Hearing Of Music In Three Spatial Dimensions

Jens Blauert

1. Introduction

All the objects of our perceptive world are distinct in time and space and have distinct properties, i.e., they appear at certain times, in certain positions, and are equipped with qualities specific to them. What is valid for all the percepts, consequently, also applies to what one can hear - namely, what we shall call the auditory events in the following. People who listen to a speaker have, for example, an auditory event speaker which is perceptually present during a certain period of time, in a certain region in space, and posesses certain properties such as specific loudness, pitch, timbre, etc.

In the framework of this essay our primary interest is the spatial distinctness of auditory percepts, in other words, the positions and spatial extensions of the auditory events. We want to explain how these positions and extensions depend on certain features of the sound signals which are received and processed by the human auditory system. These dependencies, i.e., the facts of spatial hearing, are of great practical importance for music, among other things in the following contexts.

- For composition and performing technique: Phenomena of spatial hearing are increasingly taken into consideration.
- For room acoustics: The acoustical conditions of the concert hall with its consequences for spatial hearing are a deciding factor for the artistic effect of music.
- For electroacoustics: The spatial features of the auditory events of a listener to a recording or life broadcast should correspond in a defined manner to those at the recording end.

In the following we have collected some selected physical and psychoacoustical principles of spatial hearing which are of particular importance for music. Further details and extensive bibliographies can be found, for example, in Mills (1972), Durlach & Colburn (1978), Cremer & Müller (1978/82), Gatehouse (1981), and Blauert (1983).

2. On the reception of sound by human hearing

Expressed in a simplified form, the human auditory system can be considered to be a computer with two input ports. The input signals are sound signals which reach the left and right eardrums. On their way to the eardrums, these signals pass through the external ears. First they are diffracted and partly shadowed off by the skull, then they enter the pinna-earcanal system, where their spectrum is modified by resonances in a specific way. Thus, spectral modifications are superimposed on the incoming sound signals. These linear distortions have been provided to be particularly important for spatial hearing. Proof is provided by an experiment shown schematically in Fig. 1.

The sound signals from the eardrums of a subject are recorded by means of probe microphones in the earcanals, and then recorded. Next they are played back to the same subject over earphones after appropriate spectral equalization, so that the original sound signals are reproduced as accurately to the original as possible (Fig. 1a). The positions and timbre of the auditory events during playback will then correspond almost exactly to those produced during the recording session.

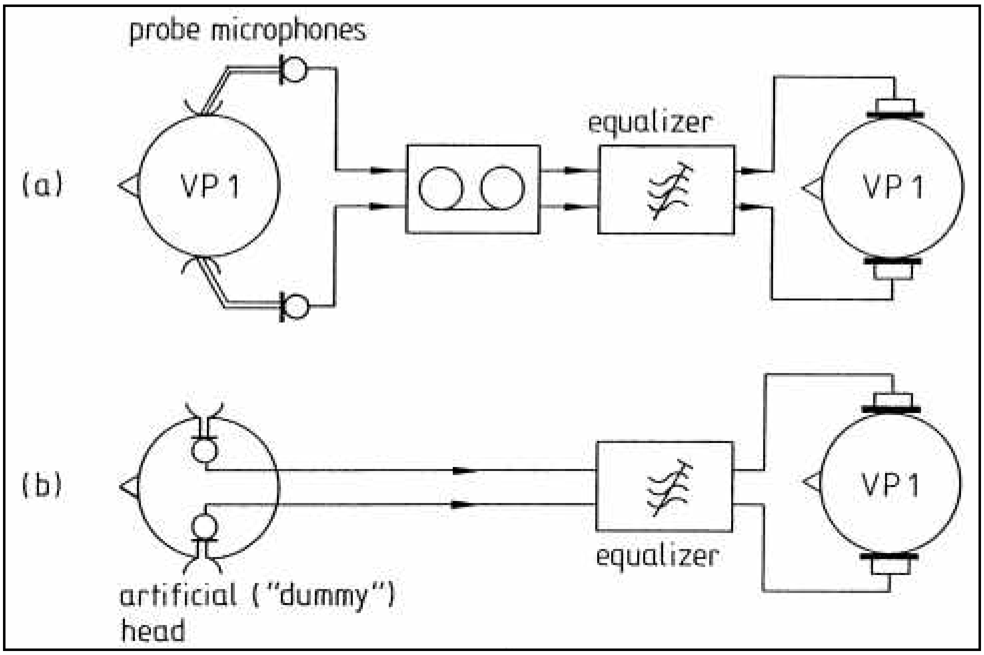

Fig. 1
Principle of head-related electroacoustical transmission (a) Recording with probe microphones on a real head (b) Recording with an artificial head

In Fig. 1b a modified experimental set-up is shown. The signals here are obtained from an acoustically similar imitation of a human head - a so-called artificial head. If the outer ears of this dummy are not shaped exactly like those of the real subject, there will be errors in the transmission of the positions of the auditory events. A more exact analysis of the linear distortions which the sound signals experience when passing the external ears (incl. the skull) shows that these distortions depend in a characteristic way on the directions of incidence of the sounds and the distances to the sources of sound. The external ears, thus, become a special coding machine which transforms directions and distances into spectral information. The auditory system, on the other hand, is in a position to decode this information contained in the ear signals concerning the directions and distances of the sound sources and to take these into consideration when forming the directions of the auditory events. One can distinguish two categories of features of the input signals to the two ears which are analysed by the auditory system:

- Monaural signal features: One ear is sufficient to receive these.
- Interaural signal features: These are features which touch upon differences between the ear signals. Two ears are necessary for their reception.

The relevant monaural features are contained in the shape of the energy spectra of the signals at the two ears. The most relevant interaural features are interaural differences of the levels and arrival times of the two input signals to the two ears. Measured values for those features are, e.g., documented in Blauert (1983).

3. SPATIAL HEARING WITH A SINGLE SOUND SOURCE

The simplest case of sound radiation is when there is only one single sound source and no sound reflection, e.g., from walls. To begin with, we consider the special case of the sound source happening to be in the median plane, this is the plane which vertically cuts through the middle of the head. If one uses the approximation case that the human head is symmetrical, then in this case both ear signals will be identical, i.e., the interaural differences will equal zero. Differences in the directions of sound incidence inside the median plane therefore mainly have an effect on the monaural ear-signal features. An important psychoacoustical relationship that plays a role here will be explained in Fig. 2.

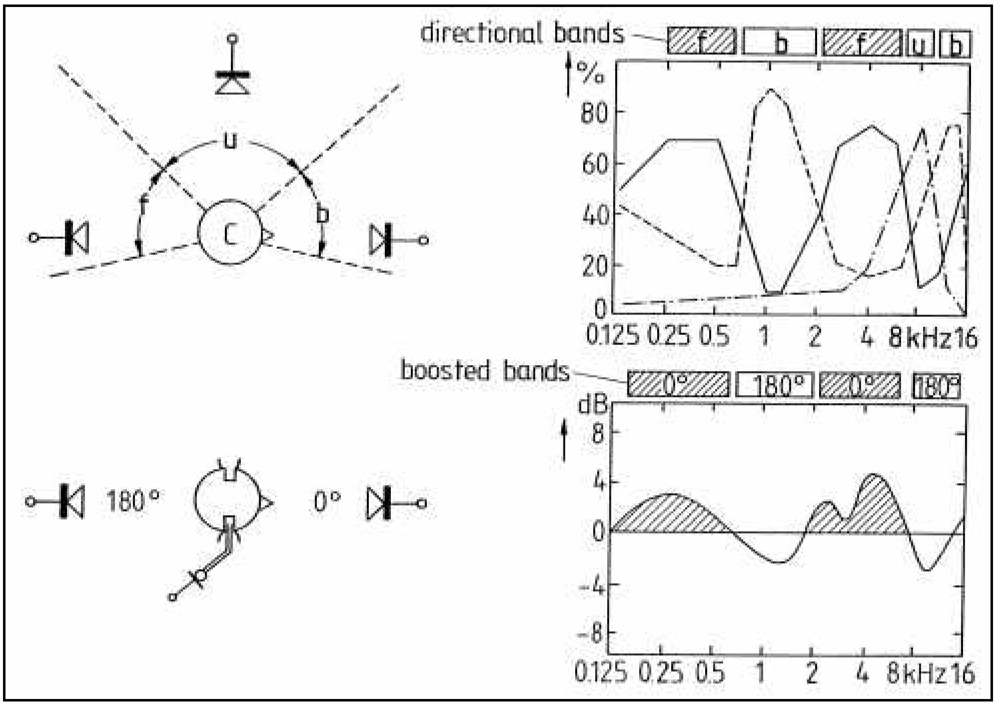

Fig. 2
For the formation of the perceived direction in cases of sound incidence from the median plane. Above: spectral bands determining the direction. Below: bands in which the external ear enhances the relative amount of signal power

In the upper half this figure describes an experiment in which the subjects were stimulated with 1/3-octave-wide bands of noise from different directions of the median plane. In this case the auditory events also appear in the median plane. It has been established in this context that the direction of sound incidence is unimportant for the perceived direction - the direction of the auditory events is decided only by the center frequency of the narrow-band noise. In the upper panel of the diagram, the estimated probability is noted with which the auditory events will occur in the median plane in sector b (behind), a (above) and f (front). The spectral bands which determine the perceived directions are schematically summarized on the upper edge of the diagram (directional bands).

In the lower panel of Fig. 2 the difference in the power spectra at the ear which results for frontal sound incidence when normalized with respect to incidence from behind is represented. The raised spectral regions in each case have been schematically presented above the diagram. These boosted bands fit the directional bands insofar as the irradiation from behind produces a rise in the directional rear - bands when irradiated from behind. A corresponding coordination is produced when irradiation takes place from the front and - not shown here - from above.

We know from our own experience and from many measurements that when irradiated with signals in our everyday life, e.g., music or speech, the auditory events generally arise in the directions of the sound sources. This also applies to the median plane. Because of our knowledge of the mutual fit of directional bands and boosted bands the following explanation becomes apparent. When the sound signals to the ears pass the external ear they are always raised or attenuated in specific spectral regions depending on the direction of sound incidence. The usual congruence of the directions of sound incidence and of the perceived directions is produced by relatively enhancing the signal power in the appropriate class of directional bands.

What has been explained here for the directions of sound incidence from the median plane is important for yet another phenomenon: Monaural features, particularly the shape of the level spectra of the signals at the eardrum, also play an important role in forming the directions of auditory events outside the median plane - and for their distances.

For the spatial hearing of music there is the following important consequence. The shape of the power spectrum of the sound signals of a sound source - of a musical instrument or a singing voice - is not only of importance for the timbre of the sound heard. In addition the spectrum is significant for spatial perception. The following rules can be applied in this connection.

- Signals with strong components in the range of approx. 2 to 7 kHz are perceived as being particularly present and forward oriented.
- Signals with strong parts around 1 kHz lead to auditory events that sound rather diffuse in space and sometimes even sound as if they are coming from behind the listener.
- Signals with strong components around 8 kHz lead to auditory events above the horizontal plane under a greater or smaller angle of elevation.

As has already been noted, interaural features of the ear signals are essential as well for directing the auditory events - besides the monaural ones. These interaural features - especially interaural level differences and interaural time differences - determine the lateral deviation of the auditory events from the median plane. With the help of interaural features we can recognize lateral deviation of sound sources of only 1 degree from the median plane. In comparison, deviation of sound sources within the median plane, which are mainly recognized by monaural ear-signal features, can only be perceived when they exceed approximately 10 degrees

4. SPATIAL HEARING WITH MULTIPLE SOUND SOURCES

The case of only one sound source in reflection-free surroundings hardly ever occurs in practice. There are almost always surfaces present to reflect the sound, especially in enclosed spaces. In many cases one can describe the sound thrown back as a wave field of one or more virtually mirroring sound sources which radiate the same signal as the primary sound source. Primary signals and reflection signals are then completely correlated. For simplicity's sake, the spatial effects that can be observed on several correlated sources will be explained using the example of two sources. A generalization on more than two sources presents no difficulties. The following effects can be observed with two correlated sources: Summming localization, the presedence effect (formerly also called the law of the first wave front) and the echo effect.

- Summing localization appears when there is only one auditory event whose position depends on both sound sources.
- The presedence effect (law of the first wave front) applies when only one auditory event whose position is determined by one sound source alone is apparent.
- The echo effect is produced when in addition to the first auditory event a second one is heard. This appears as an echo of the first. The position of the primary auditory event is determined by the sound source which radiates first, the position of the echo by the later source.

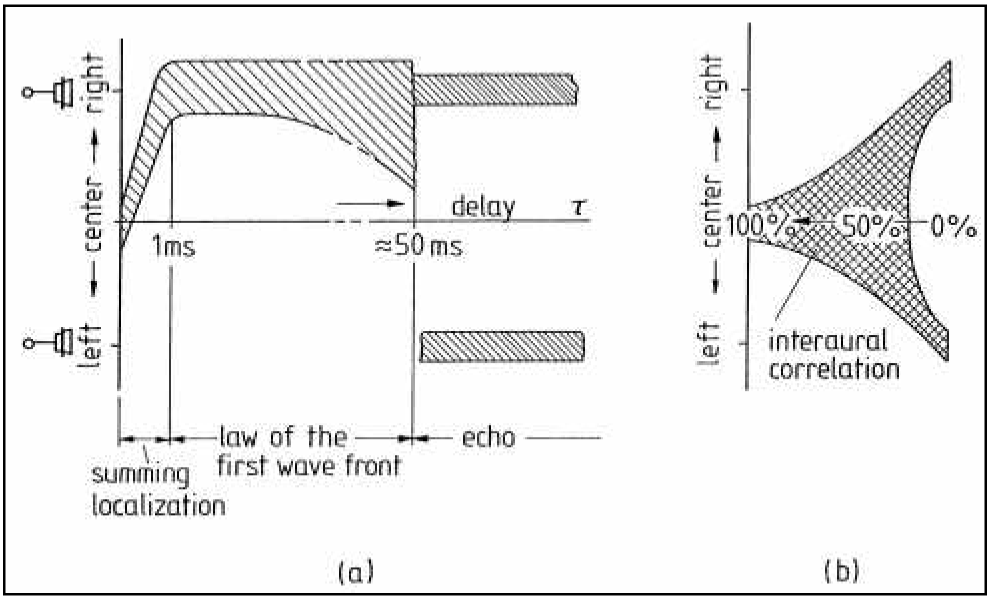

Fig. 3
Perceived directions and spatial extensions when listening to two loudspeakers in a standard-stereo arrangement. (a) two fully correlated loudspeaker signals, with one delayed. (b) no delay, but differing correlation

In Fig. 3a these effects are explained. As an example we take two loudspeakers in a standard-stereo arrangement as sources. The left-hand loudspeaker sends the same signal with the same strength as the right-hand one, but with an inter-loudspeaker delay. In the diagram the perceived direction is plotted as a function of this delay.

With zero delay there is an auditory event exactly in the middle between the two loudspeakers summing localization. When the delay increases to about 1ms, the percept drifts off to the earlier loudspeaker. For delay times above approx. 1ms the region of the presedence effect comes into play. The perceived direction there is only determined through the first-sounding loudspeaker; this simulation persists up to delay times of approx. 50 to 80ms. After further raising the inter-loudspeaker delay, the auditory event splits into two parts, a primary one in the direction of the right-hand loudspeaker and an echo in the direction of the left-hand one. The threshold value for the appearance of an echo depends on the type of signal. In most cases it is higher in music than in speech. Both effects, summing localization and the presedence effect are technically applied in room- and electroacoustics. The appearance of echoes is generally undesirable. Note also that below echo threshold the spatial extention of the auditory event increases with delay in a set pattern.

With summing localization it is possible to generate auditory events at positions at which there are no sound sources. Thus, with two-channel stereohony, auditory events of hearing can be produced anywhere in the sector between the two loudspeakers and - using a special procedure - even outside. More than two loudspeakers can be used, but there are, however, some loudspeaker positions which are not favorable. Summing localization can also have a disturbing effect. The presence of reflecting surfaces causing slightly delayed reflections may lead to disturbing deflections of the auditory events away from the original sources (image shift).

The presedence effect is of fundamental importance for acoustics. It is because of this that the auditory event is perceived in the direction of the original sound source and the reflected sound does not disturb this orientation. This does not mean, however, that the reflected sound would be completely masked. As a rule it contributes to the loudness and can for example raise the intelligibility of speech. In electroacoustics this is exploited in that amplified sound is radiated with a delay. The loudness is then raised, yet the percept still appears at the position of the original sound source.

Fig.3a dealt with the example of two completely correlated loudspeaker signals. Fig.3b now shows us what effect a variation of the correlation can have. Two signals with a variable correlation of between 0% and 100% can be created, e.g., by an appropriate mixing of noise from three sources. Concerning the features of the auditory event the following can be observed. The auditory event which is sharply localized and narrowly limited in space when there is a complete correlation of the loudspeaker signals, continues to spread out in space with a decreasing correlation, and with only slight correlation fills the whole sector between the loudspeakers. With completely uncorrelated signals two independently perceived auditory events, one in the direction of each loudspeaker, are finally achieved.

The most important effects dealt with in this section can be summarized as follows. With several sound sources that radiate correlated signals - e.g., several loudspeakers or an original sound source and its mirroring sound sources - the single sources are not as a rule perceived as separate auditory events. Often only one, whose position is determined by the first wave arriving at the listener, is perceived. When the correlation of the radiated sound signals decreases, the spatial expansion of the auditory event increases. With several uncorrelated radiated signals there are mostly several distinct auditory events. The latter situation is given, for example, when listening to an orchestra, where the individual instruments are to be seen as sources weakly correlated to each other. It is important for the listener to be able to distinguish the single instruments within the global auditory picture. The orchestra should sound transparent. It is also said that the individual instruments should not mask each other. In principle we can state that signals from various sources least mask each other when the components of the input signals to the listeners' two ears coming from separate sources clearly differ from each other in their interaural features. Thus, if the sources' azimuths are different, the masking will be less.

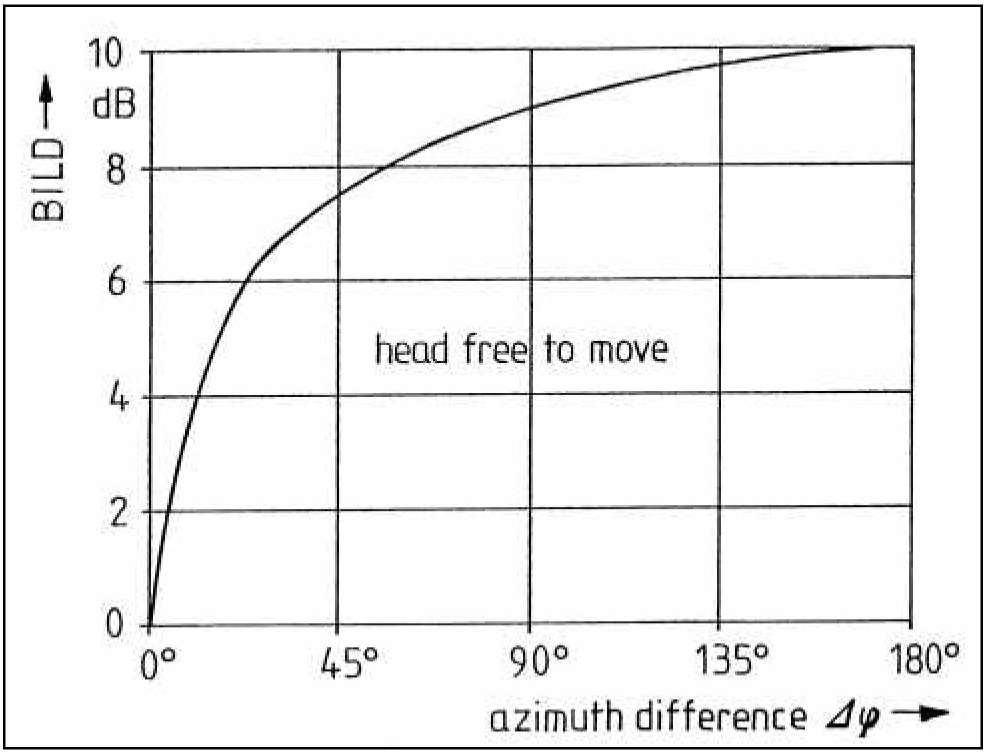

Fig. 4
Intelligibility-level increase due to release of masking caused by directional separation of two sources in the horizontal plane. Reference sound: speech at normal speed from a frontal direction. Masking sound: broad-band noise from different directions of the horizontal plane

Fig.4 schematizes measured results which deal with this connection. A loudspeaker radiating continuous speech at a 60-decibel level was hung up in front of a subject in a free sound field. A second loudspeaker radiating broad-band noise was placed in a swinging position in a horizontal circle around the listener. First it was brought into a frontal position and the noise level was adjusted so that the syllable intelligibility sank to 60%. When the second loudspeaker was swung to the side, the syllable intelligibility rose again. In order to let it sink to 60% again, the noise level had again to be raised. The necessary level rise (BILD = Binaural Intelligibility-Level Difference) came to a maximum of 10dB in this particular experiment. In the presence of reflecting sound lower values would be obtained, however, approximately 5dB are still typical in an enclosed space.

The following practical lessons can be learnt from this. In order to obtain a transparent hearing pattern, the separate sources must be arranged under distinctly different azimuths with respect to the listeners. This also means that the perceived directions of the auditory events will clearly be different. Creating an auditory perspective therefore also means creating a better auditory transparence. Release from masking by means of separation in space depends on the evaluation of interaural ear-signal features and thus on two functioning ears. In electroacoustical recording and broadcasting, several (at least two) channels are necessary to transmit these features.

5. AUDITORY SPACIOUSNESS

The last part of this article is devoted to a phenomenon of spatial hearing which has recently been intensively examined and discussed with regard to the design and evaluation of concert halls, namely, (auditory) spaciousness. Fig.5 shows the effect using an example of an orchestra in a concert hall. When the orchestra plays at a low volume, listeners experience auditory events which are spatially limited to the visual outline of the orchestra. When the volume is increases it is observed in many concert halls that the auditory events increase in spatial extension - particularly to the sides, in depth, and upwards. It has been known for a long time that listeners to music prefer sound fields in which the spaciousness effect can be experienced. This has been confirmed by surveys in recent years in which the components of preference judgements in the evaluation of concert halls have been analysed.

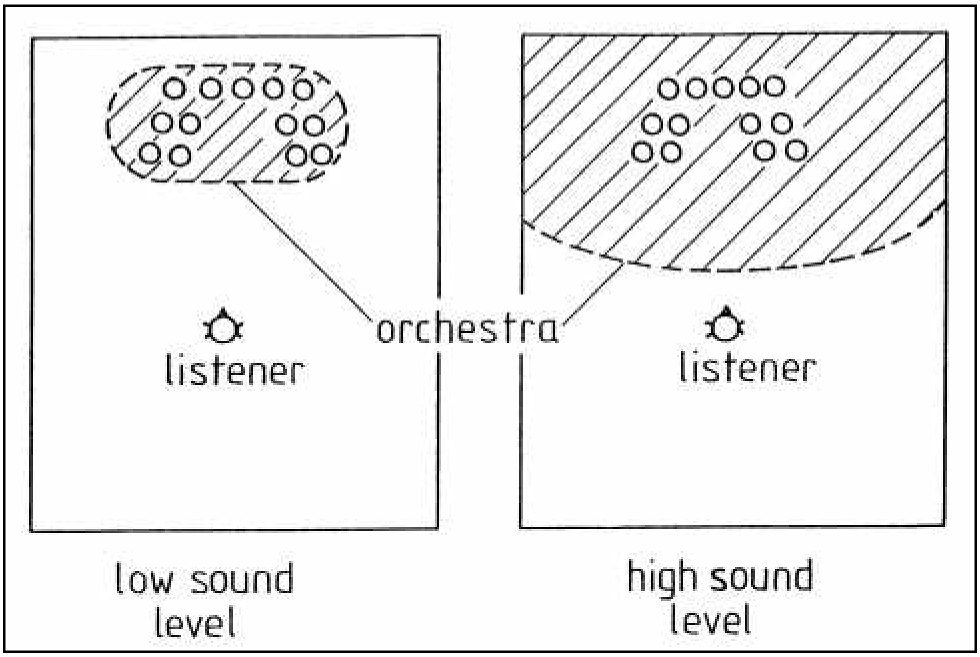

Fig. 5
To explain the effect of auditory spaciousness

We have already mentioned that the extension of the region in space filled by the auditory events depends on the correlation of the source signals and, hence, on the interaural correlation of the signals at the listener's ears. It is worthwhile considering how the ear signals can become interaurally decorrelated in enclosed spaces in the desirable manner - which is obviously the condition for spaciousness. The sound field in enclosed spaces is different from a free field because in addition to the direct sound, reflected sound also reaches the ears. The fundamental physical connections can already be understood by using the most simple example of one direct sound and one (delayed) reflection, each received from different directions. Direct and reflected sound superimpose themselves at each ear of the listener in a different manner, which leads to differing interferences. Thus we obtain an interaural transfer function with comb-filter characteristics, i.e., with pronounced peaks and troughs of the level curve plotted v.s. frequency. As a result, a dissimilarity in the ear signals is produced which the auditory system interprets as a decorrelation, and which is perceived as a spatial extension of the auditory events. In the case of a single reflection, the spatial effect appears in the region of validity of the presedence effect (law of the first wave front, see Fig.3a), i.e., for reflections delayed between approx. 2ms and 50 to 80ms. However, only the range from 10ms is really effective, since shorter delay times can lead to disturbing distortions in timbre (so-called coloration). Spaciousness can still be perceived when the reflection level lies within a range of 20dB below that of the direct sound. Under the presedence effect the main center of gravity of the auditory event remains in the arrival direction of the first wave front. Thus, in spite of the expansion of the auditory event, the impression that the source of primary sound determines the perceived direction remains. The spatial effect is still partially retained after reflection delays higher than the echo threshold. Yet, because of perceptual disturbances due to the echo, this region of delays is generally not useful in room acoustics.

We began our description of spaciousness with the simplest case - that of one direct sound and a single reflection. In real rooms this situation is mostly more complicated. After the direct sound the listener is reached by up to six reflections of the first order from walls, ceiling and floors, followed by reflections of a higher order (multiple reflections) which succeed each other in decreasing intervals. The later reflections follow each other more and more closely - in the end they can no longer be measured separately. They form a diffuse reverberation field which declines exponentially after the sound source has died away. For the desired spaciousness in halls where music is performed, early reflections of a low order reaching the listener from lateral directions are the most important. The later reflections, including the reverberation, contribute in principle to the desired spaciousness, but they must be carefully dimensioned for two reasons.

- Late reflections, particularly those with a comparatively high degree of energy, can lead to disturbing echoes being perceived.
- A too strong and too long reverberation leads to a smearing of the sound signals in time and thus to a loss of sharpness in articulation and clarity.

However, the early lateral reflections not only create the desired spatial effect but also make a positive contribution to the sharpness of articulation and clarity.

One property of spaciousness which has not been sufficiently studied up to now, is its dependence on the sound level. This is also the reason why exact quantitive planning of spaciousness is not yet possible. Spaciousness increases sharply with increasing sound level, which can be explained by the nonlinear loudness characteristic of the auditory system, but may also be affected by the fact that by increasing the level, an increasing amount of reflected sound arrives at levels above the masked threshold, i.e., becomes audible regardless of the background noise.

The extent to which early lateral reflections reach the listener in a room - assuming that no electroacoustical measures have been taken - depends on the size and arrangement of the reflecting surfaces of the room. Since the side walls of a room are generally the most important surfaces for creating lateral reflections, the shape of the room takes on considerable importance. In the classical concert halls with a rectangular ground-plan, lateral reflections were achieved with parallel side walls which were not too far apart (maximum approx. 28m).

With equally good viewing conditions, one can accommodate more listeners in fan-shaped and arena-shaped halls than in rectangular ones. Consequently, in recent times many halls have been built with varying ground-plans. Their acoustic quality, however, has often been disappointing because only slight lateral energy has reached the listener in large areas of the room. These problems can be solved, for example, by installing appropriately dimensioned and arranged large sound reflectors in the hall so that they provide sufficient lateral energy at the listeners' seats. Such reflectors can also be made mobile, e.g., in multi-purpose halls. Fig.6 shows a recently constructed example. The hall shown, with an acoustically unfavorable plan (square 40m x 40m), can still yield thoroughly acceptable musical performances if the reflectors are swung out.

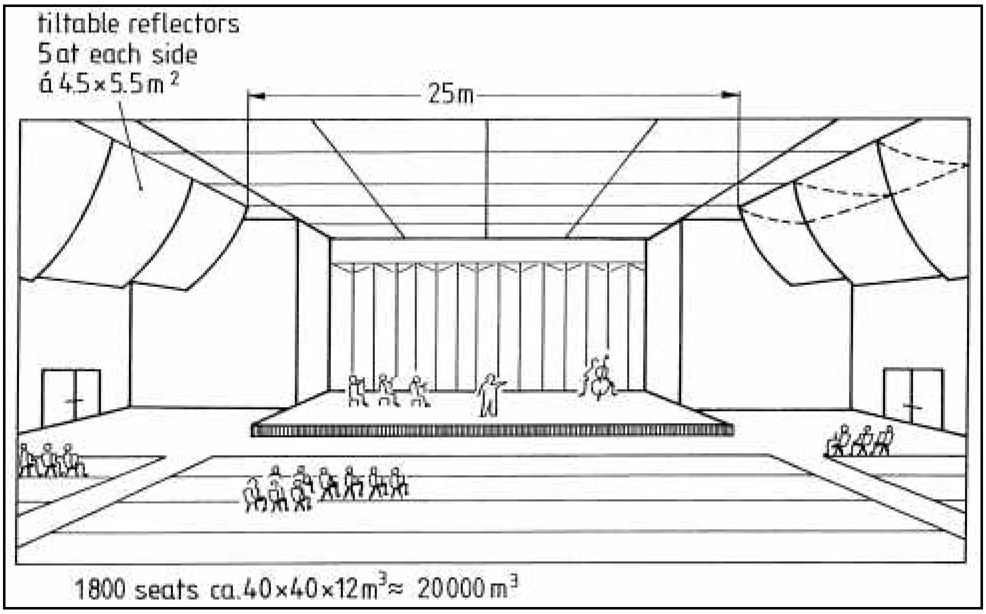

Fig. 6
Jupiter Hall Neuss - a multi-purpose hall with swinging (mobile) ceiling reflectors for creating early lateral reflections.

Up to now in room acoustics, when choosing the dimension and arrangement of reflectors, it has generally been assumed, due to lack of data, that as much early lateral energy as possible must be directed towards the listeners - without the actual sequence of reflections playing any role with respect to the creation of spatial effect. Yet, it should not be concluded from this that different sequences of reflections with the same overall energy are perceptually equal. For example, the time interval between direct sound and the first reflections thrown back is particularly important for the listeners' impression of the size of the room.

Psychoacoustical research has shown that auditory spaciousness is by no means a one-dimensional feature of the auditory event (e.g., Blauert & Lindemann, 1987). Subjects can differentiate between two components of spatial effect. The one component becomes perceptually dominant whenever the early lateral reflections contain considerable energy above 3000 Hz. This is chiefly described as width of the auditory event. The second component rests on the low frequency parts of early lateral reflections and leads to a greater depth of the auditory events up to the impression of being enveloped. These results will have to be considered in future acoustical design.

Besides purely acoustical methods for creating early lateral reflections, modern electroacoustics offer many possibilities for creating spatial effect which can even be variable. These electroacoustical methods will be increasingly used in the future because of their flexibility, even in halls where so-called serious music is performed.

References

BLAUERT, J., (1983), Spatial hearing - the psychacoustics of human sound localization , The MIT Press, Cambridge, MA

BLAUERT, J. & Lindemann, W., (1986), Auditory spaciousness: Some further psychoacoustic analyses, J. Acoust. Soc. Amer. 80, 533-541

CREMER, L. & Mueller, H.A., (1978/82), Die wissenschaftlichen Grundlagen der Raumakustik. S. Hirzel Verlag, Stuttgart. Engl. transl. by T.J. Schultz, Applied Science Publ. Essex, GB, 1982

DURLACH, N.I. & COLBURN, H.S., (1978), Binaural phenomena. In: E.C. Carterette & M.P. Friedman (eds.), Handbook of perception, vol.4, Academic Press, New York, 365-466

Gatehouse, R.W. (ed.), (1982), Localization of sound: Theory and application, The Amphora Press, Groton, CT

MILLS, A.W., (1972), Auditory localization. In: J. Tobias (ed.), Foundations of modern auditory theory, vol.2, Academic Press, New York, 301-345