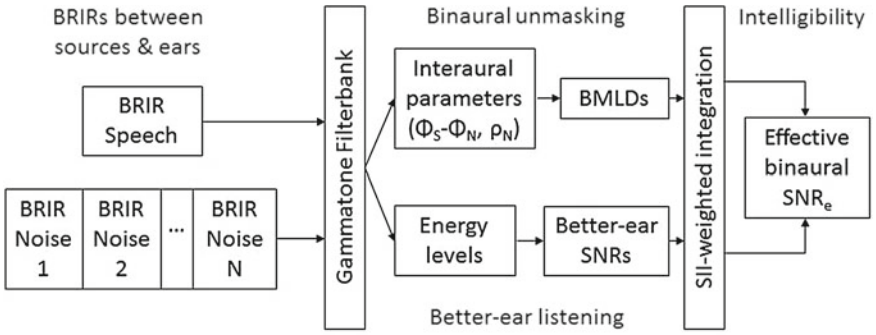

Fig. 1 Schematic illustration of the binaural intelligibility model. Φ_S and Φ_N... interaural phase differences of speech and noise, ρ_N... interaural coherence of the noise, BMLD ... binaural masking level difference

processes will need to be considered. However, notwithstanding this limitation, the model can make interesting predictions about the effects of room design and layout on communication.

1.3 Anatomy of the Binaural-Intelligibility Model

As noted above, the binaural model is based upon additive contributions from better-ear listening and binaural unmasking—Fig. 1. The model takes as input *binaural room impulse responses*, BRIRs, between the listening location and each of the sound-source locations. Its output is an *effective signal-to-noise ratio*, $\mathrm{SNR_e}$, that takes these processes into account. The remainder of this section describes how the

BRIRs are to be prepared and processed in order to generate the SNR_e and may not be of interest to the non-technical reader, who can skip to Sect. 1.4.

BRIRs may be generated by an acoustic model of a virtual room using suitable acoustic modeling software[1] or they may be recorded in a real room using an acoustic manikin. Where multiple noise sources are present, the impulse responses for all these sources are concatenated into one long impulse response. Concatenation has the effect of summing the frequency-dependent energy of each contributing impulse response, and generating an averaged cross-correlation function. It may seem intuitively reasonable to add together the BRIRs, just as one would add together different masking noises. However, summing directly the BRIRs would result in spectral distortion due to mutual interference, which does not occur when summing statistically independent interfering noises that have been convolved with those BRIRs. Only in the particular case of different sound sources, such as loudspeakers, driven by the same acoustic waveform, should the BRIRs be summed, to take into account the interference between these correlated sound sources at the ears.

The impulse responses for speech and noise(s) are separately filtered into different frequency channels, which are processed independently. The two contributions to intelligibility from binaural hearing are then modeled, namely, *better-ear listening* and *binaural unmasking*.

Better-ear listening simply reflects listeners' ability to pick up sound from the ear with the better signal-to-noise ratio. Interaural differences in SNR can occur as a result of head shadow, where the masking noise is occluded at one ear by the head, and also of room coloration, where frequency-dependent room absorption and complex interference between multiple room reflections creates different spectral distortions at each ear. Within each frequency channel the SNRs in dB at each ear are derived from the relative total energies in the filtered noise and speech BRIRs at that ear. The higher SNR of the two is selected as the better-ear SNR for that frequency.

Binaural unmasking is a psychoacoustic phenomenon in which the brain exploits the differences in interaural phase between signal and noise sources in order to improve detection or identification of the signal. These differences in phase are caused by differences in path distance to each ear. The size of the improvement is known as the *binaural masking level difference*, BMLD. The predicted BMLD is calculated within each frequency channel, of center frequency, ω_0. In order to predict speech intelligibility, the filtered BRIRs for speech and noise are separately cross-correlated. The speech and noise interaural phases, Φ_S and Φ_N, and the noise interaural coherence, ρ_N, are extracted from the resulting cross-correlation functions. These values are then used in the following equation, based on equalization-cancellation theory [15].

$$BMLD = 10\log_{10}\left[\frac{k - \cos(\phi_S - \phi_N)}{k - \rho_N}\right] \tag{1}$$

where, $k = (1 + \sigma_\varepsilon^2)\exp(\omega_0^2\sigma_\delta^2)$, $\sigma_\varepsilon = 0.25$, and $\sigma_\delta = 105\,\mu s$.

[1] For example, Odeon or Catt Acoustic.

Following the principle that binaural processing can only improve performance over what is possible based on listening with one ear, the BMLD is reset to zero if it has a negative value.

The better-ear listening and binaural unmasking components are each frequency weighted by the importance function for different frequencies in the *speech intelligibility index* [16], SII, and are assumed to make additive contributions to the *effective signal-to-noise ratio*, SNR_e, in decibels. This value is not intelligibility *per se*, because this would depend upon the nature of the speech materials and the integrity of the listeners' auditory systems, but making assumptions about these, one can go on to derive an intelligibility prediction through the SII [15]. The SNR_e can be used to predict differences in speech reception threshold across different listening situations; any resulting increase in SNR_e should give rise to an improvement (decrease) in SRT of equal magnitude. The SNR_e incorporates both the physical signal-to-noise ratio at that location and the benefits of binaural listening.

1.4 Suitability of the Binaural Model to Architectural Acoustics

In architectural acoustics, the effect of a room is fully described by the impulse responses between the positions of sound sources and receivers, for example, stage and seating area. Because the binaural model described above works directly with binaural room impulse responses as inputs, it can very easily be used in connection with room simulation software producing such impulse responses as output, or with acoustical measurements of impulse responses in real rooms. The only requirement is that these impulse responses should be binaural.

Because the model manipulates short impulse responses rather than the long source signals used by other models [7, 9], it produces fast and non-stochastic predictions, avoiding the averaging of predictions over several source signals. Thanks to its resulting computational efficiency, it can be used to draw intelligibility maps of rooms. Such maps were obtained by simulating the listener at different positions in a room containing a speaker and multiple noise sources [11]. The resulting spatial representations offer visualization of the space accessible to a listener who would wish to maintain a given level of intelligibility while moving within the room. Other types of representation can be computed—as illustrated later in this chapter.

Another advantage of the model is its modularity. The contributions of better-ear listening and binaural unmasking are computed independently in each frequency channel. The two contributions of binaural hearing can be considered separately, monaural listening can be simulated, and some frequency regions can be "deactivated". This would allow for specific forms of hearing impairment to be taken into account to guide technical applications directed towards the listener, such as by using directional microphones on hearing aids, or environmental policies concerning room design. For example, as of today, binaurally implanted cochlear implantees benefit from better-ear listening but not binaural unmasking [17], because current implants usually encode the temporal envelope of incoming sounds but not the temporal

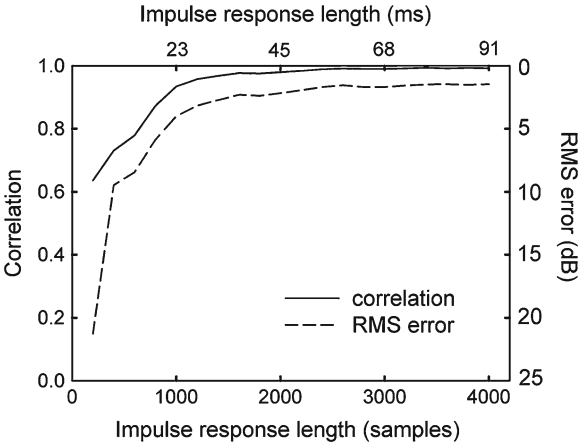

Fig. 2 Correlation between observed and predicted SRTs, and the RMS error of the prediction, plotted as a function of impulse-response length for the set of conditions examined by Beutelmann and Brand [7]

fine structure. Room intelligibility maps involving monaural or binaural listening, and without binaural unmasking, indicated where listeners can stand without losing understanding [11]. This might prove a useful tool towards predicting *room accessibility* for hearing-impaired listeners—see Sect. 2.6.

A key issue for practical implementation is the length of impulse response necessary to obtain an accurate prediction. Because the predictions of binaural-intelligibility by the model depend on the exact spatial configuration, it may be necessary to make many predictions for different listening positions and for different potential configurations of speech and masking-noise sources. Each prediction would require generation of BRIRs between each of the sources and the listening position. Many BRIRs may therefore be required. The calculation time for predicted BRIRs grows exponentially with the length of the BRIR, so the potential computational explosion may be contained by using the shortest BRIRs necessary for an accurate result.

To examine this, the effect of impulse response length on the accuracy of prediction was evaluated, using real-room impulse responses and corresponding SRT data collected by Beutelmann and Brand [7]. These 1.5-s impulse responses, originally 65,536 samples long, were collected from two different rooms, an office and a large cafeteria. As noted above, the model predicted the SRTs measured by Beutelmann and Brand quite accurately using their impulse responses. The correlation between observed and predicted SRTs was 0.99. In order to examine the effect of impulse response length, their data were modeled with those impulse responses truncated to lengths, between 200 and 4000 samples, that is, between 4.5 and 91 ms.

Figure 2 shows the correlations between observed and predicted SRTs as a function of impulse-response length, as well as the RMS error. It can be seen that long impulse

responses are not necessary, with performance reaching asymptote at around 3000 samples, that is, 70 ms. The cafeteria and office rooms in question have reverberation times, T_{60}, of 1.3 and 0.6 s, respectively, yet only the first 70 ms of reverberation is needed for an accurate prediction of intelligibility. This may be explained by the fact that, in each case, 96 % of the energy in each of the impulse responses occurred within the first 70 ms.

2 Applications to Architectural Design

The binaural model is suitable for answering a number of questions about the acoustic design of spaces in which listeners contend with background noise, such as classrooms, restaurants, cafeterias, railway stations and foyer areas. For the purposes of this chapter an acoustic model of a virtual restaurant is used as an example case, and the predictions of the binaural model will be explored for some simple design choices.

The room-acoustic model employed here was an image-source model [18] restricted to simple rectangular boxes. As noted above, commercial software could produce more accurate modeling of the room acoustics. Consequently, the acoustic model contained no representation of the furniture or the occupants and all sound sources were omnidirectional. On the other hand, the receiver characteristics of the listener are quite accurately modeled, because the acoustic wave fronts arriving at the listeners' heads are represented in the BRIRs by suitably delayed and scaled head-related impulse responses. Each head-related impulse response is selected from a database for the azimuth and elevation of that acoustic ray at the head position. The head-related impulse responses used were recordings made at the Massachusetts Institute of Technology [19] from a KEMAR manikin [20]. Although a more sophisticated room simulation would be preferable for practical applications, the present implementation has the advantage that all the resources for the simulation are in the public domain and the simplicity of the layout allows direct assessment of the principal room parameters. The aim was to demonstrate how the binaural-intelligibility model can be useful in architectural acoustics and to draw out some preliminary conclusions on the influence of these room parameters.

In order to examine these parameters a simple restaurant layout was developed, which included most of the critical factors one might expect to encounter in real life—see Fig. 3. The simulated restaurant contained nine *tables for two* in a regular 3×3 grid. Each table served to define two potential source/receiver locations, each being 1.2 m above the floor. The restaurant thus included pairs of source/receiver locations that had walls to the side, that is, tables #2 and 8, and others that had walls at one end, namely, tables #1, 3, 4, 6, 7, and 9, and also a pair that was surrounded by other sources—table #5. The room was 6.4 m square, the default ceiling height 2.5 m. The table positions were distributed evenly at 1.6-m intervals, with the source/receiver pairs separated by 0.7 m. The tables all had the common orientation shown in Fig. 3. Walls, ceiling and floor had controllable frequency-independent

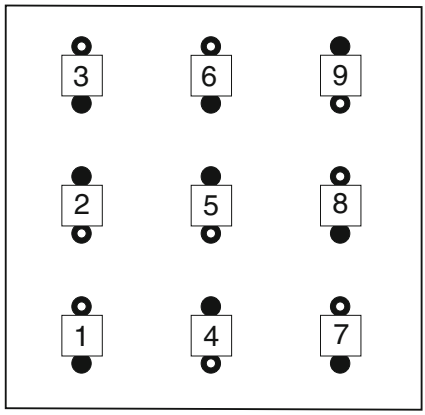

Fig. 3 Restaurant layout: Each rectangle represents a notional table, across which two diners (*black circles*) may wish to talk. In the model, one diner at each table (*with a white spot*) is nominated as the default location of a noise source

absorption coefficients. Across all simulations, there was one masking source at each table, which was selected at random, indicated by black spots with white dots in Fig. 3.

2.1 Effect of Seating Position in a Restaurant

How should one pick a good seat in a restaurant? Ideally one should be able to hear other individuals at the same table clearly. The model can make predictions of the variations in speech intelligibility across different tables and also within a given table. If one were able to answer such a question in a real restaurant, it would be possible to advise those who require better listening conditions, such as hearing-impaired listeners, to use particular seats. It may also be possible to tailor the acoustic treatment of the room to iron out such variations and provide a consistent acoustic experience across the entire space.

This question was addressed by looking at SNR_e across the different seats in the virtual restaurant. The absorption coefficients of the walls were set to 0.7, that of the floor to 0.1 and that of the ceiling to 0.9. Figure 4 shows the predicted SNR_e for each diner in the room, represented by the size of the corresponding black spot. The size of the spot is related to the SNR_e in dB. One can see that tables in the corner of the room are more favorable than those elsewhere, and that those placed between other tables, namely, the three middle tables, fare worse than those which are aligned with the wall. There are some local modifications to this pattern caused by the particular configuration of noise sources. For instance, one of the diners at table #8 has an interfering noise source immediately behind, that is, on table #7. This decreases the local SNR_e—see Fig. 4.

Using suitable acoustic modeling software, similar evaluations could be made in more complex acoustic spaces, such as alcoves, balconies, etc. The effects of different

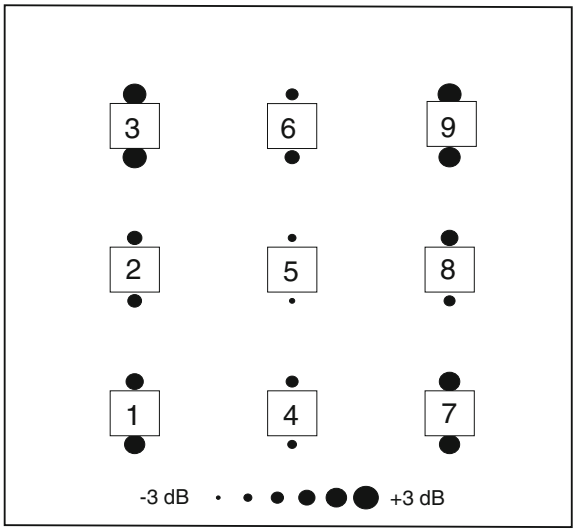

Fig. 4 SNR_e at each seating position in the virtual restaurant. The diameter of the *black spots* is proportional to SNR_e in dB

assumed configurations of masking noises could be addressed by averaging over a number of different random selections.

2.2 Effect of Head Rotation

As the head is turned horizontally, the relative directions of all sound sources are rotated around the head. Although SRM relies on *differences* in the directions of the target speech and the masking noise(s), the model indicates that changing all source directions together in this way can change the benefit to intelligibility. It is most often assumed that listeners directly face their interlocutor during a conversation, but this is not necessarily the case. In fact, observation of any busy social event will reveal that many people engaged in a conversation have their heads at an angle to each other. It is not currently clear whether this behavior is deliberate or whether it is related at all to optimizing speech intelligibility. Nonetheless, it is instructive to examine the potential impact.

If listeners do orient their heads in order to improve intelligibility, there must clearly be some limit to this behavior. It would be rude to turn one's back, eye contact may occasionally be required and lip-reading, which most listeners use unconsciously to improve intelligibility in noise [21], requires sight of the speaker's face. Counter-rotation of the eyes can be used to some degree in order to maintain sight of one's interlocutor, but it seems unlikely that such a sidelong posture would be practical beyond a head-turn of about 30°. Research on gaze control [22] indicates that, when fixating a target, observers make an initial eye turn of up to about 40°. Some observers will follow this movement with a head turn, which reduces the eye displacement down

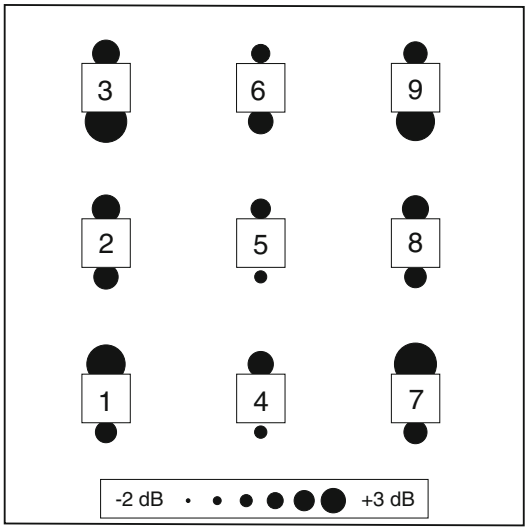

Fig. 5 As Fig. 4, but assuming that listeners have oriented their heads to the optimum angle for speech intelligibility within a range of ±30°

to 20° or so, while others will maintain a 40° eye displacement. The effect on the situation used in Sect. 2.1, of an optimized head turn of up to 30° was therefore evaluated.

Figure 5 shows revised values of SNR_e after the listeners have made optimal head turns. It can be seen that SNR_e has improved substantially in all cases. The mean improvement is 2.5 dB with values for individual listening positions ranging up to 5.3 dB. In addition to this general improvement, one can see a change in the pattern of results compared to Fig. 4, where no head rotation was assumed. Once head orientation is taken into account, the seats facing the wall at the four corner tables have a clear advantage over other locations. In each case, the optimal head orientation is to turn away from the side wall, such that the interlocutor on the other side of the table is to one side of the head and other sources in the room are on the other side of the head. The ear that is turned towards the interlocutor is thus maximally isolated by head shadow from the sources of masking noise and enjoys an improved SNR_e. It is also noticeable that local variations due to the configuration of masking-noise sources are also less evident; for the most part, SNR_e for each seat is similar to that for mirror-image locations across the room.

2.3 Effect of Ceiling Height

Many people have an intuitive sense that high ceilings contribute to a poor acoustic. However, there are good reasons to believe that this intuition is false and that high ceilings are actually beneficial. Their benefits may come from two acoustical factors. First a higher ceiling will increase the total absorbent area of the room, due

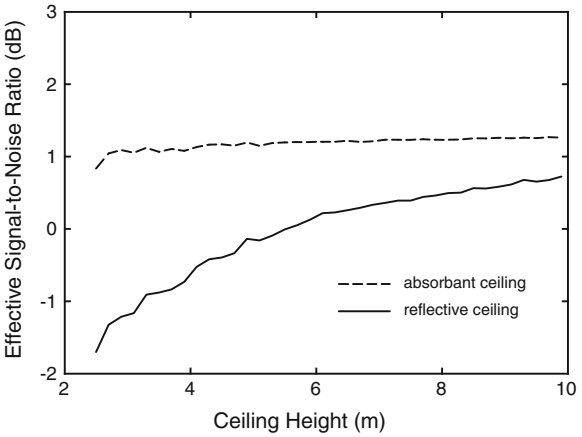

Fig. 6 Effective signal-to-noise ratio averaged across tables as a function of ceiling height for reflective and absorbent ceilings. Absorption coefficients were 0.9 and 0.1, respectively

to absorbent surfaces provided by the additional wall height. Such an increase in absorption reduces the total amount of reverberant energy within the space. Second, a high ceiling increases the volume of the room. This means that the reverberant energy spreads throughout a larger space, thus reducing the energy density. The prediction model can be used to simulate a range of different ceiling heights and determine the overall effect of these different processes on intelligibility.

To this end, a dining couple on each table in the restaurant was modeled. There were eight masking-noise sources, as in the configuration from Fig. 3. The SNR_e without head rotation was then evaluated as a function of ceiling height between 2.5 and 10 m, in 0.2-m steps. Other parameters were again similar to those of Fig. 4.

As can be seen from the solid line in Fig. 6, the SNR_e increases with the height of this reflective ceiling, indicating that a high ceiling provides easier communication to people in a noisy room. Once the ceiling was raised by 5 m, there was a 2.4 dB mean improvement in SNR_e. Across different tables, improvements ranged from 1.7 to 3.1 dB. For comparison, a similar level of benefit could be obtained with an acoustic ceiling that increases the ceiling absorption coefficient from 0.1 to 0.9, but the dashed line shows that only 0.4 dB of improvement would occur if the height of an absorbent ceiling was raised by 5 m, with half of this change occurring in the first 20 cm. To place both these effects in context, a totally anechoic room would only increase the signal-to-noise ratio by a further 2.5 dB.

It can be seen that intuitive impressions of ceiling height as a negative factor in room design are misleading. High ceilings are good. However, intuition is not the only false friend, here. T_{60} is generally used as a measure of how reverberant a room is; a larger T_{60} is usually considered a measure of a "more reverberant" room, which is generally assumed to result in lower intelligibility. Consistent with this association, when the absorption coefficients of the room boundaries are low-

ered, the corresponding increase in T_{60} is accompanied by a *decrease* in predicted intelligibility [6]. However, the Sabine equation, which can be used to estimate T_{60} from only the volume, V, and the effective absorbent area of a room, $\overline{\alpha}S$,—that is, the total surface area, S, times the mean absorption coefficient, $\overline{\alpha}$—shows that T_{60} is proportional to the volume, V, but inversely proportional to the surface area, S, that is,

$$T_{60} \approx 0.163 \frac{V}{\overline{\alpha}S} \text{ [s/m]} . \tag{2}$$

Now, since the volume to surface area ratio of any object or space increases with its dimensions, if the average absorption is held constant, T_{60} will increase with the room dimensions, including the ceiling height. Volume to surface area ratio will increase even if only the ceiling height is changed. Consequently, T_{60} can also be associated with an *increase* in speech intelligibility when ceiling height alone, or room volume in general, is manipulated. This fact is well illustrated by Beutelmann and Brand's data [7], which show consistently lower SRTs in their cafeteria environment with a T_{60} of 1.3 s, than in the office environment with a T_{60} of 0.6 s. In isolation, T_{60} is, therefore, a fairly useless measure of room quality for speech intelligibility unless room volume is factored out in some way. In BB93, there is little cognizance of room volume in the recommended T_{60} targets; particularly, a spacious classroom with a high ceiling would be over-treated in order to meet the specification, while a smaller than average classroom with a low ceiling would be under-treated.

2.4 Effect of Absorber Placement

It is most common to provide acoustic treatment to a ceiling. However, the benefits of binaural hearing depend upon the interaural differences produced by spatial separation of different sound sources. Since the ears are usually on the same horizontal plane, these interaural differences tend to be reduced by lateral reflections. Consequently, one might expect that designs which selectively reduce lateral reflections would generally provide greater benefit. Moreover, first-order ceiling reflections tend to reinforce interaural differences, because they come from the same azimuth. Thus, it may be better to place acoustic absorbers on the wall rather than the ceiling.

In order to quantify the potential benefit of laterally placed absorbers, two versions of the restaurant have been created with different absorber placements but the same overall T_{60} of 385 ms, as determined by the Sabine equation. These two rooms had identical floors with an absorption coefficient of 0.07. For the room with a reflective ceiling, the walls had an absorption coefficient of 0.6 and the ceiling an absorption coefficient of 0.06. For the room with an absorptive ceiling, these numbers were 0.05 and 0.9, respectively.

These configurations were tested by calculating the SNR_e for each diner, assuming that they were listening to the diner across the table with their heads fixed and that masking-noise sources were present at all other default locations for masking noise.

The mean benefit of absorptive walls compared to an absorptive ceiling was 0.7 dB. This benefit was entirely driven by better-ear listening. In contrast, the benefit of binaural unmasking fluctuated erratically around the value of 0.5 dB from one table to the next.

The advantage of wall absorbers should not, however, mislead one to thinking that ceiling treatment is ineffective. As shown in Fig. 6, ceiling treatment is always better than no ceiling treatment when the ceiling is high, but the benefit, here, is only 0.35 dB. For a high ceiling therefore, it would be particularly important to consider treating other surfaces. An equivalent change to the floor, for instance, perhaps by adding carpeting, would improve SNR_e by 1.75 dB.

2.5 Effect of Table Orientation

As noted above, optimum head orientation can substantially assist listeners in background noise, but such orientation is limited to, perhaps, $\pm 30°$ by the need to maintain visual contact with one's interlocutor. This limitation leaves open the possibility that diners may be assisted in reaching beneficial head positions/orientations by turning the whole table by 90°. In other words, might it be possible to use the model to derive an optimal table layout?

In order to investigate this possibility, the restaurant scenario described in Sect. 2.2 has been re-evaluated including optimal head rotations of up to 30°, but with some tables in different orientations. In each simulation, SNR_e was calculated for each pair of diners with eight masking-noise sources randomly distributed across the remaining tables. The results from twenty different random distributions were averaged. Due to the number of seats, head orientations and masker distributions considered, this analysis was quite time consuming. There are 2^8 unique permutations of table rotations to be considered, so it was necessary to concentrate on just a few interesting alternatives to the regular layout used above. Two layout strategies rotated the tables that were found to be most difficult in the analysis of Sect. 2.2. In one case, only the central table was rotated. In a second case all three of the tables down the centre of the room were rotated. In a third strategy, the case of rotating every second table throughout the room was considered.

The results showed that all three alternative strategies showed some benefit over a regular layout, but the benefits were fairly small. Rotating only the middle table, #5, or rotating every second table, that is, #2, 4, 6 and 8, improved the mean SNR_e by only 0.07 dB. Rotating the three middle tables, #4, 5 and 6—see Fig. 7, yielded a more noticeable mean improvement of 0.3 dB. Moreover, it is noteworthy that this option reduced somewhat the variability in SNR_e across different tables. In this scenario, large improvements of >2 dB were predicted for the diners on tables #4 and 6 who previously had their backs to the wall. None of these interventions produced significant benefit for the diners on table #5, however, and the standard deviation in SNR_e across seats was only reduced from 2 to 1.9 dB.

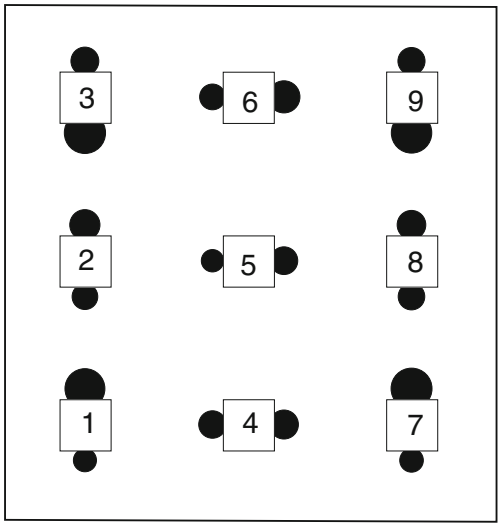

Fig. 7 Alternative table layout providing improved effective signal-to-noise ratios. The diameters of the *black spots* are proportional to the SNR_e in dB

2.6 Effect of Room Occupancy

Intelligibility worsens as a room fills up with people. How many people should a room be designed to accept? This has been termed the *acoustical capacity* of the room [23]. One can look at this question using the restaurant simulation. For a couple at each table, a given number of noise sources were distributed at random across the other eight tables. SNR_e of 20 such random distributions was then averaged. No head rotation was assumed.

Figure 8 shows that, unsurprisingly, SNR_e should fall with increasing room occupancy. The critical issue is the level of the SNR_e. Even when there is a noise source at every other table, and listeners are making no use of head orientation, the SNR_e falls no lower than -1.1 dB. Speech understanding in noise becomes impossible below about -3 dB, so this room seems to be acceptable for the assumed table layout.

It should be noted, however, that this analysis takes no account of the Lombard effect [24]. As the level of background noise increases, people instinctively start to raise their voices in order to be heard. As a result, the sound level in a room tends to increase with increased occupancy level more rapidly than would be expected from the number of sources present. Effectively, each doubling in the number of speakers tends to produce an increase of 6 dB in the ambient noise level rather than the expected 3 dB [23, 25]. Because all voices in the room are increasing together, this increase in vocal output and ambient noise level has no effect upon the SNR_e. Consequently, the effectiveness of communication is only disrupted to the extent that auditory processing is impaired by elevated sound levels [16, 26]. However, it also has an effect on the experience of the diners. People do not want to be shouting to

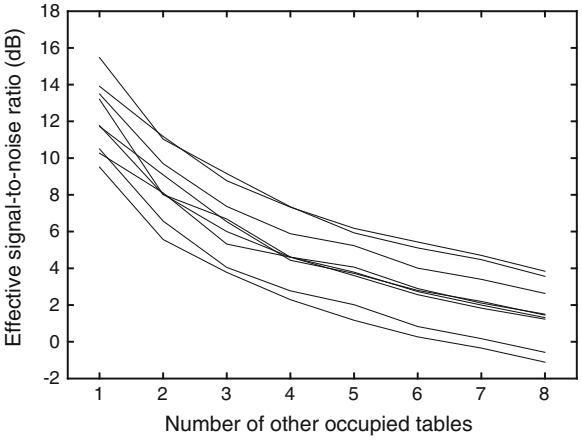

Fig. 8 Effective signal-to-noise ratio as a function of the number of occupied tables for each of the nine tables in the restaurant

make themselves heard. A separate analysis of the impact of room occupancy on vocal effort would therefore be advised [27].

2.7 Towards Predicting Room Accessibility

Accessibility of public spaces to those with disabilities is an increasingly important aspect of public policy. Architects now need to consider not only whether normally hearing listeners will be able to communicate effectively in a given acoustic space, but also whether the hearing-impaired listeners or non-native listeners will be able to do so. The level of intelligibility corresponding to a given signal-to-noise ratio is dependent on hearing and comprehension abilities. To ensure the same level of understanding, hearing-impaired listeners and cochlear implantees, for example, will require a better ratio than normally-hearing listeners.

The problem is a difficult one to address in a precise way, because hearing loss is a very individual disability. Different listeners will have different patterns of loss across frequency and the different etiologies of hearing losses have different consequences for speech understanding in noise. Moreover, there are currently gaps in our understanding of how a given hearing impairment leads to a given elevation in SRTs, which make it difficult to produce an accurate predictive model. Nonetheless, some notable successes have been achieved. Beutelmann and Brand [7] simulated cochlear hearing loss in their model by assuming that any elevation in pure-tone threshold was equivalent to an increased effective noise floor at that frequency and Culling et al. [17] modeled unilateral cochlear-implant patients simply by running their model in monaural mode, and assuming that each patient had an individually reduced recep-

tive capacity. The same strategy should work for listeners with single-sided deafness, but without the need to vary receptive capacity. Some other maneuvers are possible.

Listeners with cochlear hearing loss tend to have so-called *sloping losses*. This means that their pure-tone detection thresholds increase with frequency. These listeners might be modeled by assuming that they lose information at higher frequencies. This loss in information could be represented in the model by reducing the SII weighting values for high-frequency channels.

Listeners with asymmetric hearing have different SRTs when tested monaurally with each ear. These listeners could be modeled by assuming that their better ear is the ear that has the better signal-to-noise ratio after the difference in monaural SRT has been taken into account. That is, if the left-ear SRT is 3 dB better than the right-ear SRT, the model would assume that in binaural listening situations, the listener uses the left ear until the right ear SNR is at least 3 dB better than the left ear. Culling et al. [17] used this approach in order to model SRT data from bilateral cochlear-implant users [28]. In this instance, taking account of asymmetry in this way did not improve the fit to the data compared to ignoring the asymmetry, but this may be because the asymmetries in these cochlear implant users were fairly small; a minority of cochlear implant users have very large asymmetries, for which this technique might be essential.

The predictions of the model have been explored for the case of an asymmetry in monaural SRT. Such a manipulation does not affect its predictions of binaural unmasking, but only the selection of the better ear within each frequency band. The situation described in Sect. 2.2 was modeled, including the listeners' option to make a head turn of up to 30°, but assuming that each listener's right ear had a monaural SRT that was elevated by 10 dB with respect to their left ear. This has no effect when the better physical SNR is at the ear with the better monaural SRT, but when it is on the other side, it may require the listener to attend to the speech with the ear that has the poorer physical SNR. This inevitably has an impact on SNR_e. One would therefore expect only certain seating positions in the restaurant to be affected.

Consistent with this expectation, it turned out that, although the average predicted elevation in SRT was 1.2 dB, the effect was very strongly affected by seating position. Figure 9 shows the uneven distribution of those deficits. Essentially, in those seating positions where a deficit is visible in Fig. 9, it is approximately 3 dB. There are much smaller deficits distributed over the other positions.

The distribution in Fig. 9 can be understood in terms of the spatial distribution of speech and noise sources with respect to each listening position. For instance, for the listener experiencing a problem on table #9, a good right ear would allow them to rotate their head to the left and create a situation in which the target voice is to their right while all the noise sources are on their left. Since their right ear is impaired, they are less successful in following this strategy.

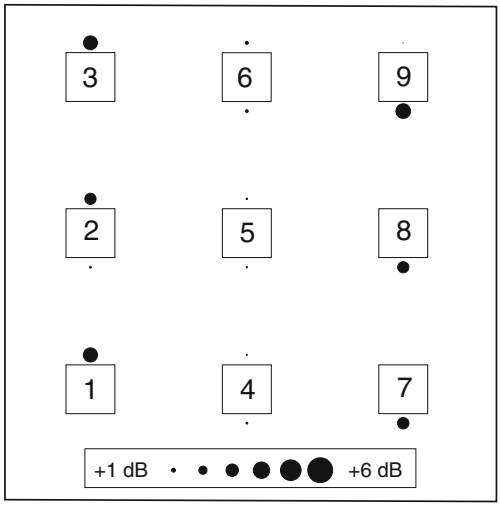

Fig. 9 Deficit in effective signal-to-noise ratio experienced by a listener, whose right ear has a monaural speech reception threshold that is elevated by 10 dB with respect to that for the left ear. The diameters of the *black spots* are proportional to the decrease in effective signal-to-noise ratio

3 Limitations of the Simulations and Further Developments of the Model

The simulations described above illustrate the power of the model to provide insights into the effects of room parameters on speech intelligibility in complex listening situations. A simple room geometry was used. An advantage of this approach is that general principles regarding such things as the effect of ceiling height can be addressed without confounding influences of uncontrolled room parameters. For any practical application, however, one would want to model the specific geometry of a room in order to evaluate the exact effect of making a design change in a specific project. The aim of this chapter was to illustrate the potential applications of the binaural intelligibility model to support the design of social interaction spaces.

3.1 Room Simulations

In order to draw conclusions regarding specific architectural designs, more sophisticated room simulations or real-room measurements need to be used to produce the BRIRs. The room simulations used here only considered the simplest room geometry, without taking into account the strong frequency dependence of room-materials absorption, the diffusion properties of these materials, or the directivity of different sounds sources. The binaural model can be used with any type of BRIRs, and it can only benefit from the use of more realistic BRIRs, be it measured or simulated, that take these acoustic phenomena into account.

The simulations used in this chapter only considered sources with the same sound level and long-term spectrum. The application of the binaural model is not limited

to these situations. Sources at different sound levels can be modeled by scaling their respective BRIR to the appropriate relative level; absolute source levels are not relevant, only differences in level between sources are. Sources with different spectra can also be modeled by appropriate filtering of their BRIRs. If the sources all have the same average spectrum, no filtering is required. In the case of multiple masking noises, concatenation of the scaled and/or filtered BRIRs would have the effect of summing the frequency-dependent energy of each contributing BRIR and generating an averaged cross-correlation function, weighted according to the energy in each BRIR.

3.2 Model Developments

The binaural model can accurately predict speech intelligibility against any number of stationary noise maskers, in any spatial distribution within a room and for any orientation of the listener. However, because it does not take into account the potential temporal smearing of target speech in very reverberant environments, this model can only predict intelligibility of target speech sufficiently close to the listener, at positions where the direct-to-reverberant ratio is not too low and segregation from sources of masking noise is the overriding factor for intelligibility. It needs to be extended to take into account this direct effect of reverberation on target speech, as has been done in a revision of the Oldenburg model [29]. Because the model works directly with BRIRs, it offers the opportunity to separate the early and late reflections within the BRIRs, so that temporal smearing can be modeled following the concept of useful-to-detrimental ratio [3, 30], in which the early reflections of the speech are regarded as useful because they reinforce the direct sound, while the late reflections are regarded as detrimental and effectively a part of the noise.

A model that intends to completely describe cocktail-party situations in rooms needs to handle competing speech sources and so to predict the segregation mechanisms associated with the temporal envelope modulations and the periodicity of speech. Fundamental frequency, F0, differences facilitate segregation of competing voices [31, 32], but reverberation is detrimental to segregation by F0 differences where F0 is non-stationary [33, 34] as in the case of normal intonated speech. Modulations in the temporal envelope of the masking noise allow one to hear the speech better during the moments when the speech-to-noise ratio is higher [35, 36], so-called *listening in the gaps* or *dip listening*, and this ability is impaired by reverberation which reduces modulations [8, 37], filling in *gaps* of the masker.

Restaurant simulation has been used to test the overall implications of these effects empirically [38]. SRTs were measured as a function of the number of masking sources, where those sources were either speech or continuous speech-shaped noise, and where the room was either reverberant or anechoic. The predictions of the model were accurate for the speech shaped noise, but speech maskers are less effective than noise. That is, SRTs were lower, when there was only a single masking voice, especially in anechoic conditions. On the other hand, SRTs were a few dB higher

for speech maskers than for noise when there was more than one masking voice. The advantage for a single masking voice may be attributed to some combination of *F0-difference processing* and *dip listening*, while the disadvantage for multiple masking voices appears to be some form of informational masking. It seems likely that in these multiple-masker cases, both dip listening and F0-difference processing are markedly less effective, although their precise role is, as yet, unclear.

The binaural model has recently been adapted to take dip listening into account, thus providing intelligibility predictions in the presence of speech-modulated noises [39]. However, this modified version of the model does not work directly on BRIRs, it requires the signals produced by the sources in the rooms as inputs. Following an approach proposed by Rhebergen and Versfeld [40] and then Beutelmann et al. [8], it consists of applying the stationary model to short time frames of the speech and noise waveforms, and then averaging the predictions over time. This signal-based approach would need to be adapted to be applied to the model based on BRIRs. For example, signal statistics could be associated with the BRIRs as model inputs, because BRIRs do not contain information about signal modulations. The advantage of having separated inputs for room and source information is that one might be able to simply update signal statistics to make predictions for different speech materials without requiring the actual signals in rooms.

4 Conclusions

The modeling presented here has clear limitations, both in terms of the sophistication of the acoustic model employed and the generality of the predictions to more structured masking sources, notably speech. Nonetheless, it captures aspects of the listening task which have hitherto been ignored in the acoustic assessment of rooms. It has been demonstrated in this chapter that the binaural model is markedly better suited to the prediction of intelligibility in rooms than the measures of reverberation time which are generally used. This modeling has also provided novel insights, such as the relative ineffectiveness of acoustically treating a high ceiling, which may well be general.

The limitations of the acoustic model could, for example, be addressed by using impulse responses generated by commercial room-simulation software. Since the binaural model is simple and computationally efficient, it could easily be incorporated into existing software in order to produce maps of the effective signal-to-noise ratio across a room or predictions for particular spatial configurations as in the simulated restaurant. Work continues on gaining sufficient understanding of human hearing to accurately predict the effects of dip listening and exploitation of F0 differences. It is as yet unclear whether they play a significant role in the complex listening situations with multiple maskers, for which the binaural model is designed.

Acknowledgments Work supported by U.K. Engineering and Physical Sciences Research Council. The authors thank their two external reviewers for valuable suggestions.

References

1. M. Lavandier, J. F. Culling (2008) Speech segregation in rooms: monaural, binaural and interacting effects of reverberation on target and interferer. *J. Acoust. Soc. Am.* 113:2237–2248
2. T. Houtgast, H. J. M. Steeneken (1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* 77:1069–1077
3. J. S. Bradley (1986) Predictors of speech intelligibility in rooms. *J. Acoust. Soc. Am.* 80:837–845
4. M. L. Hawley, R. Y. Litovsky, J. F. Culling (2004) The benefit of binaural hearing in a cocktail party: Effect of location and type of masker. *J. Acoust. Soc. Am.* 115:833–843
5. A. W. Bronkhorst, R. Plomp (1988) The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J. Acoust. Soc. Am.* 83:1508–1516
6. R. Plomp (1976) Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acustica* 34:200–211
7. R. Beutelmann, T. Brand (2006) Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 120:31–342
8. R. Beutelmann, T. Brand, B. Kollmeier (2010) Revision, extension and evaluation of a binaural speech intelligibility model. *J. Acoust. Soc. Am.* 127:2479–2497
9. M. Lavandier, J. F. Culling (2010) Prediction of binaural speech intelligibility against noise in rooms. *J. Acoust. Soc. Am.* 127:387–399
10. S. Jelfs, M. Lavandier, J. F. Culling, (2011) Revision and validation of a binaural model for speech intelligibility in noise. *Hear. Res.* 275:96–104
11. M. Lavandier, S. Jelfs, J. F. Culling, A. J. Watkins, A. P. Raimond, S. J. Makin (2012) Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources. *J. Acoust. Soc. Am.* 131:218–231
12. A. W. Bronkhorst, R. Plomp (1992) Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *J. Acoust. Soc. Am.* 92:3132–3139
13. J. Peissig, B. Kollmeier (1997) Directivity of binaural noise reduction in spatial multiple-source arrangements for normal and impaired listeners. *J. Acoust. Soc. Am.* 101:1660–1670
14. J. F. Culling, M. L. Hawley, R. Y. Litovsky (2004) The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *J. Acoust. Soc. Am.* 116:1057–1065
15. N. I. Durlach Binaural signal detection: Equalization and cancellation theory. In J. Tobias (Ed) Foundations of Modern Auditory Theory, Vol. 2. Academic, New York, pp. 371–462 (1972)
16. ANSI (1997) Methods for calculation of the speech intelligibility index. ANSI S3.5-1997, American National Standards Institute, New York
17. J. F. Culling, S. Jelfs, A. Talbert, J. A. Grange, S. S. Backhouse (2012) The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise. *Ear Hear.* 33:673–682
18. J. B. Allen, D. A. Berkley (1979) Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.* 65:943–950
19. W. G. Gardner, K. D. Martin (1995) HRTF measurements of a KEMAR. *J. Acoust. Soc. Am.* 97:3907–3908
20. M. D. Burkhard, R. M. Sachs (1975) Anthropometric manikin for acoustic research. *J. Acoust. Soc. Am.* 58:214–222
21. A. MacCleod, Q. Summerfield (1987) Quantifying the contribution of vision to speech perception in noise. Br. *J. Audiol.* 21:131–142
22. J. E. Goldring, M. C. Dorris, B. D. Corneil, P. A. Ballantyne, D. P. Munoz (1996) Combined eye-head gaze shifts to visual and auditory targets in humans. *Exp. Brain. Res.* 111:68–78
23. J. H. Rindel (2012) Acoustical capacity as a means of noise control in eating establishments. Joint Baltic-Nordic Acoustics Meeting, Odense, Denmark
24. H. Lane, B. Tranel (1971) The Lombard sign and the role of hearing in speech. *J. Sp. Hear. Res.* 14:677–709

25. M. B. Gardner (1971) Factors Affecting Individual and Group Levels in Verbal Communication. *J. Audio Eng. Soc.* 19:560–569
26. B. C. J. Moore, B. R. Glasberg (1987) Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hear. Res.* 28:209–225
27. J. H. Rindel, C. L. Christensen, A. C. Gade (2012) Dynamic sound source for simulating the Lombard effect in room acoustic modeling software. Inter Noise 2012, New York
28. P. C. Loizou, Y. Hu, R. Litovsky, G. Yu, R. Peters, J. Lake, P. Roland (2009) Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *J. Acoust. Soc. Am.* 125:372–383
29. J. Rennies, T. Brand, B. Kollmeier (2011) Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet. *J. Acoust. Soc. Am.* 130:2999–3012
30. J. P. A. Lochner, J. F. Burger (1964) The influence of reflections on auditorium acoustics. *J. Sound Vib.* 1:426–454
31. J. P. L. Brokx, S. G. Nooteboom (199 2) Intonation and the perceptual separation of simultaneous voices. *J. Phonetics* 10:23–36
32. J. F. Culling, C. J. Darwin (1993) Perceptual separation of concurrent vowels: within and across formant grouping by F0. *J. Acoust. Soc. Am.* 93:3454–3467
33. J. F. Culling, Q. Summerfield, D. H. Marshall (1994) Effects of simulated reverberation on binaural cues and fundamental frequency differences for separating concurrent vowels. *Speech Comm.* 14:71–96
34. J. F. Culling, K. I. Hodder, C. Y. Toh (2003) Effects of reverberation on perceptual segregation of competing voices. *J. Acoust. Soc. Am.* 114:2871–2876
35. A. W. Bronkhorst, R. Plomp (1992) Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *J. Acoust. Soc. Am.* 92:3132–3139
36. J. M. Festen, R. Plomp (1990) Effects of fluctuating noise and interfering speech on the speech-reception SRT for impaired and normal hearing. *J. Acoust. Soc. Am.* 88:1725–1736
37. A. W. Bronkhorst, R. Plomp (1990) A clinical test for the assessment of binaural speech perception in noise. *Audiology* 29:275–285
38. J. F. Culling (in press) Energetic and informational masking in a simulated restaurant environment. in Moore, B C J, Carlyon R P, Gockel H, Patterson R D, Winter I M (eds) Basic Aspects of Hearing: Physiology and Perception (Springer, New York)
39. B. Collin, M. Lavandier (under review) Binaural speech intelligibility in rooms with variations in spatial location of sources and depth of modulation of noise interferers. *J. Acoust. Soc. Am.*
40. K. S. Rhebergen, N. J. Versfeld (2005) A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *J. Acoust. Soc. Am.* 117:2181–2192