80th AES Convention, Montreux, March 1986

RECORDING CLASSICAL MUSIC: HANDLING OF MAIN MICROPHONE AND SPOT MICROPHONES, A NEW APPROACH TO WELL-TRIED MICROPHONE TECHNICQUE

by Günther Theile
Institut für Rundfunktechnik GmbH (IRT), München, FRG

1. Introduction

Listening in a concert hall is essentially a result of the room impression. The sound engineer has the difficult task of conveying to the listener the acoustic ambience of the musical presentation with the assistance of the stereophonic recording/reproduction systems available to him.

The most common systems use loudspeakers in the standard stereophonic arrangement. They all apply a basic phenomenon of spatial hearing, i.e. the creation of so-called phantom sound source between the loudspeakers. The phantom sound source is being understood as an "fictitious sound source" because auditory events /1/*) occur at positions where no real sound source is to be found. But the "fictitions sound source" between the loudspeakers implies the assumption, which does not hold true: it can not be assumed that the phantom sound source represents a (fictitious) "substitute sound source", which produces the same ear signal characteristics as the two coherently radiating stereo loudspeakers /2/.

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

These theories have been described in various studies /3/, /4/, /5/, /6/, /7/, /8/. A detailed bibliography is presented in /1/. However, their areas of validity are limited to directional hearing (distance hearing has not been taken into consideration), and even very often only to directional hearing in the horizontal plane. The main shortcoming of those studies lies in the fact that the regularities were measured "frequency-dependent", i.e. that they

/2

---

*) Many terms used in this paper are transfered from /1/.

were found in narrow-band signals. The importance of the spectral characteristics for the localization was unknown, or at least it was not taken notice of, so that the validity of the "summing localization theories" was never critically investigated for broad-band signals.

As an example, the following phenomenon should be mentioned, which has been studied just a few years ago: In summing localization, each ear input signal is derived from at least two loudspeaker signals that are displaced in time with respect to each other. Consequently each ear input signal shows comb-filter effects (see **figur 1**). The auditory event, however, has almost the same timbre as if only one of the two loudspeakers were driven; that is, it has the same timbre as if no superposition of differently delayed signals occured.

Figure 2 presents the comb-filter effect in the ear input signals, which can be measured in the case of inter-loudspeaker delay t ("delay stereophony"). The lateral displacement of the phantom sound source (towards the right loudspeakers in **fig. 2**) depends on t. It can be observed that the timbre of the phantom sound source, as well as the distance, elevation and focus of localization do not vary adequately if t is varried in the range of 0 ... 400 us.

As a result a **"spectral objection to summing localization"** was formulated /2/, /10/. It stated that the monaural and interaural spectral attributes of broad band ear input signals, which result from the superimposition of the two sound fields produced by the stereo loudspeakers, are not compatible with the perception of

- timbre
- distance
- elevation
- focus of localization

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

The validity of the summing localization theories is severely limited. Therefore these theories could be helpful neither in understanding the function of the hearing sense for the localization in a superimposed sound

/3

field /2/ nor in applying the phantom sound source phenomenon to stereophonic and reproduction technique /9/. (State and trends of development are described and a detailed bibliography in this field is presented in /9/).

2. How does the auditory perspective originate in stereophony?

The "spectral objection to summing localization" provided the motivation for the development of a complete localization model. The validity of which includes phantom sound source as well as real sound sources. It is known as "association model" and was presented during the 11. Convention of Sound Engineers in Berlin in 1978. It was described in detail in /2/, /10/, /11/ and provides a uniform premise for the uncontradicted explanation of important phenomena of spatial hearing. As it will be shown in this paper, it is helpful in understanding the generation of phantom sound sources and in applying the phenomenon to stereophonic imaging of auditory perspectives.

2.1 The Association Model

The association model is based on the assumption that association phenomena constitute a basic principle of sensory stimulation perception. On that basis, auditive spatial perception fundamentally results from two separate processes. Each of these two processes occurs by means of an associatively guided pattern recognition: a momentary stimulus, derived from a sound source, initially induces a "location association" and secondly a "Gestalt association". The characteristic feature of the localization model resides in the two-stage processing of the stimulus. It contains not only the stage of processing that determines the localization, but also the stage that determines the Gestalt ("Gestalt" is a term used in perception research. Here it means the property of the sound source, e.g. timbre). Both stages must be gone through for the stimulus to achieve a form of perception. Accordingly, both processes jointly determine in every instance the properties of the auditory event.

Fig. 3 shows a block diagram of the model. The spatial transmission system in hearing consists of the outer ear, represented by transfer function M, and of the localizing stage, represented by the transfer function M, and of the localizing stage, represented by the transfer function M^{-1}, inverse to M. The dependence of ear-signals on the source location is indicated as a form of coding of spatial information (M), which enables the spatial information (M^{-1}) to be decoded. The operation can be described as that of an adaptive filter with a transmission function M^{-1} which, as a result of the associative recognition of the model, is in inverse relation to M, the transmission function of the outer ear, which depends on the location of the source. Details have been described in /2/. The four important characteristics for stereophony of the two processing stages are:

SELECTION OF LOCALIZATION STIMULUS
(location determining stage)

If possible, the acoustical stimulus should be construed as localization stimulus in the location-determining stage **(fig. 3).** That means that the stimulus of immediate interest will be compared to stimulus patterns which have been learned and which are associated with certain sound source locations. The localization stimulus has been defined in /2/:

Localization Stimulus

Adequate broad-band acoustical stimulus or parts of acoustical stimulus at the eardrums form together a localization stimulus, if they can be associated with a single sound source location as regards the time and spectral characteristics.

Thus the location-determining stage is able to identify the individual locations of sound sources also in the sound field of several sound sources, with certain limitations, even then if the sound sources are coherent (loudspeaker stereophony). In the sound field of several sound sources (precondition: adequate broad-band spectrum) with certain limitations, even then if the sound sources are coherent (loudspeaker stereophony).

INVERSE FILTERING
(location determining stage):

The selection of localization stimulus through the location determining stage is combined with adaptive filtering, which operates inversely to the instantaneous transfer function M of the outerear. Thus the input signals, which are "linear distorted" by M, are again equalised. Only after this process they enter the higher-set "Gestalt-determining" stage, the

/5

decoded spatial information can be processed independent of the "Gestalt" of the source. Head movements, which alter the powerspectrum of the ear inputsignals, do not cause any change of timbre, due to the adaptive inverse filtering /12 /, /13 /.

STIMULUS FUSION
(Gestalt-determining stage)
If two or more sufficiently similar stimulus responses of the location-determining stage arrive at the Gestalt-determining stage at the same time, they fuse into each other. This also occurs if the assigned spatial informations are different. In this case a common, average spatial information results from the fusion, which determines the location of the resulting auditory event (phantom sound source).

STIMULUS DOMINANCE
(Gestalt-determining stage)
In so far as the in the Gestalt-determining stage fusing stimuli demonstrate different intensity or arrive at different times, those stimuli which are stronger compared to the weaker ones, or leading compared to the lagging ones, are given more importance. The characteristics of this "stimulus dominance" can be measured via "lateralization experiments", e.g. trading experiments (survey in /1/).

2.1.1 The principle "dominating localization stimulus"

These four basic properties of the model specify special understanding of the hearing phenomena which occur in the superimposed sound field. Especially such phenomena which depend on the relationship of the two loudspeaker signals (loudspeaker stereophony), are thereby surprisingly vividly and completely explainable. It will be shown below that this new comprehension is helpful in being able to optimally assess and implement stereophonic recording/reproducing systems.

To begin with: How can the creation of a phantom sound source be explained? The following idea results from the four basic properties of the model (here only the loudspeaker stereophony is taken into consideration and

/6

especially the common reproduction via two loudspeakers in standard stereophonic arrangement):

Reproduction of mono signals:

- Each of the two loudspeakers creates a localization stimulus. As a result of the "selection of localization stimulus" and the "inverse filtering" the information about both of the loudspeaker locations and the information about both of the loudspeaker signals reach the Gestalt-determining stage separately (see **fig. 1**). Due to the coherence of both of the loudspeaker signals they can not be differentiated in the Gestalt-determining stage. Thus they fuse together and this results in a fused auditory event ("stimulus fusion").
- The information about both of the loudspeaker losations merge together because of the "stimulus fusion"; this results in a single location of auditory event between the loudspeakers (location of the phantom sound source).

Reproduction of stereo signals

- If the loudspeaker signals demonstrate intensity differences L (intensity stereophony), then the concurrent localization stimuli exist with corresponding varying intensity. That is why they are not assigned the same importance in the stimulus fusion of the Gestalt-determining stage: This results in a lateral displacement of the phantom sound source location into the direction of that loudspeaker which produces the dominating localization stimulus ("stimulus dominance").
- If the loudspeaker signals demonstrate time differences t (delay stereophony) (or if the listener is not situated at exactly the axis of symmetry), then the localization stimuli of identical intensity do not exist concurrently. Thus the localization to first arrive dominates in the stimulus fusion, the following stimulus appears weaker. This results in a lateral displacement in the direction of the "earlier" loudspeaker ("stimulus dominance").
- If the loudspeaker signals demonstrate differences in intensity as well as in time ("mixed stereophony"), then the L- t-combination causes lateral displacement additively. Combinations in the same direction strengthen the lateral displacement phantom sound source, and weaken it if in opposite directions.

Thus the phantom sound source location in the association model is not determined by the resulting summing signals to the ears but rather simply by the characteristics of both of the localization stimuli caused by the stereo loudspeakers. The sense of hearing does not evaluate the resulting interaural differences in intensity and/or time but rather directly the differences in intensity and/or time of the loudspeaker signals (as a result of the "selection of localization stimulus" and "inverse filtering").

This fundamental distinction of the processing of the ear input signals demonstrates **fig. 4**: In the assoziation model **(fig. 4a)** the location determining stage is able to select the localization stimuli A' and B', whereas in summing localization models **(fig. 4b)** the summation input signal at left and right ear is processed in its entirety. In the contrary of the ear input signals A_L+B_L and B_R+A_R, that depend on the interloudspeaker signal difference (A-B), has influence on properties (e.g. timbre of the auditory event).

The known regularities for the creation of the phantom sound source location in intensity and time stereophony can be understood accordingly as regularities for the creation of a "location middle value". The creation of the middle value of the loudspeaker locations occurs in dependence upon whether the loudspeaker signals demonstrate differences in intensity and/or time.

The facts can be more precisely presented:
In the "Gestalt-determining stage" of the sense of hearing all of the parts of the signal which "belong together" fuse with each other. Parts of the signal which "do not belong together" are differentiated here. The Gestalt-determining-stage differentiates for example tenor from soprano in reproduction over mono loudspeakers. In stereo production that occurs by the functioning of the location-determing stage independent of the location of the participating stereo loudspeaker. Precisely those loudspeaker sound events fuse with each other in loudspeaker stereophony, which each of the loudspeakers would individually produce. All of the parts which can not be differentiated are fused with each other: "tenor to the left loudspeaker" with "tenor to the right loudspeaker" and "soprano to the left loudspeaker" with "soprano to the right loudspeaker". If varying difference in intensity and/or time become effective during this fusion, then the following stereophonic image emerges: tenor and soprano are in different locations between the loudspeakers.

/8

The effect of differences in intensity or time one the location of phantom sound source has been measured numerous times (survey in /1/); the regularities are quite well known and used in practice in many ways. They can be generally understood as "regularities of the dominating localization stimulus"; differences in intensity and time of the loudspeaker signals influence the dominance of the localization stimulus during the fusion. Details thereof are described in /2/, /11/.

2.1.2 Narrow band sound sources

The selection of localization stimulus is possible in the model by processing the cross and auto correlation and comparison of the current correlation pattern with stored patterns /2/. That is why the described principle of the dominating localization stimulus can be evaluated. Thus if both loudspeakers are emitting narrow band signals, the the stimuli in the location-determining stage can not be differentiated: Summing localization occurs. In this case the localization of the phantom sound source is very frequency in particular in time stereophony, somewhat less so in intensity stereophony. The localization curves (the phantom sound source direction as a function of the difference of intensity or time of the loudspeaker signals) for the narrow band signals differ strongly from the localization curves which apply to broad band signals and stereophony /1/, /2/.

Fortunately almost all normal sound sources have a sufficient band width, so that the selection of localization stimuli can occur. The localization of the phantom sound source is then, for example, hardly independent upon the "pitch" of a musik instrument. However, in some cases, if music instruments with a minimal bandwidth are involved, a transition from "selection of localization stimulus" to "summing localization" can occur. The localization of the phantom sound source is then "dependent on the pitch"; stability and localization focus of the location of the phantom sound source are impaired.

2.2 Distance of auditory events, "spatial depth"

Generally the localization of a sound source requires that the perception of direction as well as the perception of distance is adequately assured; that is the case only when it has an adequate band width. The requirement is especially true for the localization of a phantom sound source, because the simultaneous localization stimuli which both loudspeakers produce have to be

/9

differentiated in the location-determining stage. Thus the information about distances acquired by the localization stimuli fuse together - not influenced by the comb filter effect on the ears - to the **averaged distance** of the phantom sound source.

It follows: The distance of the phantom sound source is fundamentally the same as the distance of the loudspeakers. A presentation of "spatial depth" can only be achieved by simulating distances. The sound engineer must deliberately employ certain phenomena of spatial hearing in order to produce, for example, an auditory perspective characterized by perception of "depth". This is directly comparable to the presentation of visual perspectives: there is only actually the distance of the figure or screen **(fig. 5)** - that represents the distance of the loudspeakers - but a perception of depth is simulated, because phenomena of spatial vision are employed. Parallel lines join in the distance, equal sizes and distances shrink.

The spatial depth in the stereophonic auditory pattern can be simulated with three different "stereophonic imaging elements":

1. The direct part of the sound field (the "first wave front") contains information of distance, because the sense of hearing as a result of experience associates the relationship of level/spectrum of the signal with the relationship volume/tone/distance of the sound source (details in /14/, /15/).

2. The relationship direct part/indirect part of the sound field contains important information of distance, which the sense of hearing evaluates based on experience. This information gives the auditory pattern a spatial perspective - similar to the perspective presentation in **figure 4** (details in /14/, /15/).

3. For natural hearing the reverberation leads to ear signals, which show a decreasing correlation with the frequency (caused by the separation of the ears in a diffuse sound field). This characteristic supports greatly the perception of spatial depth and has not bee consistently observed in commonly used intensity stereophony.

/10

A stereophonic auditory pattern with spatial depth thus this results because - especially through the choice of the microphone location - each of the two loudspeaker signals contains the characteristics according to points 1) and 2). The spatial depth simulated by 1) and 2) is also perceived by corresponding mono reproduction. The stereophonic side information only leads to lateral expansion of the auditory pattern, because the differences of intensity and time of the loudspeaker signals cause this according to chapter 2.1.1.

As opposed to this the reverberation correlation of the point 3 is an explicity stereophonic imaging element of simulation of the spatial depth. It disappears completely in mono reproduction and the uncorrelated parts contained in the stereo signal remain ineffective (even alter the reverberation balance unfavorably). In the case of stereo reproduction these parts can not form any phantom sound sources due to the absence of a correlation of these parts. In "correct" stereophonic imaging only the first wave front of the concert hall should lead to localization - the reverberation should lead to inter-loudspeaker-signal correlation similar to the natural interaural correlation in the concert hall. This is important for the simulation of spatial depth. - It can already be stated that a **coincidence microphone,** that does not produce signals with natural interaural correlation in a diffuse sound field, has unfavorable characteristics as a main microphone. The "spatial parts" of the intensity stereo signal can only be evaluated to a limited degree by the sense of hearing for the perception of spatial depth.

That is why one tries to add corresponding uncorrelated parts through the "room microphone" or through artificial reverberation. But that has one disadvantage: The "spatial parts" of the coincidence microphone are "unnecessarily" present; they become superposed by the added uncorrelated parts, however, the effect is not favorable as regards depth perception and clarity.

From this point of view the **polymicrophony,** which does not require a main microphone provides better results. The individual microphones in the near field of the sound sources contain few indirect sound and only produce the first wave front (good clarity), which is important for the localization. But this technique also has a disadvantage in imaging the distance and depth: The unnatural presence and timbre, which result from the small recording

/11

distance, contradict the intended simulation of depth (in point 1). The spatial characteristics of the auditory pattern are mostly unsatisfactory.

1.2. L/ t - 8/14 equivalence

Lateral displacements of the phantom sound source, which are caused dependent upon the level difference L and/or upon the time difference t of the coherent loudspeaker signal parts, was measured by many authors. The localisation curve (L) runs about linear in the range = 0 to = ± 20° for broad band signals, the increase is in this range - depending upon the signal - approximately **2.1 ... 2.5 degrees/dB.** The localisation curve (t) also runs about linear in the range of smaller displacements - the increase lies between **3 ... 5 degrees/100 us,** depending upon the signal.

Regarding the **focus of localization** differences in intensity and time have about the same effect in the range of small displacements (to about = ± 15°). However, there are differences in the range of large displacements: an increasing L leads to an increase, an increasing t (approximately t ± 300 us) leads to a decrease of the focus of localization. Large values for L (approximately L ± 20 dB) lead to the **focus of localization** for real sound sources (loudspeakers under = ± 30°). Large values for t (approximately t > ± 500 us) lead to impairment of the localization. Differences in time in the range approximately t = 2 ms ..... 30 ms lead to the fact that the lagging localization stimulus does not have any effect (precedence-effect (see /1/, "law of the first localization stimulus" /2/).

Small differences in time t then also function like small differences in level L. This L/ t-equivalence factor can be observed in **fig. 6;** 1 dB level difference influences the direction of the phantom sound source the same as 60 us time difference in. The **equivalence factor** is **60 us/dB.** The equivalence is lost for large signal differences.

Does this equivalence apply if the differences of the loudspeaker signals consist of differences in level as well as differences in time? The regularities for the phantom sound source direction with a combination of L and t (mixed stereophony) are not yet known in full detail. **Fig. 7** shows in this regard measurement results (measured in non-reflecting room, 10 test subjects, test signal: Gausian curve modulated white noise, impulse width 50 ms). The result is:

/12

1. The localization curve (L) is shifted by = 10° by a time difference t = 200 us working in the same sense.

2. The standard deviation with mixed stereophonic signals is - at least under these conditions - smaller than with pure intensity or time stereophonic signals.

Similar results have been measured for the signal differences of level and time resulting from a dummy-head microphone. The phantom sound source direction evidently results - within certain limits - from the sum of the individual effects of L and t. The given L/ t-equivalence given in **fig. 6** also applies to L/ t-combinations in the same sense.

It is (L, t) = (L) + (t)

At this stage it must be pointed out, that the reference to combinations in opposite sense is not practically applicable. Corresponding measurements have shown that even small L/ t values ("trading values" / /) in opposite sense sharply reduce the focus of localisation. Furthermore, reference is only made here to measurement results, which show that the L/ t-equivalence according to **Figure 6** does not apply to the signal spectrum above 2 kHz.

The effect of L/ t-equivalence (same sense, mixed stereophony) is not explainable by summing localization, but rather with "dominating localization stimulus". Corresponding investigations as well as practical experience show that, as contains to **directional imaging** and **focus of localization** the mixed stereophony is surely at least as efficient as intensity stereophony and considerably more efficient than time stereophony.

In order to achieve an optimum imaging of ambience and spatial depth as well as a minimum localization blur of phantom sound sources the microphone should produce level differences as well as time differences (same sense). The ratio of both signal difference values should be in the range of the L/ t-equivalence (about 60 us/dB). A microphone with this properties is called **"equivalence microphone"**. An optimum equivalence microphone utilizes both of the possible mechanisms for lateral displacing the phantom sound source - each of them to the same degree of effect.

/13

2.4 Result

The adaptive filtering function M^{-1}, inverse to the transfer function M of the outer ear, causes an important consequence: Because $M.M^{-1} = 1$ (see **fig. 3**) is true also in the case of two coherent radiating stereo loudspeakers, the inter-loudspeaker signal relationship effects directly the stereophonic image between the loudspeakers. It follows that the inter-loudspeaker signal relationship (e.g. difference in intensity and in time correlation) should be similar to the interaural signal relationship in the case of natural hearing, in order to create a natural stereophonic image.

This new knowledge, deduced from the association model and contradictory to summing localization models, enables an assessment of the performance of main microphones and spot microphone techniques, as regards the achievable imaging of depth. In particular consequences for the optimisation of the main/spot microphone systems can be derived, as presented below.

3. The performance of a main microphone

The main/spot microphone systems shall be investigated more closely during the recording situation "orchestra in the concert hall". The question of to what extent the new knowledge can be carried over to other recording situations is specifically left out of consideration here.

As regards the choic and positioning of the main microphone, the following is the initial result:

- The desirable imaging of the spatial depth of the orchestra can only result optimally if
  - a) the microphone placement is carefully chosen in order to reproduce the relationship direct/indirect parts as well as the spectra of the parts "room and distance justified" (finding the optimum distance between microphone and orchestra).
  - b) the microphone capsules are spaced approximately with "ear distance", so that the microphone signals represent the natural degree of correlation in the diffuse sound field of the hall.

- Because of the regularities of T/ t-equivalence it makes sense to have a spacing of the microphones, that is just large enough, so that the lateral instrument groups of the orchestra - which the microphone should

/14

just sufficiently pick up - produce a time difference of $t = 350$ us. Larger time differences are unfavourable because a sufficiently effective level difference must exist in the sense of high focus of localization (see chapter 2.3).

As compared to coincident microphones (XY, MS stereophony), which do not make a satisfactory imaging of the spatial depth possible and as compared to spaced microphones (omnidirectional, A/B-stereophony), which do not assure a satisfactory focus of localization, both can be expected of an equivalence microphone (e.g. ORTF, OSS, dummy head).

In the course of studies of "loudspeaker compatibility of dummy head signals" in various listening comparisons the performance of the dummy head microphone was surprising. The auditory pattern demonstrated an impressive ambience, (especially as regards depth graduation) as compared to the usual intensity stereo, and a high spatial transparency with sharp stereo images.

In order to compare the performance of the various equivalence microphones among themselves, but also to the coincidence and A/B microphones, various recordings were conducted in six concert halls (Philharmonie / Berlin, Old Opera / Frankfurt, Hercules Hall / Munich, Large Broadcasting Hall SWF / Stuttgart, Large Broadcasting Hall and production studios WDR/Köln) and with various positions of the main microphones. The microphone signals (including diverse spot microphones - see section 4) were recorded in parallel on a multi-track machine (24-track-digital). They serve as test material for investigating various questions. First results will be presented during the ..... Convention of Sound Engineers, November 1986.

3.1 Directional imaging

Figure 8 reproduces a spatial constellation for a symphony orchestra, as an example. The directions of sound incidence at the main microphone (approx. 4 m behing the conductor) are shown.

Listening tests (recordings in the Hercules Hall with diverse main microphones, loudspeaker reproduction in "living room") show results, which are summarised in **fig. 9.** They correspond well with the localization values presented in section 2.2. The results presented in **fig. 10** can be better

/15

assessed. With the dummy head (KU81), as well as with the OSS-microphone (see **fig. 9**), a reasonable distribution of image position across the sound stage is obtained.

Figure 11 shows the corresponding directional imaging characteristics for various microphones. Evidently the differences in level and time of a equivalence microphone should be adjusted to that a direction of sound incidence at the microphone = 30° results to the direction of phantom sound source = 20° approximately.

3.2 Interfering reflection from the floor

Listening tests have shown that by careful selection of the height of the main microphone the sound quality can be clearly improved. In many cases a height of approximately 3 m above the stage has proven favorable.

It is shown in **fig. 12** that a relatively small floor surface area in front of the conductor between the stringled instruments (see **fig. 8**) can cause interfering reflections. The delay of the reflections are an average 3 ms. A possible interfering effect could be eliminated with simple means. Corresponding PZM-equivalence microphones are technically feasible, but would have other disadvantages.

3.3 Mono-compatability

The differences in time t of the microphone signals result with mono-reproduction in principle due to the comb filter effect in tone color effects. For this reason the A/B-stereophony is not considered practical. However, **fig. 13** shows that the comb filter effect in the mono signal of a microphone (here ORTF-capsule separation as A/B 17 cm) is reduced due to the difference in level.

Figure 14 shows the mono signal of a dummy head (KU81), sound incidence direction = 30°. In comparison, an A/B microphone, the capsule spacing of which (34 cm) leads to the same directional imaging like L/ t -combination of the dummy head.

It can be seen, that the difference in level of the equivalence microphones reduces the comb filter effect bus does not eliminate it. But the comb filter

/16

effect is of course not contained in the uncorrelated part of the microphone signal (originating from the diffuse sound field). Listening tests have shown, that this part masks the audibility of the comb filter effect in the mono signal of an equivalence microphone; sound quality impairment can not be detected.

3.4 Result

The performance of a main microphone can be improved if

- the distance and height of the microphone are carefully determined through listening comparisons
- neither a coincidence microphone nor a A/B microphone (capsule spacing more than 0,5 m), but rather an equivalence microphone is used (mixed stereophony).

In some cases - if the orchestra is itself properly balanced and if the acoustics of the concert hall is suitable - this main microphone is sufficient in order to create a balanced, natural auditory sensation. This is especially true for imaging of the space (directions, distances, extensions).

4. Possibilities of the spot microphone technique

In many cases kinds of imperfections of the orchestra or of the acoustics of the hall will surely not be able to be compensated through an appropriate selection of the location of the main microphone. In these cases spot microphones are unavoidable.

It is emphasized, that here the objective thereof shall be to produce as natural and balance music recordings as possible. For the benefit of this objective it is deliberately resisted here the opportunity of producing whatever "sound" is desired in as short a time as possible and as independent as possible from the orchestra and the hall. The polymicrophony, with or without a main microphone, aritificial sound, equalisation etc. is surely predestined for this.

Spot microphones should be viewed here in this regard as (often necessary) a compromise which must be accepted if the auditory pattern of the main microphone is perhaps natural, but by no means balanced: Up until now there has been no support technique which has adequately taken the perception of

/17

distance into consideration so that whatever desired distances of auditory events could be simulated with the assistance of phantom sound sources.

In the following a method will be proposed, which will reduce fundamental disadvantage of the spot signals. Through a special delay of the spot signal the aim will be to be able to largely retain the spatial depth given by the main microphone if the spot signal is admixed.

4.1 Effect of undelayed spot signals

In section 2.2 it has been shown that the stereo signal must contain certain characteristics for the simulation of the spatial depth. According to the association model these should be as similar as possible to those characteristics of the ear input signals which are contained in natural listening. Direct sound, first reflections and reverberation of the recording room at the location of the main microphone are of similar importance for loudspeaker stereophonic imaging as for natural listening. As it is already known the first reflections also comprise an important imaging element if the simulation of natural ambience is desired. However, the first reflections are mostly only used in connection with artificial reverberation.

An optimal main microphone signal contains important characteristics of the first reflections, especially the time sequence of direct sound and first reflections. The sense of hearing also uses these characteristics to identify the property and origin of the signal even if just the location determining stage "recognises" the locations of the loudspeakers.
In other words: The location determing stage recognises the locations of the loudspeakers according to section 2.1. The loudspeaker signals, carrying the information about the spatial surrounding at the main microphone, are processed in the Gestalt-determining stage. Therefore the auditory pattern between the loudspeakers contains the ambience of the recording room, and the spatial depth is a simulated one, comparable to the presentation of visual perspective (see **fig. 4**).

As a result the sense of hearing requires the important characteristics (provided for example by the time sequence of direct sound, first reflections and reverberation) for the stereophonic imaging of natural ambience. If the time sequence (pattern of brief impulses) of a stereo signal in **Figure 15** is

/18

observed from thie point of view, the following is determined:

- The (undelayed) spot microphone signals simulates the direct sound
- The main microphone simulates the reflections and the reverberation
- The original time sequence of the impulses is distorted. The first reflection comes too early, the reverberation too late, etc.

The effect of the undelayed spot signal on the stereophonic imaging of natural ambience is negative. The positive characteristics of the main microphone are lost. The supported sound source is located as a phantom sound source between the loudspeakers.

4.2 Room related spot microphone technique

On the other hand, by carefully measured delay it can be achieved that the acoustical characteristics of the recording room recorded by the main microphone be much less seriously falsified.

Figure 16 demonstrates the two additional spot signals are added to the already existent first reflections by admixing correspondingly delayed spot signals. These change the balance of the recording more or less "distance neutral". It is important that the delay be so measured so as to preserve the original time delay between direct sound and first reflection.

In **fig. 17** results of listening tests are plotted which show that the directional imaging, created by the main microphone, is not effected when a spot microphone signal is added as presented in the figure above. Neither the average values nor the standard deviation values vary when this "artificial reflection" is added (main and spot microphone signal have equal level, 12 subjects). It should be pointed out that the result is found without using a panpot, i.e. without shifting the supporting source into the direction of the supported phantom sound source.

Thus the creation of directions with the usual panpot of the mixing desk can be dispensed with because the directional imaging results by the direct sound (which is produced by the main microphone, see **fig. 16**). The method can be described as **room related spot microphone technique** and may also be called "panpot-free" spot microphone technique.

/19

Fig. 18 shows the principle of the room related spot microphone technique. The delay time which must be set consists of:

1. The compensation of path difference delay between spot and main microphone
2. The additional time delay. It should be adjusted, that the spot signal barely follows on the first reflections of the recording room, which is recorded by the main microphone (see **fig. 16**).

Figures 19 and 20 show a practical example (orchestra). The path-difference delay between spot microphones and main microphone can be valuated easily with assistance of **fig. 19.** If more than one spot microphone have to be used, certain spot microphones may require approximately the same compensation of path-difference delay, for example "horn" and "double-bass". These spot microphone signals can be combined into spot signal groups in the mixing desk, and then delayed, as **fig. 20** demonstrates.

The delay lines can be saved through the delay of the combined spot microphone signals (spot groups). However, it should be emphasized again here, that each spot microphone has negative effect on the natural ambience and the focus of localization, which is given by the main microphone (equivalence microphone). However, perhaps an additional spot microphone could even be avoided by changing the positioning of the orchestra, soloists or of the main microphone, or may be by tips to the musicians to play somewhat louder at certain points?

Bibliography

/1/ BLAUERT, J.: Spatial Hearing. The Psychophysics of Human Sound Localization. The MIT Press, Massachusuetts, 1983.

/2/ THEILE, G. (1980): Über die Lokalisation im überlagerten Schallfeld (On localization in a superimposed sound field). Dissertation, Technische Universität, Berlin.

/3/ DE BOER, K. (1940): Stereofonische Geluidsweergave (Stereophonic Sound Reproduction). Dissertation, Institute of Technology, Delft.

/4/ WARNCKE, H. (1941): Die Grundlagen der raumbezüglichen stereophonischen Übertragung im Tonfilm (The fundamentals of room-related stereophonic reproduction in sound films). Akust. Z. 6, 174-188.

/5/ MAKITA, Y. (1962): On the directional localization of sound in the stereophonic sound field. Europ. Broadcasting Union Rev. Part A, 73, 102-108.

/20

/6/ WENDT, K. (1963): Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie (Directional hearing with two superimposed sound fields in intensity- and delay-difference stereophony). Dissertation, Technische Hochschule, Aachen.

/7/ FRANSSEN, N.V. (1963): Stereophony. Philips Tech. Bibl., Eindhoven.

/8/ MERTENS, H. (1965): Directional hearing in stereophony: Theory and experimental verification. Europ. Broadcasting Union Rev. Part A, 92, 1-14.

/9/ STEINKE, G.: Stand und Entwicklungstendenzen der Stereofonie. (Stereophony: state and trends of development) Technische Mitt. des RFZ 28, 1984, 1-10 und 25-32.

/10/ THEILE, G. (1978): Westhalb ist der Kammfiltereffekt bei Summenlokalisation nicht hörbar? (Why is the comb filter effect in summing localization inaudible?). Paper presented to the 11th Conference of Audio Engineers, Berlin.

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

/12/ THEILE, G.: Sind "Klangfarbe" und "Lautstärke" vollständig determiniert durch das Schalldruckpegel - Spektrum am Trommelfell? ("Timbre" and "loudness" - are they completely determined by the spectrum of sound-pressure-level at the eardrum?) DAGA '84, VDE-Verlag, Berlin, pp. 747-752.

/13/ THEILE, G.: On the standardization of the frequency response of high quality studio headphones. 77th AES Convention, 1985, Preprint 2207.

/14/ PLENGE, G. (1972): Über das Problem der Im-Kopf-Localisation (On the problem of inside-the-head-locatedness). Acustica 26, 241-252.

/15/ Laws, P. (1972): Zum Problem des Entfernungshörens und der Im-Kopf-Lokalisiertheit von Hörereignissen (On the problem of distance hearing and the localization of auditory events inside the head). Dissertation, Technische Hochschule, Aachen.

Bildunterschriften

Fig. 1: Phantom sound source situation:
comb filter effect in the ear input signals

Fig. 2: Lateral displacement of the phantom sound source
caused by inter-loudspeaker delay Δt:
comb filter effect in the ear input signals

Fig. 3: Principal function of the association model

Fig. 4: Two stereophonic imaging theories:
principal processing of the ear input signals

Fig. 5: The distance of this picture can be compared with the distance of stereo loudspeakers. The visual perspective, which is simulated by applying phenomena of spatial vision, can be compared to the stereophonic perspective, which can be simulated by applying corresponding phenomena of spatial hearing.

Fig. 6: Equivalence of ΔL and Δt, corresponding to the
direction of phantom sound source

Fig. 7: Direction of phantom sound source $\varphi(\Delta L, \Delta t)$
(delay time Δt and level difference ΔL
in the same sense)

Fig. 8: Directions of sound incidence at the main microphone
(orchestra)

Fig. 9: Directional imaging achieved with different types of
main microphones in the concert hall

Fig. 10: Directional imaging (loudspeaker reproduction in a living room).
Type of main microphones: X/Y (135°), A/B (3,3 m), KU 81

Fig. 11: Directional imaging, characteristics of different types o main
microphone

Fig. 12: Interfering reflections from the stage

Fig. 13: Comb filter effect in the mono signal
Direction of sound incidence at stereo microphones $\Omega = 60°$

Fig. 14: Comb filter effect in the mono signal
Direction of sound incidence at the stereo microphones $\Omega = 30°$

Fig. 15: Stereo signal for brief impulses: main microphone plus spot microphone without time delay

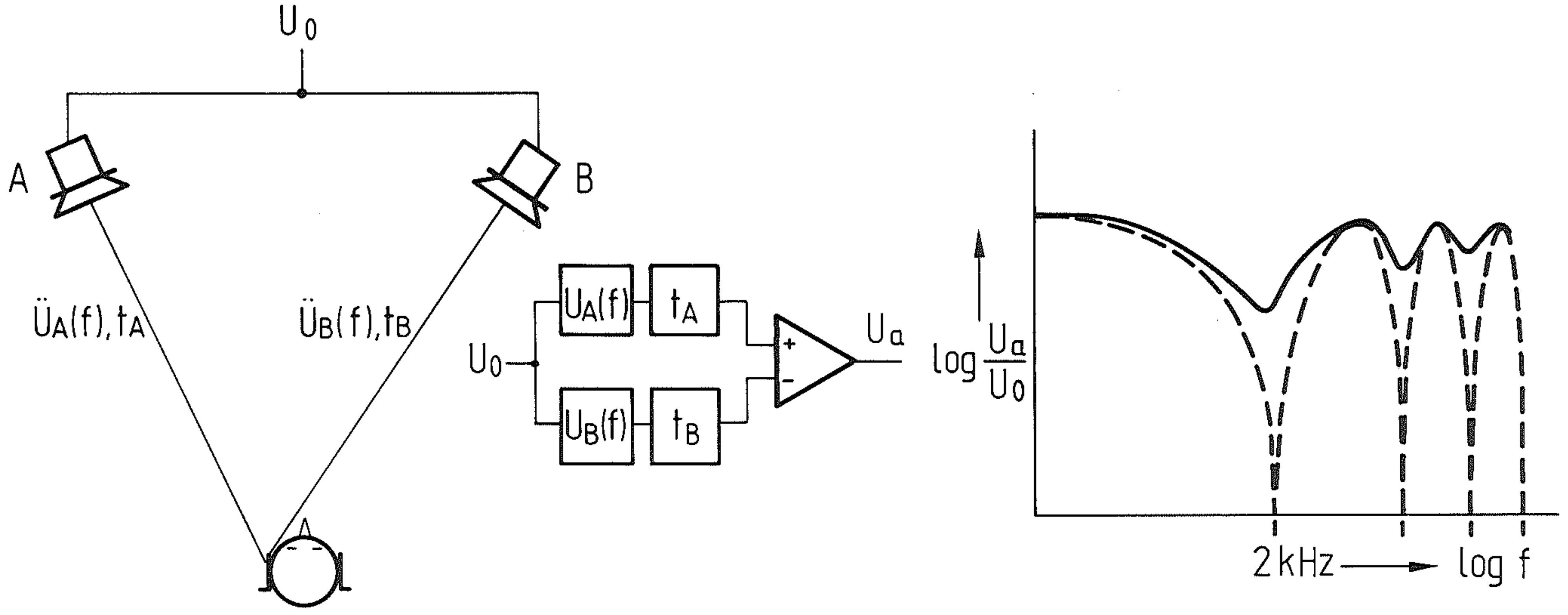
Fig. 16: Stereo signal for brief impulses: main microphone plus time delayed spot microphone

Fig. 17: Effect of delayed spot microphone on directional imaging

Fig. 18: Principle of the "room related spot microphone technique"

Fig. 19: Valuation of the compensation of path-difference delay (orchestra)

Fig. 20: Room related spot microphone technique using mixed spot microphone signals (spot groups)

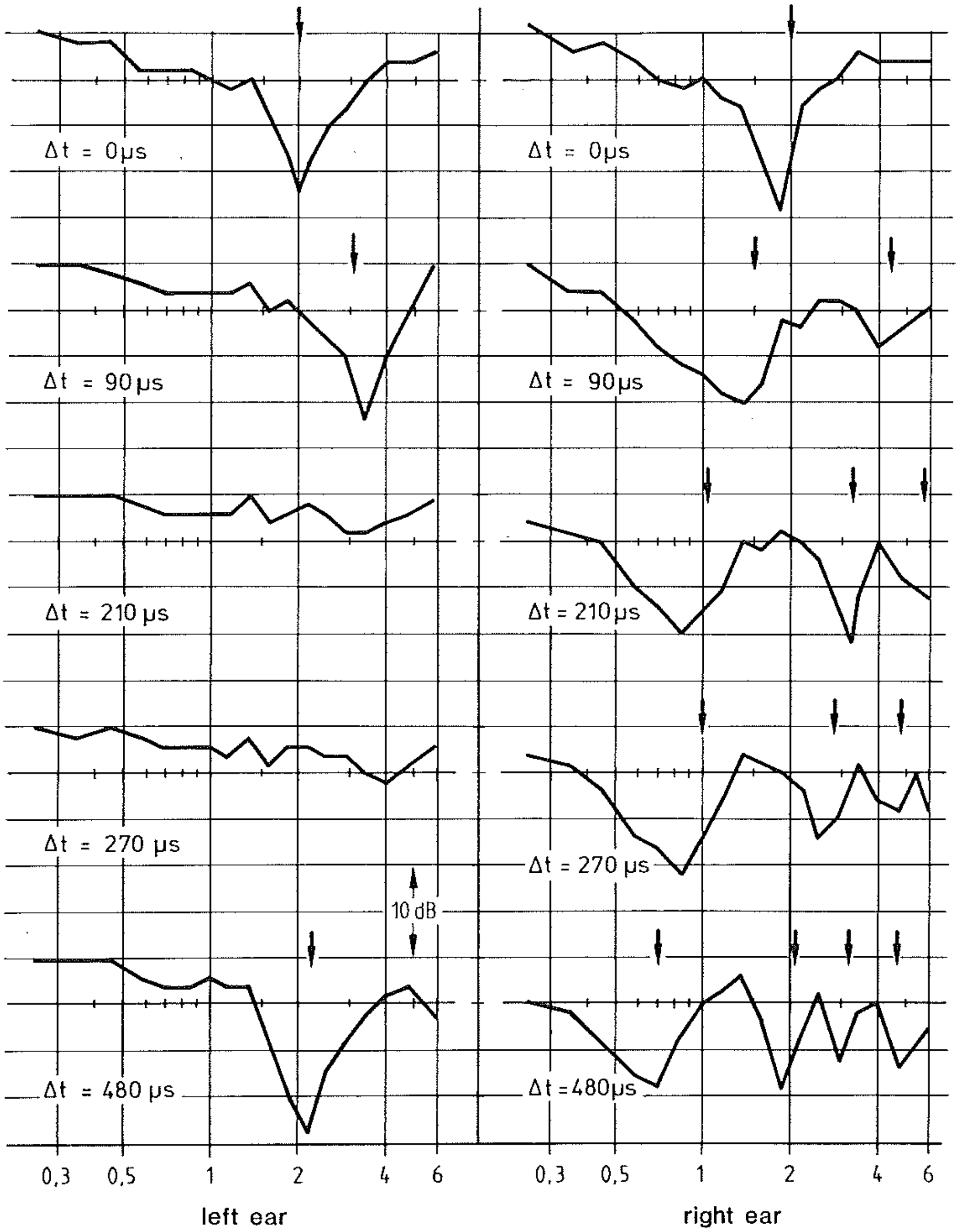


IRT

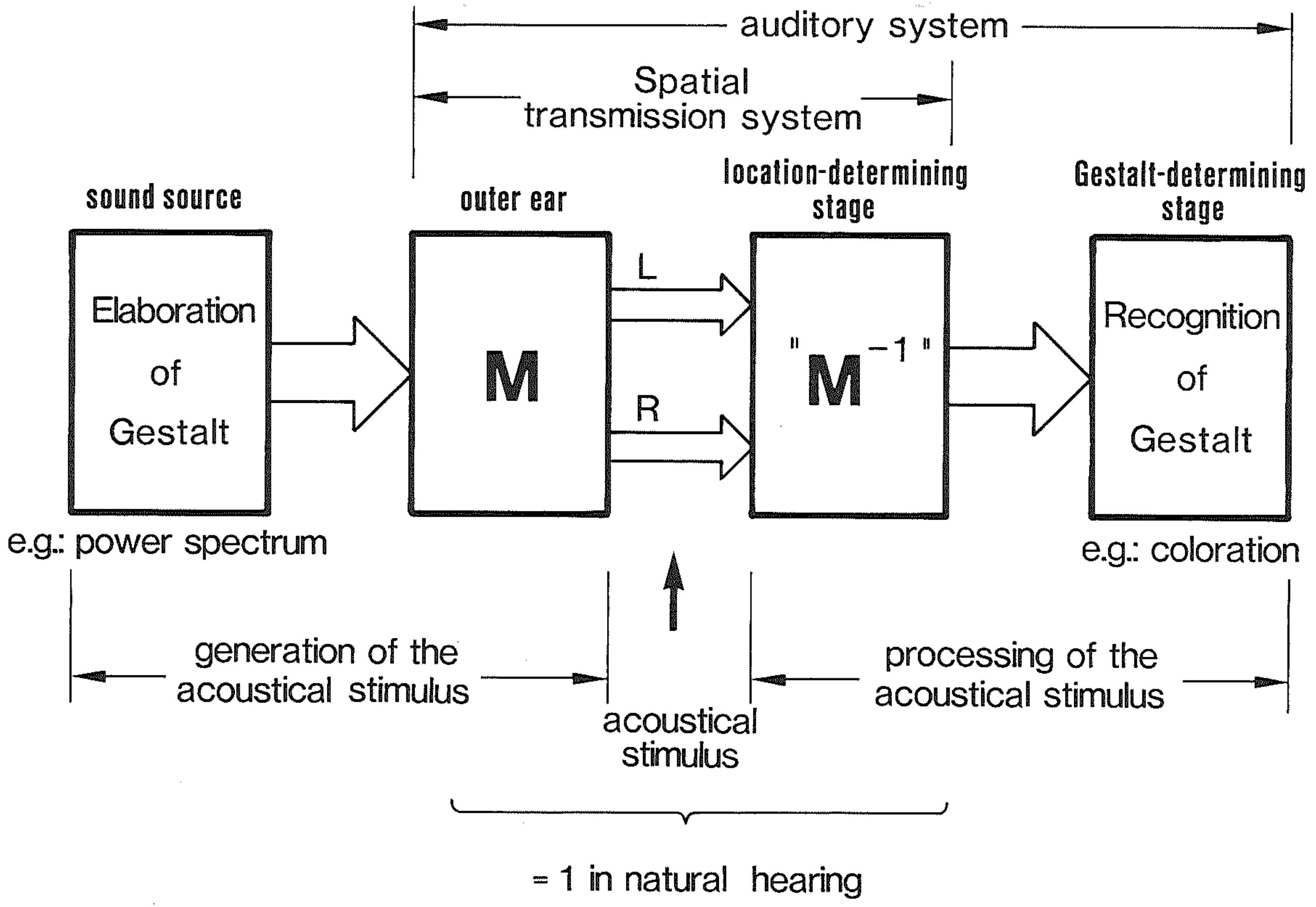
Fig. 1: Phantom sound source situation:
comb filter effect in the ear input signals

H−123a/D

3.2.86 C.Staudte



IRT	**Lateral displacement of the phantom sound source caused by inter-loudspeaker delay Δt: comb filter effect in the ear input signals**	H-121a/D **3.2.86 C.Staudte**

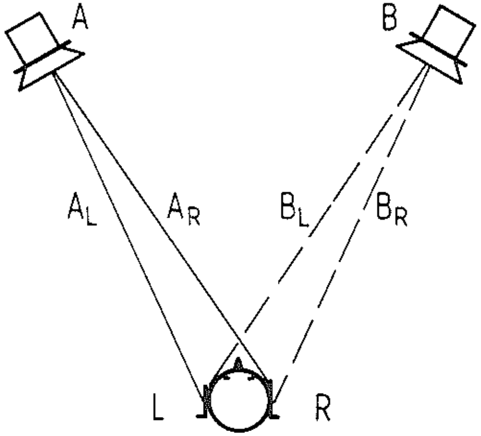


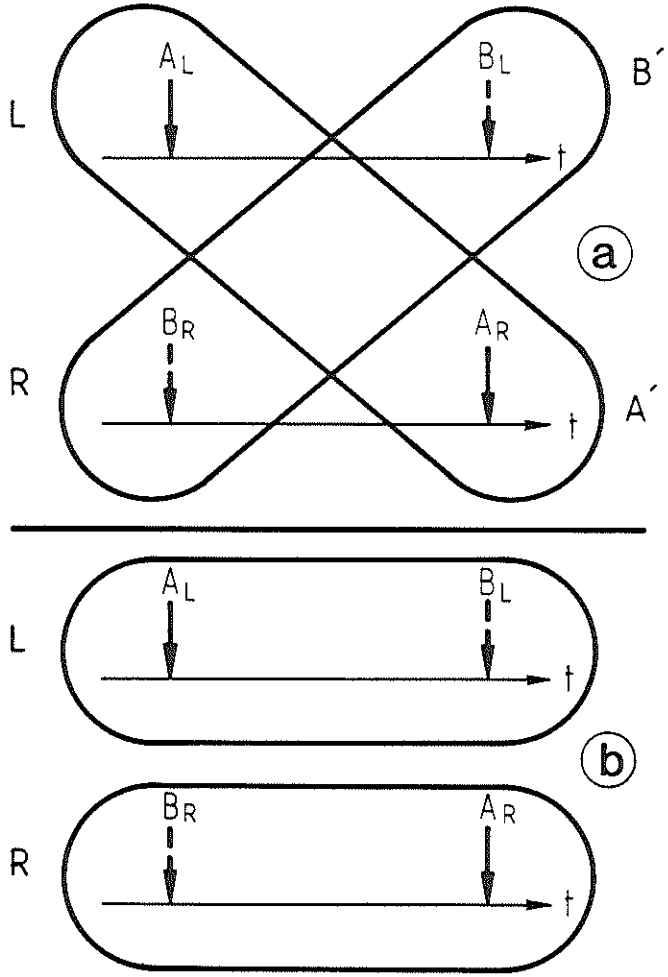
IRT

3 Principal function of the association model

H-428a/D

31.1.85 C.Staudte





ⓐ Assoziation model

ⓑ Summing localization model

IRT 4

Two stereophonic imaging theories: principal processing of the ear input signals

H-119a/D

3.2.86 C.Staudte

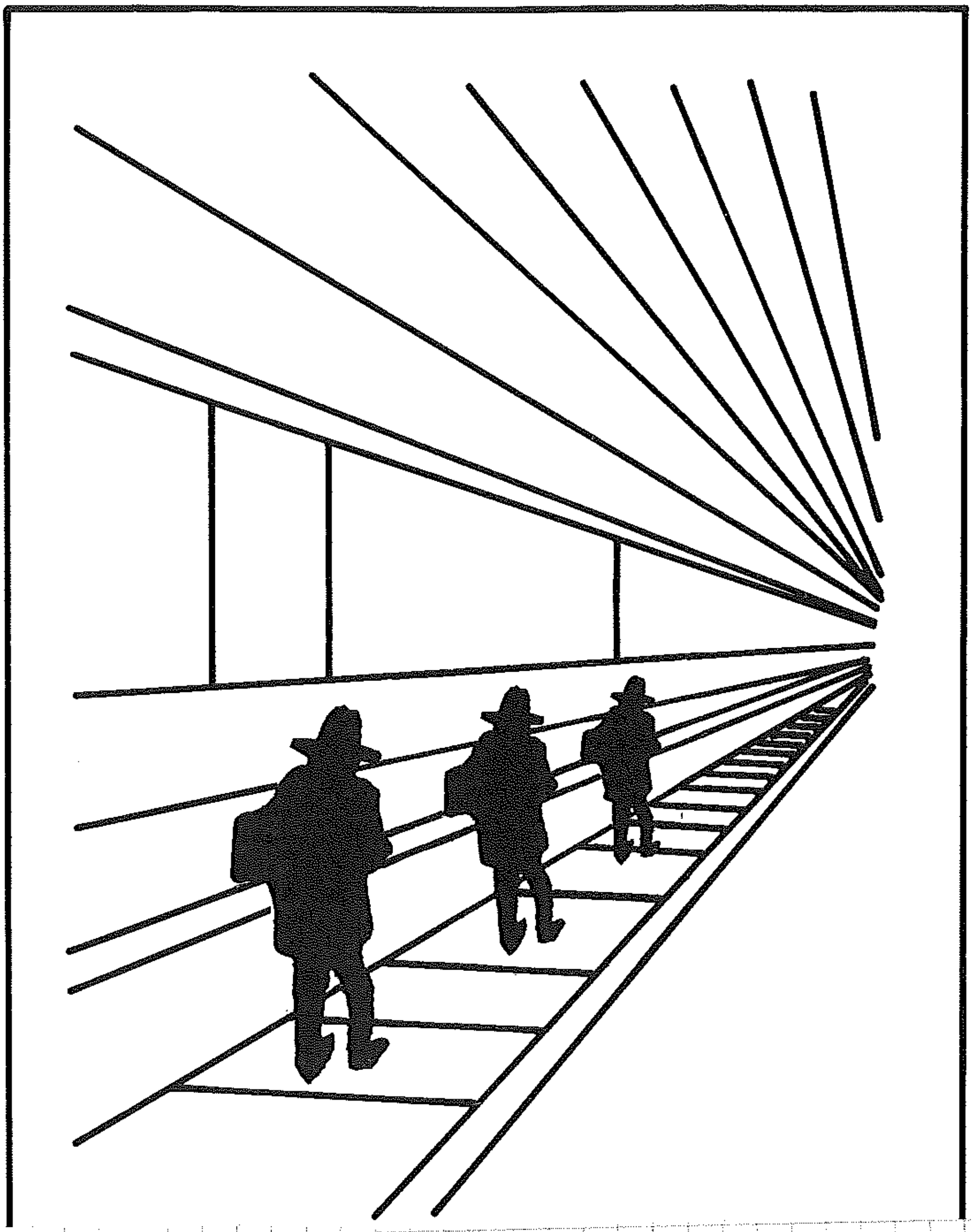
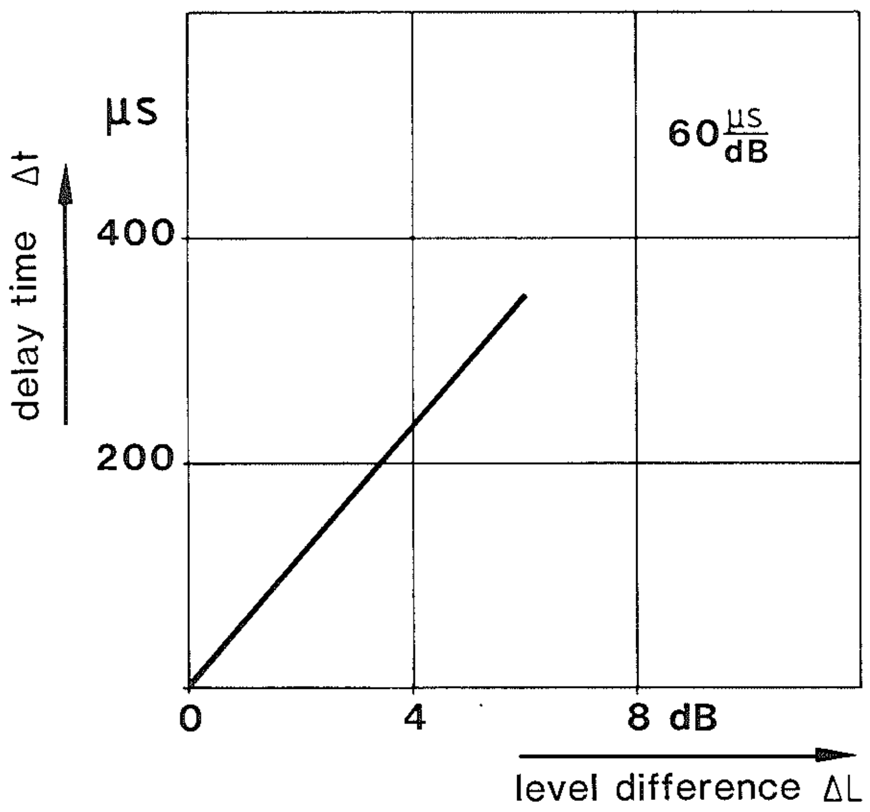


Fig. 5: The distance of this picture can be compared with the distance of stereo loudspeakers. The visual perspetive, which is simulated by applying phenomena of spatial vision, can be compared to the stereophonic perspective, which ~~has~~ can be simulated by applying corresponding phenomena of spatial hearing.



localization values : 2,4°/dB
4°/100 µs

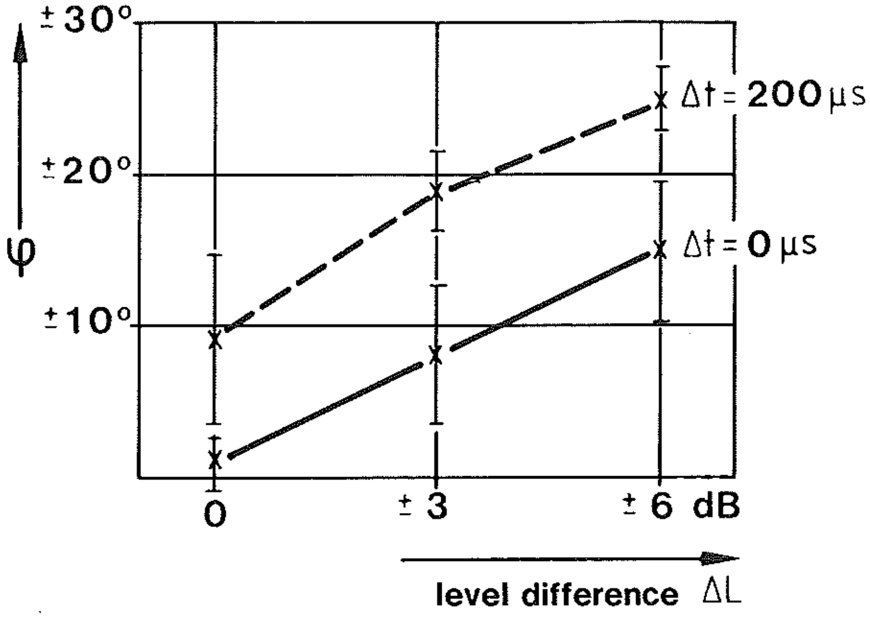
IRT

Equivalence of ΔL and Δt,
corresponding to the direction of phantom sound source

H-462a/D

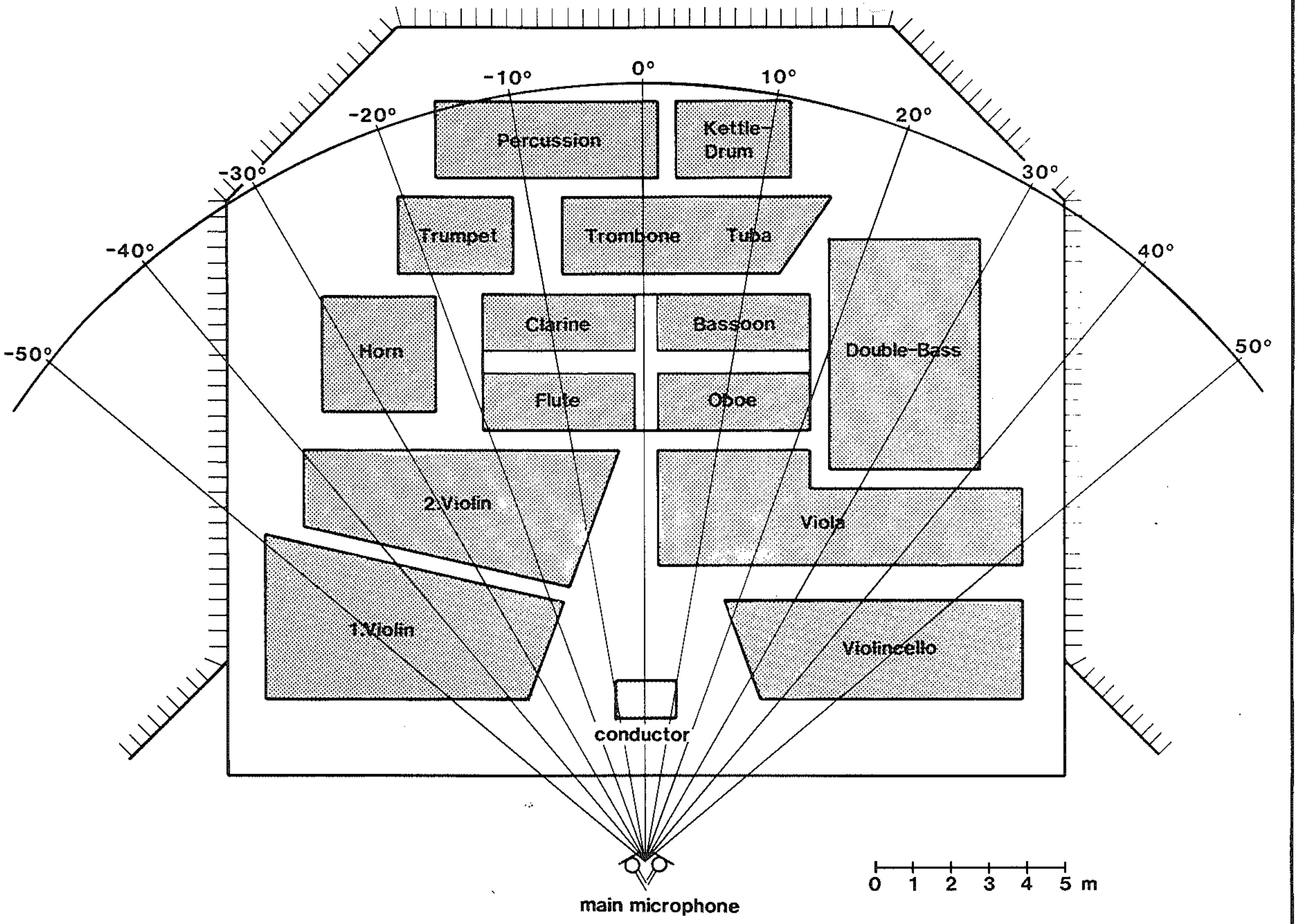
23.1.86 C.Staudte

6



IRT	**Direction of phantom sound source φ(ΔL,Δt)** **(delay time Δt and level difference ΔL in the same sense)**	**H-578a/D** **23.1.86 C.Staudte**

7



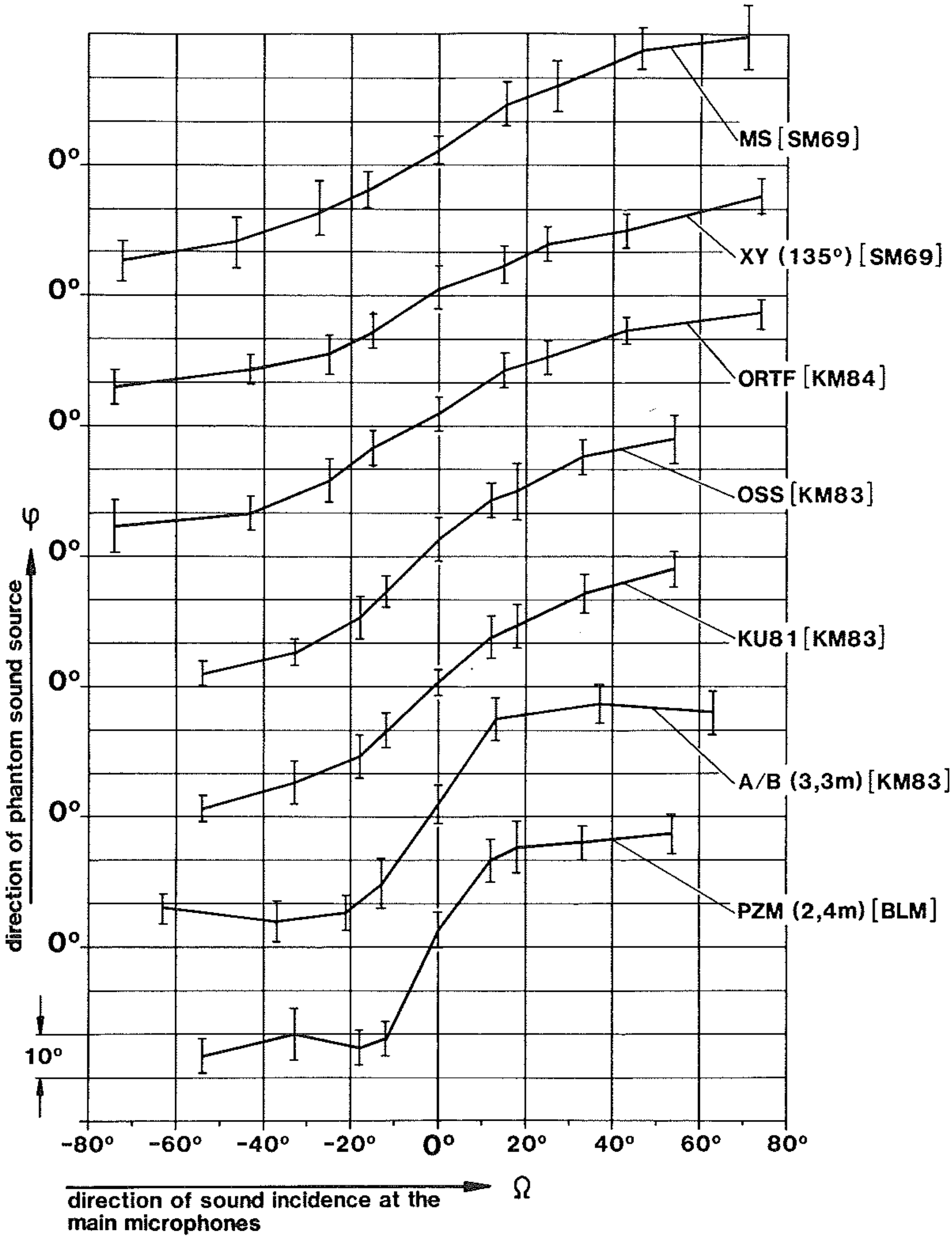
IRT

8

Directions of sound incidence at the main microphone (orchestra)

H-452a/D

24.1.86 C.Staudte



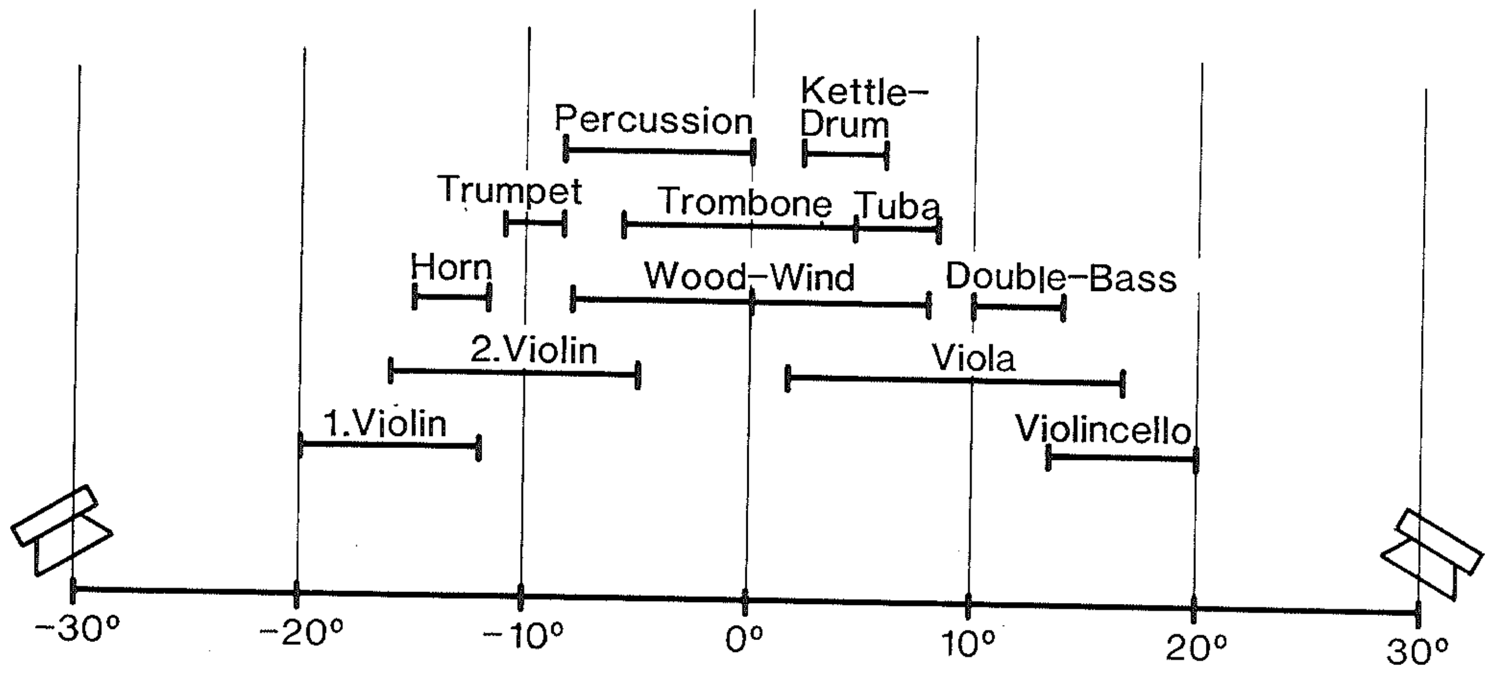
IRT

9

Directional imaging achieved with different types of main microphones in the concert hall

H-449a/D

27.1.86 C.Staudte

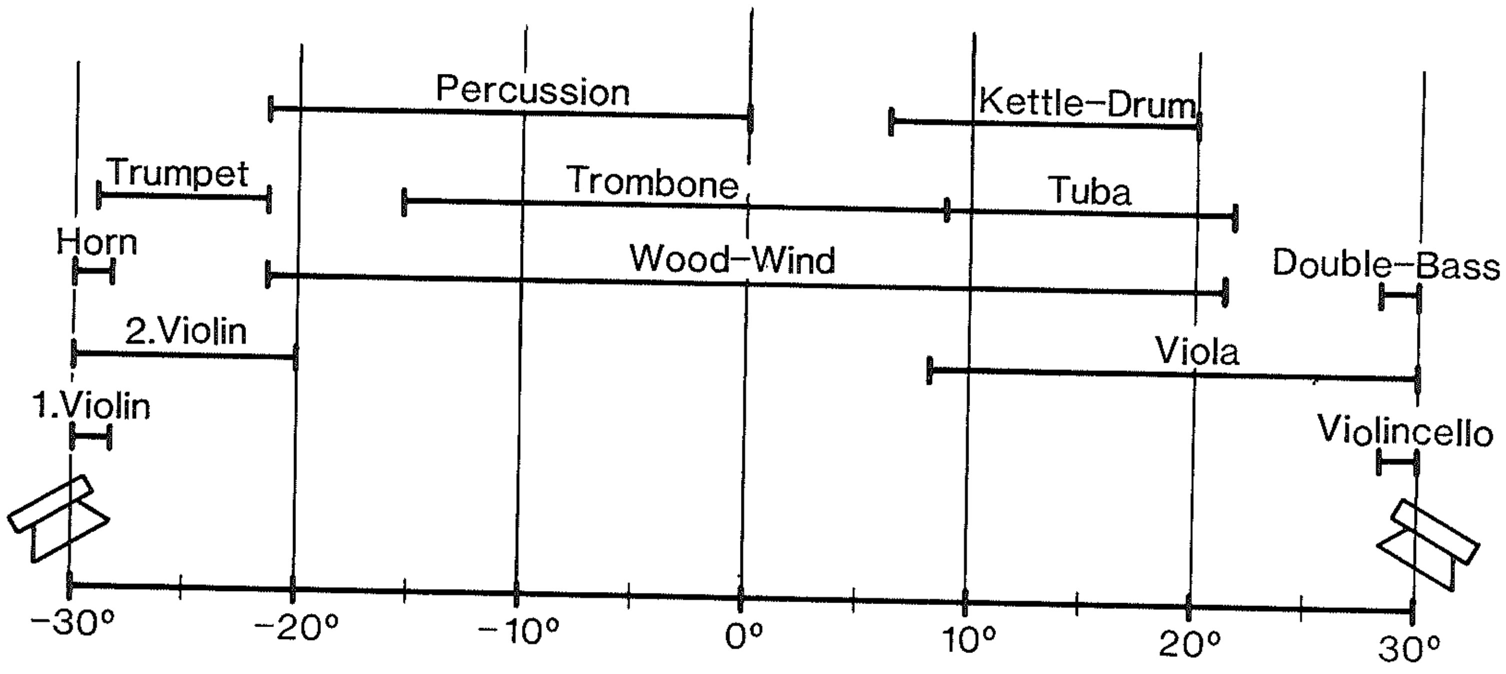


IRT

Directional imaging (loudspeaker reproduction in a living room). Type of main microphones:X/Y (135°), A/B (3,3m), KU81

H-455a/D

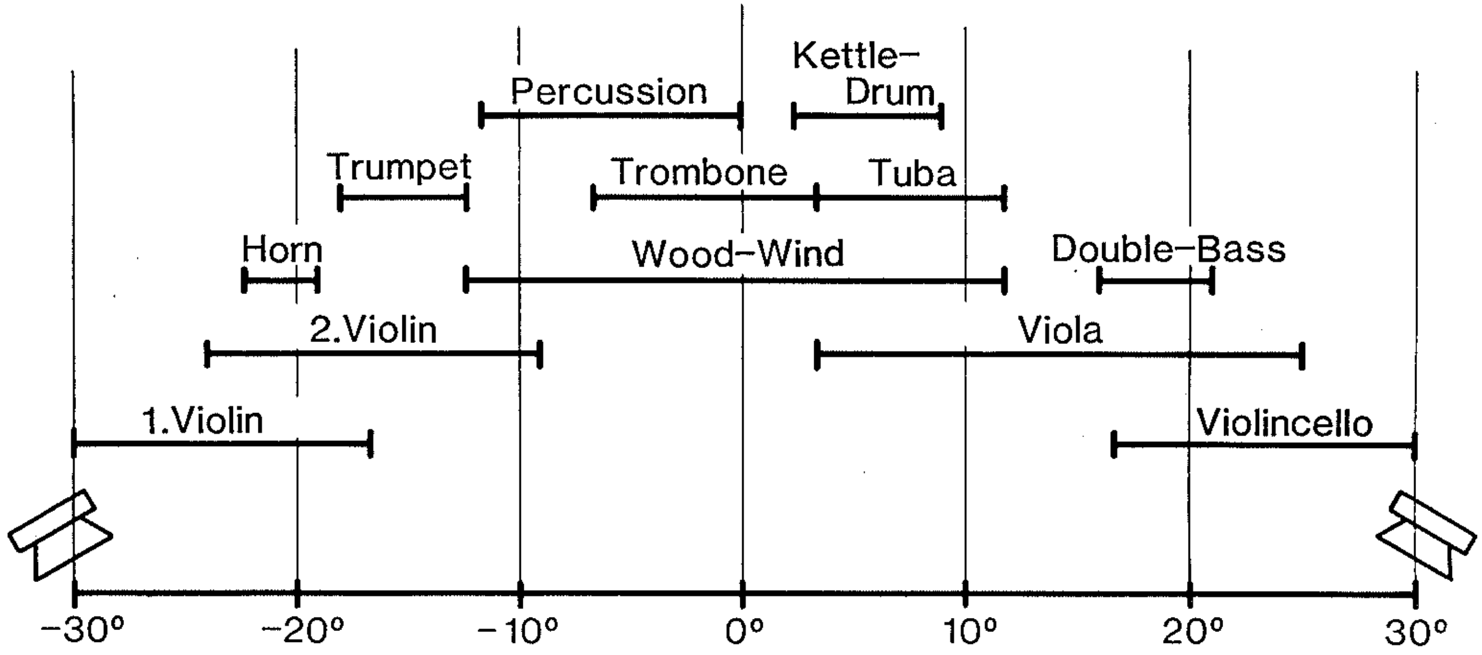
22.01.86 C.Staudte



IRT Directional imaging (loudspeaker reproduction in a living room) . Type of main microphone: A/B (3,3m)

H-454a/D

22.01.86 C.Staudte



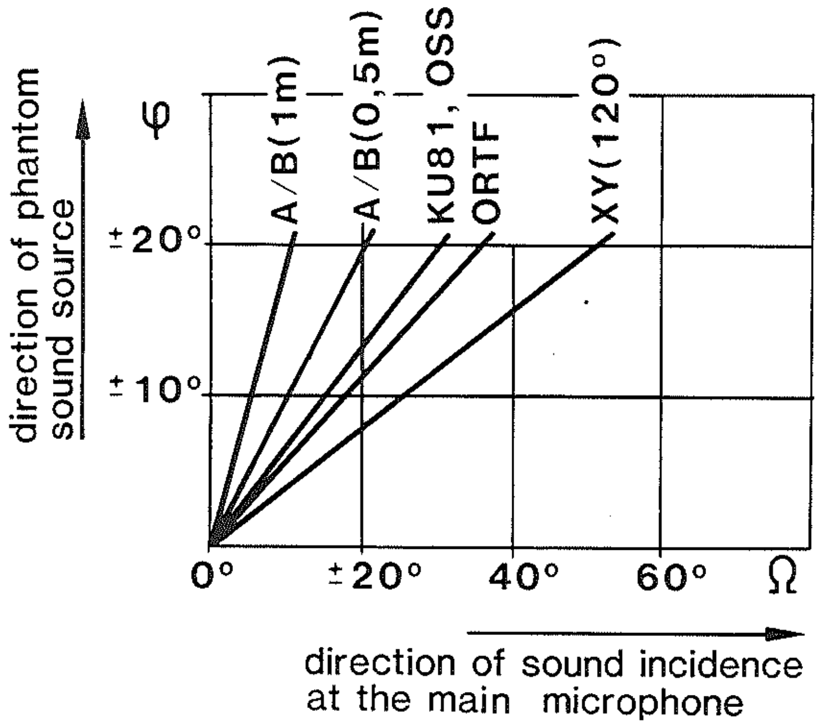
IRT

Directional imaging (loudspeaker reproduction in a living room) Type of main microphone: KU 81

H-453a/D

22.01.86 C.Staudte

10

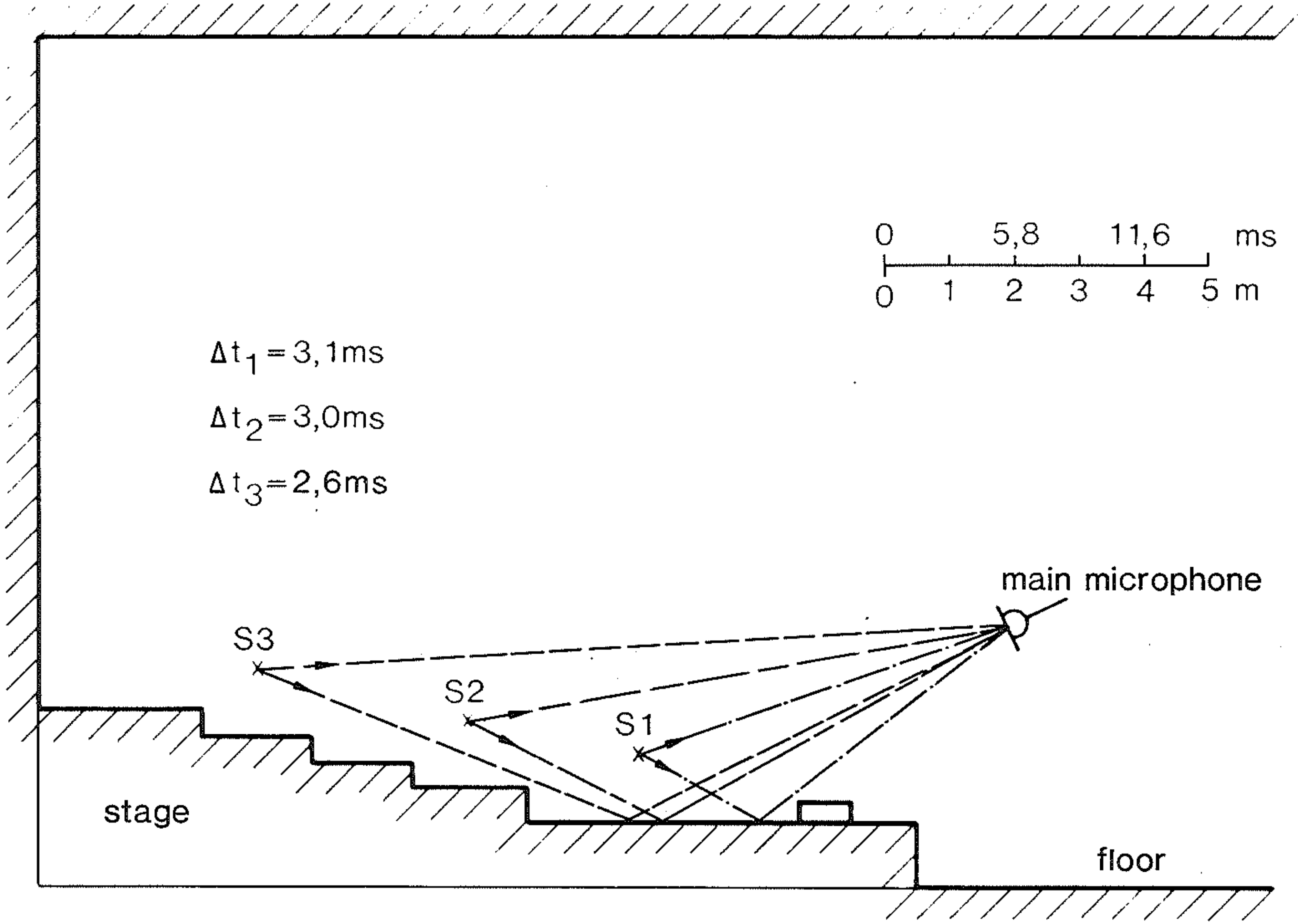


IRT

Directional imaging, characteristics of different types of main microphone

H-461a/D

23.1.86 C.Staudte

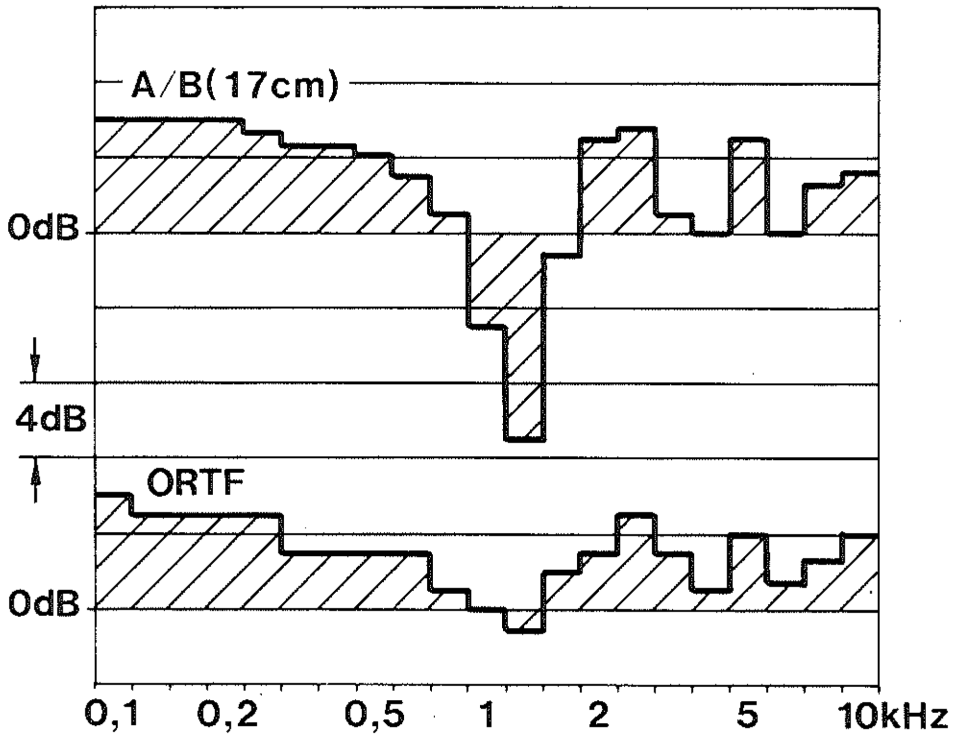


IRT

Interfering reflections from the stage

H-456a/D

23.01.86 C.Staudte



IRT

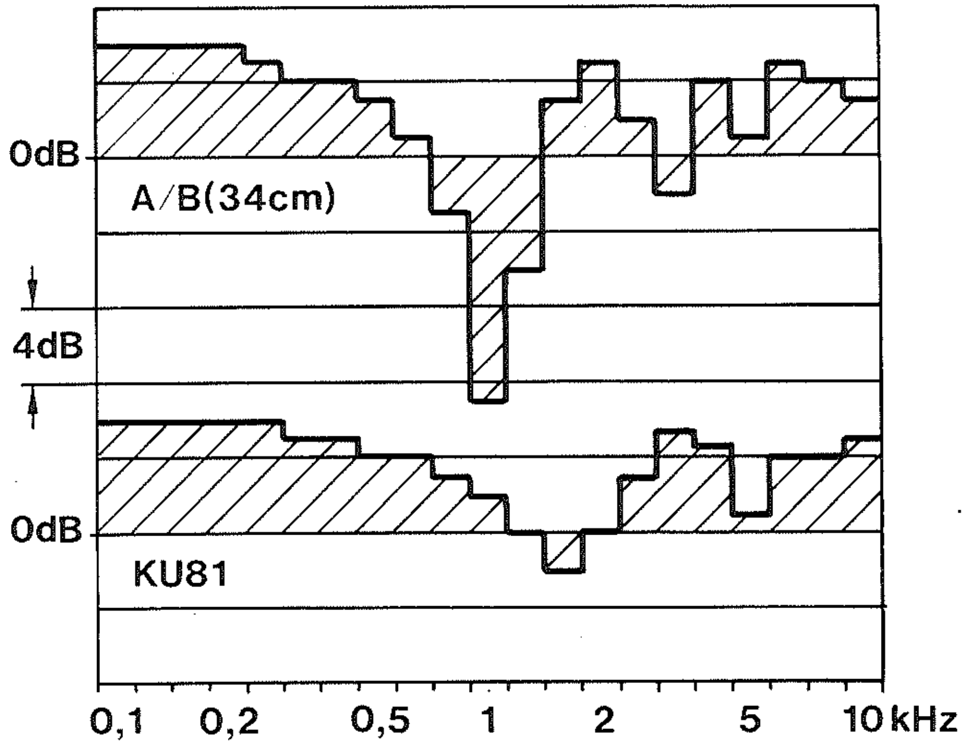
Comb filter effect in the mono signal

Direction of sound incidence at the stereo microphones $\Omega = 60°$

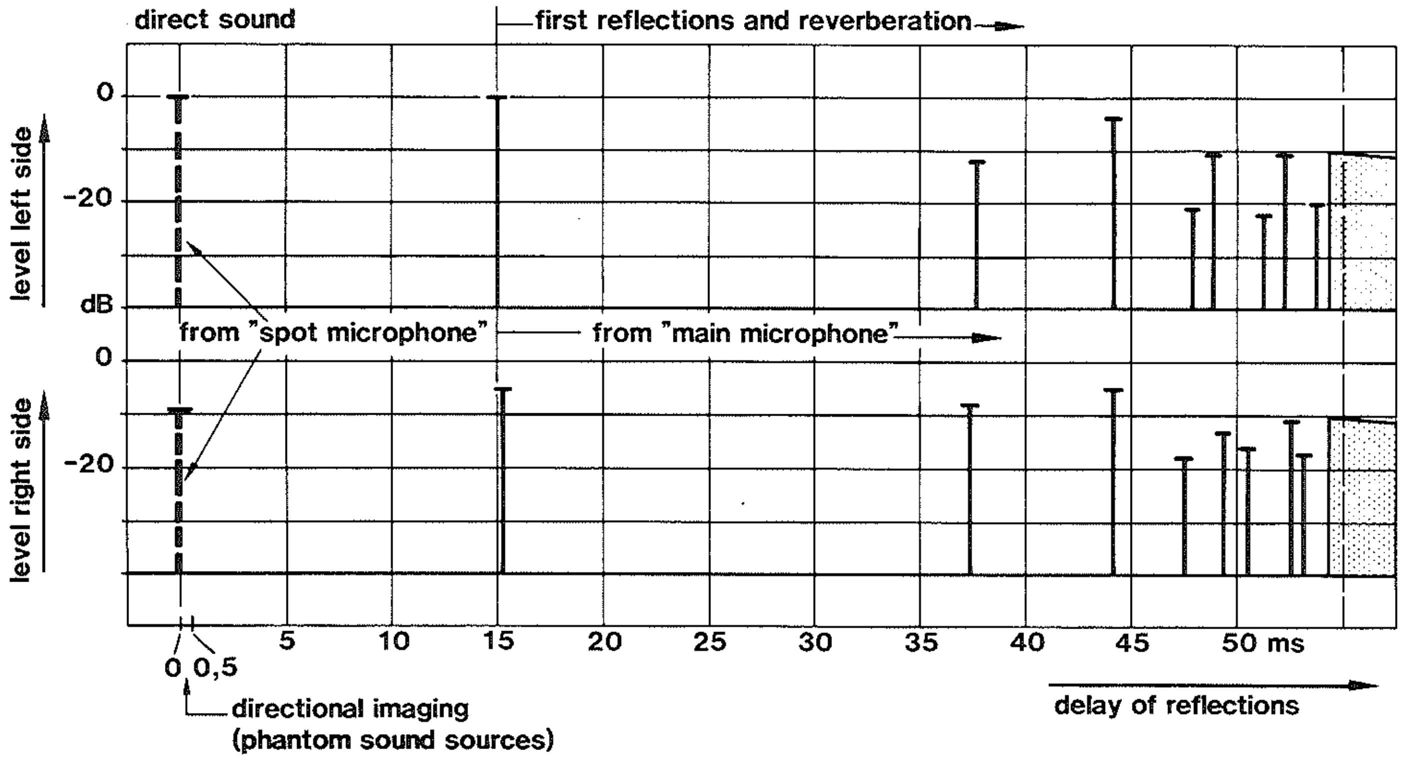
H-467a/D

23.1.86 C.Staudte

13



IRT	**Comb filter effect in the mono signal** Direction of sound incidence at the stereo microphones $\Omega = 30°$	**H-466a/D** 23.1.86 C.Staudte



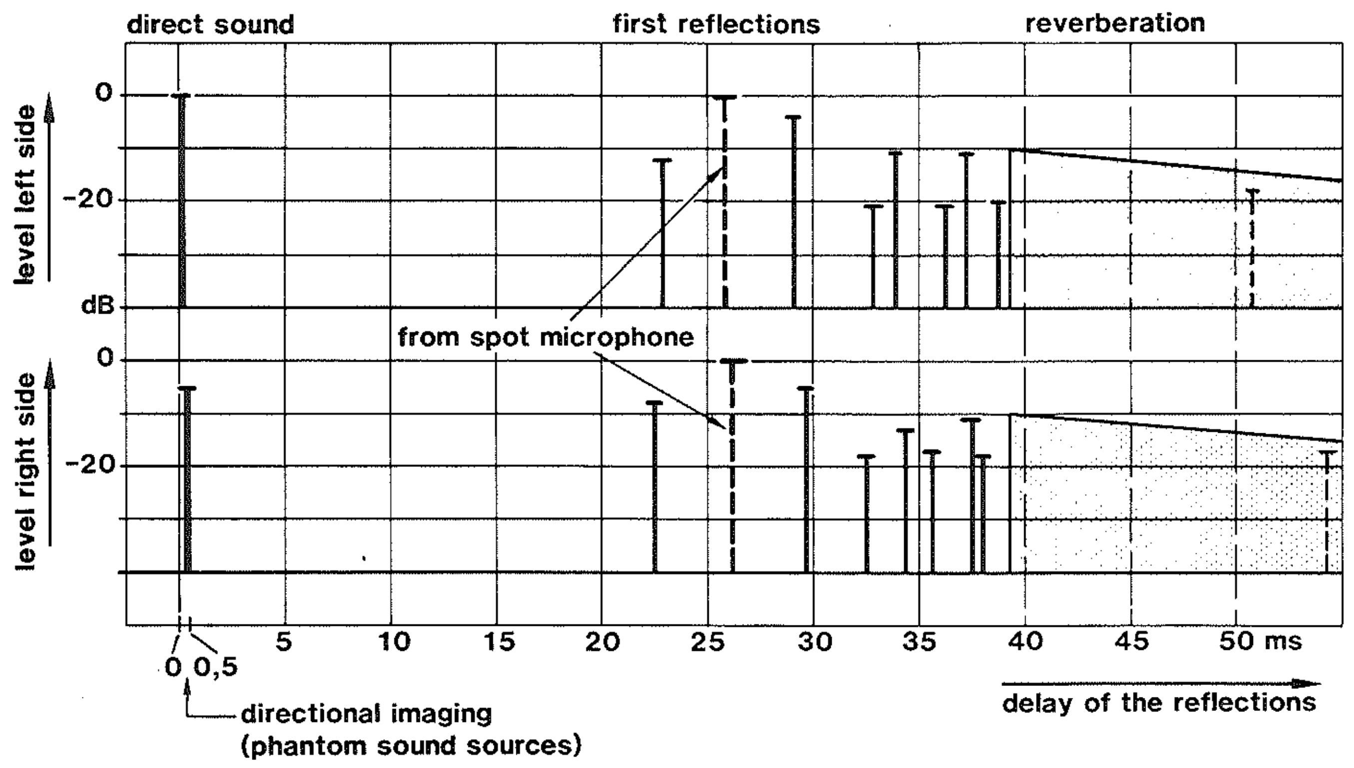
IRT

Stereo~~phonic~~ signal for brief impulses:

(main microphone plus spot microphone without time delay)

H-459a/D

23.1.86 C.Staudte



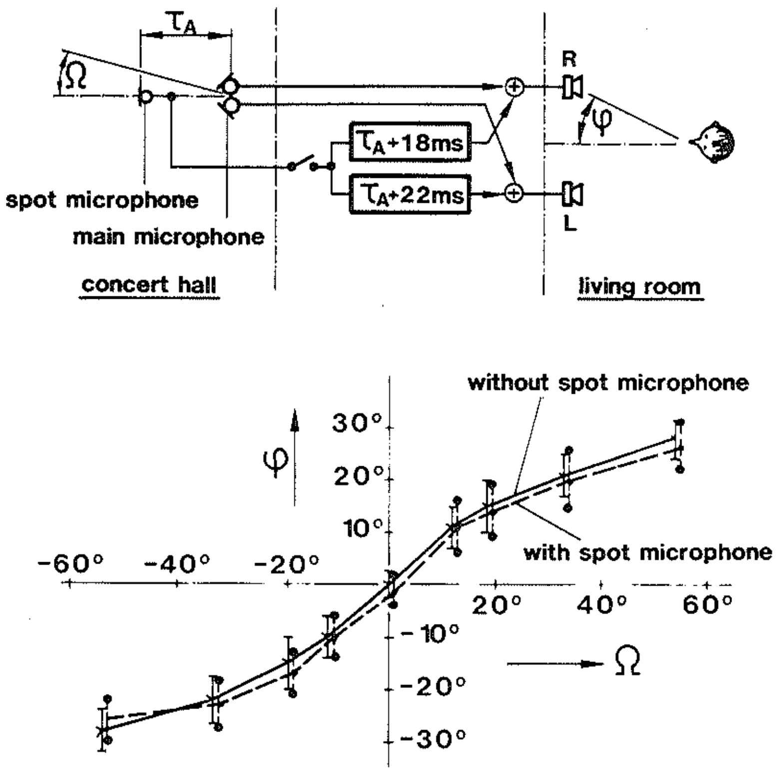
IRT

Stereo~~phonic~~ signal for brief impulses:

~~(~~main microphone plus time delayed spot microphone~~)~~

H-460a/D

28.01.86 C.Staudte



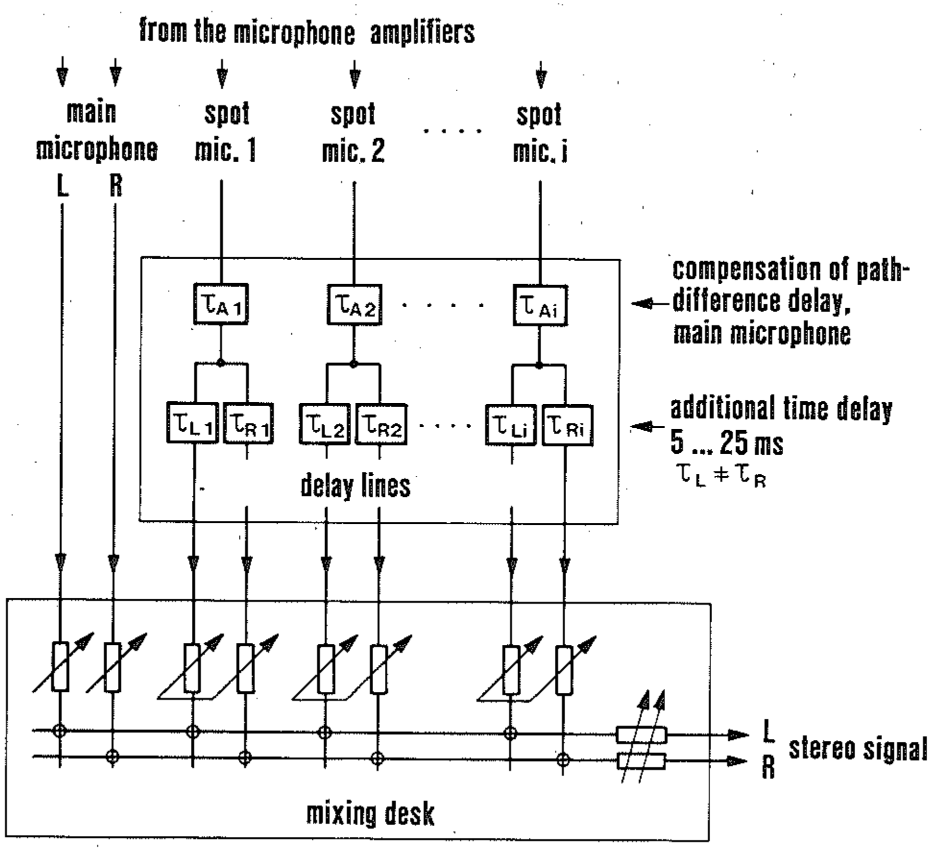
IRT

Effect of delayed spot microphone on directional imaging

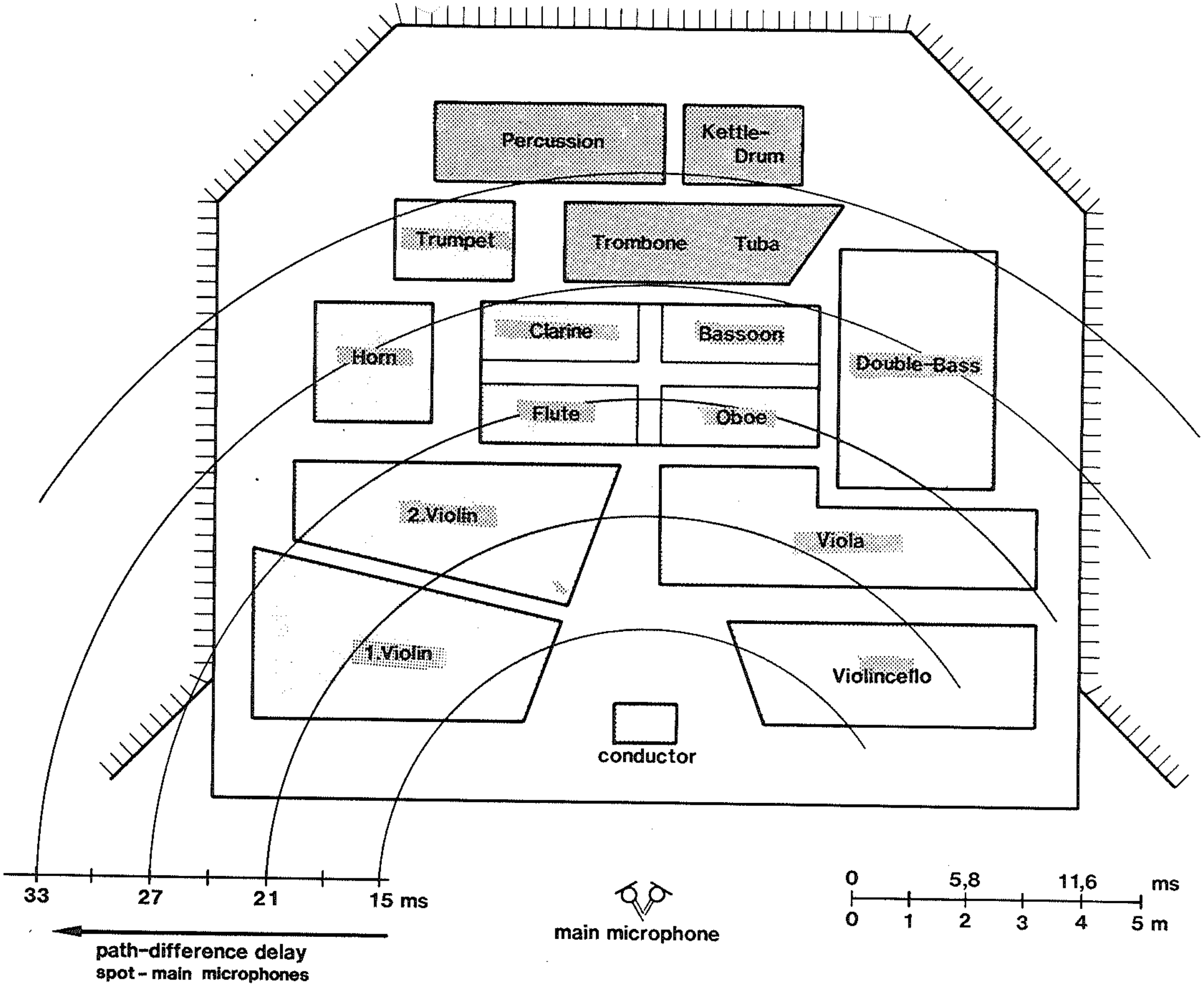
H-579a/D

27.1.86 C.Staudte

17



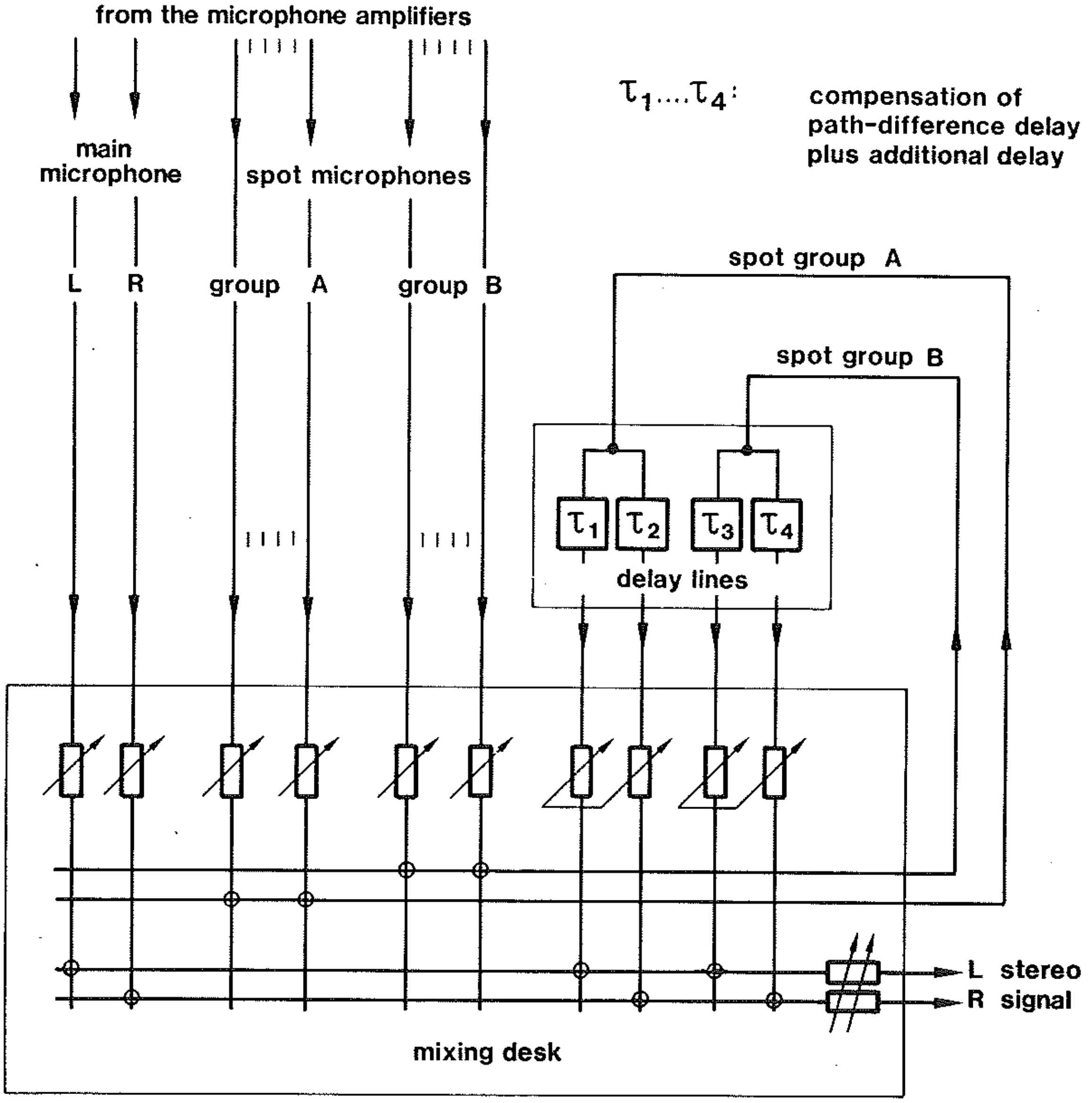
IRT	"Principle of the room related spot microphone technique"	H-447a/D 27.1.86 C.Staudte



IRT 19 Valuation of the compensation of path-difference delay (orchestra)

H-450a/D

23.1.86 C.Staudte



IRT 20

Room related spot microphone technique
using mixed spot microphone signals (spot groups)

H-448a/D

28.1.86 C.Staudte