ON THE LOCALISATION

IN THE SUPERIMPOSED SOUNDFIELD

by

Dipl.-Ing. Günther Theile

Dissertation
submitted for the award of the degree
Dr.-Ing.
Fachbereich 21 (Umwelttechnik)
Technische Universität Berlin

Examiners :
Prof. Dr.-Ing. G. Boerger
Prof. Dr.-Ing. L. Cremer
Prof. Dr. phil. G. Plenge

Date of oral examination: 30.April 1980

Berlin 1980
D83

Translated by:

Tobias Neher, PhD
t.neher@surrey.ac.uk

Content

Abstract

Theile, Günther

On the localisation in the superimposed sound field

A new localisation model, the “association model”, is introduced. It is based on the fundamental notion that localisation takes place because of auditory experience (perceptual process). The ear signals’ dependency onto the location of a sound source is interpreted as a mechanism for encoding spatial information, whereby knowledge of this mechanism enables the decoding of this spatial information. In the superimposed sound field, localisation therefore manifests itself as a process for the simultaneous decoding of different spatial information.

The model contains a “location association stage”, which performs this process. This stage precedes a second central processing stage, the “gestalt association stage”. Each of these two processing stages manifests itself in the form of a pattern selection process, which is controlled in an associative manner. A sound stimulus leads to a location association during the first and to a gestalt association during the second processing stage. The two stages determine the auditory event properties in a conjoint manner.

The rigorous differentiation of these two stimulus evaluation stages corresponds entirely to the two elementary areas of auditory experience. The received ear signals can be attributed to the two sound source characteristics of “location” and “signal”, which always occur in a pairwise fashion.

This two-stage, associative selection process in the association model thus represents a uniform approach for the consistent explanation of important phenomena of spatial hearing. In contrast to summing localisation theories, the model is in agreement with the psychoacoustic principles governing the formation of “phantom sources” with respect to their directions, distances, elevations and sound colours. Moreover, it supplies plausible explanations for the “law of the first wavefront”, the “cocktail party effect” as well as “inside-the-head locatedness”. As a consequence, the model also postulates that the human hearing system’s localisation function can neither be investigated by means of narrow-band signals, nor can it be scrutinised under lateralisation conditions.

1. Introduction

In the context of spatial hearing, BLAUERT 1974 defined the term “localisation” as the mapping of the location of an auditory event to certain characteristics of a sound event. If, based on BLAUERT 1969, the term “auditory event” is generally understood to describe the “auditory perception determined by temporal, spatial and other attributes” and the term “sound event” as the equally determined “physical aspect of the listening process”, then initially one obtains an important descriptive basis for investigations into spatial hearing.

During the course of this work, however, it has been found that as soon as more than one sound source determines the properties of the ear signals, a more precise definition of terms is required.

Under natural conditions, the listening process predominantly takes place in a superimposed sound field. Normally, a multitude of interfering sources [e.g. reflections from the floor or room boundaries (mirror sources)] but also equipollent sound sources (e.g. different speakers in the vicinity of a listener) produce ear signals that do not correspond in any way to the elementary case of a single sound source. Yet, even under such sound field conditions, the human hearing system is capable of “carrying out a very meaningful selection, ordering and grouping process” (PLENGE 1973), so that, despite the superposition, the auditory event mappings can be accomplished. In this respect, two well-known auditory phenomena are the “law of the first wavefront” (CREMER 1948) and the “cocktail party effect” (CHERRY 1953, 1954).

It appears that in the superimposed sound field the localisation mechanism relies on underlying processes that are not directly applicable to the localisation of a single sound source in a free sound field. These are all those processes that enable a separate evaluation of certain ear signal components.

What are the ear signal components that the auditory system discriminates in the superimposed sound field? Which characteristics of a sound event allow for this to happen?

The literature is full of investigations addressing these two questions. Most notably, this is the case for all studies into the phenomena of the “phantom source”, the “law of the first wavefront” and the “cocktail party effect” (see BLAUERT 1974 for references as well as the following sections of this thesis). Furthermore, investigations in the area of binaural signal detection (see BLAUERT 1974 for references as well as DOMNITZ / COLBURN 1976, 1977, HAFTER 1977, KOENIG et al. 1977, ALLEN et al. 1977, HAWKINS et al. 1978, HAFTER et al. 1979) and some studies concerning particular questions of auditory perception (e.g. see PLENGE 1973 for references as well as WETTSCHURECK 1976, MASSARO / WARNER 1977, POLLACK 1977, DANNENBRING / BREGMAN 1978, PALEFF / NICKERSON 1978, LEHRINGER 1979) are of relevance, too. The various individual phenomena of spatial hearing have often been scrutinised experimentally in great detail, and for the predominant part of the psychoacoustic principles that were thus identified there are models attempting to describe the hearing system's various functions.

The majority of these are communications engineering models that, within their areas of validity, specify relationships between certain ear signal characteristics and attributes of auditory events. However, due to their areas of validity being restricted, it has to be asked whether they can provide information improving the understanding of the function of the whole evaluation process of spatial hearing. For instance, what is the point of a model that only caters for directional hearing in the context of stereophony and that is only valid in the horizontal plane, but not for broadband signals and not for estimating the distance of an auditory event (see WENDT 1963)?

Presumably, it is precisely the relationships between the different sub-aspects of spatial hearing, which can bring about the desired comprehension of the functioning of the human hearing system. Yet, these connections are hitherto relatively unknown. To give four examples:

1. In the case of stereophony in the superimposed sound field, there arise certain auditory events in-between the loudspeakers because of the resultant ear signals ("phantom source", see Chapter 2). Although it is possible to create identical auditory events using a suitable single sound source, the two sets of ear signals will not be identical in terms of frequency content and degree of coherence (see Section 2.2). Until now, the connection between the localisation of a single sound source in the superimposed and in the free sound field has not been resolved consistently (see Section 4.2).

2. The areas of validity of the "law of the first wavefront" and of "summing localisation" are given by different time delay regions (see Section 2.1). Both phenomena arise in the superimposed sound field due to the second sound event arriving later compared to the first one. The demarcation of these two phenomena is based on different psychoacoustic principles governing the direction of an auditory event – possibly also because the first effect is of importance for listening in enclosed spaces, whilst the second effect is relevant to applications of stereophonic sound reproduction. It is assumed, however, that a function of the hearing system (applicable to listening in the superimposed sound field) can be found, which can conjointly explain the auditory event mapping for all signal delays (see Section 4.3.1).

3. Lateralisation experiments provide information about the evaluation of interaural signal differences. Based on these, one can also develop models for the

occurrence of laterally displaced auditory events. Nevertheless, there is no verified hypothesis that allows for a generalisation of these findings with respect to spatial hearing. Hitherto, the functional relationship between lateralised and localised auditory event locations is unknown (see Chapter 5).

4. The "cocktail party effect" implies that a target signal arriving from a certain direction will be masked less by an interfering signal arriving from a different direction if a subject listens binaurally rather than monaurally. Generally speaking, for both the superimposed sound field as well as headphone-based sound reproduction, the masking threshold for the target signal decreases if the target and masker signals produce auditory events at different locations. It seems that lateralisation experiments conducted in the area of binaural signal detection are clearly related to the question of which ear signal components can be discriminated by the hearing system in the superimposed sound field. Up to now, however, little is known about the significance of so-called "detection models" to the localisation process in the superimposed sound field (see Section 4.3.2).

These examples indicate that rather than investigating it "holistically", the spatial hearing mechanism of the auditory system has largely been studied in terms of demarcated sub-areas so far. Perhaps this may be explained by the efforts of communications engineers to model immediately any observed relationships between properties of sound and auditory events in the form of psychoacoustic principles, thereby necessitating a precise demarcation.

The above examples further show that especially the limits of the areas of validity of individual auditory phenomena as well as possible interrelations between these phenomena must be of particular interest, which can also be said about the findings from other scientific fields such as neurophysiology and perceptual research. In this respect, the use of associative storage technology, signal processing (e.g. see WIGSTROEM 1974, WILLWACHER 1976, KOHONEN 1977, BOHN 1978) and possibly also holography (see WESS / ROEDER 1977) may create new possibilities (see Section 3.2.1).

In this thesis, a model for the localisation process in the superimposed sound field will be described that takes into account a large number of different auditory phenomena in a *uniform* manner. It is based on the notion that localisation occurs due to a "comparison of current and learnt stimulus patterns" (PLENGE 1973). That is, the mapping of auditory events takes place as a result of auditory experience. If the ear signals' dependency onto sound source location is interpreted as a mechanism for encoding spatial information, then knowledge of this dependency can be seen as a key for decoding the spatial information. In the superimposed sound field, localisation therefore manifests itself as a process for the simultaneous decoding of different spatial information. Depending on the number and properties of the involved sound events, this process will work completely, partially or will not work at all.

What are the involved sound events? At this point, it is appropriate to define the term "sound event" precisely. So far, it is unclear how two sound events are actually to be discriminated. If, in accordance with BLAUERT 1974, one considers exclusively "the physical side of the listening process", only the effects of the head and pinnae in the sound field will be taken into account. In this case, however, two different, simultaneous sound events will only exist if they exhibit different spatial properties (source locations, propagation directions of the wavefronts etc.). With regard to the superimposed sound field, this view does not seem meaningful, since the listening process will then be substantially characterised by the discriminability of individual signal components.

Two spatially separated loudspeakers, for example, can, depending on the properties of the emitted signals, either produce two simultaneous auditory events at different locations or a single auditory event at a third location. What is more, a single loudspeaker radiating two different signals can produce two simultaneous auditory events. The evaluation processes in the hearing system, which lead to the formation of the auditory event location and gestalt, always determine the properties of the auditory events in a conjoint fashion. It seems that they influence each other. The physical side of the localisation process can only be explained with the help of sound events that cannot just be discriminated in terms of their spatial characteristics. For the remainder of this thesis, the following definitions will therefore apply:

> Sound event:
> A sound event is that part of a sound, which stems from a single sound source and which determines or influences the associated auditory event with respect to its location and gestalt.

Localisation will thus be defined as follows

> Localisation:
> Localisation is the mechanism/process that maps the location of an externalised auditory event to certain characteristics of one or more sound events.

The aim of this work is to uncover relationships between individual localisation phenomena in the superimposed sound field, as this can possibly lead to a better understanding of the functioning of our hearing system with regard to spatial perception. This objective does not appear inappropriate or unrealistic, provided that localisation is strictly regarded as the result of a perceptual process that is solely possible because of auditory experience.

2. .The "phantom source"

An important special case of localisation in the superimposed sound field is if several sound events are mapped to a common auditory event, so that its location does not correspond to the ones of the sound sources. This case is referred to as a "phantom source", since an auditory event is perceived to be at a position where no sound source is actually located.

Yet, bearing in mind the definitions given above, the term "sound source" cannot be utilised for describing an auditory event. In the literature, a phantom source is thus also interpreted as an imaginary, fictitious sound source, with the help of which one tries to describe the physical side of the localisation process. One proceeds on the assumption that, in principle, a phantom source constitutes a substitute sound source producing the same ear signal characteristics in its sound field as it would in a corresponding superimposed sound field at the same listening position.

As will be shown below, in many cases serious objections have to be raised with regard to this assumption (see Sections 2.2 and 4.2). Nevertheless, for the time being, the term "phantom source" shall be used to characterise particular sound field constellations in the usual manner. Unless specified otherwise, "phantom source" specifically refers to the sound field produced by two coherent signals that are played back over a conventional, loudspeaker-based stereo set-up (see Figure 1).

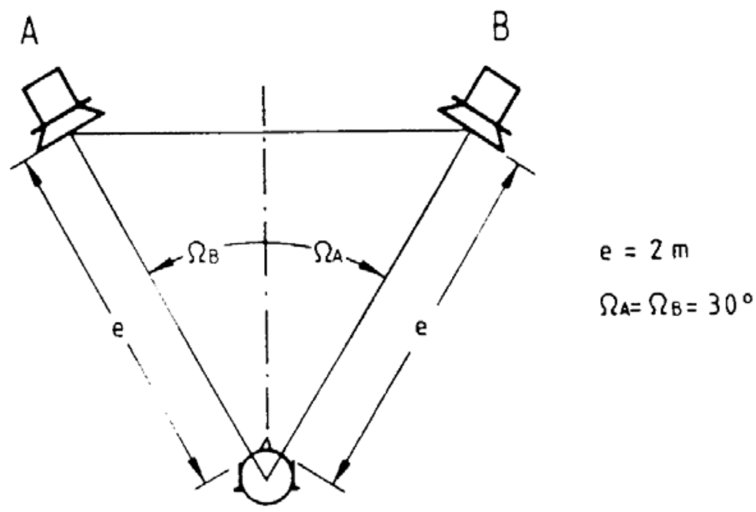


Figure 1:

Conventional, loudspeaker-based stereo set-up

This special case of localisation in the superimposed sound field is not just of particular relevance to the electroacoustic reproduction of spatially distributed sound sources. As already indicated above, important insights into the functioning of the auditory system with regard to spatial hearing can possibly be gained if individual phenomena can be integrated into a generally valid theory of auditory perception. In this context, the psychoacoustic regularities governing the phantom source phenomena appear to be especially important. These have been scrutinised and described in a multitude of studies (e.g. DE BOER 1940, WENDT 1963, BOERGER 1965, ORTMEYER 1966, DAMASKE 1969/70, BLAUERT 1970, GARDNER 1973, THEILE / PLENGE 1977, LOY 1978), and different theories have been developed for their explanation.

2.1 The hypothetical principle of summing localisation

All of the known phantom source theories have in common that they are based on the fundamental assumption that "summing localisation" (WARNCKE 1941) takes place. Summing localisation is understood to imply that the superposition of sound fields at the ears leads to summed signals, the components of which the human hearing system is unable to discriminate. For the localisation of a "phantom source" and a corresponding real source that is situated at the same location as the phantom source one therefore assumes an equivalence of the ear signal characteristics at the left and right ear, respectively. The studies dealing with summing localisation describe different theories of equivalence. However, their areas of validity are limited to directional hearing, often even to directional hearing in the horizontal plane. Examples are the well-known summing localisation theories by LEAKEY 1959, FRANSSEN 1960/63, MAKITA 1962, WENDT 1963, MERTENS 1965.

2.2 The "spectral objection" to summing localisation

The summing localisation models take into account the interaural phase, time and intensity differences that occur in the superimposed sound field, but they do not consider the resultant spectral characteristics of the ear signals. The sound field at each ear is composed of components that are delayed with respect to each other, so that each signal arriving at the two ears will be spectrally modified in accordance with the comb filter effect (see Figure 2).

Especially the most recent experiences with head-related, stereophonic recording-reproduction systems have shown that even small linear distortions can lead to inside-the-head locatedness (PLENGE 1973, NAKAMURA 1976, LAWS et al. 1976/77, BLAUERT et al. 1978). The localisation, i.e. the mapping of externalised auditory events, requires precisely the characteristic spectral attributes that, under natural listening conditions, are provided by the outer ear. This finding contradicts the summing localisation theories. The linear distortions resulting from the comb filter effects do not give rise to spectral characteristics that a substitute sound source situated at the same location as the phantom source would produce. Nonetheless, the phantom source is perceived outside the head, i.e. at approximately the average distance of the auditory event locations that the two loudspeaker signals produce individually (REICHARDT / HAUSTEIN 1968). The comb filter effect at the ears does not influence the distance perception.

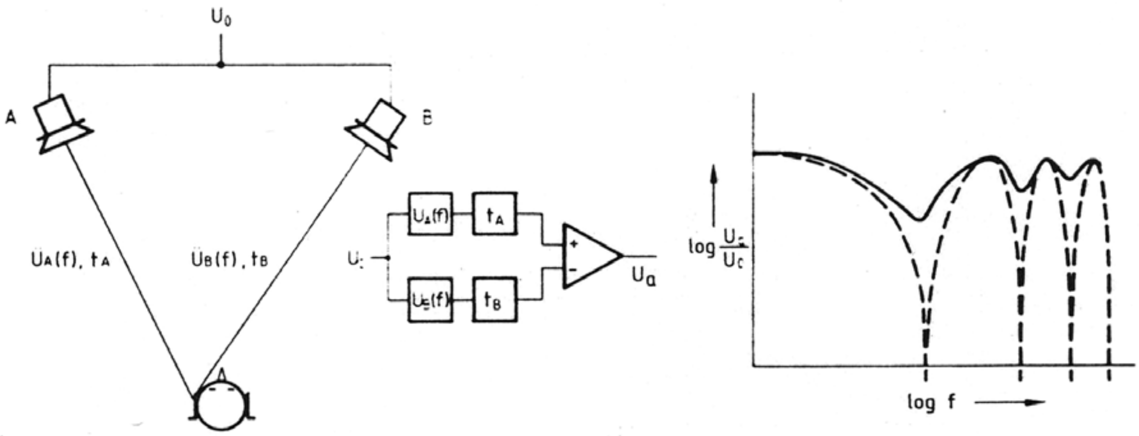


Figure 2:

Graphical illustration of the comb filter effect at the ears for the case of a phantom source. The solid line applies for $Ü_A \neq Ü_B$.

Another objection to the summing localisation principle also derives from the effective spectrum. In the median plane, characteristic frequency bands determine the direction of an auditory event (BLAUERT 1969). Accordingly, the elevation of an auditory event results from the spectral properties of the ear signals (BLOOM 1977). Notwithstanding, this fact is not generally compatible with summing localisation, as equivalent interchannel level and time differences lead to the same auditory event locations. If the loudspeakers are set up symmetrically with respect to the listener's median plane, then, depending on the applied interchannel level difference, the phantom source will move in the frontal plane at a constant distance. The same effect occurs if instead of the level difference a time difference is varied. A certain elevation angle can be achieved both with an interchannel level and time difference. Yet, in these two cases the comb filter effects at the ears lead to very different spectra, also in the crucial region around 8 kHz (see BLOOM 1977).

Thus, it seems that reservations have to be expressed with respect to the summing localisation principle, because in terms of its spectrum the phantom source may not be considered a substitute sound source.

This "spectral objection" relates to the sound colour of the phantom source, too. It can be shown that equivalent interchannel level and time differences, which lead to the same phantom source direction, also create the same phantom source sound colour, even though the two scenarios will result in different spectra. This phenomenon can easily be demonstrated in a simple listening test (Figure 3):

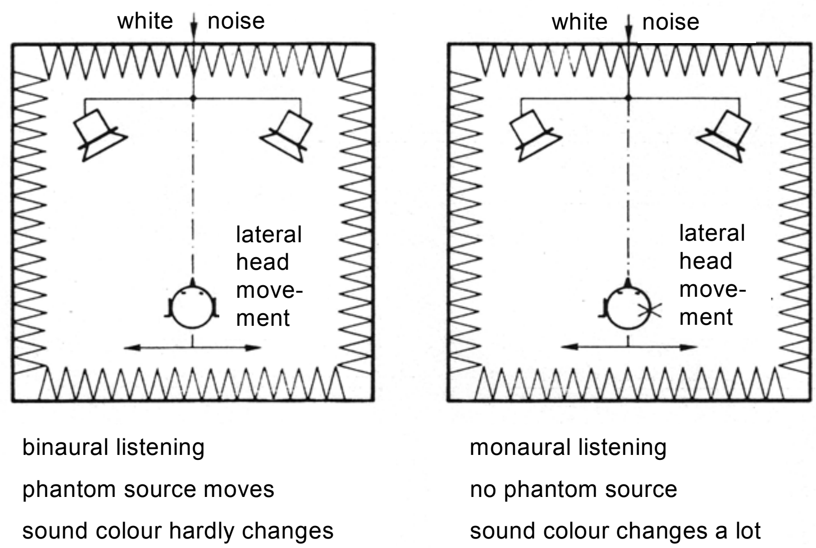


Figure 3:

Testing the suppression of sound colouration. The comb filter effect can only be detected when listening monaurally.

The listener should be positioned in front of two loudspeakers that are set up in an anechoic chamber and that radiate coherent white noise (at a suitably low level). Lateral head movements cause lateral displacements of the phantom source (time-based stereophony), whilst the sound colour hardly changes. Moving the head in the same manner whilst covering one ear will lead to clearly perceivable sound colour changes. The sound colouration, which is caused by the comb filter and which can be detected at each ear individually, disappears when listening binaurally, i.e. as soon as a phantom source arises. Its occurrence implies the suppression of sound colouration. This effect was first pointed out by THEILE 1978.

Before investigating the connection between the suppression of sound colouration and the localisability of phantom sources by means of a listening test, the linear distortions occurring at the left and right ear due to the comb filter effect shall be quantified.

2.2.1 Measurement of the comb filter effect

An artificial head (NEUMANN KU 80) is set up at the listening position of a conventional stereo layout (see Figure 1). The levels of the frequency bands (see ZWICKER / FELDKELLER 1967) of the artificial head signals in-between 0.25 and 6 kHz are measured as a function of the time delay Δt between the loudspeaker signals. The time delays Δt = 0, 90, 210, 270 and 480 µs are achieved by displacing the artificial head by 0 to 16 cm to the right. This corresponds to a shift in phantom source direction from Φ = 0° ... +25°. The occurring level changes are negligible. The results are shown in Figure 4.

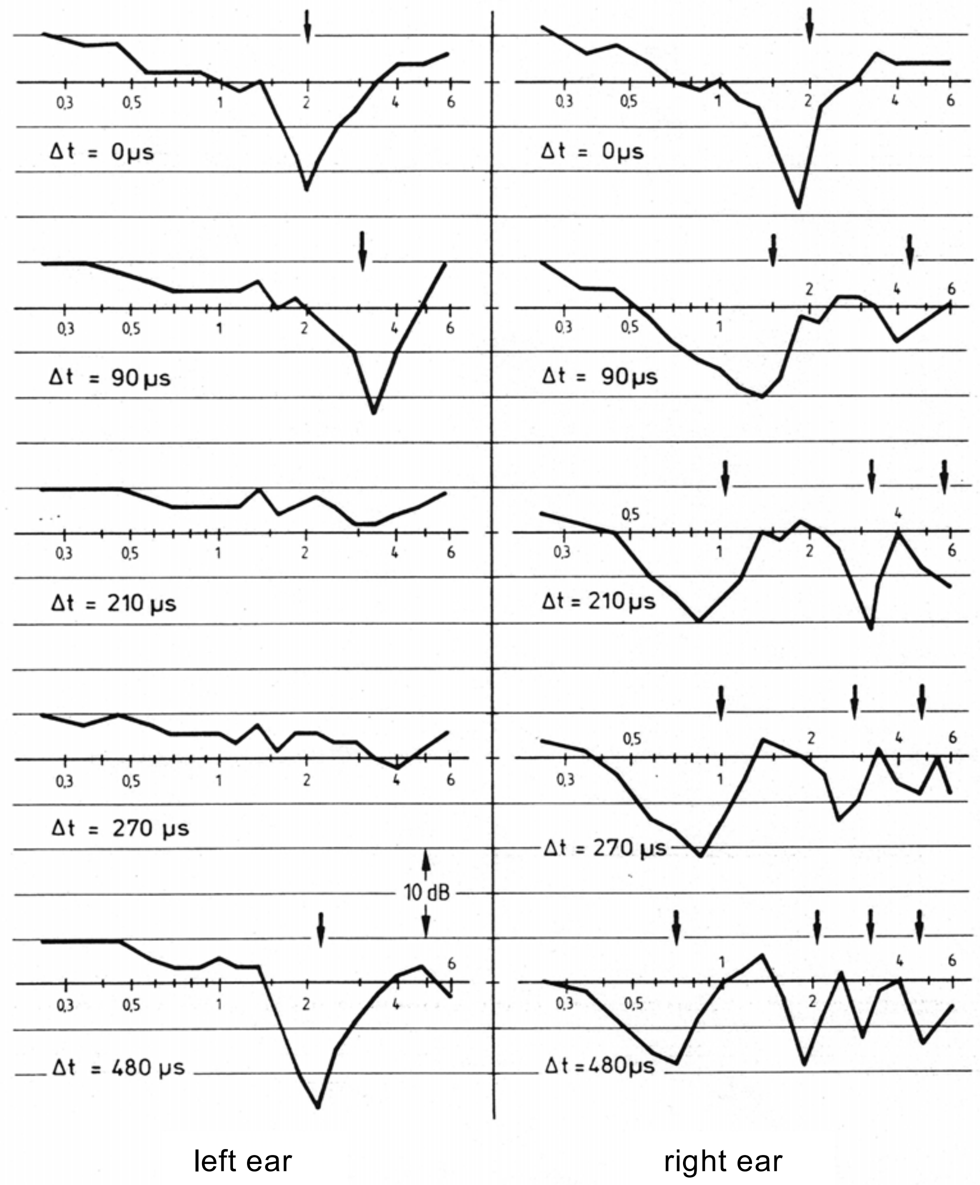


Figure 4: Comb filter effect for time-based stereophony, see text (the auditory event shifts to the right))

The measurements display the spectral changes that result at the left ear when switching on Loudspeaker B (i.e. the right loudspeaker) and the spectral changes that result at the right ear when switching on Loudspeaker A (i.e. the left loudspeaker). The transfer functions of the entire 'loudspeaker - outer ear - microphone' set-up have been eliminated. The inserted arrows mark those frequencies at which, from a calculational viewpoint, one would expect there to be a trough if one assumes that the interaural time differences are frequency-independent ($\tau_A - \tau_B \approx 250$ µs for $\Omega_A = \Omega_B = 30°$). Looking at these measurements, the following observations can be made:

1. The troughs are in the order of 10 dB or more. The boosts, especially at low frequencies, often amount to 5 dB. At mid frequencies spectral differences of about 15 dB occur, whilst at high frequencies this value is somewhat reduced due to the greater attenuation of the crosstalk components.
2. The troughs' locations in the frequency spectrum are only "ear-symmetrical" for $\Delta t = 0$, i.e. if Δt is increased up to 250 µs, the troughs will shift to the high frequencies at one ear and to the low frequencies at the other ear. For Δt up to 250 µs, the time delay resulting at one ear equals zero (in this case the left ear, which does not 'face' the phantom source). For $\Delta t > 250$ µs, the troughs appear at progressively lower frequencies at both ears.

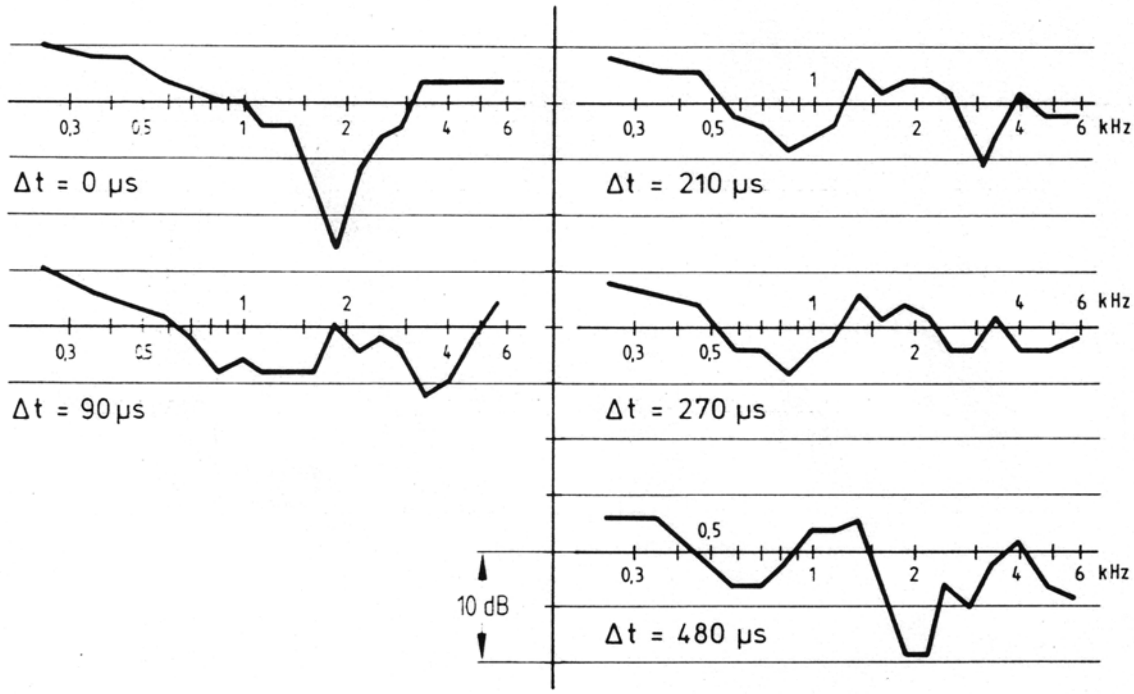


Figure 5:

The mean values of the resultant spectra at the ears, taken from Figure 4

3. It has to be ruled out that the suppression of sound colouration can be explained by means of some kind of an averaging process that the hearing system carries out on the spectra resulting at the left and right ears. Figure 5 shows the mean values of these spectra. The resultant fluctuations in the amplitude spectrum would be clearly detectable.

4. In the case of time-based stereophony, the ear signals' spectral characteristics are very different from the ones of a real sound source that, like the phantom source, moves between 0° and 30°. This is evident from comparing the measurements shown in Figure 6. This figure displays the spectral modifications that result if Loudspeaker B is moved from Ω_B = 30° to the centre, Ω_B = 0°. As the listening test to be reported in Section 2.2.2 will show, these deviations cause perceivable sound colouration in the case of monaural listening.

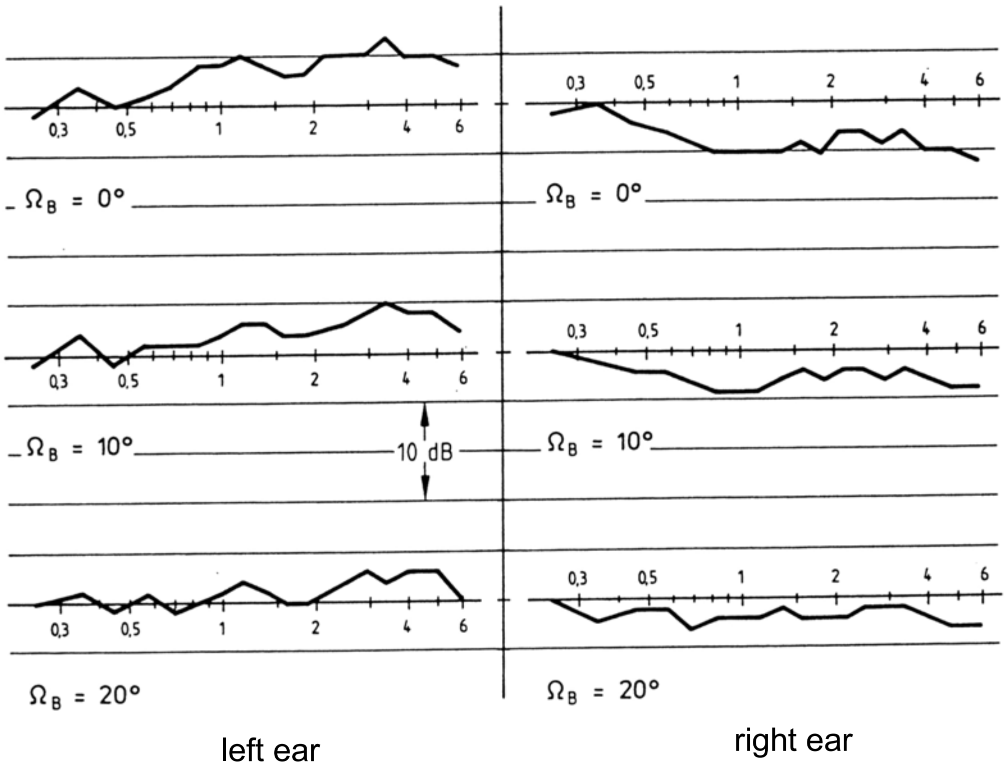


Figure 6:

Changes in the ear signals' spectra as a function of the angle of incidence Ω_B, relative to Ω_B = 30°.

5. As for the distance and elevation of the phantom source, one can see from Figure 4 that the auditory system has to interpret the signals as being "linearly distorted", because no real sound source can produce such interaural spectral differences. The

respective interaural transfer functions are not available (except when $\Delta t = 0$). This is especially evident for 200 µs < Δt < 300 µs. At the left ear troughs only occur above f = 10 kHz, whereas at the right ear four minima exist in-between 1 and 10 kHz (see Δt = 210 µs and 270 µs in Figure 4). Besides, for Δt = 250 µs ± 50 µs the location of these minima varies by 22% without the spectrum of the left ear signal changing at all. Since the transfer functions of the outer ears are relatively constant in the important frequency region (see Figure 6), neither the stable (externalised) localisation nor the elevation can be brought into agreement with the interaural spectral differences.

2.2.2 Listening test on the suppression of sound colouration

In order to ensure identical test conditions for each subject, the ear signals occurring in an anechoic chamber at the listening position of a stereophonic loudspeaker set-up are reproduced with the help of the artificial head. The loudspeakers radiate coherent white noise and the duration of the presentations ranges from 5 to 10 s. During this time interval, the artificial head is moved along the interaural axis at a rate of approx. 0.5 to 0.1 Hz and a maximum displacement amplitude of 20 cm from the centre. This movement more or less corresponds to a variation in time delay of maximally ±600 µs (time-based stereophony). As expected, when playing back the artificial head signals over headphones, this leads to large lateral displacements of the auditory event.

Keeping the pivoting movement of the artificial head, the following test signals are generated and recorded in a random order:

Group A:
The variable is the horizontal direction of the artificial head, δ. The following signals are chosen:

- A1: with δ = 0°
- A2: with δ = 30°
- A3: with δ = 60°
- A4: with δ = 90°
- A5: with δ = 180°

Group B:
The direction of the artificial head is kept constant, δ = 0°. The variables are different errors that have been introduced deliberately into the head-related recording and reproduction system. The following reproduction scenarios are chosen:

- B1: Dichotic, phase-inverted presentation of the artificial head signals
- B2: Dichotic presentation of the artificial head signals with one ear signal being delayed by Δt = 200 ms
- B3: Monotic presentation of one artificial head signal
- B4: Diotic presentation of one artificial head signal

B5: Dichotic presentation of the artificial head signals, which have been recorded without the pinnae of the artificial head

B6: Dichotic presentation of the microphone signals, which have been recorded without the head and pinnae of the artificial head

When presented over headphones, the test signals A1 … A4 and B1 … B6 create more or less stable phantom sources. To give an example, depending on the accuracy of the recording and playback stages, Signal A1 should produce an easily localisable phantom source, whilst Signal B2 will certainly result in a poor reproduction. In order to be able to compare these scenarios to the localisation conditions applicable to a real sound source, three other test signals are also included:

Group C:
No phantom source situation, but a loudspeaker positioned in the median plane. The following reproduction scenarios are chosen:

C1: Otherwise as A1

C2: Instead of the pivoting movement, the artificial head is rotated by $\delta \pm 30°$

C3: As C2, but diotic presentation of one artificial head signal

Using headphones, each test signal is presented to the subjects twice. For each presentation, the auditory event (produced by the associated test signal) has to be evaluated in terms of *one* attribute of detectability, i.e. the detectability of sound colour changes (**Attribute Kl**) or the detectability of directional changes (**Attribute Ri**). A 5-point rating scale is used:

0: not detectable
0.25: just about detectable
0.5: detectable
0.75: clearly detectable
1: very clearly detectable

Before each trial, the subjects are presented with a few test examples. This helps improve judgement consistency and prevents learning effects during the actual listening test. During this training phase, the subjects' attention is drawn either to Attribute Kl or Attribute Ri. For 11 participants this happens first for Kl, whilst 11 other participants are asked to evaluate Ri first.

Listening test results

1. The relationship between the attributes Kl and Ri for the test signal groups A and B is evident from Figure 7. Each included Kl/Ri-value constitutes the arithmetic mean of the verdicts of 11 subjects for one test signal (four mean

values per test signal). The two calculated regression lines (the first one going from Ri to Kl, the second one from Kl to Ri) demonstrate a close, linear relationship; the correlation coefficient is

$$r = -0.91.$$

The detectability of sound colour changes is inversely related to the perceptibility of directional changes.

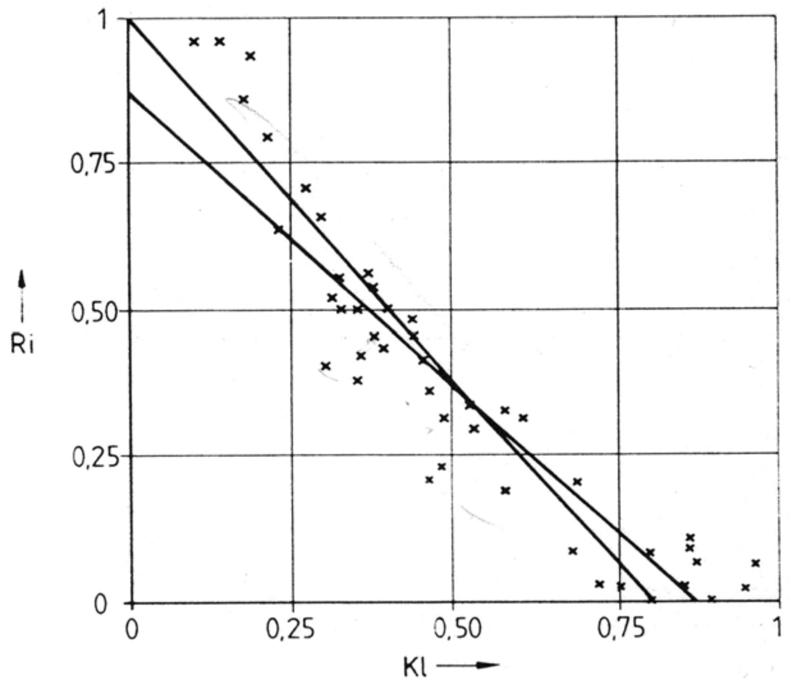


Bild 7 :

Relationship between the detectability of directional and sound colour changes

2. The attributes' dependency onto the test signals is apparent from the overview of the results given in Figure 8. The test signals have been sorted along the abscissa in such a way that an in/decreasing trend in terms of the attribute ratings results (the curves at the bottom of Figure 8 depict the corresponding standard deviations). As can clearly be seen, for all reproduction scenarios that cannot produce phantom sources (test signals A4, B4, B3, B2), the comb filter effects at the ears are clearly detectable. On the other hand, one can see that in the case of an error-free reproduction the auditory event clearly moves and that the spectral changes at the ears are hardly detectable (test signal A1).

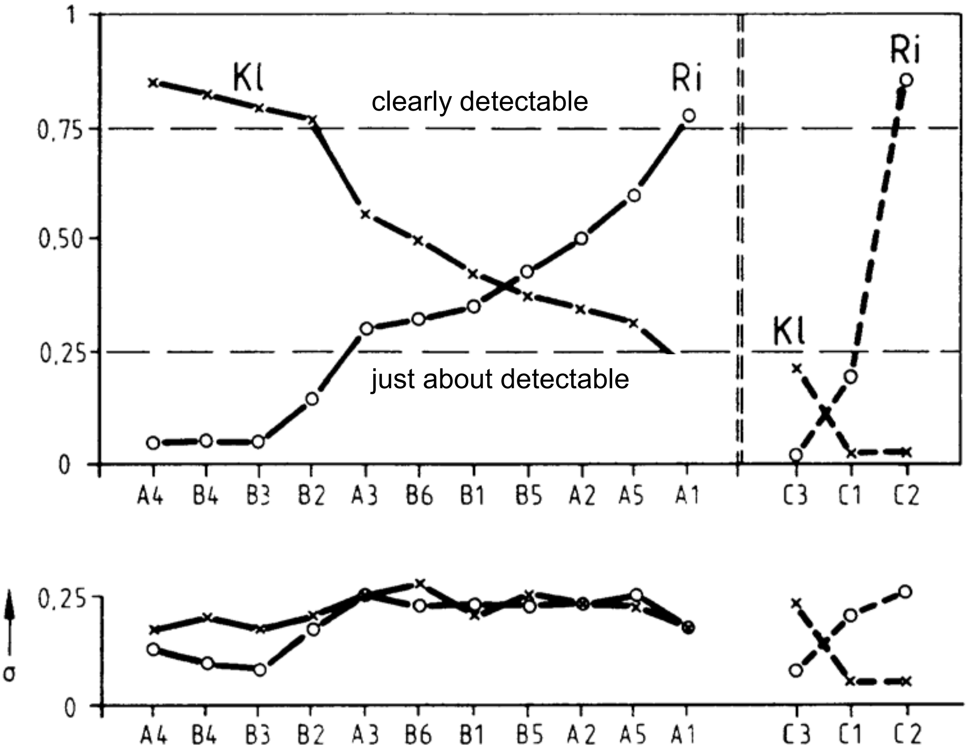


Figure 8:

Overview of results, see text

2. The attributes' dependency onto the test signals is apparent from the overview of the results given in Figure 8. The test signals have been sorted along the abscissa in such a way that an in/decreasing trend in terms of the attribute ratings results (the curves at the bottom of Figure 8 depict the corresponding standard deviations). As can clearly be seen, for all reproduction scenarios that cannot produce phantom sources (test signals A4, B4, B3, B2), the comb filter effects at the ears are clearly detectable. On the other hand, one can see that in the case of an error-free reproduction the auditory event clearly moves and that the spectral changes at the ears are hardly detectable (test signal A1).

3. Compared to test signal C2, however, it has to be noted that when reproducing the ear signals stemming from a real sound source that moves within an angular range of ±30° over the artificial head system, a smaller Ri and a larger Kl result for the phantom source compared to the real source (see also Figure 9).

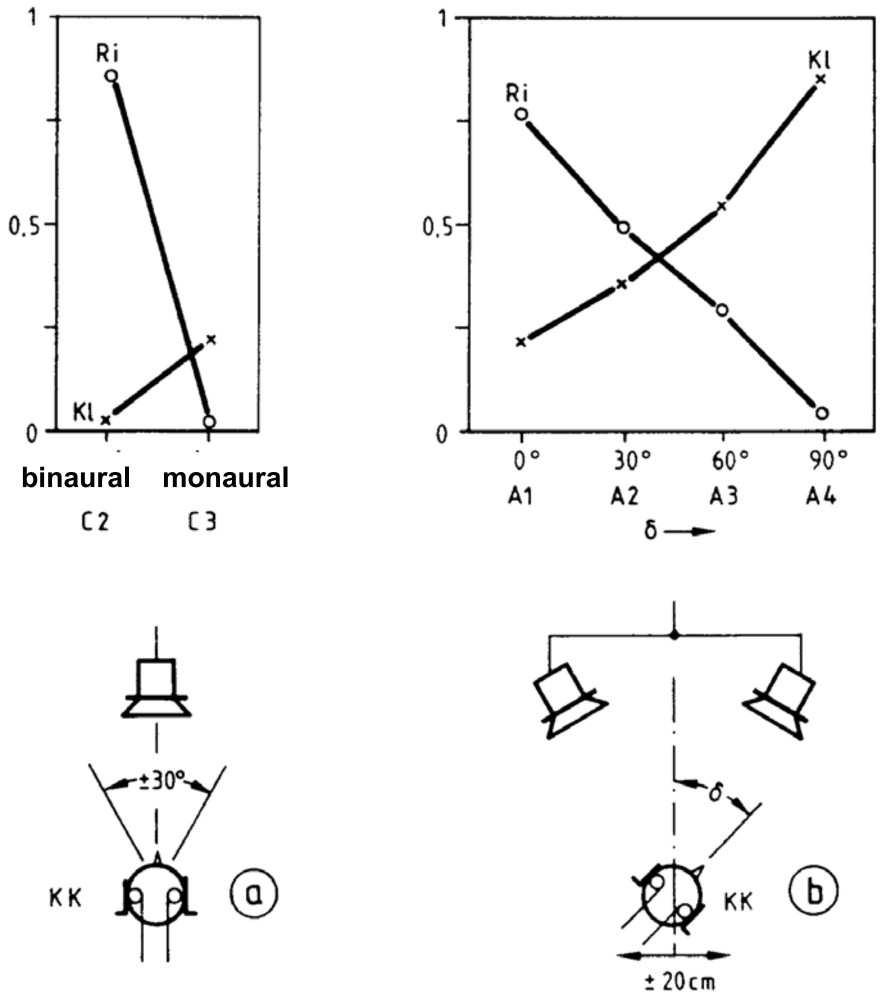


Figure 9:
Results for the test signals C2, C3 and A1 … A4

4. A perceivable sound colour change also occurs for real sound sources if the associated binaural, directional characteristics are missing (test signals C2 and C3, Figure 9a). This detectability of sound colour changes is as high as the one for the normal phantom source situation (test signals C3 and A1, Figure 9).

5. Figure 9 presents Attributes Kl and Ri as a function of the artificial head direction δ. The localisability of the phantom source reduces with increasing δ. Lateral phantom sources (δ = 90°) do not occur (e.g. see RATLIFF 1974, THEILE / PLENGE 1977), the dectability of the sound colour changes is accordingly large.

6. For the phantom source situation, a relatively large drop in the detectability of directional changes already occurs if instead of a normal artificial head system a dummy head without the pinnae is used (Figure 10). Manipulating the artificial head in this way causes negligible changes in the interaural level and time differences; the resultant changes in the ear signals' spectra on their own give rise to a greatly reduced detectability of the phantom source movement.

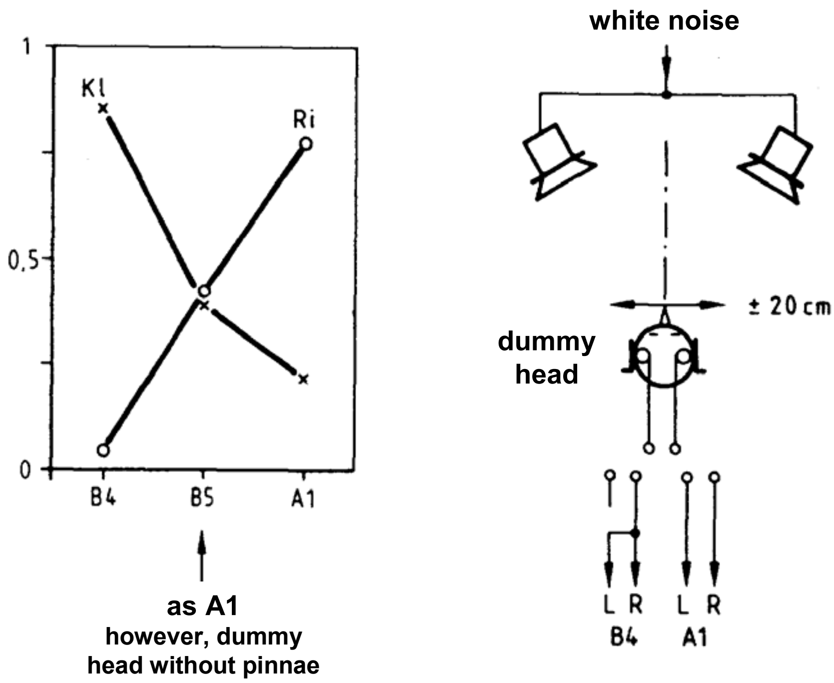


Figure 10:

Results for the test signals B4, B5, A1

Overall then, this listening test showed that:
The suppression of sound colouration in the superimposed, broadband sound field for the case of two coherent sound sources takes place to the extent to which a phantom source arises. The sound colour of the phantom source does not correspond completely to the spectral properties of the ear signals.

3. An association model for the localisation process

If the ear signals resulting from the superposition of sound fields and the ear signals stemming from an individual sound source produce identical auditory events, then in both cases broadband signals will exhibit different spectra. Assuming that the ear signals are processed as a whole, the spectral characteristics of the ear signals in the phantom source situation cannot be brought into agreement with the distance, elevation and sound colour of the auditory event.

The aforementioned spectral objection is valid if one tries to interpret summing localisation models as localisation models. Summing localisation models are not based on findings concerning localisation, but on theories of direction hearing. It is not possible to extend the summing localisation principle's limited area of validity to all phantom source phenomena. If one assumes that understanding the phantom source should ultimately improve our understanding of spatial hearing, it follows that the "summing localisation" approach has to be questioned.

A fundamental approach should be aimed for that can lead to a uniform explanation of the properties of auditory events in the superimposed sound field, i.e. an approach that takes into account the general findings with respect to localisation (psychoacoustic principles of inside-the-head locatedness, distance and direction hearing as well as the perception of simultaneous auditory events).

The spectral objection initiates the development of a complete localisation model having an area of validity that includes phantom as well as real sources.

In Section 2.2 it was found that in the phantom source situation the comb filter effects at the ears influence neither distance perception nor perceived sound colour. It is not the actual spectra at the ears that are operative, but rather those spectra that the loudspeakers produce at the ears individually.

This observation leads to the supposition that the auditory system separates the components of the summed signals occurring at the ears. If the ear signals' dependency onto sound source location is understood as a mechanism for encoding spatial information, then knowledge of this dependency can be seen as a key for decoding the spatial information. Supposing that in the superimposed sound field the simultaneous decoding of at least two types of spatial information is possible (e.g. by means of pattern recognition processes), it is conceivable that the summed signals' discrete components are processed separately.

Should this type of processing in the auditory system be considered further? As will be seen, this approach leads to a localisation model that is in agreement with many phenomena of spatial hearing, hence offering new possibilities for their explanation. Initially, this approach appears functional for two reasons:

1. It can offer an explanation for the ineffectiveness of the comb filter effect in the phantom source situation (see Section 3.2.2).
2. It agrees with the hypothesis that localisation should be seen as the consequence of a perceptual process, which arises due to auditory experience alone (see Sections 3.1.2 and 3.2.1).

Therefore, the new localisation model contains a processing stage that, as a result of auditory experience, can separate such ear signal components. These components are coupled to each other due to the effect of the head and pinnae in the superimposed sound field. In the following, these ear signal components shall be described as "localisation stimuli".

Definition:

Sufficiently broadband ear signals or ear signal components at the ear drums of the two ears together constitute a **localisation stimulus** if, based on their temporal and spectral properties, they can be mapped to a single sound event location.

The processing stage for the selection of localisation stimuli is called "location association stage". It precedes a second, higher-level central processing stage contained in the model. This so-called "gestalt association stage" comprises those processes that determine the qualitative properties of an auditory event except for its spatial characteristics (see Section 3.2 for details).

Hence, the localisation model is essentially characterised by **two-dimensional stimulus processing**. The rigorous differentiation of the two processing stages corresponds entirely to the two elementary areas of auditory experience: the received ear signals can be attributed to the two sound source characteristics of "location" and "signal", which always occur in a pairwise fashion. As a result, the auditory events occurring in the model can be traced back to the effects of a location- and gestalt-determining processing stage.

A stimulus has to pass both of these stages for it to lead to a perception. The two processes thus always determine the properties of auditory events in a conjoint fashion. Consequently, the gestalt-determining processing stage is an element of the localisation model as well.

3.1 The localisation stimulus selection

A characteristic of the model's location association stage is that it tries to interpret a received stimulus as a localisation stimulus, i.e. it automatically compares the input stimulus with stimulus patterns that have been associated with certain auditory event locations as a result of auditory experience. Only a localisation stimulus leads to localisation. Such a stimulus will be available if the characteristics of the ear signals are compatible with the auditory experience in terms of both time and spectrum. Due to this property, the human hearing system is capable of combining signal components that are characteristic of a sound event location and of forwarding them collectively (localisation stimulus selection). "Collective forwarding" implies that the localisation stimulus selection can be interpreted as a fusion process, which combines the binaural signal components of a particular sound source in an unseparable manner and which equips these components with the re-codified spatial information that leads to the perception. In the superimposed sound field, the localisation stimulus selection acts as a filter enabling the simultaneous discrimination of individual source signals.

This function of the hypothetical location association stage can probably not be modelled using operators that, based on current knowledge, are physiologically possible. Rather, the performance features of this processing stage shall be described as much as possible by means of linear system theories and, within the framework of this work, for maximally two sound events of an arbitrary spatial constellation only (see Section 4.1). In essence, a controllable filter is envisaged having a transfer function that is inversely related to the location-dependent transfer function of the outer ear (see Section 3.1.1). The identification of the currently operative transfer function takes place by means of pattern recognition processes (see Section 3.1.2 as well as PLENGE 1973).

3.1.1 The spatial decoding

The human hearing system comprises an evaluation system with two input channels. These channels are preceded by a common linear network having a transfer function M that results from the effects of the head and pinnae in the sound field. M is a function of the sound source location i that links the ear signals in a defined way ("spatial encoding"). If a source signal Q is present at the input to the matrix M(i), the linked signals appear at the inputs to the evaluation system in the form of liQ = L and riQ = R (li, ri = acoustic transmission factors for the left and right ear, respectively, and a given source location i; see BLAUERT / LAWS 1973). After the spatial decoding, response Q' appears at the allocated output i' (see Figure 11).

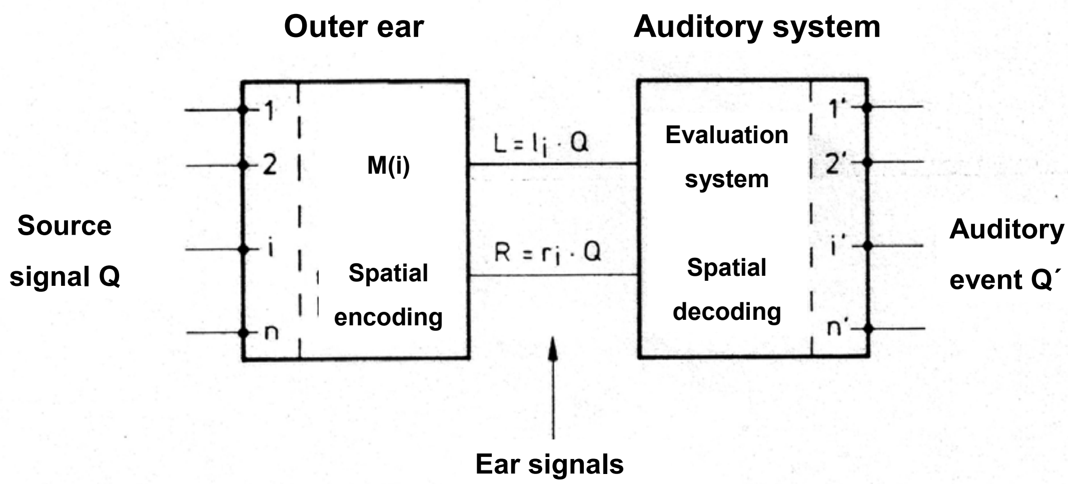


Figure 11:

The spatial transmission system

To this end, the decoding mechanism has to adapt to the sound source locations. This adaptation requires an associative process of pattern recognition.

The nature of localisation shall therefore lie in a recognition process that leads to the selection. Correspondingly, the functional principle of the location association stage contains a "pattern recognition" building block that provides the information for an optimal filter adaptation. This pattern recognition process lays the foundation for an extraction of certain signal patterns.

3.1.2 The associative pattern recognition

The assumption that certain stimuli trigger particular association processes constitutes an important hypothesis for the development of pertinent pattern recognition models. If localisation is also seen as an associative process, one obtains a plausible explanation for the assumed ability of the auditory system to discriminate the components of a superimposed sound field in the case of two coherent sound sources.

If association processes normally lead to congruent sound and auditory event locations, it has to be possible to explain localisation deviations with the help of these association processes. In fact, this does not just seem to apply to the "phantom source" phenomenon. Other localisation phenomena suggest associative processes, too. The following examples of divergencies between sound and auditory event locations arise under specific, unnatural conditions:

1. The direction of an auditory event in the median plane is directly related to the spectrum of the source signal. For narrowband signals it is independent of the signal's angle of incidence. This spectral dependency can be traced back to the linear distortions that are caused by the head and pinnae and that lead to the forming of direction-determining stimulus patterns ("directional bands", BLAUERT 1969).

2. If the source signal is known, the distance of an auditory event will be directly related to the level and spectrum of this signal (e.g. see LAWS 1972). In the case of loudspeaker reproduction in the free sound field, it is dependent on the auditory experience (PLENGE 1973). The relationship between the perceived loudness level and sound colour and the associatively determined loudness and "pitch" of the sound source is crucial.

3. In the case of a wrong adaptation to the present auditory experience, an auditory event's direction does not correspond to the angle of incidence of the source signal, e.g. after the sudden normalisation of a hearing system that was affected monaurally by an illness (operation, RÖSER 1965).

4. The location of an auditory event can be influenced through guided association, e.g. by means of accompanying acoustic or optical stimuli [e.g. KLEMM 1909 ("spatial complication"), BLAUERT 1970, PLENGE 1973, MASSARO / WARNER 1977, LEHRINGER 1979].

5. The manifold causes for inside-the-head locatedness can be conjointly explained if one assumes that localisation takes place based on a comparison of current input and learnt stimulus patterns (PLENGE 1972, 1974). Inside-the-head locatedness occurs if a stimulus cannot be mapped to a location-determining stimulus pattern. This can be prevented by means of guided association (e.g. visual information JEFFRESS / TAYLOR 1961).

Important phenomena of localisation suggest that the relationship between the location of an auditory event and the location of a sound source is determined by the association characteristics. Before developing the complete localisation model, the significance of the association principle to other fields of research and applications shall be briefly depicted. In this context, it turns out that with the help of some kind of "associative pattern recognition" essential areas of the perceptual process can be explained and also described in a greatly simplified form. Conversely, association processes related to neural processing in the central nervous system turn out not to be generally proven from a neurophysiological viewpoint.

However, from a perceptual psychologist or information theorist's perspective, association and pattern recognition mechanisms often provide the only possible explanation for certain phenomena concerning auditory (and visual) perception. This includes the occurrence of “simultaneous auditory events” when presenting different source signals over a single loudspeaker, for example. From a communications-engineering point of view, this corresponds to the reception of separate signals after transmission over a single channel with the bandwidth of one signal only. These fundamental abilities of our hearing system can easily be demonstrated and, from a communications engineering viewpoint, probably only be explained with the help of the information-reducing effect of pattern recognition mechanisms.

Definition :

Associative pattern recognition is a process that links an input pattern with a stored pattern, even if only some parts of the stored pattern are contained in the input pattern.

For a long time, this mechanism, which to this day remains hypothetical, has been the subject of interest in various fields of research, most notably so in cybernetics (see FLECHTNER 1972). In neurophysiology, one assumes the existence of a “sensory association system” for the processing of sensory stimuli in the central nervous system (e.g. CASPERS 1973); meanwhile it is considered to be virtually unexplored. The ideas observable in this context are strongly influenced by cybernetic models. Especially more recently, researchers have tried to model specific capabilities of the human brain in order to scrutinise them or to apply them for technical purposes. Apparently, associative information storage and associative information recall constitute a basic processing principle (see FUKUSHIMA 1973, POGGIO 1975, KOHONEN et al. 1976, KOHONEN / OJA 1976, KOKONEN 1977, WESS / ROEDER 1977, BOHN 1978, FUKUSHIMA / MIYAKE 1978, MURAKAMI et al. 1978).

For instance, a “model of a neural network” by WIGSTROEM 1974, “which is intended as a possible description of the cerebral cortex, consists of a network of cells assumed to be of excitatory and inhibitory type. It is shown that, under suitable conditions, the output pattern will become composed of just one major component even if the excitatory input pattern is a mixture of several patterns that were present during learning. This major component is a part of the specific output pattern that during learning became associated with the input pattern. The behaviour is obtained through a dynamic process in which the pattern separation properties of the feedback link play an important role. The model’s operation can be viewed as pattern recognition.”

Another relevant example is the work by WILLWACHER 1976 who compared the abilities of an associative storage system to functions of the human brain. The presented network is capable of imitating effects of the human brain in a greatly simplified form: parallel association (complete recall of a pattern due to input of a component of this pattern), serial association (recall of a temporal sequence of information patterns due to input of the first pattern of the sequence), classification of an unknown (not stored) information pattern, coordination of patterns from two fields of the system, association of a more probable pattern sequence, disturbance of the association process, “memory aids”, “abstraction of common characteristics”, “reversal learning”, “productive ideas”.

These and other cybernetic approaches have in common that the employed system components exhibit a behaviour analogous to the one of real neurons. Since to this day there are no detailed neurological findings regarding the structure of the human brain and the form of pattern recognition, one synthesises respective cortical effects based on the known functional modules so as to derive hypotheses regarding the functioning of the brain. The central hypothesis is that associative processes represent a basic principle of sensory stimulus processing.

3.1.3 The functional principle of the location association stage

The location association stage performs the localisation stimulus selection according to Section 3.1. The processing of this stage can be described by the effect of a location-dependent filter (see Section 3.1.1), the parameters of which are controlled as a result of associative pattern recognition as outlined in Section 3.1.2. A “binaural correlation pattern” (see Section 4.1) is chosen as a useful signal pattern for the recognition process. In Figure 13 the functional principle of the location association stage is shown.

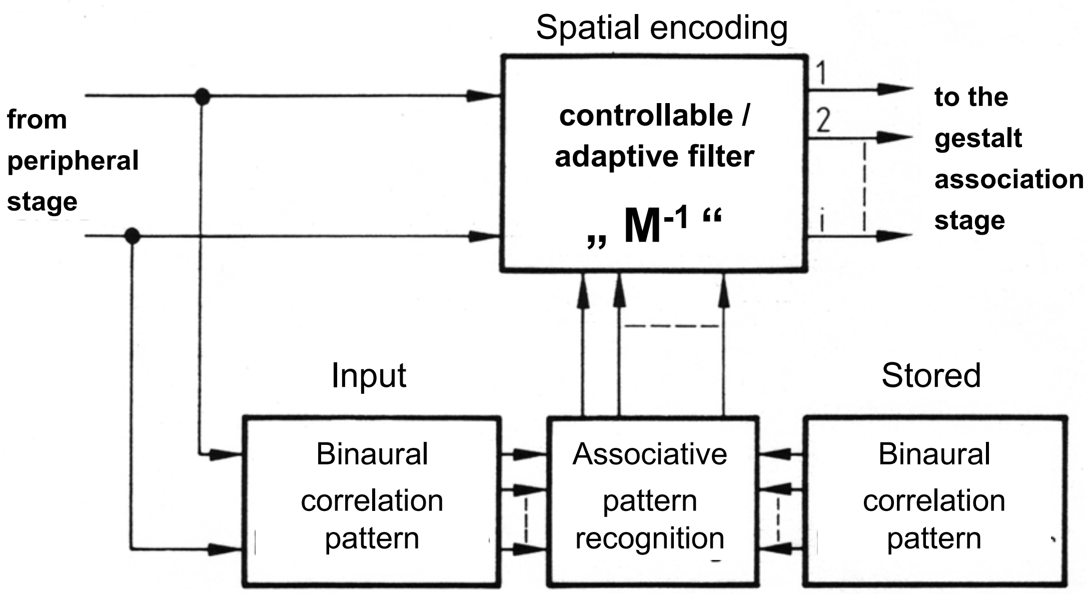


Figure 13:

Functional principle of the location association stage

It is assumed that the relationship between the in- and output signal can describe the behaviour of this hypothetical processing stage. This functional principle is not meant to replicate the inner structure of the sensory processing. Its functioning requires the materialisation of the experience processes in the past, so that the stored binaural correlation pattern is available.

A characteristic of the location association stage is its selective property. The signal coming from the peripheral stage (see Section 3.2) contains information about spatial as well as gestalt features (of one sound source). The spatial information is recognised with the help of the pattern recognition process, whilst the gestalt information is discriminated by means of the adaptive filter and then passed on to the gestalt association stage.

In other words, the impinging signal is 'freed from' the influence of the outer ear with the help of the location-dependent filter "$\mathbf{M}^{-1}$", the weighting of the source signal by means of the effective transfer function is reversed, and the source signal as well as the information obtained about the direction and distance of the source are forwarded **separately**.

3.2 The association model

The new localisation model is called "association model". Before depicting the functioning of the complete localisation model, the fundamental approach – i.e. to regard localisation as the result of a perceptual process that is solely possible because of auditory experience – shall be described in a more detail.

3.2.1 Sensational or perceptional model?

In the sensory system, the hypothetical association process corresponds to the function of a highly effective filter carrying out information reduction in-between peripheral reception and conscious perception. On the one hand, stored association patterns that are influenced by previous experience give rise to a meaningful selection of information. On the other hand, in spite of the information reduction, they allow for a sufficiently accurate recognition of the stimulus configurations of the environment (MARKO 1971, KEIDEL 1973).

If, due to lack of experience, missing adaptation or biological reasons, no associations can be formed, meaningless sensations rather than conscious perceptions will be the consequence. This phenomenon is well known in the field of neurophysiology. Deactivation of the so-called association cortices in the sensory system causes perception disorder, i.e. it impairs the perceptual function of sensory information (agnosia). For humans, agnosia can develop with respect to different sensory modalities. If, for example, the auditory association area (in the left temporal lobe) is destroyed, speech comprehension (amongst other things) is lost. Although the affected person can still hear, the meaning of the signal is hidden. It is not known, however, if the ability to localise can be cut off as well (such physiological verification of location association would be very enlightening indeed).

The significance of association processes to conscious perceptions leads to a meaningful terminological distinction between the terms "sensation" and "perception".

> Definition :
>
> **Sensations** are sensory events caused by stimuli that cannot be subdivided further and that are not affected by learning processes, conscious or subconscious interpretations. They do not arise as a result of associations.

> Definition :
>
> **Perceptions** are sensory events caused by stimuli that are mapped to the outside world as a result of sensory experience and that can therefore be influenced by learning processes, conscious and subconscious interpretations. They arise as a result of associations.

This demarcation implies that when developing perceptual models, it should always be clarified whether a sensation or perception model shall or rather can describe the process of interest.

When investigating perceptual processes, the main problem therefore lies in the necessary demarcation. This is where sources of error originate, because without knowledge of the whole process one cannot decide if or to what extent sub-areas of the perceptual process can be described separately.

With some reservations, it should be possible to consider spatial hearing in isolation from visual perception. However, how can one justify considering directional hearing in isolation from distance hearing?

Direction and distance are only co-ordinates of the location of an auditory event; direction does not exist without distance. A model for directional hearing is no perceptual model, which is why it does not inevitably describe a specific function of the auditory system with regard to spatial hearing. Likewise, investigating localisation by means of sine tones is also questionable. How should the auditory system be able to determine the distance of a sound source radiating a pure tone (see Section 4.2.4)?

In terms of the localisation model to be depicted below, it is assumed that direction and distance are inseparable qualities of spatial hearing.

It is striking that in the area of visual perception many genuine perception models are known, whereas this is hardly the case for auditory perception. From a neurophysiologist and information theorist's viewpoint this is incomprehensible, as no fundamental differences exist between the eye and ear's function as receptors of information. Nonetheless, there are two reasons for this deficiency. Firstly, due to commercial reasons a much larger interest exists for technical applications of visual perception models (data reduction by means of appropriate source coding for image transmission, e.g. replication of pattern recognition principles with the help of classification processes or adaptive filters). Secondly, pattern recognition and association principles are more clearly recognisable in visual compared to auditory perception.

At this point, two perceptual models – one from the visual and one from the acoustical domain – shall be briefly delineated, since they contain mechanisms also featured in the new localisation model.

Replicating the visual system, MARKO 1974, 1978 proposed the so-called "layer model" for pattern recognition, which he based on considerations of system theory (see Figure 14).

The "layer" is meant to symbolise a layer of neurons on which one should imagine excitation patterns. The hatching of the layer is supposed to indicate that threshold elements are present at the output, so that only suprathreshold signals are passed on. In terms of system theory, this arrangement corresponds to a multi-stage threshold logic. The non-linearity given by the threshold elements is crucial for the classification process. Within the discrete stages, local filtering is carried out (orientation-selective filter, edge detection), the subsequent non-linear distortion (e.g. maximum detection) enabling the classification. It is worth emphasising the multi-stage principle of this classification model.

The pattern characteristics are discriminated and forwarded one after the other. They complement each other after having passed through all layers, thus forming the complete pattern. This is also a feature of the association model.

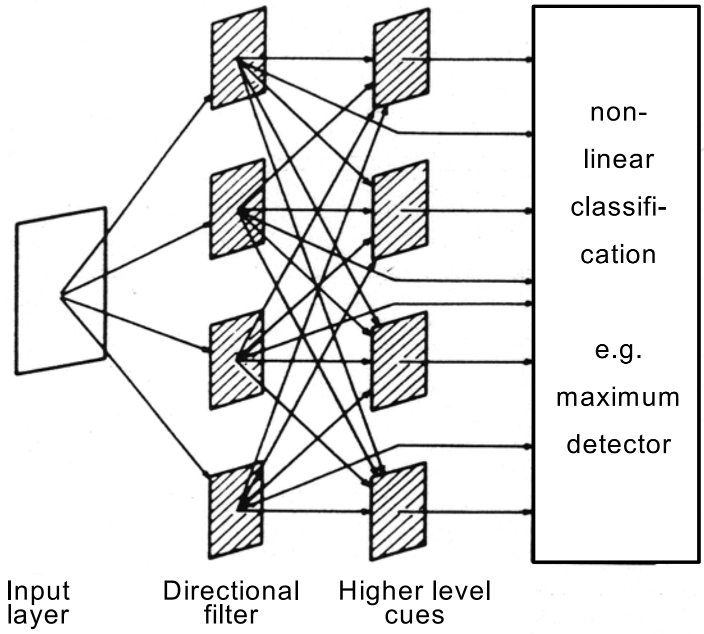


Figure 14 :

Layer model for pattern recognition with local filtering within the discrete stages and subsequent non-linear computation (after MARKO 1978)

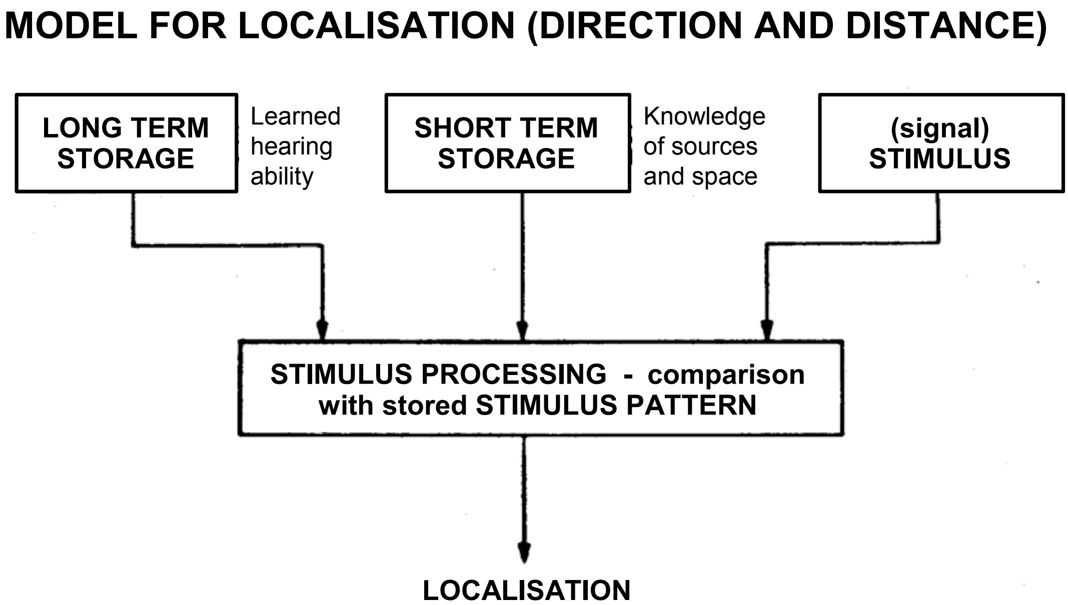


Figure 15:

Localisation model for the explanation of inside-the-head locatedness (after PLENGE 1973)

A rigorous perceptual model for auditory perception is given by a psychologically oriented localisation model that was proposed by PLENGE 1973. This model probably offers a comprehensive explanation for inside-the-head locatedness (Figure 15).

It comprises both short- and long-term memory as well as a stimulus processing mechanism that compares stored stimulus patterns to input patterns. The stimulus processing does not lead to (externalised) localisation if (1) the short-term memory does not contain or (still) contains wrong information about sound sources and their locations or (2) the stimuli are of such a kind that "they cannot be mapped to a stimulus pattern contained in long-term memory".

Remarkably, in this model localisation is traced back to a learning process, the short-term as well as the long-term memory being of particular importance. The approach "stimulus processing by means of comparison with stored stimulus patterns" is contained in the association model. The "associatively guided pattern recognition" can be seen as its further development.

3.2.2 The functioning of the association model

The above considerations lead to the following functional principle of the association model (Figure 16).

Apart from the peripheral stage that subdivides the ear signals into spectral components of approximately constant relative bandwidth with the help of filter banks (see ZWICKER / FELDKELLER 1967, DUIFHUIS 1972, BLAUERT 1974, 1978), the association model comprises the two central processing stages "location association stage" and "gestalt association stage". Each of these two processing mechanisms takes place in the form of an associatively guided pattern selection. Having passed the peripheral stage, certain ear signals give rise to a location association in the first and to a gestalt association in the second, higher-level stage. The two stages always determine the auditory event properties in a conjoint fashion.

The rigorous differentiation of these two stimulus evaluation stages corresponds entirely to the two elementary areas of auditory experience. The received ear signals can be attributed to the two sound source characteristics of "location" and "signal", which are independent of each other but always occur in a pairwise fashion.

In principle, the two central evaluation stages therefore also contain the same processing mechanisms. Similar to the location association stage's mechanism for localisation stimulus selection that was outlined above, the gestalt association stage contains a mechanism for the selection of the gestalt.

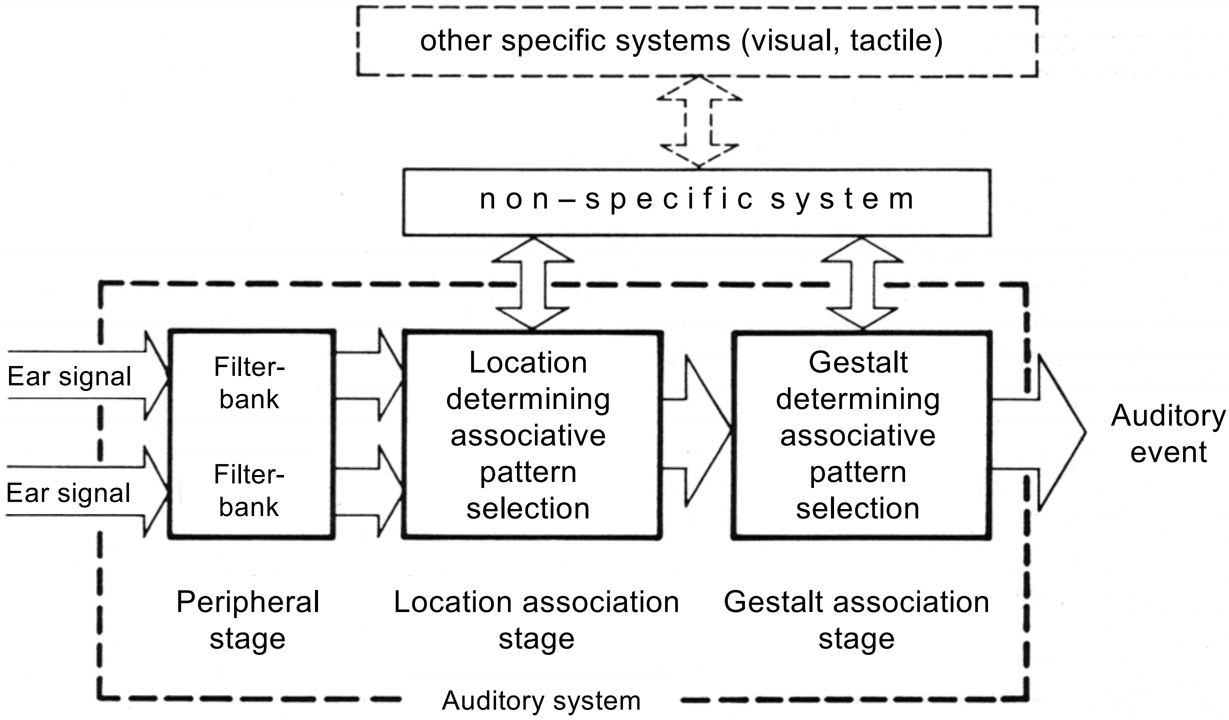


Figure 16 :

The functional principle of the association model

It is assumed that the function of the acoustic association system is influenced by a so-called "non-specific system" in terms of guided association (see Section 3.1.2). (Neurophysiological research has shown that this "reticulate system" is connected with all the areas of the cerebrum. One assumes that it is responsible for coordinating the integration of stimulations, e.g. the processing of acoustic and optical signals, as well as a consciousness- and attention-depending filtering of information, see e.g. CASPERS 1973).

The functioning of the location association stage was already depicted in Section 3.1.3. A characteristic of the gestalt association stage is that it processes a received stimulus independently of its spatial information. This stage represents all mechanisms that are necessary for the source- or gestalt-dependent perception of a stimulus. Its sub-areas are responsible for a multitude of different auditory phenomena. These include mechanisms for signal fusion as well as the recognition and evaluation of music and speech.

The psychoacoustic regularities of binaural signal detection are important to the functioning of the association model (BMLD and BILD, see BLAUERT 1974, pp. 206), which are attributed to the gestalt association stage (see Chapter 5). Such models (e.g. the "accumulation model" by SCHENKEL 1967, the "EC model" by

DURLACH 1963, 1972 or the “correlation model” by OSMAN 1971) therefore follow the location association stage.

The model for pitch perception of complex sounds by TERHARDT 1972 agrees especially well with the associative processing of the gestalt association stage. It is based on the assumption that

> “… the human hearing system does not inherently know the difference between harmonic and non-harmonic sounds. If the hearing system is ‘in a natural state’, every sound is made up of a complicated combination of spectral pitches and perceived sound pressure fluctuations. … With the help of acquired knowledge of pitch relations between harmonically related sounds and the perceived frequency of the sound pressure fluctuations, the hearing system allocates a tonal meaning to each presented sound. This tonal meaning is perceived more or less distinctively as ‘virtual pitch’.” (TERHARDT 1972)

The association model is compatible with this model, too. The source signals, which are discriminated by the effect of the location association stage, contain all spectral information. The perception (see definition on page 29) of sound colour can occur in terms of both proposed models.

At this point, a localisation phenomenon shall be pointed out that so far has not been explicitly investigated. The sound colour of an auditory event turns out to be largely independent of the location of a sound event, even though the spectra of the ear signals are strongly location-dependent (e.g. see BLAUERT 1974). To illustrate, a listener slowly turning by 180° in front of a loudspeaker that is located in an anechoic chamber and that radiates white noise hardly perceives changes in sound colour when listening binaurally.

This phenomenon can easily be explained by means of the effect of the location association stage. As a result of the localisation stimulus selection, only the discriminated source signal reaches the gestalt association stage. It is not until here that the process for determining sound colour takes place.

What effect does the model have in the superimposed sound field? To begin with, this shall be illustrated for the phantom source situation.

A useful way of depicting the effect is by means of “impulse diagrams”, which depict the temporal, binaural characteristics in a simple manner. The well-known hypothesis that the auditory system initially evaluates each ear signal similar to an auto-correlation analysis (e.g. LICKLIDER 1951) but separately for narrow frequency bands (e.g. ZWICKER / FELDKELLER 1967, DUIFHUIS 1972, BLAUERT 1974, 1978) justifies this illustration method. The relative positions of the maxima reflect the interaural time difference for each frequency band (see Section 4.1). For the simplified case of the interaural time differences being frequency-independent, an impulse pair AL / AR shall represent the temporal characteristics of a discriminable, sufficiently broadband sound source A (Figure 17).

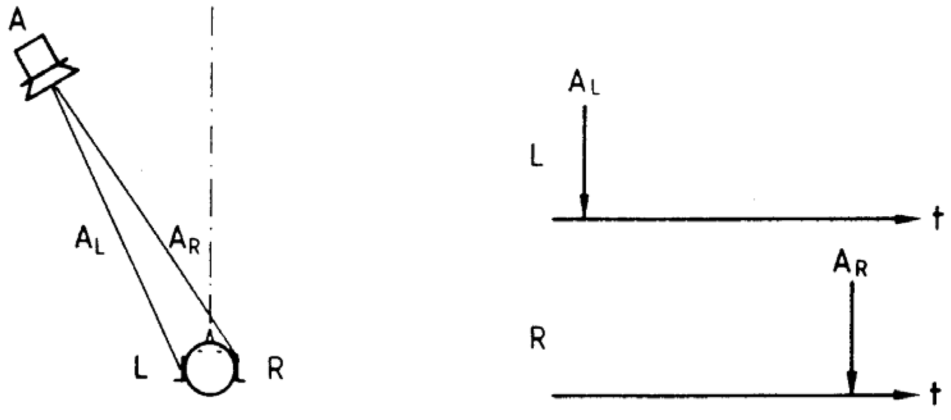


Figure 17 :

Illustration of the temporal, binaural characteristics of a sound event

In accordance with the hypothetical association process of the hearing system, the diagram has to be interpreted as follows. If the ear signals featuring these temporal properties contain appropriate spectral characteristics, they will cause a location association. Appropriate temporal and spectral characteristics are available if they can be attributed to a particular sound source positioned at a certain location.

In the phantom source situation the association model could therefore have the following effect (a discussion will follow in Chapter 4):

Based on associatively guided pattern recognition, the location association stage carries out the localisation stimulus selection. It supplies the gestalt association stage with two separate stimulus responses A' and B', which contain the source signal characteristics of the sound sources A and B, respectively. In the gestalt association stage the stimulus responses A' and B' are subjected to a process that leads to the gestalt of the auditory event being determined. It is not until this point that the two stimulus responses are fused to a single stimulus response. This is because the (source-specific) spatial information is not evaluated here, but rather the (identical) information about the source signals. As long as the loudspeakers radiate sufficiently similar signals, complete fusion occurs in this processing stage. Hence, there is only one auditory event and consequently only one common auditory event location – the location of the phantom source.

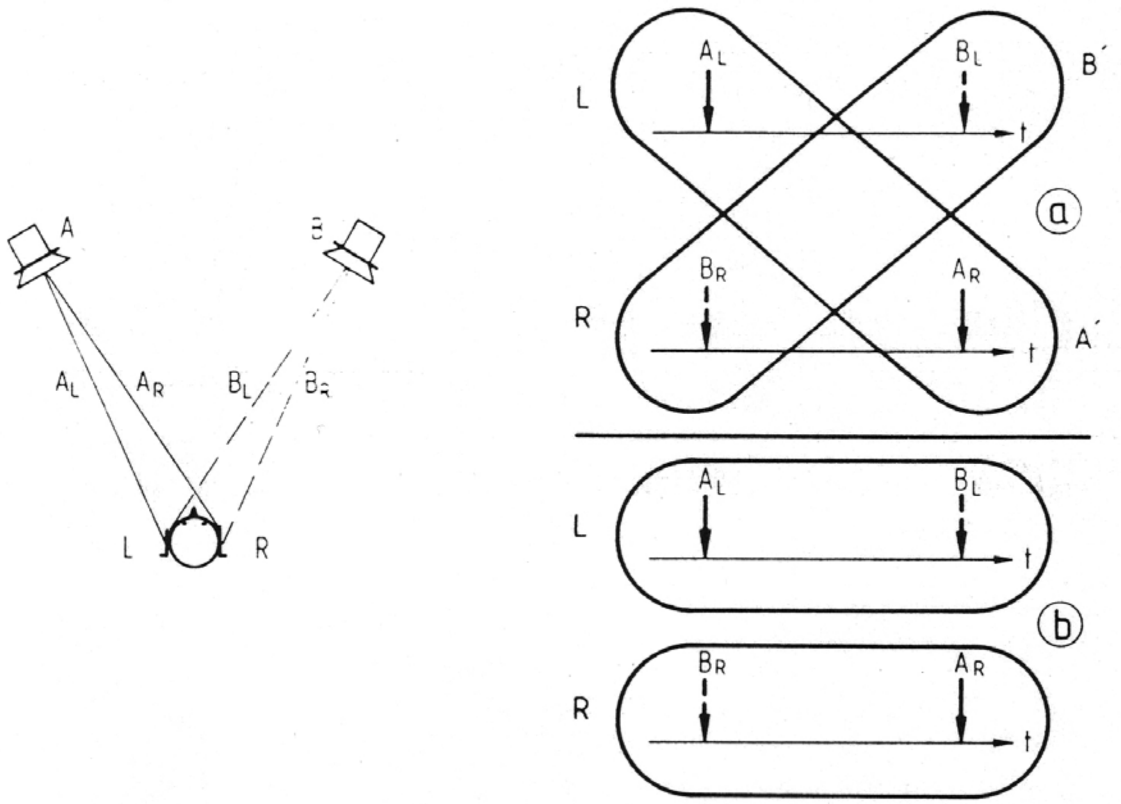


Figure 18 :

Evaluation of the ear signals in a phantom source situation
a) association principle
b) summing localisation principle

Using the hypothetical location association, the hearing system is able to retrieve those components from the superimposed sound fields AL + BL and BR + AR that each stem from one loudspeaker. Due to their suitable temporal and spectral characteristics, the components AL / AR and BL / BR each cause a location association. This is contrasted by the summing localisation principle shown in Figure 18 b, the difference to the association principle being clearly visible. The summed signals AL + BL and BR + AR are evaluated; they contain the spectral changes due to the present comb filter (see Section 2.2.1).

It appears that the association model offers a plausible explanation for the ineffectiveness of the comb filter effect in the phantom source situation – presumably, this is the only possible explanation.

In the next chapter, the association model shall be discussed in terms of specific auditory phenomena occurring for particular constellations of the superimposed sound field.

4. The localisation process in the superimposed sound field – Discussion of the association model

The localisation model described in the previous chapter states that, due to its association properties, the auditory system is capable of discriminating at least two sound events simultaneously – even if they differ with respect to their spatial qualities only. The hypothetical demarcation of the location and gestalt association stages as well as the assumption that both stages carry out a specific pattern selection as a result of associative signal processing form a possible basis for the explanation of our hearing system's capabilities in the superimposed sound field.

If it is assumed that the two association processes have different effects only because of different stored patterns, then, within certain limits (see Section 4.1), the abilities of the location association stage can be inferred from the abilities of the gestalt association stage. To give an example, the ability of the gestalt association stage to discriminate simultaneously two (or more) sound events with identical spatial properties corresponds to the hypothetical ability of the location association stage to discriminate simultaneously two (or more) sound events having identical gestalt properties.

From an information theorist's point of view, the limits of the localisation stimulus selection are reached if the information reduction, which is achieved during associative signal processing, is used in an optimal way. Within the framework of this work, however, the limits shall only be further investigated with respect to the simultaneous discrimination of **two** particular sound events of an arbitrary spatial constellation.

4.1 Limits of the localisation stimulus selection

As was pointed out in Section 3.1, the hypothetical localisation stimulus selection in the sound field produced by two sound sources can be understood as an adaptive filter. This filter performs the spatial decoding according to Equations (5) and (6), provided that the associated pattern recognition process can supply the information necessary for controlling the filter.

It has to be verified now if and with what restrictions a localisation stimulus selection can be assumed. For the purpose of determining these limits, the two established mechanisms of 'adaptive filter' and 'associative pattern recognition' shall be utilised.

A prerequisite for the spatial decoding is that an adequate pattern recognition can take place. That is why a useful signal pattern needs to be specified first, with the help of which the possibility of pattern recognition can be demonstrated. In conformity with many well-known theories of auditory perception, the hearing system is interpreted as a correlation receiver in this context (see LICKLIDER 1951, SAYERS / CHERRY 1957, SCHREIBER 1965, GRUBER 1967, OSMAN 1971, SCHROEDER 1975, BLAUERT 1974, 1978).

It is presumed that to begin with the ear signals are subdivided into spectral components of approximately constant relative bandwidth by means of a filter bank (see ZWICKER / FELDKELLER 1967, DUIFHUIS 1972, BLAUERT 1974). The signals appearing at the filters' outputs are then subjected to a short-term auto-correlation analysis of the form

$$\Phi(\tau) = \overline{x(t) \cdot x(t-\tau)} = \frac{1}{T}\int_0^T x(t) \cdot x\ (t-\tau)dt$$

as well as to a short-term cross-correlation analysis of the form

$$\Phi_{LR}(\tau) = \overline{x_L(t) \cdot x_R(t-\tau)} = \frac{1}{T}\int_0^T x_L(t) \cdot x_R(t-\tau)dt$$

so that an analysable pattern is present in the various functions of τ. This pattern shall be referred to as the **"binaural correlation pattern"**.

The binaural correlation pattern supplies monaural as well as interaural temporal and spectral information. A sound source located at a certain distance produces maxima in the binaural correlation pattern. Their relative locations, heights and widths can thus determine the location of the auditory event. It is possible to determine the discriminability of different location-determining patterns that are simultaneously present in the superimposed sound field, which shall be described by means of the discriminability of the maxima within the time delay region τ.

If, for example, one computes the auto-correlation function of an ear signal stemming from a phantom source situation featuring two loudspeakers (conventional stereo set-up) radiating coherent, band-limited white noise (Δf = 10 kHz), one will find two distinct maxima separated by τ_S = 250 µs, as illustrated in Figure 19 (only the positive τ-axis is shown).

The minimal separation of these maxima necessary for their discrimination shall be defined as the resolution limit τ_{Smin}. τ_{Smin} decreases as a function of the width $\Delta\tau$ of these maxima. Due to

$$\Delta\tau = \frac{1}{2\Delta \mathrm{f}}$$

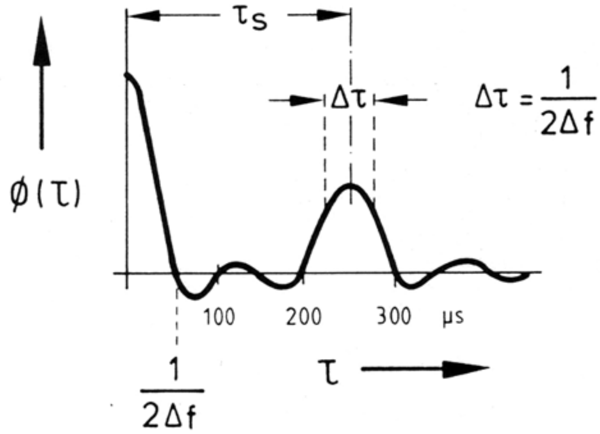


Bild 19 :

Auto-correlation function of an ear signal, conventional stereo set-up, loudspeaker signals: band-limited noise f0 = 10 kHz

a minimal signal bandwidth exists that is required for an adequate pattern recognition. Supposing that two maxima can still be discriminated if their separation τ_S is not smaller than their width $\Delta\tau$, i.e. if

$$\tau_{Smin} = \Delta\tau$$

then for the phantom source situation with τ_S = 250 µs and for signal bandwidths below

$$\Delta f = \frac{1}{2\tau_S} = \frac{1}{2 \cdot 250\mu s} = 2kHz$$

erroneous pattern recognition would result. As yet, however, the assumed resolution limit does not take into account the effects of associative signal processing. To the extent to which hidden components of a current input pattern can be recognised with the help of the stored binaural, location-determining pattern (see Section 3.1.2), localisation stimulus selection is also possible for correspondingly narrow bandwidths.

Another limitation of the effectiveness of the localisation stimulus selection derives from the maximum delay time τ_{max} of the correlation processes. The values given in the literature vary a lot and lie in-between about 2 and 20 ms. BLAUERT 1974 concludes the existence of a maximal delay time, since the lateral displacement of an auditory event disappears if a certain interaural phase delay is exceeded (BLODGETT et al. 1956).

Assuming that a τ_{max} exists, coherent signal parts can only lead to different localisation stimuli if their time difference is $\Delta t < \tau_{max}$. Two signals with a time difference greater than τ_{max} will be judged to be completely incoherent by the auditory system. An auto-correlation analysis of the sum of two such signals would reveal only one maximum. In this case, spatial decoding is no longer possible, provided that the signals are sufficiently similar.

This hypothesis can easily be verified by modifying the experiment described in Section 2.2 and shown in Figure 3. Using an interchannel time difference of, say, 3 ms, sound colour changes (as caused by lateral head movements) are also perceivable when listening binaurally. It appears that τ_{max} is smaller than 3 ms, so that both localisation stimuli coincide in the τ-region and are thus inseparable. Consequently, the comb filter effects occurring at the ears are audible (see Section 4.3.1).

For the case of a sufficiently broadband signal, a third limitation of the effectiveness of the localisation stimulus selection has to be specified, i.e. for spectral components below a critical frequency f_K.

In Section 3.1.1 it was stated that under certain conditions an inverse matrix "M^{-1}" will exist, which reverses the interaural crosstalk ("linking") of the ear signals that is caused by the matrix M completely. A prerequisite for this is not just the associative pattern recognition process described in Section 3.1.2, which supplies the information necessary for controlling the matrix "M^{-1}". Rather, the third limitation becomes evident when scrutinising the matrix "M^{-1}" more closely.

Presupposing a decoder matrix that is optimally adapted to the sound source (see Figure 11), the function of this matrix "M^{-1}" can be described numerically. It can be shown that the system does not give an unambiguous response for low frequencies, because the acoustical **transmission factors** will approach 1 for frequencies below the critical frequency f_K.

This is in line with the finding that the low-frequency components of a sound event contribute very little to the formation of an auditory event's location. For instance, in the case of loudspeaker reproduction the location of an auditory event will not depend on the location of the woofer if the crossover frequency is below about 250 Hz (see Section 4.2.2).

To summarise, it is assumed that the two localisation stimuli that stem from the loudspeakers of a stereo set-up and that give rise to a "phantom source" can be discriminated. However, the following restrictions apply:

1. For a given bandwidth Δf, the time difference between the superimposed signals must not fall short of a critical value Δt. Conversely, for a given Δt the bandwidth must not be less than a critical value Δf. For a conventional stereo set-up, $\Delta f < 2kHz$ (estimate).

2. The time difference Δt between the superimposed signals must not exceed a certain value.

3. Below a critical frequency f_K localisation stimulus selection is no longer possible; $f_K < 500$ Hz (estimate).

4.2 Two coherent sound sources

Within the qualitatively specified area of validity, a localisation stimulus selection is assumed for the phantom source situation. It shall now be demonstrated that based on this hypothesis the spectral and temporal characteristics of ear signals can be brought into agreement with the direction, distance, elevation, width and sound colour of an auditory event.

The limits of the localisation stimulus selection also have to agree with the observable phenomena of auditory events if the association model shall describe the auditory system's localisation function in a superimposed sound field produced by two coherent sources. As will be shown, this is actually the case for a number of different phantom source phenomena. These include:

1. the localisation in the case of low frequencies
2. the sound colour of the phantom source
3. the localisation of lateral phantom sources
4. the auditory events in the case of narrow-band signals.

In the following sections, the association model will be discussed with regard to different well-known phantom source phenomena and the results of the listening test described in Section 2.2.2.

4.2.1 The elevation and distance of the phantom source

It is possible that the phantom source location is not determined by the summed signals occurring at the ears, but rather by the localisation stimuli that the loudspeaker signals produce separately. Since the fusion process takes place after the spatial decoding has been completed, the phantom source location is given by the means of the distances and directions of the discriminated source locations. For this reason, distance perception is not influenced by the comb filter effect, and neither do the linear distortions (see Section 2.2.1) cause inside-the-head locatedness, nor do they prevent front-back reversals (THEILE 1975).

The elevation effect can be attributed to the localisation stimulus selection, too. If two loudspeakers (arranged symmetrically with respect to the median plane) are moved on a circle in the horizontal plane with the listener being positioned in the centre, the auditory event location in the median plane will move with the discriminated average source distance. The mean of the distance and direction lies on a circle in the median plane. The elevation angle results from sound colour properties that, due to the localisation stimulus selection, are not influenced by the comb filter effect.

This is proven by the fact outlined in Section 2.2 that equivalent interchannel level and time differences (loudspeakers positioned at the listener's sides along the interaural axis) do not lead to different elevations of the phantom source, even though spectral differences will occur in the crucial frequency region around 8 kHz (see BLOOM 1977). If the listener moves the head sidewards by a few centimetres in the direction of one loudspeaker, the resultant change of the (elevation-salient) spectrum will be very different from the change that a head rotation in front of a substitute (real) sound source would produce.

The spectrum salient to perceived elevation is not the resulting spectrum of the ear signals, but the average spectrum of the two ear signals that the loudspeakers produce separately. If, as a result of the localisation stimulus selection, the distance e and horizontal direction φ of the phantom source have been determined, it will be located on a circle with the radius e. The plane of this circle will be perpendicular to the horizontal plane and will exhibit the direction φ.

The position on the circle is determined by the average spectrum, i.e. the phantom source will be perceived to be at the location at which its sound colour characteristics match those of a substitute sound source as closely as possible.

This notion explains the measurements made by DAMASKE 1969/70 who determined perceived elevation as a function of the angle of the stereo base $\Delta\Omega = 2|\Omega_A| = 2|\Omega_B|$ using band-limited noise (0.65 to 4.5 kHz) as the test signal. In principle, he found that as $\Delta\Omega \rightarrow 180°$ the elevation angle $\varepsilon \rightarrow 90°$, even though the direction-determining frequency band was not contained in the noise signal. Regrettably, DAMASKE did not report the associated distances of the auditory events. The result contradicts the summing localisation principle.

Furthermore, the association model predicts a dependency of the elevation and distance of a phantom source onto the bandwidth of the signal. If the bandwidth is below a critical value, localisation stimulus selection is not possible. In this case, elevation and distance perception do no longer result from the spectra that the loudspeakers produce individually.

For informal verification purposes, a simple experiment can be executed. Coherent, low-pass filtered noise with f_o = 500 Hz is played back in an anechoic chamber over a loudspeaker-based stereo set-up. The subject stands in front of the loudspeakers, $\Delta\Omega \approx 150°$. The auditory event is perceived close to the head. As expected, the distance information is destroyed (a single loudspeaker would lead to an adequate auditory event distance), the elevation angle is zero. The result is the same if the subject moves backwards. The auditory event is characterised by a modest width.

If the high-frequency content of the noise signal is suddenly switched in, an interesting effect occurs. At first, a second, high-frequency auditory event appears in addition to the low-frequency one, which is clearly separated in space, i.e. it seems to be above the first one and farther away. Shortly afterwards, the two auditory events fuse. The low-frequency event moves upwards, the resultant auditory event location being given by the position of the high-frequency event.

The association model also explains this phenomenon, which until now has not been described. Switching in the high-frequency components leads to a greatly improved pattern recognition, because the signal bandwidth is increased. Hence, the characteristics of the discriminated localisation stimuli determine the distance and elevation of the phantom source. Although the components are not discriminable at low frequencies (see Section 4.1), they are attributed to the phantom source. It seems that this due to the association processes.

4.2.2 The sound colour of the phantom source

The non-discriminable low-frequency content also influences the sound colour of the phantom source. In the case of complete spatial decoding, the sound colour is determined by the average spectrum of the two ear signals that the two loudspeakers produce separately. A comparison of the sound colour of a phantom and a real source (both having the same location in the median plane and the same loudness for a white noise test signal) reveals that at low frequencies the loudspeaker signals cannot be sufficiently discriminated to suppress completely the level increase evident below about 500 Hz (approx. +5 dB, see Section 2.2.1, Figures 4 and 5). Each of the two discriminated localisation stimuli exhibits the level increase at low frequencies. The sound colour of the phantom source is ‘darker’ than the one of the real sound source. This can be seen as another confirmation for the validity of the association model.

The model shall now be discussed further with regard to the results of the listening test outlined in Section 2.2.2. The aim of the experiment was to prove that the comb filter effect evident at the ears does not influence the sound colour of the phantom source as much as is stipulated by the summing localisation theories.

To recap, an inverse relationship was measured between the perceptibility of changes in sound colour and direction of the auditory event (as caused by laterally displacing the artificial head positioned in front of the phantom source configuration) and the inclusion of more or less serious errors in the reproduction set-up (test signals B1 … B6).

The association model predicts this finding. The localisation stimulus selection requires an adequate pattern recognition, which the different reproduction errors handicapped to a greater or lesser degree. The magnitude of this localisation stimulus ‘destruction’ determines the audibility of the spectral changes in the ear signals.

The result for test signal B5 (see Section 2.2.2, Item 6, Figure 10) clearly shows that the introduced reproduction errors lead to a reduced discriminability of the localisation stimuli. Due to the absence of the pinnae, hardly any changes in the interaural time and level relationships are produced; essentially, the ear signals are linearly distorted. These linear distortions reduce the perceivability of directional changes. This can be explained by means of a gradual destruction of the localisation stimuli, even if the resultant ear signals lead to an inside-the-head locatedness already. Neither in a localisation experiment (one sound source) nor in a lateralisation experiment can linear distortions of such a magnitude impair the perceptibility of the auditory event’s direction or displacement.

It can be shown that the degree of interaural coherence is not strongly influenced by the pinnae. In a phantom source situation, this measure will be small and, in principle, would lead to an increased width of the auditory event (see JEFFRESS/ BLODGETT/ DEATHERAGE 1962). It is only the spectral effects of the pinnae that provide the hearing system with the possibility to separate clearly two localisation stimuli, so that despite the small interaural degree of coherence a defined auditory event location can occur.

At this point, it is interesting to note that even if the artificial head system is free from reproduction errors, the localisation stimulus selection does not appear to work completely either. Compared to a corresponding rotation of the artificial head in front of a real sound source (Item 3, Figure 9, Test signals A1 and C2), the perceivability of directional changes is smaller. Based on this finding, one could deduce that the employed dummy head system causes sound imaging errors. However, more listening tests would have to be carried out to be able to confirm that this is the case. The reproduction of a phantom source by means of artificial head signals requires a very high degree of accuracy of the ear signals, because in this case the localisation stimulus selection will be more difficult than in a natural superimposed sound field.

In the phantom source situation the pinnae are especially important. It can be shown that their influence onto the sound colour of a phantom source is larger compared to their influence onto the sound colour of a real source.

Therefore, the suppression of sound colouration is dependent on the localisation stimulus selection. If this mechanism is impaired, the resultant spectra will have the effect postulated by the summing localisation principle. The transition between impaired and unimpaired localisation stimulus selection is continuous.

4.2.3 Lateral phantom sources

A destruction of the localisation stimulus selection does not just occur as a result of the destruction of the localisation stimuli (e.g. for monaural listening), but also in those cases for which the two stimuli do not sufficiently differ in the binaural correlation pattern. For broadband signals this is the case if the time difference Δt between the superimposed signals is either too large or too small (see Section 4.1).

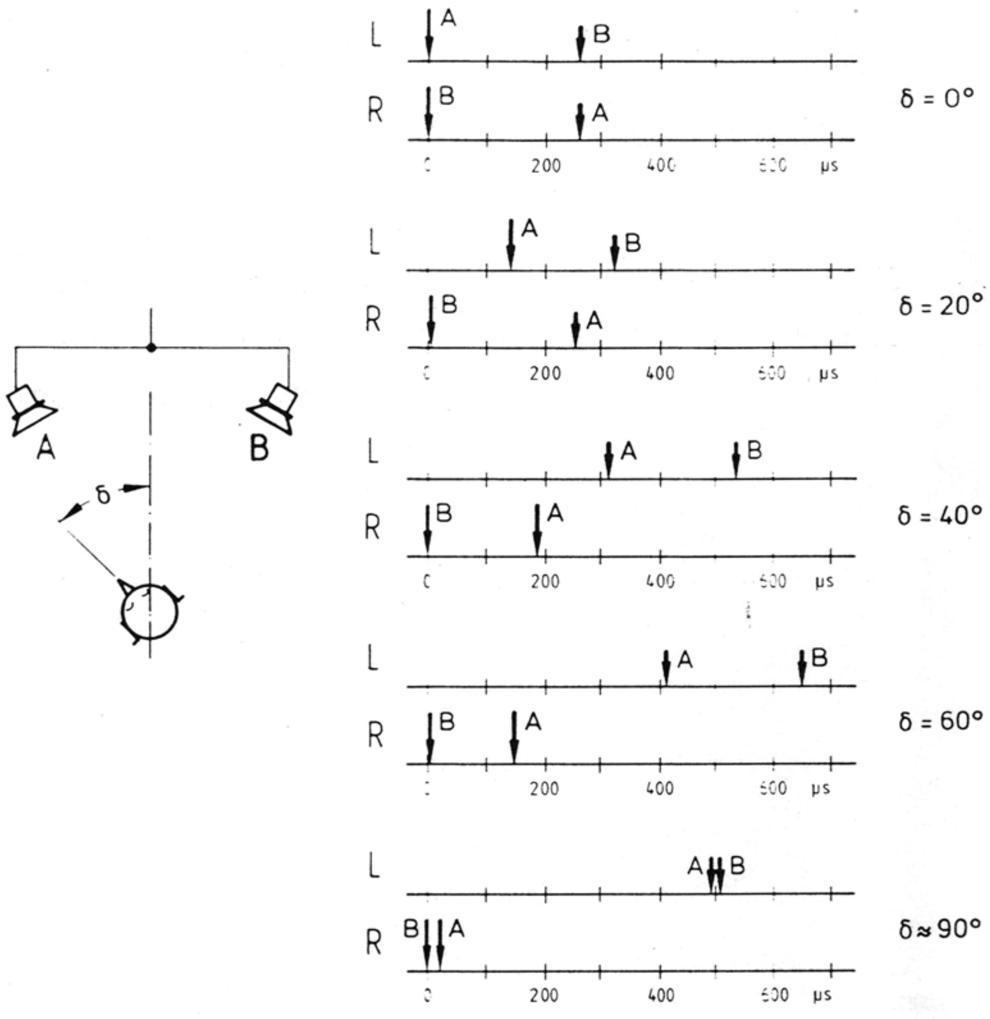


Bild 20:

The interaural time characteristics for a rotation of the head in the superimposed sound field produced by a conventional stereo set-up

For example, too small a time difference Δt will result for a loudspeaker arrangement that is symmetrical with respect to the frontal plane (Figure 20, $\delta = 90°$). The two maxima at each ear will appear at the same time; the localisation stimuli are not discriminable. Consequently, there will be no suppression of sound colouration, which was confirmed experimentally in Section 2.2.2 (Item 5, Figure 9b, Test signal A4). Furthermore, from Figure 9b it can be seen that the perceptibility of sound colouration increases as $\delta \rightarrow 90°$. This is in line with the decrease in Δt that is evident from Figure 20.

In THEILE / PLENGE 1977 phantom source direction was investigated as a function of δ. The results showed that:

1. For $\delta = 90°$, the auditory event is located either at the frontal or the rear loudspeaker, depending on the level ratio of the loudspeaker signals. In principle, lateral phantom sources do not arise. This is explained by the localisation model if one assumes that for equal signal levels the localisation stimuli cannot be discriminated due to identical interaural time differences, but that for unequal levels one or the other spectrum will dominate.

2. The focus of the phantom source greatly decreases for $\delta > 60°$. The time differences between the superimposed signals are of the magnitude $\Delta t < 200$ µs (see Figure 20). It seems that this is where the discrimination threshold of the localisation stimuli lies.

3. As $\delta \rightarrow 90°$, the direction of the phantom source does not remain in the centre of the loudspeaker base, but shifts by maximally 10° in the direction of $\varphi = 0°$. This effect is in agreement with the fact that for a laterally displaced loudspeaker base the localisation stimuli will not be simultaneously available. Of course, a localisation stimulus can only be discriminated if it is available in binaural form. This will happen later for the signal radiated by the loudspeaker closer to the frontal plane (Signal B in Figure 20). Thus, the model explains the displacement from the centre of the loudspeaker base by means of a time difference between the localisation stimuli. This time difference will have a maximum value of about 200 µs, which, according to measurements made by WENDT 1963, corresponds to a change of $\varphi = 10° \ldots 15°$ in phantom source direction (see Section 4.3.1).

4.2.4 Auditory events in the case of narrow-band signals

The association model is based on a pattern recognition process that is made possible by associative stimulus processing. The localisation process thus requires a sound source that allows for a sufficiently robust distance perception (see Section 3.2.1). This is the case if spectral characteristics of a current stimulus can be analysed. Associative stimulus processing necessitates a sufficiently broadband stimulus.

In the phantom source situation, a localisation stimulus selection only takes place if a certain minimum signal bandwidth is available. This bandwidth derives from treating the hearing system as a correlation receiver; it is likely to reduce if the operation of associative pattern recognition is assumed.

Hence, if the two loudspeakers radiate narrow-band signals, “summing localisation” is an accurate description of the listening conditions. From the localisation model it can be inferred that an investigation of the phantom source with the help of pure tones does not allow conclusions to be drawn about the functioning of the auditory system when presented with broadband signals. Even the application of such findings to the sound imaging principles of stereophony is questionable.

An example for this is the FRANSSEN effect (FRANSSEN 1960). Loudspeaker A is driven with a tone having a duration of several seconds and an exponential attack and decay [Signal $u_A(t)$]. Loudspeaker B is fed with Signal $u_B(t)$, which has such properties that the sum $u_A(t) + u_B(t)$ would result in a rectangular amplitude envelope. Loudspeaker B therefore reproduces the (broadband) onset, whereas Loudspeaker A reproduces the (narrow-band) tone that is free from on- and offsets. It is found that the auditory event location is exclusively determined by the onset. This is in conformity with the association model if it is assumed that the tone does not constitute a localisation stimulus, but rather that it is attributed to the localisation stimulus already available because of missing distance information. In this respect, WENDT 1964 found for stereophony that tones with very rapid on- and offsets behave like broadband signals. A relationship between these tones and the auditory event locations, which WENDT determined for continuous and impulsive tones, cannot be observed.

There is no uniform summing localisation theory that allows correct predictions to be made for continuous tones (or Gaussian tones, see BOERGER 1965) as well as broadband signals. This is not possible because in the case of tones only the lateral displacements (as caused by the resulting interaural phase differences) will be interpreted as the auditory event’s direction (due to preconditioning of subjects, see e.g. JEFFRESS / TAYLOR 1961, PLENGE 1973). In the case of broadband signals, the hearing system initially discriminates two localisation stimuli, which are subjected to a fusion process subsequently.

At this point, an important finding announces itself. For a fixed head located in the far field, distant sources radiating pure tones do not supply any information about their distance. The resultant interaural phase and level relationships are equivalent to those of a corresponding headphone-based reproduction set-up.

In both cases, the location association stage of the localisation model does not supply spatial information, i.e. the ear signals pass it without being affected by it (see Figure 16). The interplay of interaural time and level relationships is thus liable to the same psychoacoustic regularities. These regularities, however, can be traced back to the effect of the gestalt association stage (see Chapter 5).

4.2.5 Auditory events in the case of out-of-phase signals

The special case of auditory events created by out-of-phase signals provides further confirmation for the validity of the association model. This sound field constellation has been studied several times for narrow- as well as broadband signals (e.g. see GARDNER 1969, PLENGE 1972, MATSUDAIRA / FUKAMI 1973, LOY 1978). Based on the work by SANDEL / FEDDERSEN / JEFFRESS 1955, LOY 1978 calculated the phantom source direction for pure tones by deriving interaural phase delays from the resultant interaural phase differences. This was also the approach taken by LEAKEY 1959, WENDT 1963, 1965 and others. For frequencies below 2 kHz, this results in auditory event directions $\varphi > 30°$, i.e. auditory events outside the loudspeaker base (as frequency decreases $\varphi \rightarrow 90°$). Whilst LOY could confirm his predictions experimentally for Gaussian tones, a very large spread of the responses could be observed that increased as the interchannel level difference $\Delta L \rightarrow 0$. For $\Delta L = 0$, no directional judgements were possible.

Besides, LOY did not provide any information about the distances of the auditory events. According to the association model, Gaussian tones do not give rise to a location association – neither in the sound field of a real source, nor in the phantom source situation. Instead, the interaural signal differences only lead to a lateral displacement, which will be interpreted as the auditory event direction because of the experimental conditions (e.g. the subject's knowledge that the sound source is a certain distance away, evaluation of the direction of the auditory event rather than its location).

Yet, it comes as a surprise that these results agree with those obtained by other authors who used broadband test signals for their studies. GARDNER 1969 as well as PLENGE 1972 found that the auditory event will be perceived to be in the vicinity of the listener's head if the loudspeaker signals have equal levels ($\Delta L = 0$). If a level difference exists ($\Delta L \neq 0$), the location of the centre of the auditory event will shift as a function of ΔL to maximally the loudspeaker distance in a direction outside the loudspeaker base, i.e. more or less according to the results reported by LOY 1978. A significant localisation blur is characteristic of this particular listening condition.

Using speech and music recordings as the test signals, MATSUDAIRA / FUKAMI 1973 measured phantom source direction as a function of the interchannel phase difference Θ. They reported that up to $\Theta \approx 90°$ the localisation blur was low and that the direction of the auditory event $\varphi \rightarrow 25°$ as $\Theta \rightarrow 90°$. For even larger phase differences (i.e. $\Theta \rightarrow 180°$) the localisation blur increased dramatically. However, the medians of the auditory event directions approached zero, $\varphi \rightarrow 0°$. Auditory events outside the loudspeaker base did not arise. These effects can only be observed if the signals contain spectral components below about 2 kHz (BLAUERT 1974). The low-frequency components determine the displacement of the auditory event. It seems that a fusion of the discriminated localisation stimuli does not take place (see Section 4.2.1). The localisation stimulus selection takes place in the location association stage (see Section 3.2.2) and thus is independent of the phase relationship between the source signals. It is not until the gestalt association stage that the relationship between the source signals takes effect.

An increasing phase difference $\Theta \rightarrow 180°$ leads to the two high-frequency localisation stimuli progressively causing separate auditory events. The effect is similar to the sound image produced by incoherent source signals (see DAMASKE 1967/68). The sound image disintegrates more and more and localisation blur increases. For $\Theta = 180°$ and $\Delta L = 0$, the "locatedness is almost completely undefined" (MATSUDAIRA / FUKAMI 1973). The centres of the auditory events are close to the listener's head (low-frequency components) and at the positions of the loudspeakers (high-frequency components). For a constellation that is asymmetric with respect to the median plane ($\Delta L \neq 0$), a high-frequency auditory event dominates over the second one. The direction of the centre of the auditory event is determined by the resultant low-frequency component.

Thus, it follows that lateralisation experiments do not allow inferences to be made about the location association stage. It is striking that for both headphone and free field reproduction the source signals' relationship gives rise to similar auditory event characteristics. As will be illustrated in Chapter 5, this does not just apply to the coherence and phase relationships, but also to the level and time differences between the signals.

In this section, the phase relationship is especially important. It appears that out-of-phase signals lead to a discriminability in the gestalt association stage that is dependent on the spectrum of the signals. This is in agreement with the psychoacoustic principles known from signal detection studies, which are predominantly verified in lateralisation experiments. The BMLD (for a definition see BLAUERT 1974, pp. 206) for $N_\pi S_0$ or $N_0 S_\pi$ has a magnitude of about 12 dB, which, in the context of the association model, can be interpreted as the ability of the gestalt association stage to discriminate out-of-phase stimulus responses of the location association stage. In contrast, this differentiation does not seem possible for in-phase stimulus responses. These conditions correspond to the observation made by SAYERS 1964 that in a lateralisation experiment interaural phase differences around $\Theta = 180°$ can create two auditory events.

4.3 Two incoherent sound sources

In order to specify qualitatively the limits of the localisation stimulus selection, the binaural correlation pattern was utilised. In view of the discriminability of the maxima within the time delay region τ, a possible discriminability of the source signals was deduced. In the case of two coherent sound sources, this approach can explain a limited area of validity of the localisation stimulus selection. For two *incoherent* sources, however, it does not provide a sufficient basis for the spatial decoding. To illustrate, whilst two speakers positioned at different locations produce two auditory events, they only create one maximum within each region of the binaural correlation pattern.

It seems that the auditory system features much more effective processes for the localisation stimulus selection in the superimposed sound field. It is assumed that important selective properties of the hearing system, including those of the location and gestalt association stages, can be traced back to the effect of the hypothetical associative stimulus processing. Hence, the occurrence of "simultaneous auditory events" when presenting different signals over a single loudspeaker is the result of the stimulus processing taking place in the gestalt association stage. Accordingly, "simultaneous location associations" can be triggered due to similar mechanisms being at work in the location association stage.

Both stages enable the simultaneous recognition of different patterns, since "simultaneous auditory events" can be characterised not just by different auditory event gestalts, but also by different auditory event locations. This is even possible for two source signals having identical amplitude envelopes if their spectra are sufficiently far apart ("double localisation" of two Gaussian tones, see BOERGER 1965). The likelihood for this effect to occur is especially high if the source signals exhibit uncorrelated amplitude envelopes (cocktail party effect).

The area of validity of the association model in the superimposed sound field is given by the discriminability of the source signals. In the case of coherent sound sources, the point at which the localisation stimuli can still be discriminated depends on the signal spectrum (see Section 4.1) as well as the resulting time difference Δt between the superimposed signals. For incoherent sources, this discrimination threshold shall be defined by the "maximum gestalt resolution" of the hearing system, which is given when the sound events are spatially congruent.

As a result of the location association stage, every auditory event occurring in the superimposed sound field contains some associated spatial information. However, discriminable localisation stimuli are not just available for relatively dissimilar source signals, but, under certain conditions (see Section 4.1), also for spatially separated sources that emit identical signals. This is an important statement made by the localisation model.

4.3.1 The "law of the first localisation stimulus"

Accepting the correctness of the assumption made in Section 4.1 that the hearing system judges two coherent signals as being incoherent if their relative time difference exceeds a value $\Delta t = \tau_{max}$, an interesting relationship between the localisation stimulus selection and the "law of the first wavefront" (CREMER 1948) emerges.

For a conventional stereo-up, a phantom source shifts from $\varphi = 0°$ to $\varphi = 30°$ if the time difference between two broadband loudspeaker signals is increased from zero to about 600 µs. The association model could explain this phenomenon (time- as well as level-based stereophony) by means of psychoacoustic principles of the gestalt association stage. The localisation stimulus arriving at the gestalt association stage first has a greater weight compared to the second stimulus (the equivalent for level-based stereophony would be the localisation stimulus with the higher level). Despite their identity and relative time delay, the localisation stimuli can be discriminated, since each of them is present in the binaural correlation pattern in a complete and discriminable form (see Section 4.1).

Yet, a further increase in the interchannel time difference leads to an exceedance of the maximal time delay τ_{max}. For stationary broadband signals (continuous noise), this causes a disruption of the localisation stimulus selection, which manifests itself in the form of a reduced suppression of the comb filter effect, for example. In this particular sound field constellation, the law of the first wavefront cannot be observed in accordance with the association model. Analysable wavefronts that would allow for a localisation stimulus selection of the impinging sound components do not exist.

In contrast, for non-stationary impulsive signals (clicks, speech, impulsive tones) an increase in the interchannel time difference has a different effect. In the association model, evaluation of the amplitude envelope ensures that the primary and the delayed sound (reflection) can be discriminated as localisation stimuli. According to a hypothetical function of the gestalt association stage, the primary localisation stimulus determines the auditory event. It does this even more so the larger the time difference between the arriving localisation stimuli gets. Only when a time difference of about 10 … 30 ms is exceeded will the subsequent localisation stimulus gain in perceptual weight. Beyond the echo threshold (for a definition see BLAUERT 1974), it will be perceived as a separate auditory event.

It appears that the "law of the first wavefront" can be interpreted as the **"law of the first localisation stimulus"**.

Moreover, the model states that for time differences between 0 and 10 … 30 ms no fundamental difference exists in the stimulus evaluation. Rather, it is assumed that the psychoacoustic regularities of both time-based stereophony and the law of the first wavefront can be traced back to a time-dependent evaluation of the localisation stimuli that arrive one after the other.

Of course, the law of the first localisation stimulus is only valid if a localisation stimulus selection can take place. Therefore, the sound field conditions discussed in Sections 4.1 and 4.3 have to be satisfied. In particular, the need for sufficiently broadband signals also applies in the case of the law of the first wavefront. This is clearly demonstrated by an investigation by BLAUERT / COBBEN 1978 who measured the auditory event direction for narrow- and broadband impulses as a function of the interchannel time difference (see Figure 21).

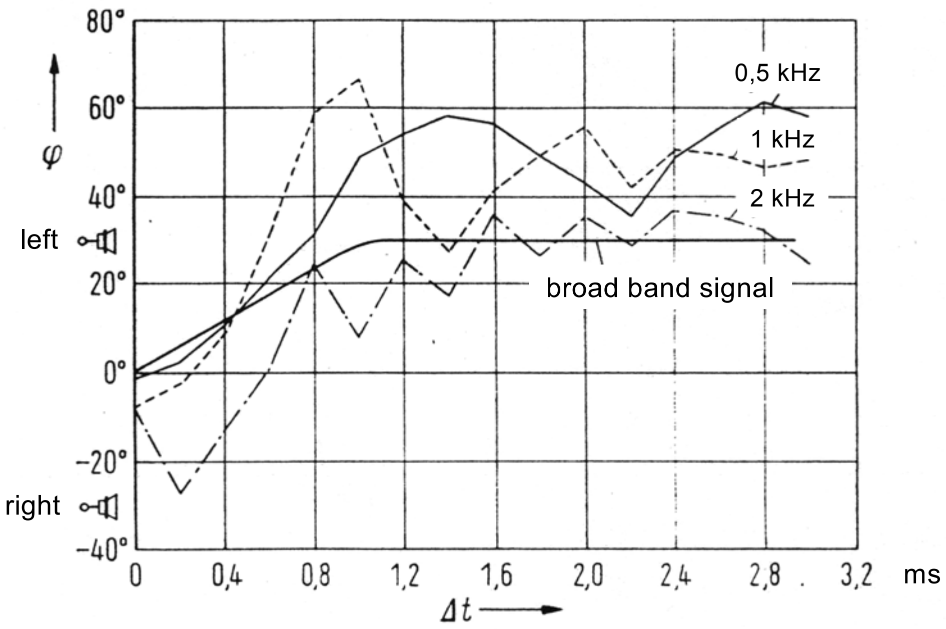


Bild 21:

Auditory event direction φ as a function of the time delay Δt of the right loudspeaker signal (after BLAUERT / COBBEN 1978)

Neither is there agreement between the localisation curves of the low-frequency narrow-band signals and the curve measured for the broadband signal within the region of time-based stereophony, nor can any agreement be certified for larger time differences. From the periodicity of these curves it can be seen that within the whole time delay region summing localisation has taken place, i.e. that the resultant interaural signal differences have determined the direction of the auditory event. In conformity with the association model, the law of the first wavefront is not valid for low-frequency narrow-band signals. This "anomaly of the law of the first wavefront" (BLAUERT / COBBEN 1978) is equivalent to the respective anomalies of time-based stereophony (see WENDT 1963). Both phenomena arise due to insufficient signal bandwidths.

Particular support for the assumed "law of the first localisation stimulus" is provided by the fact that the principles of time-based stereophony as well as the law of the first wavefront also apply if the first signal is solely sent to one and the delayed signal exclusively to the other ear using headphones. As will be shown in Chapter 5, in the case of dichotic presentation the stimuli pass the location association stage without being affected by it. As long as an adequate spatial decoding takes place (see Section 4, Figure 12), identical stimulus responses (A' and B', respectively) will appear at the inputs to the gestalt association stage for loudspeaker as well as headphone reproduction. Since the "law of the first localisation stimulus" is attributed to the time-dependent evaluation of stimuli arriving at the gestalt association stage one after the other, its area of validity has to include dichotic headphone reproduction. Specifically, this applies to the following phenomena:

1. For both loudspeaker and headphone reproduction, the maximum lateral displacement of the auditory event occurs for time differences in-between 655 and 800 µs (see TOOLE / SAYERS 1965, WENDT 1963).

2. For clicks, the localisation blur and the lateralisation blur of a phantom source positioned in the median plane is about 20 µs (see KLEMM 1920, WENDT 1963, HALL 1964).

3. For both modes of reproduction, the law of the first wavefront is valid in the case of speech signals and time delays in-between 0.8 … 20 ms (the echo threshold is used for the threshold definition, see BLAUERT 1974).

4. In both cases, the echo threshold is strongly signal-dependent. The steeper the on- and offsets of the signals, the smaller the time delay at which the auditory event splits into a primary auditory event and an echo.

5. If the level of the reflection is increased relative to the one of the primary sound, the time delay needed for an echo to occur decreases in both cases. If the reflection level is lowered, the opposite effect occurs (see BABKOFF/ SUTTON 1966, BLAUERT 1974).

For impulsive broadband signals, the association model thus predicts that the auditory events arising due to interchannel time differences can be explained with the help of the "law of the first localisation stimulus". This means that the stimulus responses arriving at the gestalt association stage first will dominate the processing carried out in this stage, whilst the later responses will be suppressed. The model further states that time-based stereophony is also possible for stationary broadband signals, whereas the law of the first wavefront can only be verified for signals having distinctly transient amplitude envelopes. Thirdly, it proclaims that the effects caused by signal delays should in principle also be observable for headphone reproduction, provided that an adequate localisability of the individual sound source is possible.

4.3.2 The cocktail party effect

The selective properties of both the location and gestalt association stage enable the simultaneous discrimination of different patterns. Depending on the source locations and the source signal characteristics, the associated location and gestalt associations arise. Supposing that each selection stage can discriminate at least two patterns simultaneously, a general approach results that can also be used for explaining the cocktail party effect by means of the association model.

A phantom source disappears as soon as the two loudspeaker signals become sufficiently dissimilar, i.e. as soon as they produce different auditory event gestalts. This is the general, natural case of a superimposed sound field. Normally, two sound sources do not just give rise to two different location associations, but also to two different gestalt associations. The two resultant auditory events therefore occur after a two-stage selection from which the largest possible resolution derives. If only the gestalt association stage contributes to the auditory event resolution because the two loudspeakers are located at the same place, the auditory event discrimination decreases to its lower limit (resolution limit of the gestalt association stage, see Section 4.3). The significance of the gestalt association stage to spatial hearing is evident here, too. Simultaneous auditory event locations can only occur if the triggering stimulus patterns also produce different auditory event gestalts.

Conversely, the cocktail party effect is due to the effect of the location association stage. Under certain conditions, simultaneous auditory event gestalts can occur only if the triggering stimulus patterns also act as localisation stimuli. To illustrate, a particular voice cannot be ‘extracted’ from a large group of speakers if one ear is blocked, because the location association stage will not be able to contribute to the selection process. The localisation stimulus selection is a useful pre-selection process, which precedes the higher-level pattern recognition. Each of the two processing stages discriminates the patterns according to different characteristics that are independent of each other. The resulting resolution of the different patterns then ensures that the auditory system can discriminate individual auditory events. On these grounds, concentrating on one auditory event could lead to an increased resolution of a particular pattern as a result of guided association (association guidance by consciousness).

According to this delineation, the localisation process in the superimposed sound field is a two-stage process that performs the mapping of the auditory events because of auditory experience. A fundamental example of auditory experience is that the received ear signals can be attributed to a certain source location and to a certain source signal. These two source properties are independent of each other, but they only occur in a pairwise fashion. The association model takes this auditory experience into account: the occurring auditory events can be traced back to the effect of a location and a gestalt association stage. The two stages work independently of each other; they always determine the properties of the auditory events in a conjoint fashion.

The capabilities of our hearing system are based on two elementary areas of auditory experience. The conjoint effect of these areas with regard to spatial hearing in the superimposed sound field is especially well illustrated by the cocktail party effect.

5. A consequence of the association model

The categorical distinction between a location- and a gestalt-determining processing stage in the auditory system enables an enhanced understanding of auditory phenomena occurring in the superimposed sound field. Particularly those phenomena that depend on the source signals' relationship (phantom source, law of the first wavefront, cocktail party effect) can be explained by the association model as being the result of this two-dimensional processing. The location association stage discriminates individual source signals because of the available spatial information (see Section 3.1 and Chapter 4). The stimulus responses of this stage are then subjected to the gestalt-determining processing carried out in the next stage. It is not until here that the relationship of the source signals takes effect. Provided the involved sources give rise to discriminable and simultaneous localisation stimuli, the regularities of the gestalt association stage can be investigated for the case of two fixed source locations by varying the source signal relationships, even if these influence the location of the auditory event.

The phantom source situation in particular showed that the location of an auditory event (direction, distance and elevation) cannot be attributed to a processing mechanism that amounts to the localisation of a corresponding substitute sound source. The association model is in agreement with this fact, since the location-determining processing mechanism does not necessarily determine the auditory event location. Rather, in the model the location associations only lead to auditory events if they are evaluated in conjunction with associated gestalt associations. Without such auditory event gestalts there will be no auditory event location.

Hence, different processing mechanisms of the auditory system can, in principle, lead to identical auditory event locations. This depends on the ear signals and always happens if different location and gestalt associations lead to the same localisation. For instance, if driven with particular signals, two loudspeakers arranged in the conventional manner can produce exactly the same ear signals as a single real sound source (e.g. TRADIS method, DAMASKE / MELLERT 1969/79). The auditory event location arises due to *one* location association, the fictitious sound source being a **substitute sound source**. However, two different loudspeaker signals can also produce a sound field that, in accordance with the actual sound source locations, gives rise to ***two*** location associations whilst still leading to the same auditory event location as in the first case. In this case, the fictitious sound source is a phantom source.

This statement by the association model does not correspond to the literature's understanding of the functioning of the auditory system with respect to spatial hearing. Up to now, it was assumed that the extremely complex localisation process should be unambiguously describable if at least the spatial properties of the auditory event have been clearly determined. This view is incorrect, as different ear signals can produce the same spatial properties as soon as more than one sound source influences the formation of the auditory event. It was further assumed that the localisation process could be attributed to the effect of a location-determining processing stage in the auditory system. This view is not compelling and seems unlikely, since in this case the location-determining processing stage would interpret certain ear signals differently.

In contrast, the association model specifies the following relationships with regard to spatial hearing:

1. Under certain conditions, a single sound source causes one location association in the location-determining stage, which completely determines the auditory event location on its own.

2. Under certain conditions, two sound sources that are different and independent of each other in terms of their locations and signals cause two location associations in the location-determining and two gestalt associations in the gestalt-determining stage. Similar to the sound events, the auditory events are "de-coupled" with respect to their locations and gestalts, so that the localisation is only due to the effect of the location association stage.

3. Under certain conditions, two sound sources that only differ in terms of their locations and that are coupled in terms of their signals still cause two location associations in the location-determining stage. However, in the gestalt association stage no de-coupling of the coupled source signals can take place. Since an auditory event is always determined by both processing stages, coupled auditory events arise as a result of their common gestalt properties. Identical source signals therefore lead to a complete fusing of the auditory events and thus also to a fusion of the auditory event locations. In this case, the localisation can be traced back to the effects of the location as well as the gestalt association stage.

Generally speaking, the localisation process can only be explained by means of the function of a location-determining processing stage if just *one* localisation stimulus contributes to the formation of an auditory event. A localisation stimulus will be available if sufficiently broadband binaural ear signals are coupled to each other due to the effects of the head and pinnae in the sound field only.

5.1 Lateralisation – The loudspeaker distance zero

Binaural ear signals or ear signal components that do not constitute a localisation stimulus but are coupled to each other in an arbitrary way are not subject to the psychoacoustic regularities of a location-determining processing stage. This fact is especially evident from lateralisation experiments. Indeed, it can be illustrated even more clearly if the effects of the two processing stages in the association model are scrutinised.

The location association stage precedes the gestalt association stage. This means that the auditory system is equipped with the ability to select (and discriminate) localisation stimuli when presented with multiple different stimuli. In the superimposed sound field, the localisation stimulus selection acts as a filter that enables the discrimination of individual source signals (see Chapter 4), so that a possible coupling of the source signals does not take place until the gestalt association stage. In a phantom source situation, for example, the location association stage prevents summing localisation. The localisation stimulus selection ensures that – in spite of the superposition of the sound fields – the same signals are processed in the gestalt association stage as in the case of headphone reproduction.

Lateralisation experiments can only provide information about the function of the gestalt association stage, because, independent of the source distances, the two source signals are discriminated and forwarded to the gestalt association stage separately. Hence, as a basic principle, lateralisation experiments do not allow for any conclusions to be drawn about the functioning of the location association stage. They can only illustrate psychoacoustic regularities of the phantom source ("phantom source inside the head"). A "substitute sound source inside the head" does not exist.

The relevance of lateralisation experiments to spatial hearing is hitherto unknown. It is assumed that the localisation process can be split into sub-components, which can be scrutinised individually using headphones. In this context, the following statement by BLAUERT 1974 (pp. 131) shall serve as a representative example of the currently accepted understanding evident in the literature:

> "The formation of lateral auditory events requires different ear signals. Specific types of ear signal differences are created due to diffraction, shadowing and resonance effects occurring at the head and pinnae. These very subtle signal differences can be studied in terms of their "effective" components in lateralisation experiments using headphones."

With regard to summing localisation, it is stated (pp. 169):

> "It stands to reason to study the influence of individual signal components onto the direction of an auditory event in more detail. This is feasible by presenting synthetically created impulse groups over headphones, for example. The individual signal components can then be varied at will."

It is presumed that the "evaluation of different ear signals" (see Section 2.4 in BLAUERT 1974), which the auditory system carries out as part of the localisation process, can be investigated by means of two sound sources that are in sufficient proximity of the ears. Admittedly, the resultant inside- or near-the-head locatedness is considered formally by describing the lateralisation experiment as "an experiment with deficient distance perception". Nevertheless, there is no evidence for the assumption prevailing in the literature that the results from these lateralisation experiments and hence the associated hypotheses concerning the formation of lateral auditory events "can also be applied to spatial hearing in a free sound field" (BLAUERT 1974).

The association model explains the "deficient distance perception" in the case of lateralisation experiments with the help of the "loudspeaker distance zero"[1]. The sound source locations "at the ears" also produce localisation stimuli, which in turn cause two corresponding location associations (see Section 4.2.1). Due to the missing sound field superposition this happens particularly easily (the ear signal in the case of monotic presentation is preserved in the case of dichotic presentation).

Since the signals are only fused when entering the gestalt association stage, the lateral displacements of auditory events resulting from level and time differences between the headphone signals must not be compared with direction hearing relevant to the localisation of a single real sound source.

In general, the evaluation of different ear signals, which the auditory system carries out when localising a single sound source, can only be investigated under localisation conditions.

5.2 Time-intensity equivalence in the case of artificial head signals

The localisation of a single sound source takes place by means of an evaluation of interaural signal differences. According to BLAUERT 1974, one can differentiate between two types of interaural differences: interaural time and level differences. As to the relative significance of these two types of signal characteristics and their intermodal effect, the association model postulates that only localisation experiments can provide relevant information with respect to this issue.

---

[1] **Definition**: A sound or auditory event with a distance of zero is located at the head. In contrast to LAWS 1972 and others who measured the distance from the centre of the head (origin of the co-ordinate system), the association model stipulates that a distance of zero is only possible at the receiver's boundaries. A sound event inside the head does not exist. Hence, an auditory event inside the head cannot arise due to a localisation process; it contradicts the auditory experience. An "auditory event without distance" is characteristic of inside-the-head locatedness..

The lateralisation experiments that have commonly been carried out to measure this time-level equivalence and that so far have utilised broadband signals and sound pressure levels in-between 40 and 80 dB, have led to an average equivalence factor of about 50 µs/dB ± 25 µs/dB (mean and standard deviation according to results published by V.BEKESY 1959, DEATHERAGE / HIRSCH 1959, DAVID / GUTTMAN / V. BERGEIJK 1959, HARRIS 1960, FRANSSEN 1960, HAFTER / JEFFRESS 1968). When comparing this value with those for level- and time-based stereophony, then based on measurements made by DE BOER 1940 (approx. 30 µs/dB), WENDT 1963 (on average approx. 60 µs/dB) and this author (approx. 45 µs/dB) one arrives at a mean value of 45 µs/dB. The two equivalence factors have a similar magnitude and their range is high. As already highlighted by WENDT 1963, this similarity also hints at the possibility that for both modes of reproduction the source signal relationships are perceptually salient because of the same processing mechanisms.

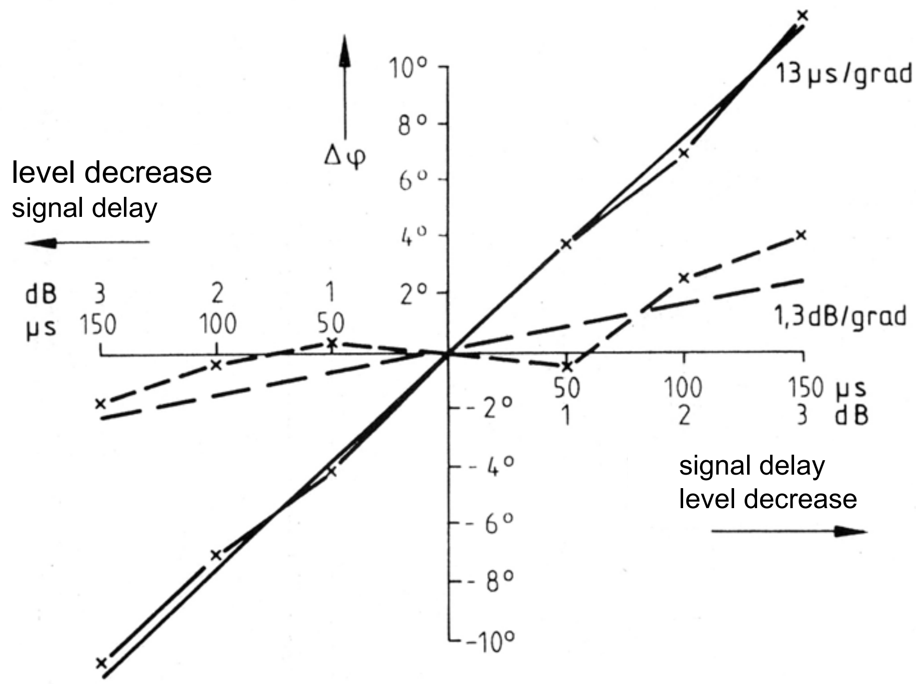
Yet, similar equivalence values obtained from both loudspeaker- and headphone-based experiments so far do not confirm the assumption that the localisation process is liable to other psychoacoustic principles than those revealed by lateralisation experiments. If these equivalence values are meant to shed some light on this issue, they will have to be compared with the equivalence values that actually apply to the localisation of a real or substitute sound source. However, no such information was found when searching the literature. It seems that hitherto equivalence values have not been determined in localisation experiments. Presumably, apart from experimental difficulties, the reason for this is that the relevance of lateralisation experiments to the localisation process has never been questioned.

In view of these findings, a listening test was conducted in order to determine the equivalence value by means of artificial head signals. More specifically, the artificial head signals were presented over headphones and the shift in auditory event direction as caused by delaying or attenuating the level of one ear signal was measured. These measurements were made for two artificial head signals:

a) A speaker in an anechoic chamber at a distance of 2 m from the artificial head and at an angle of incidence of $\Omega = 135°$

b) A speaker in an anechoic chamber at a distance of 2 m from the artificial head and at an angle of incidence of $\Omega = 180°$

In both cases, the speaker was positioned behind the dummy head (NEUMANN), resulting in better localisation compared to the frontal imaging area. Nonetheless, the results can also be applied to the situations $\Omega = 45°$ and $\Omega = 0°$. In the case of a), the additionally included ear signal differences were 0, 50, 100, 150 µs and 0, 1, 2, 3 dB, both at the contra- and ipsilateral ear. In the case of b), they occurred at one ear only. The test signals were presented in a random fashion. A total of 20 subjects participated in the experiment. For each test signal, four subjects were asked to judge the direction of the auditory event.



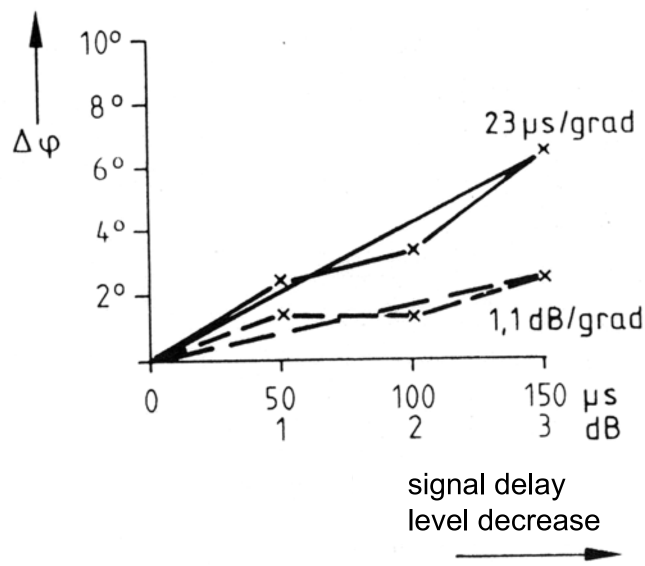


Bild 22:

Shift in auditory event direction for the case of artificial head signals and the delay or attenuation of one ear signal

Above: Ω = 45° (135°), case a)

Below: Ω = 0° (180°), case b)

Figure 22 displays the arithmetic means of the shifts $\Delta\varphi$ in the auditory event directions. They have been plotted as a function of the time delay or level attenuation respectively, which were inserted into one ear signal. The standard deviations lie in-between 12° and 16° in the case of a) and in-between 2° and 5° in the case of b).

The straight lines that have been included depict the average gradients of the curves. Based on these results, the following equivalence factors are obtained:

10 µs/dB in the case of a)

21 µs/dB in the case of b).

Admittedly, this investigation is problematic in that the insertion of an additional ear signal difference may largely destroy an original localisation stimulus. For this reason, the differences were chosen to be maximally 150 µs and 3 dB, respectively. Despite these restrictions, one cannot be completely sure if and to what degree the localisation experiment turned into a lateralisation experiment. For the purpose of this investigation, the employed artificial head system cannot ensure a sufficiently high degree of accuracy of the reproduction of a speaker located in an anechoic chamber. The inter-subject distance judgements vary too much and, on average, are too close to the head. However, there is still no meaningful and verifiable demarcation of localisation and lateralisation. To make such a demarcation possible, the nature of these two processes needs to be specified in detail first.

Nevertheless, it seems that based on the results of this experiment the following conclusion can be drawn. The relative significance of the interaural time and level differences to the localisation of a sound source differs from the one applicable to lateralisation. Direction hearing occurs primarily due to an evaluation of the temporal characteristics (in this case time differences). The lateral displacements of the auditory events in the case of lateralisation are subject to different psychoacoustic regularities – at least from a quantitative point of view.

For b) (sound source located in the median plane), however, the obtained equivalence factor is twice as large as for a) (sound source at $\Omega = 135°$). This agrees with the fact that for artificial head signals the localisation in the median plane is more difficult and unstable than the one of a source signal having a lateral angle of incidence. The larger equivalence factor for b) can be explained by means of a more severe disturbance of the localisation stimulus as caused by the additionally inserted ear signal differences.

A further increase in the inserted ear signal differences would therefore mean that the regularities pertinent to lateralisation would become more and more relevant, i.e. that the equivalence factor would get bigger. However, the results of some informal experiments cannot confirm this unreservedly. In particular, further attenuation of the level of one ear signal caused the auditory event to split into two parts in many cases.

It seems that the first part was determined by the temporal information, i.e. whilst the location of this auditory event could hardly be influenced by level changes, time delay variations had a large effect. The second part was located at the ear that was presented with the signal having an unnaturally high intensity. This may indicate an explanation for the occurrence of a “time image” and an “intensity image” evident from many lateralisation trading experiments (see WHITWORTH / JEFFRESS 1961, HAFTER / JEFFRESS 1968). It is worth noting that the results published by these authors lead to an average equivalence factor of 10 µs/dB for the “time image” and 50 µs/dB for the “intensity image”.

6. Summary

According to the association model presented in the preceding chapters, the functioning of the auditory system with respect to spatial hearing is due to two different processing mechanisms. Each of these two processing mechanisms manifests itself in the form of an associatively guided pattern selection. A current stimulus stemming from a sufficiently broadband sound source gives rise to a location association in the first and to a gestalt association in the second, higher-level processing stage because of auditory experience. Although the two stages work independently of each other, they always determine the properties of one or multiple simultaneous auditory events in a conjoint manner.

The rigorous differentiation of these two stimulus evaluation stages corresponds entirely to the two elementary areas of auditory experience. The received ear signals can be attributed to the two sound source characteristics of “location” and “signal”, which are independent of each other but always occur in a pairwise fashion. Therefore, the presented association model is in agreement with many phenomena related to localisation in the superimposed sound field. It thus offers new approaches for the explanation of some important auditory phenomena. These can be summarised as follows:

1) In the case of stereophony, auditory events occur in the superimposed sound field that could equally be produced by means of certain single sound sources located in a free sound field. Even though the auditory events will be identical in these two cases, their respective ear signals can never be identical. There will be relatively large differences in the ear signals’ characteristics, both in terms of frequency spectrum and interaural degree of coherence. This leads to the conclusion that summing localisation does not take place. The “phantom source” cannot be interpreted as a substitute sound source. Rather, it has to be assumed that due to different source locations the auditory system can discriminate the source signals (effect of the location association stage). After the spatial decoding the stimuli are fused, because the loudspeakers radiate sufficiently similar signals (effect of the gestalt association stage).

2) The areas of validity of the “law of the first wavefront” and of “summing localisation” are defined by different time delay regions. Both phenomena can be traced back to the time-dependent evaluation of stimulus responses of the location association stage, which arrive at the gestalt association stage one after the other. In the superimposed sound field, the location association stage acts as a filter enabling the discrimination of the source signals, so that the source signal relationships are only evaluated in the following gestalt association stage. Two source signals exhibiting different time delays result in two non-simultaneous localisation stimuli. All of the resultant regularities applicable to the mapped auditory event locations are summarised by the “law of the first localisation stimulus”.

3) The “cocktail party effect” implies that a target signal arriving from a certain direction will be masked less by an interfering signal arriving from a different direction when listening binaurally rather than monaurally. This phenomenon can be traced back to the effect of the location association stage. Normally, two sound sources do not just give rise to two different location associations, but also to two different gestalt associations. The two resultant auditory events therefore occur after a two-stage selection from which the largest possible resolution derives. When listening monaurally, the selection effect of the location association stage is lost at least partly, because the available stimulus patterns will be incomplete. The conjoint effect of the two processing stages, which are determined by the elementary areas of auditory experience, is especially well illustrated by the cocktail party effect.

4) Lateralisation experiments provide information about the evaluation of interaural signal differences. However, they can only provide information about the function of the gestalt association stage, because, independent of the source distances, the two source signals are discriminated and forwarded to the gestalt association stage separately. Hence, as a basic principle, lateralisation experiments do not allow for any conclusions to be drawn about the functioning of the auditory system when localising a single sound source. Rather, they illustrate psychoacoustic regularities of a “phantom source inside the head” (loudspeaker distance zero). Assuming an adapted hearing system, a “substitute sound source inside the head” does not exist. In general, the evaluation of different ear signals, which the auditory system carries out as part of the localisation of a sound source, cannot be investigated by means of two sound sources that are in sufficient proximity of the ears. Headphone-based listening tests are listening tests with two sound sources, except when artificial head signals are being presented (in which case a substitute sound source does exist).

5) The functioning of the hearing system with respect to the localisation of a sound source can only be investigated under localisation conditions. A prerequisite for this is that the sound event exhibits a sufficiently broadband frequency spectrum. The perceptual process leading to the localisation can only take place if the spectral characteristics allow for a mapping of the auditory event distance.

The association model traces localisation to a process for the selection of a localisation stimulus. A localisation stimulus will exist if sufficiently broadband ear signals can be mapped to a single sound event location in terms of their temporal and spectral properties. Under certain conditions, at least two localisation stimuli can be discriminated simultaneously in the superimposed sound field. Two discriminable localisation stimuli will lead to a single auditory event location, both in a phantom source situation and in a lateralisation experiment.

Bibliography

ALLEN, J.B., BERKLEY, D.A., BLAUERT, J. (1977): Multimicrophone signal-processing technique to remove room reverberation from speech signals. J. Acoust. Soc. Amer. 62, 912 - 915.

BABKOFF, H., SUTTON, S. (1966): End point of lateralization of dichotic clicks. J. Acoust. Soc. Amer. 39, 87 - 102.

von BEKESY, G. (1930): Zur Theorie des Hörens : über das Richtungshören bei einer Zeitdifferenz oder Lautstärkeungleichheit der beidseitigen Schalleinwirkungen. Phys. Z. 31, 824 -838 und 857 - 868.

BLAUERT, J. (1969): Untersuchungen zum Richtungshören in der Medianeben bei fixiertem Kopf. Dissertation Techn. Hochschule Aachen.

BLAUERT, J. (1970): Ein Versuch zum Richtungshören bei gleichzeitiger optischer Information. Acustica 23, 118 - 119.

BLAUERT, J. (1970): Zur Trägheit des Richtungshörens bei Laufzeit- und Intensitätsstereofonie. Acustica 23, 287 - 293.

BLAUERT, J., LAWS, P. (1973): Verfahren zur orts- und klanggetreuen Simulation von Lautsprecherbeschallungen mit Hilfe von Kopfhörern. Acustica 29, 273 -277.

BLAUERT, J. (1974): Räumliches Hören. Hirzel-Verlag Stuttgart.

Blauert, J., COBBEN, W. (1978): Some consideration of binaural cross correlation analysis. Acustica 39, 96 - 103.

BLAUERT, J., MELLERT, V., PLATTE, H.-J., LAWS, P., POULSEN, T., GOTTLOB, D., PLENGE, G. (1978): Wissenschaftliche Grundlagen der kopfbezogenen Stereofonie. Vorkolloquium zur DAGA, Bochum.

Blauert, J. (1978): Neuere Ergebnisse zum räumlichen Hören. Vortrag vor der Rheinisch-Westfälischen Akademie der Wissenschaften (im Druck).

BLODGETT, H. C., WILBANKS, W. A., JEFFRESS, L. A. (1956): Effect of large interaural time differences upon the judgement of sidedness. J. Acoust. Soc. Amer. 28, 639 - 643.

BLOOM, P.J. (1977): Creating source elevation illusions by spectral manipulation. J. Audio Eng. Soc. 25, 560 - 565.

de BOER, K. (1940): Plastische Klangwiedergabe. Philips techn. Rdsch. 5, 107 -115.

BOERGER, G. (1965): Die Lokalisation von Gaußtönen. Dissertation Techn. Universität Berlin.

BOHN, G. (1973): A structure for associative inforrnation processing. Biol. Cybernetics 29, 193 -200.

BOWER, J. S. (1975): The subjective effects of interchannel phase-shifts on the stereophonic image localization of wideband audio signals. BBC Research Department Report BBC RD 27.

CASPERS, H. (1973): Zentralnervensystem. In : Kurzgefaßtes Lehrbuch der Physiologie, Hrsg. W. D. KEIDEL. Georg-Thieme-Verlag Stuttgart, 467 - 524.

CHERRY, E.C. (1953): Some experiments on the recognition of speech with one and with two ears. J. Acoust. Soc. Amer. 25, 975 - 979.

CHERRY, E.C., TAYLOR, W.K. (1954): Some further experiments upon recognition of speech, with one and with two ears. J. Acoust. Soc. Amer. 26, 554 - 559.

CREMER, L. (1948): Die wissenschaftlichen Grundlagen der Raumakustik, Bd. 1. Hirzel-Verlag Stuttgart.

DAMASKE, P. (1967/68): Subjektive Untersuchung von Schallfeldern. Acustica 19, 199 - 213.

DAMASKE, P., MELLERT, V. (1969/70): Ein Verfahren zur richtungstreuen Schallabbildung des oberen Halbraumes über zwei Lautsprecher. Acustica 22, 154 - 162.

DAMASKE, P. (1969/70): Richtungsabhängigkeit von Spektrum und Korrelationsfunktionen der an den Ohren empfangenen Signale. Acustica 22, 191 - 204.

DANILENKO, L. (1967): Binaurales Hören im nichtstationären, diffusen Schallfeld. Dissertation Techn. Hochschule Aachen.

DANNENBRING, G.L., BREGMAN, A.S. (1978): Streaming vs. fusion of sinus oidal components of complex tones. Perception & Psychophysics 24, 369 - 376.

DAVID, E.E., GUTTMAN, N., van BERGEJK, W.A. (1959): Binaural interaction of high-frequency complex stimuli. J. Acoust. Soc. Amer. 31, 774 - 782.

DEATHERAGE, B.H., HIRSCH, 1.J. (1959): Auditory localization of clicks. Journal Acoust. Soc. Amer. 31, 486 - 492.

DEATHERAGE, B. H. (1966): Examination of binaural interaction. J. Acoust. Soc. Amer. 39, 232 -249.

DOMNITZ, R.H., COLBURN, H.S. (1976): Anaalysis of binaural detection models for dependence on interaural target parameter. J. Acoust. Soc. Amer. 59, 598 - 601

DOMNITZ, R.H., COLBURN, H.S. (1977): Lateral position and interaural discrimination. J. Acoust. Soc. Amer. 61, 1586 - 1598.

DUIFHUIS, H. (1972): Perceptual analysis of sound. Dissertation Techn. Hochschule Eindhoven.

DURLACH, N.I. (1963): Equalization and cancellation theory of binaural masking level differences. J. Acoust. Soc. Amer. 35, 1206 - 1218.

DURLACH, N.I. (1972): Binaural signal detection Equalization and cancellation theory. In : TOBIAS, J. V. (Hrsg.) : Foundation of modern auditory theory. Academic Press, New York, Bd. 2, 369 - 462.

ENGEL, W. (1966): Eine neue Methode zur Entkopplung mehrfach geregelter Systeme. Dissertation Techn. Hochschule München.

FLECHTNER, H.J. (1972): Grundbegriffe der Kybernetik. Hirzel-Verlag Stuttgart.

FRANK, H. (1970): Kybernetik, Brücke zwischen den Wissenschaften. Umschau-Verlag Frankfurt.

FRANSSEN, N.V. (1960): Some considerations of the mechanism of direction hearing. Dissertation Techn. Hochschule Delft

FRANSSEN, N. V. (1963): Stereophonie. Philips techn. Bibl., Eindhoven.

FUKUSHIMA, K. (1973): A model of associative memory in the brain. Kybernetik 12, 58 - 63.

FUKUSHIMA, K., MIYAKE, S. (1978) : A self-organizing network with a function of associative memory : Feedback-type cognitron. Biol. Cybernetics 28, 201 208.

GARDNER, M.B. (1969): Image fusion, broadening and displacement in sound localization. J. Acoust. Soc. Amer. 46, 339 - 349.

GARDNER, M. (1973): Some single- and multiple-source localization effects. Journal Audio Eng. Soc. 21, 430 - 437.

GRUBER, J. (1967): Hörversuche mit moduliertem Rauschen unterschiedlicher interauraler Korrelation. Dissertation Technische Universität Berlin.

HAFTER, E.R., JEFFRESS, L.A. (1968): Two image lateralizations of tones and clicks. J. acoust. Soc. Amer. 44, 563- 569.

HAFTER, E.R. (1977): Lateralization model and the role of time-intensity tradings in binaural masking : Can the data be explained by a time-only hypothesis? J. Acoust. Soc. Amer. 62, 633 - 635.

HAFTER, E.R., DYE, R.H., GILKEY, R.H. (1979): Lateralization of tonal signals which have neither onsets nor offsets. J. acoust. Soc. Amer. 65, 471 -477.

HALL, J.L. (1964): Minimum detectable change in interaural time or intensity difference for brief impulsive stimuli. J. Acoust. Soc. Amer. 36, 2411 - 2413.

HARRIS, G. G. (1960): Binaural interaction of impulse stimuli and pure tones. J. acoust. Soc. Amer. 32, 685 - 692.

HAWKINS, D. B., YOUNG, L. L., PARKER, C. A. (1978) : An investigation of the interaural time difference threshold for speech. Perception & Physophysics 24, 168 - 170.

JEFFRESS, L.A., TAYLOR, R.W. (1961): Lateralisation versus localization. Journal Acoust. Soc. Amer. 33, 482 - 483.

JEFFRESS, L.A., BLODGETT, H.C., DEATHERAGE, B.H. (1962): Effect of interaural correlation on the precision of centering a noise. J. Acoust. Soc. Amer. 34, 1122 - 1123.

KEIDEL, W.D. (1973) : Informationsverarbeitung. In: Kurzgefaßtes Lehrbuch der Physiologie, Hrsg. W. D. KEIDEL. Georg-Thieme-Verlag Stuttgart.

KLEMM, O. (1909): Lokalisation von Sinneseindrücken bei disparaten Nebenreizen. Psychol. Stud. 5, 73 - 162.

KLEMM, O. (1920): Untersuchungen über die Lokalisation von Schallreizen IV : Über den Einfluß des binauralen Zeitunterschiedes auf die Lokalisation. Arch. ges. Psychol. 40, 117 - 145.

KOENIG, A.H., ALLEN, J.B., BERKLEY, D.A. (1977): Determination of masking-level differences in a reverberant environment. J. Acoust. Soc. Amer. 61, 1374 - 1376.

KOHONEN, T., REUKALA, E., MAEKISARA, K., VAINIO, L. (1976): Associative recall of images. Biol. Cybernetics 22, 159 - 168.

KOHONEN, T., OJA, E. (1976): Fast adaptive formation of orthogonalizing filters and associative memory in recurrent network of neuron-like elements. Biol. Cybernetics 21, 85 - 95.

KOHONEN, T. (1977): Associative memory. Berlin - Heidelberg - New York Springer-Verlag.

LAWS, P. (1972): Zum Problem des Entfernungshörens und der Im-Kopf-Lokalisiertheit von Hörereignissen. Dissertation Technische Hochschule Aachen.

LAWS, P., BLAUERT, J., PLATTE, H-J. (1976/1977): Anmerkungen zur stereofonen kopfbezogenen Übertragungstechnik. Acustica 36, 45 - 47.

LEAKEY, D.M. (1959): Some measurement on the effects of interchannel intensity and time differences in two channel sound systems. J. Acoust. Soc. Amer. 31, 977 - 986.

LEHRINGER, F. (1979) : Ober die Beeinflussung der Hörereignisrichtung durch das Sehereignis bei der Personendarstellung in Telekonferenzen. Acustica 42, 240 - 248.

LICKLIDER, J.C.R. (1951): A duplex theory of pitch perception. Experientia 7, 128 - 134.

LOY, P. (1978): Phantomschallquellenbildung bei zwei in der Phase gedrehten Signalen. Diplomarbeit Techn. Universität Berlin.

MAKITA, Y. (1962): On the directional localization of sound in the stereophonic sound field. Europ. Broadcasting Union Rev. Part A, 73, 102 - 108.

MARKO, H. (1971): Ein Funktionsmodell für die Aufnahme, Speicherung und Erzeugung von Informationen im Nervensystem. Röntgen-Blätter 24.

MARKO, H. (1974): A biological approach to pattern recognition. IEEE Transactions on systems, man and cybernetics, vol. SMC - 4, 34 - 39.

MARKO, H. (1978): Modelle der visuellen Wahrnehmung. NTG-Fachberichte 65, 19 - 29.

MASSARO, D., WARNER, D. (1977): Dividing attention between auditory and visual perception. Perception & Psychophysics 21, 569 - 573.

MATSUDAIRA, T. K., FUKAMI, T. (1973): Phace difference and sound image localization. J. Audio Eng. Soc. 21, 792 - 797.

MERTENS, H. (1965): Directional hearing in stereophony theory and experimental verification. Europ. Broadcasting Union Rev. Part A, 92, 1 - 14.

MILLS, A. W. (1972): "Auditory localization", Chapter 8 of : Foundations of modern auditory theory (ed. J.V. TOBIAS) vol. II, Academic Press, New York.

MURAKAMI, K., AKAISHI, S., AIBARA, T. (1978): On optimal associative recall by an incomplete key. Biol. Cybernetics 30, 95 - 97.

NAKAMURA, S., KOMORI, S., KOYASU, M. (1976): On the importance of the individual shape of pinnae for the construction of a recording dummy head. J. Acoust. Soc. Japan 32, 618 - 620.

ORTMEYER, W. (1966): Über die Lokalisierung von Schallquellen bei Zweikanal-stereophonie. Hochfrequenztechn. u. Elektroakustik 75, 77 - 87.

ORTMEYER, W. (1966): Schallfelduntersuchungen bei Zweikanalstereophonie. Hochfrequenztechn. u. Elektroakustik 75, 137 - 145.

OSMAN, E. (1971): A correlation model of binaural masking level differences. J. Acoust. Soc. Amer. 50, 1494 - 1511.

PALEF, S. R., NICKERSON, R.B. (1978): Representing auditory space. Perception & Psychophysics 23, 445 - 450.

PLENGE, G. (1972): über das Problem der Im-Kopf-Lokalisation. Acustica 26, 241 - 252.

PLENGE, G. (1973): über das Problem der intracranialen Ortung von Schallquellen bei der akustischen Wahrnehmung des Menschen. Habilitationsschrift TU-Berlin.

PLENGE, G. (1974): On the differences between localization and lateralization. J. Acoust. Soc. Amer. 56, 944 - 951.

PLENGE, G., TILSE, U. (1975): Cocktail-Party-Effekt with and without conflicting visual cues. 50th AES-Convention, London.

POGIO, T. (1975): On optimal nonlinear associative recall. Biol. Cybernetics 19, 201 - 209.

POLLACK, 1. (1977): Continutation of auditory frequency gradients across temporal breaks : The auditory Poggendorff. Perception & Psychophysics 21, 563 - 568.

RATLIFF, P. A. (1974): Properties of hearing related to quadrophonic reproduction. BBC Research Department Report BBC RD 38.

REICHARDT, W., HAUSTEIN, B.-G. (1968): Zur Ursache des Effektes der Im-Kopf-Lokalisation. Hochfrequenztechn. u. Elektroakustik 77, 183 - 189.

ROESER, D. (1965): Schallrichtungsbestimmung bei krankhaft verändertem Gehör. Dissertation Techn. Hochschule Aachen.

SAKAMOTO, N., GOTOH, T., SHIMBO, M., KOGURE, T. (1977): Some considerations of an advanced stereophonic reproducting system. 19th Intern. Congr. on Acoustics, Madrid.

SANDEL, T.T., TEAS, D.C., FEDDERSEN, W.E., JEFFRESS, L.A. (1955): Localization of sound from single and paired sources. J. Acoust. Soc. Amer. 27, 842 - 852.

McSAYERS, B.A. (1964): Acoustic-image lateralization judgement with binaural tones. J. Acoust. Soc. Amer. 36, 923 - 926.

SCHENKEL, K.D. (1967): Accumulation theory of binaural-masked thresholds. J. Acoust. Soc. Amer. 41, 20 - 30.

SCHREIBER, L. (1965): Hörtheorien im Spiegel elektrischer Analogien. Frequenz 19, 255 - 261.

SCHROEDER, M. (1975): Models of hearing. Proc. of the IEEE 63, 1332 - 1350.

TERHARDT, E. (1972): Zur Tonhöhenwahrnehmung von Klängen I und II. Acustica 26, 173 - 186 und 187 - 199.

THEILE, G. (1975): Seitliche Phantomschallquellen. Diplomarbeit Technische Universität Berlin.

THEILE, G., PLENGE, G. (1977): Localization of lateral phantom sources. 9. Audio Eng. Soc. 25, 196 - 200.

THEILE, G. (1978): Weshalb ist der Kammfiltereffekt bei Summenlokalisation nicht hörbar? Vortrag Tonmeistertagung Berlin.

TOOLE, F.E., McSAYERS, B.A. (1965): Lateralization judgements and the nature of binaural acoustic images. J. Acoust. Soc. Amer. 37, 319 - 324.

WARNCKE, H. (1941): Die Grundlagen der raumbezüglichen stereophonischen Übertragung im Tonfilm. Akust. Z. 6, 174 - 188.

WENDT, K. (1963): Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitätsund Laufzeitstereophonie. Dissertation Techn. Hochschule Aachen.

WENDT, K. (1964): Das Richtungshören bei Zweikanal-Stereophonie. Rundfunktechn. Mitt. 8, 171 179.

WESS, 0., ROEDER, U. (1977): A holographic model for associativ memory chains. Bio. Cybernetics 27, 89 - 98

WETTSCHUREK, R. (1976): Über die Abhängigkeit raumakustischer Wahrnehmungen von der Lautstärke. Dissertation Techn. Hochschule Berlin.

WHITWORTH, R.H., JEFFRESS, L.A. (1961): Time versus intensity in the lokalization of tones. J. Acoust. Soc. Amer. 33, 925 - 929.

WIGSTROEM, H. (1974): A model of a neural-network with recurrent inhibition Kybernetik 16, 103 -112.

WILLWACHER, G. (1976): Fähigkeiten eines assoziativen Speichersystems im Vergleich zu Gehirnfunktionen. Biol. Cybernetics 24, 181 - 198.

ZWICKER, E., FELDKELLER, R. (1967): Das Ohr als Nachrichtenempfänger. Hirzel-Verlag Stuttgart.