PERCEPTION OF
REPRODUCED SOUND 1987
ISBN 87-982562-1-1

EDITORS
SØREN BECH
O. JUHL PEDERSEN

ON THE STEREOPHONIC IMAGING OF NATURAL SPATIAL PERSPECTIVE VIA LOUDSPEAKERS : THEORY

Günther Theile

Institut für Rundfunktechnik
Floriansmühlstraße 60
D-8000 München 45, F.R.Germany

ABSTRACT

The basic attributes relating to the auditory spatial impression are known and can be used for evaluation room acoustics of concert halls. Are they useful for evaluating reproduced sound? Principal considerations are presented, which are based on the association model. It is concluded that the stereophonic signals, not the resulting ear input signals in the case of loudspeaker reproduction, should include interaural attributes of listener's ear signals in the concert hall.

INTRODUCTION

In room acoustics the socalled "auditory spatial impression" is of particular interest. This term contains specific perceptual attributes of auditory events that are correlated to reflected sound in a room, e.g. concert hall, or living room. According to introduced terminology (see e.g. /1/, /2/, /3/, /4/) the formation of an auditory spatial impression is based on two primary perceptual attributes: One attribute is "reverberance", a characteristic "temporal slurring of auditory events" /3/, resulting from late reflections and reverberation. The other attribute is "auditory spaciousness", a characteristic "spatial spreading of auditory events" /3/, resulting from early lateral reflections (5 ... 80 ms delay related to direct sound). Based on these temporal and spatial characteristics of the auditory events the listener spontaneously perceives the acoustical properties of the actual room.

A large number of investigations have attempted to identify the auditory components that determine how listeners judge the quality of concert halls. More recent investigations demonstrated that spaciousness is a dominant attribute for getting "good spatial impression" (e.g. /3/, /5/, /6/, /7/). Auditory events, or in other words, the "apparent sources" are desired to fill larger amounts of spaces than in a free sound field under comparable conditions.

The dependence of spaciousness on delay time, level, angle of incidence, and spectrum of the early lateral reflections has been investigated. Also, the overall level of the direct sound and of the reflections has been found to be of central importance /7/, /8/. Various investigators attempted to define an index of spaciousness as a function of objective parameters of the sound field (e.g. /1/, /7/). The index depends on either the interaural degree of coherence k or the ratio of lateral sound energy to frontal or total energy ariving at the listener /4/. Thus, important characteristics of a concert hall are attempted to be measurable by defining an adequate index of spaciousness.

In summary, the basic physical quality parameters relating to the auditory spatial impression are known and can be used for evaluating and optimizing the room acoustics of concert halls (see /4/, /9/).

Are these quality parameters also useful for evaluating and optimizing reproduced sound? It is evident that this is true when using dummy-head stereophonic technique, because ideally listener's ear input signals are identical in both cases natural listening in the concert hall and presentation of dummy-head signals. However, when using conventional two-channel loudspeaker stereophony, there arise fundamental questions referred to the problem of imaging natural auditory spaciousness or natural spatial impression:

- Can natural spatial impression originate in loudspeaker stereophony?
- Which attributes of loudspeaker's input signals and/or listener's ear input signals are correlated with natural spatial stereophonic image?
- Which kind of microphone technique is favourable for spatial imaging?
- Which influence of listening room's response on the spatial stereophonic image is advantageous, which influence is disadvantageous?
- What can be concluded concerning the desirable acoustical characteristics of listening rooms, loudspeaker directivity index, and positions of loudspeakers?

At present, these questions are being studied. In this paper principal considerations on the first two questions are presented, which are based on the association model (10/, /11/, /12/). Particular emphasis is put on some theoretical fundamentals derived from the model.

OBJECTION TO SUMMING LOCALIZATION THEORIES

For stereophonic imaging, using common microphone/mixing technique and standard stereo loudspeaker arrangements, the so-called phantom sound source phenomenon is applied (details and references are given in /4/). The phantom sound source is being understood as an "fictitious sound source" because auditory events occur at positions where no real sound source is to be found. But the "fictitious sound source" between the loudspeakers implies the assumption, which does not hold true: it can not be assumed that the phantom sound source represents a (fictitious) "substitute sound source", which produces the same ear signal characteristics as the two coherently radiating stereo loudspeakers /10/.

This, however, was precisely the general hypothesis of the "summing localization therories": "Summing localization" is to be understood as the concept that summing signals result from the superimposed sound field to the ears, the components of which cannot be separated by the sense of hearing. It has been supposed that, during the localization of a phantom sound source, the sense of hearing recognises and evaluates the same attributes of the resulting ear signals as during the localization of an equivalent real sound source which is at the same location as the perceived phantom sound source.

These theories have been described in various studies, a detailed bibliography is presented in /4/. However, their validities are limited to directional hearing (distance hearing has not been taken into consideration), and even very often only to directional hearing in the horizontal plane. The main shortcoming of those studies lies in the fact that the regularities were measured "frequencydependent", i.e. that they were found in narrow-band signals. The importance of the spectral characteristics for the localization was unknown, or at least it was not taken notice of, so that the validity of the "summing localization theories" was never critically investigated for broad-band signals.

As an example, the following phenomenon should be mentioned, which has been studied in /10/: In summing localization, each ear input signal is derived from at least two loudspeaker signals those are displaced in time with respect to each other. Consequently, each ear input signal shows a comb-filter effect. The auditory event, however, has almost the same timbre as if only one of the two loudspeakers were driven; that is, the timbre is not correlated with the spectral attributes of listener's ear signals.

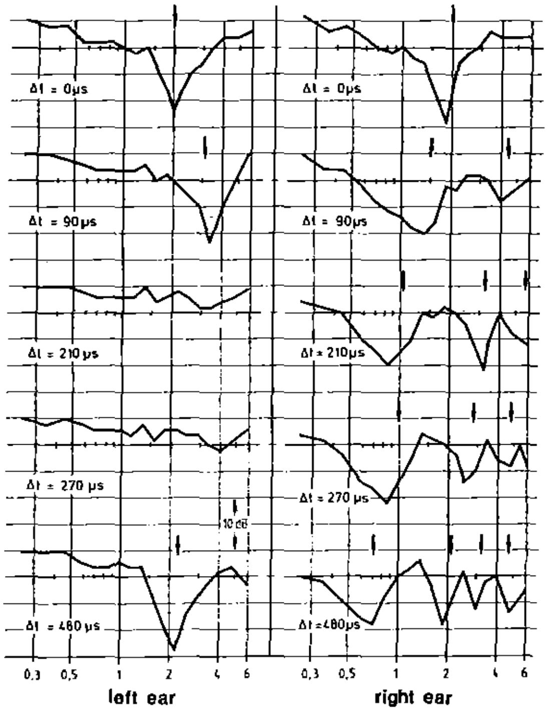

Fig. 1:
Lateral displacement of the phantom sound source caused by inter-channel time difference Δt: comb filter effect in the ear input signals

Figure 1 presents the comb-filter effect in the ear input signals, which can be measured in the case of inter-channel time difference Δt. The lateral displacement of the phantom sound source (towards the right loudspeaker in **fig. 1**) depends on Δt. It can be observed that the timbre of the phantom sound source, as well as the distance, elevation and width do not vary adequately if Δt is varried in the range of 0 ... 300 µs, or the head is moved to the side in the range of 10 cm /10/.

As a result an objection to summing localization was formulated /10/, /11/. It states that the monaural and interaural spectral attributes of broad band ear input signals, which result from the superimposition of the two sound fields produced by the stereo loudspeakers, are not compatible with the perception of

timbre - distance - elevation - width

of the phantom sound source, if it is assumed that the ear input signals are processed in their entirety.

The validity of the summing localization theories is severely limited. Therefore these theories could be helpful neither in understanding the function of the hearing sense for localization in a superimposed sound field /10/ nor in applying the phantom sound source phenomenon to stereophonic recording and reproduction technique /13/. (State and trends of development of stereophony are discussed and a detailed bibliography in this field is presented in /13/).

This statement is contradictory to commonly established understanding of today. On basis of summing localization theories it is even today tried to assess stereophonic techniques (see e.g. /14/, /15/). As a recent example, Lipshitz has concluded that coincidence microphone techniques are most advantageous for getting a natural spatial impression /14/:

> "I believe that spaced-microphone recording techniques are fundamentally flawed, although highly regarded in some quarters, and that coincident-microphone recordings are the correct way to go".

His arguments are based on an analysis of the resulting interaural characteristics of listener's ear signal, according to the principle of summing localization:

> "The level and time (or phase) differences at the listener's ears are <u>not</u> the same as those at the loudspeakers ... It is important that, as far as possible, the two loudspeaker signals combine at the listener's ears to produce cues which are compatible with natural hearing."

Unfortunately, "natural" interaural attributes of resulting broadband ear signals are still claimed. This, however, is not possible by using conventional stereophonic techniques. An "authentic" spatial impression can only be reproduced by using head-related (dummy-head) stereophonic techniques (see e.g. /11/, /13/, /16/). Summing localization theories do not differentiate head-related techniques (reproduction of head-related, authentic distances of auditory events) from "loudspeaker-related" techniques (simulation of auditive perspective between the stereo loudspeakers, see section 2.2). Thus, instead of summing localization theories, a more general localization theorie is necessary, the validity of which includes broadband phantom sound sources as well as real sound sources, and which provides a uniform explanation of fundamental phenomena of spatial hearing. It is represented by the "association model" /10/, /11/.

THE ASSOCIATION MODEL

This model is based on the assumption that auditive spatial perception fundamentally results from two separate processes. Each of these two processes occurs by means of an associatively guided pattern recognition: a stimulus, resulting from a sound source, initially induces a "location association" and secondly a "Gestalt association". ("Gestalt" is a term used in perception research. Here it means the property of the sound source, e.g. timbre). The characteristic feature of the localization model resides in the two-stage processing of the stimulus.

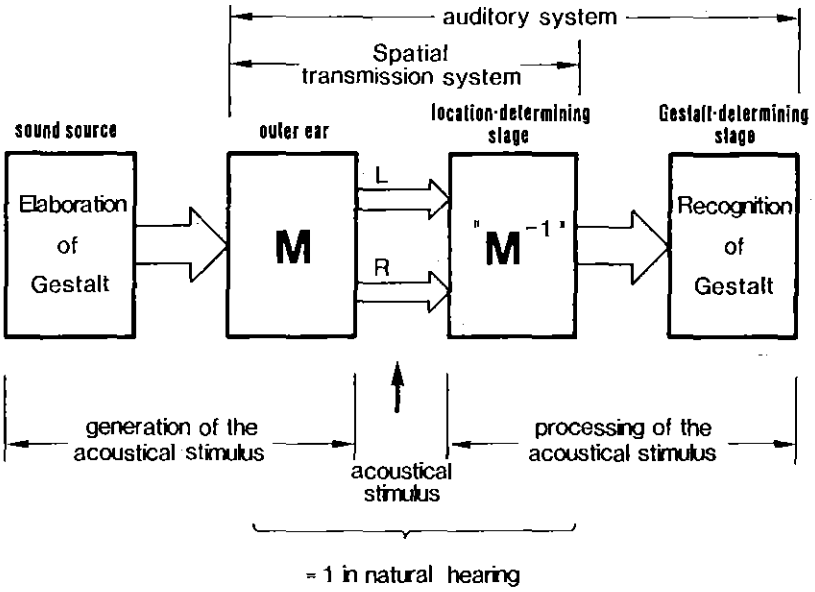

Fig. 2: Principal function of the association model

Fig. 2 shows the principal function of the model. The auditory system consists of the outer ear, represented by transfer function M, and of the location-determining stage, represented by the transfer function M^{-1}, inverse to M. The dependence of ear-signals on the source location is indicated as a form of coding of spatial information (M), which enables the spatial information (M^{-1}) to be decoded. The operation can be described as that of an adaptive filter with a transmission function M^{-1} which, as a result of the associative recognition of the model, is in inverse relation to M, the transmission function of the outer ear, which depends on the location of the source.

The inverse filtering M^{-1} frees ear signals from any influence by the outer ear before the source signal reaches the Gestalt-determining stage. This means that, in natural hearing, spectral features caused by the directivity of the outer ear are apprehended in such a way that they do not occur as tone color defects **(Fig. 2)**. The association model indicates that the perception of timbre is independent of the source location. This timbre phenomenon can be verified readily, but has been described in publications only very recently /10, /16/, and it is called "invariability of timbre". The auditory system identifies the location and property (e.g. timbre) of the sound source, and that is why timbre is not totally determined by the power spectrum of the ear signals. This fundamental attribute of spatial hearing is represented in the model by the term $M \cdot M^{-1} = 1$.

However, M^{-1} filtering only occurs when the effect of the outer ear in the formation of the ear signals is "recognized". This inverse filtering occurs normally in "natural listening", that is, when ear signals of sufficiently broad band present the corresponding outer ear features. Since the inverse filtering process is assumed to be based on an spontaneous comparison of actual pattern of auto- and cross-correlation functions (resulting from the incoming stimulus) with a set of stored auto- and cross-correlation patterns (which has been learned), the auditory system is able to recognize the location of unknown sound sources. Details of this associatively guided pattern recognition, preconditions and consequences, are given in /10/ and /11/.

In the case of stereophony, the auditory system identifies the locations of the two loudspeakers as well as the properties of the loudspeaker signals. In particular, the inter-channel relationship of the stereo signals is recognized. If the loudspeaker signals are coherent, the two stimulus responses of the locationdetermining stage fuse into each other at the Gestalt-determining stage. An average spatial information results from the fusion, which determines the location of the resulting auditory event (phantom sound source between the loudspeaker).

Perceived direction of phantom sound source

The directional information results directly from the relationship of the stereophonic loudspeaker signals: Because of inverse filtering M^{-1}, the stereophonic signal relationship is present in the relationship of the two stimulus responses of the location-determining stage. For instance, if differences in intensity and/or time of arrival becomes effective during the fusion, lateral displacement of auditory event results (for details see /10/, /11/). Thus, corresponding data (lateral displacement versus intensity or time difference) can be measured via "lateralization experiments" (headphone reproduction) and in the same way and with corresponding resultsvia stereo loudspeaker reproduction /10/.

The direction of the phantom sound source in the association model is not determined by the resulting summing signals at the ears but rather simply by the relationship of stereo loudspeaker signals.

Perceived distance of phantom sound source

Since the distances of the two loudspeakers are asssumed to be identified by the location-determining stage (preconditions are discussed in /10/), an average distance information results from the fusion process in the Gestalt-determining stage. In other words, the distance of the phantom sound source is equal to the (average) distance of the stereo loudspeakers. It is not determined by the resulting summing signals at the ears or influenced by the comb filter effect described above.

STEREOPHONIC IMAGING OF SPATIAL PERSPECTIVE

The fundamental statement of the association model is: The auditory system recognizes the effect of the outer ear on the input signal. Thus, it is able to identify the properties of the sound source. Discriminating both location and properties of the sound source from the ear signal's attributes is a spontaneous effort of the auditory system. In the case of loudspeaker stereophony, this happens simutaneously for the two loudspeakers, consequently the locations of the loudspeakers just as the inter-channel signal relationship is identified, resulting in a corresponding stereophonic image between the loudspeakers.

It follows: The distance of the phantom sound source is fundamentally the same as the distance of the loudspeakers. A presentation of "spatial perspective" can only be achieved by simulating distances. The sound engineer must deliberately employ certain phenomena of spatial hearing in order to produce, for example, an auditory image characterized by perception of "spatial perspective" (of the concert hall, not listening room!). This is directly comparable to the presentation of visual perspectives: there is only actually the distance of the figure or screen (fig. 3) - that represents the distance of the loudspeakers - but a perception of spatial perspective is simulated, because phenomena of spatial vision are employed. Parallel lines join in the distance, equal sizes and distances shrink.

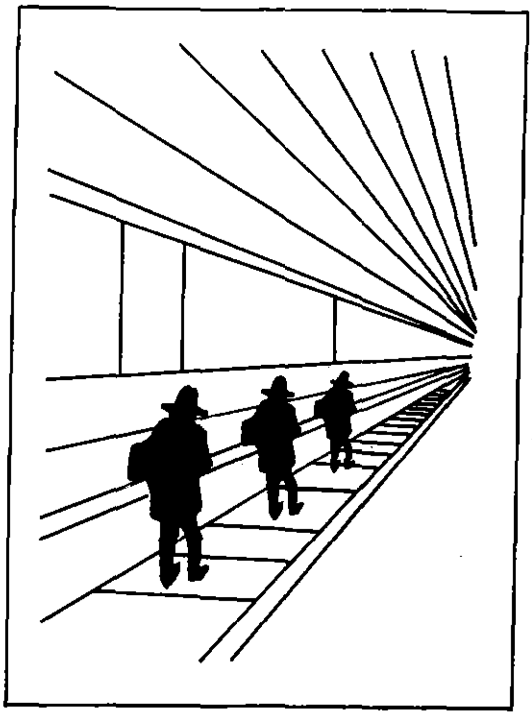

Fig. 3: The distance of this picture can be compared with the distance of stereo loudspeakers. The visual perspective, which is simulated by applying phenomena of spatial vision, can be compared to the stereophonic perspective, which can be simulated by applying corresponding phenomena of spatial hearing.

The spatial perspective in the stereophonic image can be simulated with different "imaging elements", which are deduced from natural hearing:

1. The direct part of the original sound field (the "first wave front") contains information of distance, because the auditory system as a result of experience associates the relationship of level/spectrum of the signal with the relationship volume/timbre/distance of the sound source (details in /4/, /17/, /18/).
2. In the median plan, the relationship direct part/indirect part of the sound field contains important information of distance, which the auditory system evaluates, based on experience. This information gives the auditory event a spatial perspective (details in /4/, /17/, /18/).
3. The reverberation and lateral reflections lead to ear signals, which cause auditory spatial impression, as described in section 1.

As regards the perception of distance (point 1 and 2), it seems to be evident to introduce the corresponding physical attributes into the signals to be reproduced. It is well known and common practice to **simulate** distance by choice of a proper microphone location or by artificial aftertreatment of the microphone signal, using reverberation device, delay lines, equalizers and mixing desk. These two imaging elements are not specific stereophonic elements, they are also effective in the case of monophonic representation. However, they demonstrate what "imaging of spatial perspective" means: Although the auditory system identifies the distance of the (mono) loudspeaker (because of the response of the listening room) a fictive distance is percepted, that is, the distance of auditory event is **simulated,** but no reality.

As regards the perception of spatial impression (point 3), the corresponding physical attributes can be employed only in the case of stereophonic representation.

The relevant imaging element is of specific stereophonic nature. Following the new understanding about nature of phantom sound sources presented above and ignoring summing localization theories, stereophonic imaging of spatial perspective is based on introducing corresponding physical attributes of the ear signals (which correlate with phenomena of natural spatial hearing) into the stereophonic signals. This is contradictory to the consequence of summing localization theories, which attempt to introduce them into the resulting ear signals of the listener.

Thus, in the case of conventional loudspeaker stereophony, the relevant attributes are introduced into the stereophonic signals in order to **simulate** spatial perspective, that is to say, the perceived reverberance and spaciousness are no auditive reality (if presentation of authentic spatial impression is desired, the use of head-related stereophony ist absolutely necessary). In other words: In the case of conventional loudspeaker stereophony, the spatial perspective is due to the inter-channel properties of loudspeaker input signals.

CONCLUSIONS

Two basic requirements can be concluded for getting optimum fidelity of spatial perspective:

1. The inter-channel relationship of the stereophonic loudspeaker input signals should be as similar as possible to the interaural signal relationship in the case of natural hearing.
2. The properties of loudspeakers and listening room should enable the auditory system to identify the inter-channel relationship of the stereophonic signals as exact as possible.

This result is of basic interest for evaluating and optimizing stereophonic recording and reproduction technique, in particular, for giving answer to the three open questions stated in section 1. Practical studies have been carried out.

Literature

/1/ Reichardt, W.; Lehmann, U.: Raumeindruck als Oberbegriff von Räumlichkeit und Halligkeit. Erläuterung des Raumeindrucksmaßes R. (The spatial impression as a broad concept including spaciousness and reverberance. Defining the index of spatial impression R)

/2/ Kuhl, W.: Räumlichkeit als eine Komponente des Höreindrucks (Spaciousness as a component of the auditory impression)

/3/ Blauert, J.; Lindemann, W.: Auditory spaciousness. Some further psychoacoustic analysis. J. Acoust. Soc. Amer. 80, 1986, 533.

/4/ Blauert, J.: Spatial Hearing - The Psychophysics of Human Sound Localization. The MIT-Press, Cambridge, Mass.

/5/ Schroeder, M.R.; Gottlob, D.; Siebrasse, K.F.: Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters. J. Acoust. Soc. Amer. 56, 1974, 1195-1201.

/6/ Plenge, G.; Lehmann, R.; Wettschurek, R.; Wilkens, H.: New methods in architectural investigations to evaluate the acoustic qualities of concert halls. J. Acoust. Soc. Amer. 57, 1975, 1291-1299.

/7/ Barron, M.; Marshall, H.A.: Spatial Impression Due to Early Lateral Reflections in Concert Halls. J. Sound Vibr. 77, 1981, 211-232.

/8/ Wettschurek, R.: Über die Abhängigkeit raumakustischer Wahrnehmungen von der Lautstärke. (On the dependence of spatial acoustic perceptions on loudness) Dissertation Techn. Universität Berlin, 1976.

/9/ Cremer, L.; Müller, H.A.: Principles and applications of room-acoustics, vol. 1, Applied Science Publishers, Essex, England, 1982.

/10/ Theile, G.: Über die Lokalisation im überlagerten Schallfeld (On localization in a superimposed sound field). Dissertation, Technische Universität Berlin, 1980.

/11/ Theile, G.: Zur Theorie der optimalen Wiedergabe von stereophonen Signalen über Lautsprecher und Kopfhörer (On the theory of the optimal reproduction of stereophonic signals over loudspeakers and headphones). Rundfunktechn. Mitt. 25, 1981, 155-169.

/12/ Theile, G.: On the standardization of the frequency response of high quality studio headphones. J. Augio Eng. Soc. 34, 1986, 956-969.

/13/ Steinke, G.: Stand und Entwicklungstendenzen der Stereofonie. (Stereophony: state and trends of development) Technische Mitt. des RFZ 28, 1984, 1-10 und 25-32.

/14/ Lipshitz, S.P.: Stereo Microphone Techniques ... Are the Purists wrong? J. Audio Eng. Soc. 34, 1986, 716-744.

/15/ Bennett, J.C.; Barker, K.; Edeko, F.O.: A New Approach to the Assessment of Stereophonic Sound System Performance. J. Audio Eng. Soc. 33, 1985, 314-321.

/16/ Platte, M.-J.: Zur Bedeutung der Außenohrübertragungseigenschaften für den Nachrichtenempfänger "menschliches Gehör" (On the meaning of external ear transfer properties for the receiver of information known as the "human auditory system"). Dissertation, Techn. Hochsch. Aachen, 1979.

/17/ Laws, P.: Zum Problem des Entfernungshörens und der Im-Kopf-Lokalisiertheit von Hörereignissen (On the problem of distance hearing and the localization of auditory events inside the head). Dissertation, Techn. Hochsch. Aachen, 1972.

/18/ Plenge, G.: On the difference between localization and lateralization. J. Acoust. Soc. Amer. 56, 1974, 944-951.