MULTICHANNEL NATURAL MUSIC RECORDING BASED ON PSYCHOACOUSTIC PRINCIPLES

Günther Theile, IRT, Germany

theile@irt.de

1. INTRODUCTION

It has been recognized for many years that conventional two-channel loudspeaker stereophony has serious limitations, and that additional stereophonic channels are desirable with regard to providing an improved stereophonic image for a number of listeners who are not positioned at the ideal reference point of the loudspeaker arrangement.

The universal 3/2-stereo format according to Recommendation ITU-R BS 775-1 [1] provides an additional center channel and two surround channels, completing the left and right stereo channels, thereby offering enhanced quality of the stereophonic presentation not only in case of audio-only applications but also for applications with accompanying picture (see SMPTE RP 173 [2]). The format should be able to provide easy program exchange with film sound and - at the same time – to overcome some of the weakness of conventional two-channel stereophonic systems when they reproduce music. For example, the new format should be more satisfactory for a "purist" music-lover when he is listening to DVD-Audio or Super-Audio-CD of a classical music concert without the accompanying picture.

This contribution concentrates on the goal of improving pure music reproduction. Psychoacoustic principles are used to determine both the possibilities and the limits of the new stereo format. A brief overview is given in **TABLE 1**. It illustrates that the possibilities of stereophonic imaging are limited in a number of parameters, while the 3/2 stereo format enables the sound engineer to exploit binaural cues more effectively than is possible with two-channel stereo, and thus to create a new dimension of enveloping atmosphere and spatial impression. The more accurately the psychoacoustic principles are understood and taken into account from the technical and artistic points of view, the more successful and convincing the reproduction will be. This is particularly true in all cases where optimum “naturalness” of the stereophonic presentation is desired.

	2/0-Stereo	**3/2-Stereo**	**Dummy-head**
Horizontal direction	*+30°...-30°*	*+30°...-30°.* *surround effects*	*surround* *(instable front)*
Elevation	*not possible*	*constraints?*	*possible*
Depth	*simulated*	*constraints?*	*possible*
Near-head distance	*not possible*	*no?*	*possible*
Spatial impression	*simulated*	*possible*	*possible*
Enveloping sources	*not possible*	*constraints?*	*possible*

TABLE 1: Imaging performance of stereophonic systems

The 3/2-stereo format enhances the possibilities of conventional two-channel loudspeaker stereophony. However, it is a compromise, and it has limits with its ability to present direction and distance, particularly when it is compared to dummy-head stereophony. On the other hand, it is worth taking into account that technologies such as dummy-head stereophony do not provide much room for creative sound design.

What does optimum naturalness mean? The simplest answer would be: the reproduced sound image must be nearly identical to the original sound image. This definition appears to be problematic because identity can definitely not be required, in principle, as a goal for optimising the stereophonic technique. Identity may conceivably be appropriate for dummy-head stereophony or wavefield synthesis, or perhaps for the reproduction of a speaker's voice through loudspeakers, but it is appropriate only to a limited extent for the reproduction of the sound of a large orchestra through loudspeakers. Artistic intentions of the sound engineer, aesthetic irregularities in the orchestra, poor recording conditions in the concert hall, as well as the necessity of creating a sound mix "suitable for a living room" with respect to practical constraints (poor listening conditions, reduced dynamic, downward compatibility) – all cause loudspeaker stereophony to deviate from identity.

The desired natural stereophonic image should therefore meet two requirements: it should be satisfying aesthetically and at the same time it should match the tonal and spatial properties of the original sound. Both requirements will undoubtedly be contradictory in many situations. However, a more flexible stereophonic recording technique will allow a more successful optimisation by the sound engineer.

It should be mentioned here that the enhanced possibilities of the 3/2-sterero format (see **TABLE 1**) are based on the supplemental surround channels. The center channel is not intended to provide additional room for creative sound design. As regards stereophony, the primary purpose of the center channel is to increase directional stability (to broaden the listening area) by using the two 30° imaging sectors L-C and C-R instead of one 60° imaging sector L-R, as well as to improve “clarity” and “sound colour” of the image in the center area. Consequences regarding microphone and recording methods are discussed in **CHAPTER 3**.

The term “surround” originates from the movie industry where it is used mainly to represent the acoustic environment and directional effects outside the picture. Surround loudspeakers are defined in this context as well as loudspeakers outside the frontal stereophonic imaging plane. This does not imply that the aim is to provide a full surround imaging plane giving unlimited directional imaging of arbitrary events. The three basic applications of the surround channels LS and RS are presentation of space, atmosphere, and effects, completing or supporting the frontal stereophonic image.

Precedent (direct) sound

Located sound sources → directional perception

Non-precedent (indirect) sound

Reflections / reverberation → spatial depth
Reflections / reverberation → spatial impression

Environmental (non-reflected) sound

Ambient non-located sound sources → enveloping atmosphere

TABLE 2: Three types of 3/2-stereophonic sound

Localisation of (phantom) sources, perception of spatial depth and spatial impression as well as perception of enveloping atmosphere (e.g. applause) should be distinguished and understood as phenomena of spatial hearing, each of them governed by specific laws and needing adequate microphone configurations and mixing methods.

The new 3/2-stereo format can offer improved stereophonic presentation with regard to imaging parameters such as localisation (in particular perception of depth), spatial impression, enveloping atmosphere (e.g. applause), see **TABLE 2**. The related psychoacoustic principles should be understood as phenomena of spatial hearing governed by specific laws and thus requiring suitable types, configurations and locations of microphones, as well as distinct handling of delay, interchannel correlation and level balancing of direct / indirect sound.

Correspondingly designed natural recording methods are suggested to be the adequate basis to take maximum advantage of five stereophonic channels. In the following chapters basic considerations are presented. They are applied to the classical recording concept “main microphone / spot microphones / room microphone”. However, it should be realised that the pros and cons of this concept on one hand and poly-microphony on the other are not discussed in this contribution. Rather, the psychocoustic principals can also be applied to poly-microphony. Modern mixing consoles could include tools for the synthesis of arbitrary natural spatial impressions on the same basis.

2. THREE-DIMENSIONAL SPACE IMAGING

2.1 Auditory spatial impression

The spontaneously perceived auditory spatial impression which is caused by the actual or reproduced indirect sound of a room comprises two attributes of the auditory event. The first is “reverberance” described in [3] as a temporal slurring of auditory events caused by late reflections and reverberation /[1]. The second is “auditory spaciousness”, which denotes a characteristic spatial spreading of the auditory events [3], caused by the early lateral sound which reaches the listener´s ears from lateral directions about 10 to 80 ms after the direct sound /[2] (the optimum delay of the early lateral reflections is in the range 15...25 ms).

The early lateral sound induces an interaural decorrelation of the two ear input signals of a form which is specific to the particular room, and hence a particular auditory spaciousness. The dependence of spaciousness on delay time, level, angle of incidence, and spectrum of the early lateral reflections has been investigated(see e.g. [4], [5]). The ratio of lateral reflected energy to the energy of the direct sound as well as the delay of reflected energy has been found to be relevant cues [5]. Also, the

---

/[1] According to [5] reverberance can be understood as the perception of a “background sound stream” which is primarily perceived in the time gaps between “foreground sound events”. Only in the case of continuous foreground sound streams the image is broadened by the reflected (spatially diffuse) sound.

/[2] The effect of spatial spreading of auditory events due to lateral reflections depends on their rise-time [5]. If the events have a short rise-time compared to the arrival of major lateral reflections they will be perceived as sharply localised. If the rise-time is slow the images will be broadened.

overall level of the direct sound and of the reflections has been found to be of central importance [4]. In other studies it is suggested that it is the pattern of hall reflections themself which causes listener preference, rather than the resulting low level of interaural correlation.

Experimental results reported in [4] show that the amount of spatial impression depends on the angle of sound incidence of lateral reflection. The results are of practical importance because they demonstrate that reflections from the side are the most effective way of achieving spaciousness. In contrast, early reflections in the media plane are disadvantageous.

It is an interesting phenomenon that the natural spatial impression can be achieved with loudspeakers located exclusively in the horizontal plane. There are two reasons. Firstly, reflected diffuse sound energy in the listening room (suitable directivity of surround loudspeakers as well as appropriately reflecting walls and ceilings) has an enveloping effect. Secondly, the upper hemisphere is subjectively involved, due to our spatial hearing experience (“auditory association”, see [29]), and to the precedence effect [3]. The term “three-dimensional space imaging” reflects this fact and is justified for this reason. On the other hand it is clear that additional channels above the listeners can improve the quality of reproduced three-dimensional spatial sound (see e.g. [30], [31], [32]).

2.2 Stereophonic representation

In conventional two-channel loudspeaker stereophony, the impression of space necessarily has to be provided exclusively as a two-dimensional spatial perspective created by the two front loudspeakers. This can be optimised by applying phenomena of spatial hearing, for example introducing the natural pattern of reflections into the stereophonic signal, however, the principal result is a “perspective picture in the simulation plane” between the loudspeakers [6]. It is comparable with spatial visual imaging: The distance of a picture corresponds to the distance of the stereophonic imaging plane. In the picture, a visual perspective is simulated by applying phenomena of spatial vision. In both cases the simulation of perspective has the effect of creating a natural two-dimensional image of a three-dimensional space.

In the case of the 3/2-format, the listener's acoustic environment can be tailored by the use of additional surround loudspeakers. A largely natural, real **spatial impression** can be produced by reproducing reflections and reverberation through loudspeakers outside the frontal stereophonic imaging area, notably to the side of, and behind, the listening area. The principles are applied in numerous realisations on the basis of fundamental research into the room acoustics of concert halls, with the common aim of creating lateral sound in the listening room.

The psychoacoustic laws governing the perception of auditory spaciousness in the concert hall, as described above, can be taken as the basis for the generation of surround sound. To achieve a natural sound impression it is important to avoid localisation of the surround loudspeakers. This can be ensured most effectively during recording or mixing by delaying the surround signals with respect to the front signals. The delayed surround signals act like lateral reflections from the hall and the "law of the first wave-front" [3] is effective. At the same time, the loudspeakers distributed in different parts of the listening room create a naturally-diffuse sound field in the listening area. Consequently, there is a physical imitation of early reflections and reverberation, giving a natural spatial impression. The listener not only perceives a spatial perspective in the front imaging plane - he actually feels included in the acoustic event.

Moreover, the reproduction of lateral reflections may lead to a perception of **spatial depth**. It was found earlier in the context of the so-called room-related balancing technique [6], [7] that the presentation of distances in the simulation plane is successful even if only one left and right lateral reflection can be imitated. The stereophonic quality changes from a simulated to a real impression of spatial depth if the lateral reflections are delivered by surround loudspeakers and actually arrive at the listener from lateral directions.

2.2.1 Spatial design using delay

Natural imaging of spatial impression and depth requires careful layout of the delay situation, that is to say according the principles of the room-related balancing technique [6], [7]. The basic concept is indicated in Fig. 1, showing the case “orchestra, main microphone plus spot microphones”, as an example, and assuming at first that a suitable three-channel main microphone exists. It should be emphasised here that the concept can be applied accordingly to any form of poly-microphony.

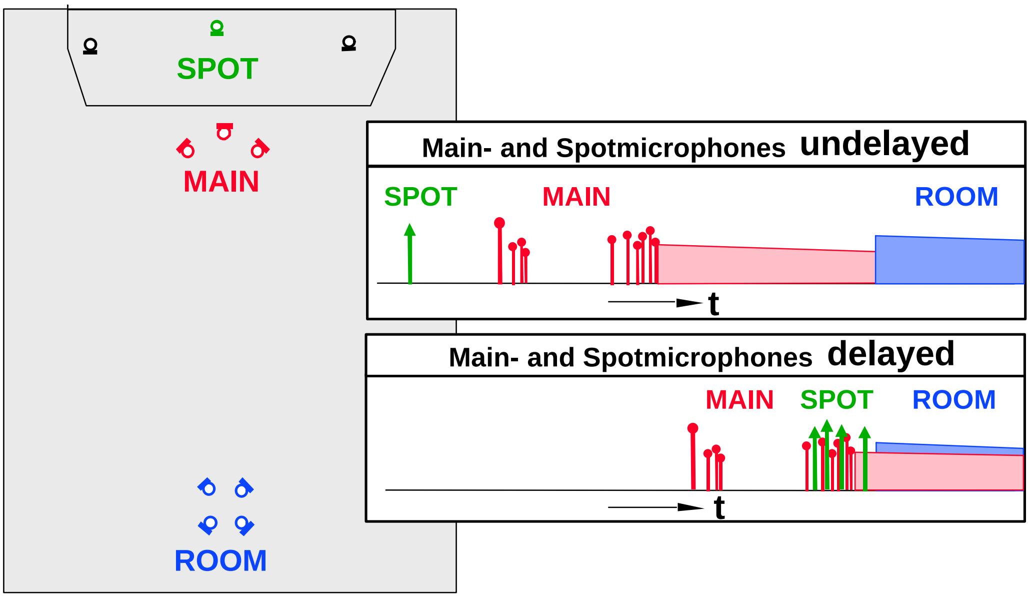


FIG. 1: Delay design according to the natural room response

The concept of room-related balancing supports the naturalness of spatial impression. It is proposed to design the delay situation in accordance with the original pattern of reflections in the concert hall. This concept is not restricted to the main / spot microphone method but is possible with any kind of microphony. Details regarding main and room microphone are discussed in **CHAPTER 3**.

In the case of undelayed microphone signals (**FIG. 1**, upper situation) the signal picked up by a spot microphone is reproduced earlier than the corresponding main-microphone signal. Thus the ear interprets the spot-microphone signal as the direct sound, and favourable imaging characteristics of the main microphone are lost. Such recordings sound unnatural, flat, without spatial depth. The cause for it is the Precedence Effect [3] (**FIG. 2**). The spatial attributes of a auditory event are in principle determined by the sound arriving first at the ear. Thus in the sound mix the spot microphone signal arrives first, and therefore the characteristic of the added spot microphone signal is relevant for the stereophonic quality of the recording.

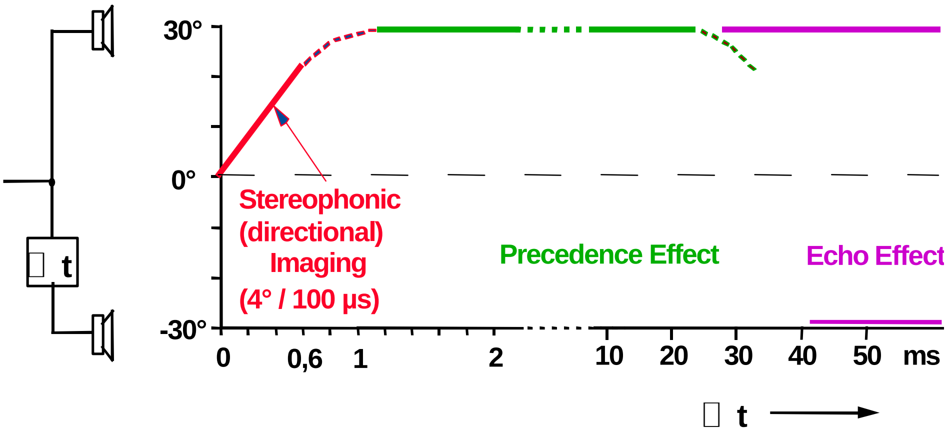


FIG. 2: **Interchannel delay inducing phantom source shift, precedence effect, or echo**

It is common practice to moderate this space-disturbing effect by artificial reverberation or by compensating the delay of the main-microphone signal. However, those techniques are not satisfying, because a pure delay compensation leads to "notching" effects, which are particularly disturbing when the musicians move about near the spot microphone. In order to avoid this negative effect, and to preserve the perception of spatial perspective due to the main-microphone signal, the spot-microphone signal should be delayed much more than necessary for the compensation, so as to fall within the region of the early reflections (“arrival-time gap” [3], [17] of about 15...25 ms, see **TABLE 3** and **FIG. 4**). In a comprehensive study of music halls around the world [18] it was found that in the superior halls the onset of the indirect (impulsive) sound followed the direct sound by about 20 ms. This arrival time gap subconsciously gives the listener the auditory spatial impression, or the sense of the size of the space.

The room-related balancing technique causes that the temporal reflection pattern of direct sound and early reflections, which is given by the main microphone, remains authentic (**FIG. 1,** situation below), and thus the spatial quality of the stereophonic recording remains natural. The spot microphone signal contributes nevertheless to the desired sound balancing effect (increased loudness, transparency, etc).

Moreover, **FIG. 1** shows that not only the spot microphone signal is delayed with respect to the main microphone, but additionally these two with respect to the room microphone. That is necessary for the avoidance of echo effects whenever the distance between main and room microphone is larger than about 10 m (according to approx. 30 ms).

2.2.2 Time-balancing proposal

Experience has shown that the careful layout of delay has an extremely high importance. It is therefore suggested to prepare a detailed delay plan for each recording, including each of the microphones involved in the mix. An example shows **TABLE 3** (again the case “orchestra, main microphone plus spot microphones”): The delay values are referred to the time base t = 0 ms, as indicated in **FIG. 3**.

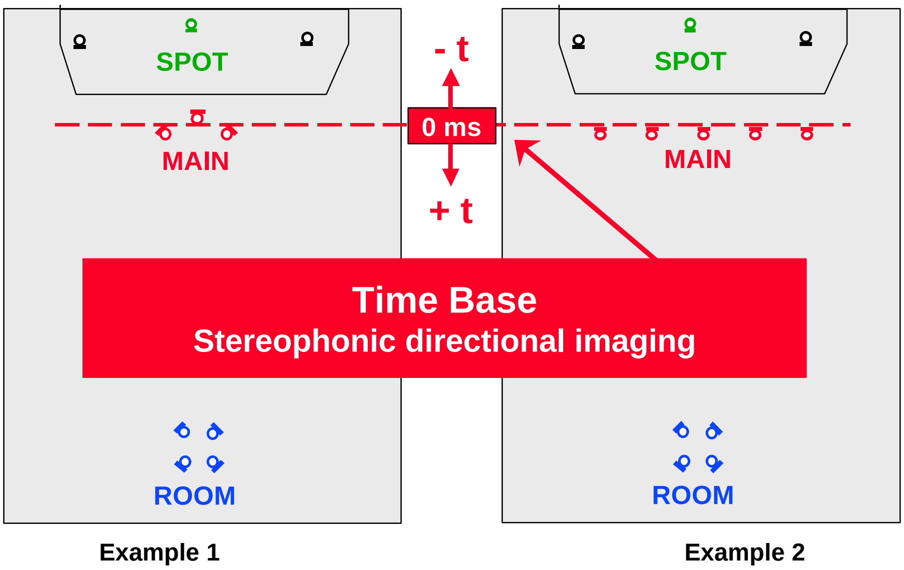


FIG. 3: Room-related time balancing: Setting the time base

A common time base should be set in any microphone configuration. On this basis the delay of each of the microphone can be designed according to the natural pattern (see e.g. **FIG. 1** and **TABLE 3**).

The basis of a delay plan is an exact conception of the pattern of reflections to be reproduced (**FIG. 4)**. It determines the temporal order and the spatial allocation of direct sound and early reflections for the reference point ("sweet spot") of the listening area. (see **FIG. 1,** situation below). It must be decided for each part of the orchestra, which microphone should be "responsible" for the direct sound and used for directional imaging. This determines the reference "zero" at the time axis. In the example used in **FIG. 1** and **TABLE 3** this is the main microphone for the complete orchestra /[3]. All further microphones involved supply either leading or lagging signals (column 2 of **TABLE 3**).

Microphone	Lead / Lag (+ / - ms)	Compensation (ms)		Arrival-time Gap (ms)	Comp. + Gap (ms)	**Resulting Delay (ms)**	Initial Direction
Main L	*Time Base*		*0*	*0*	*0*	*- 35*	-
Main C	*Time Base*		*0*	*0*	*0*	*- 35*	-
Main R	*Time Base*		*0*	*0*	*0*	*- 35*	-
Spot A	*+ 25*	Refl. 1:	*- 25*	*- 12*	*- 37*	*- 72*	*- 30°*
		Refl. 2:	*- 25*	*- 9*	*- 34*	*- 69*	*+ 30°*
		Refl. 3:	*- 25*	*- 17*	*- 42*	*- 77*	*- 110°*
		Refl. 4:	*- 25*	*- 20*	*- 45*	*- 80*	*+ 110°*
Spot B	*+ 35*	Refl. 1:	*- 35*	*- 15*	*- 50*	*- 85*	*- 30°*
		Refl. 2:	*- 35*	*- 18*	*- 53*	*- 88*	*+ 30°*
		Refl. 3:	*- 35*	*- 21*	*- 56*	*- 91*	*- 110°*
		Refl. 4:	*- 35*	*- 18*	*- 53*	*- 88*	*+ 110°*
Spot C	*+ 45*	Refl. 1:	*- 45*	*- 17*	*- 62*	*- 97*	*- 30°*
		Refl. 2:	*- 45*	*- 11*	*- 56*	*- 91*	*+ 30°*
		Refl. 3:	*- 45*	*- 19*	*- 64*	*- 99*	*- 110°*
		Refl. 4:	*- 45*	*- 23*	*- 68*	*- 103*	*+ 110°*
Room L	*- 60*		*+ 60*	*- 25*	*+ 35*	*0*	*- 30°*
Room R	*- 60*		*+ 60*	*- 25*	*+ 35*	*0*	*+ 30°*
Room LS	*- 60*		*+ 60*	*- 27*	*+ 33*	*- 2*	*- 110°*
Room RS	*- 60*		*+ 60*	*- 27*	*+ 33*	*- 2*	*+ 110°*

1 m ↔ 3 ms / 1 ms ↔ 0,33 m

TABLE 3: Proposed delay plan for practical recording application (Example 1)

The delay plan corresponds to the situation shown in **FIG. 1**. In this example three spot microphones A, B, and C are used. From each spot microphone signal (at least) four "early reflections" are derived, each of them having an individual arrival-time gap (column 5) and an individual initial direction (column 8), see also **FIG. 4**. the individual time-arrival gaps are chosen according to the real situation in the concert hall. Column 6 displays totals of the compensating delays and arrival-time gaps. Column 7 finally contains the total delays with consideration of the distance of the room microphone. – As a result, the energy of each spot microphone is distributed in terms of time (column 5) and space (stereo channels, column 8), in accordance with the natural reflection pattern of the concert hall. The delay plan does not contain level adjustments. The sound engineer can vary the level balance within a wide range, without changing the perception of direction and depth.

/3 Note that for time-balancing only the direct sound ("prime sound", determining the directional image) rather than the indirect sound is considered (see **Section 3.4**)

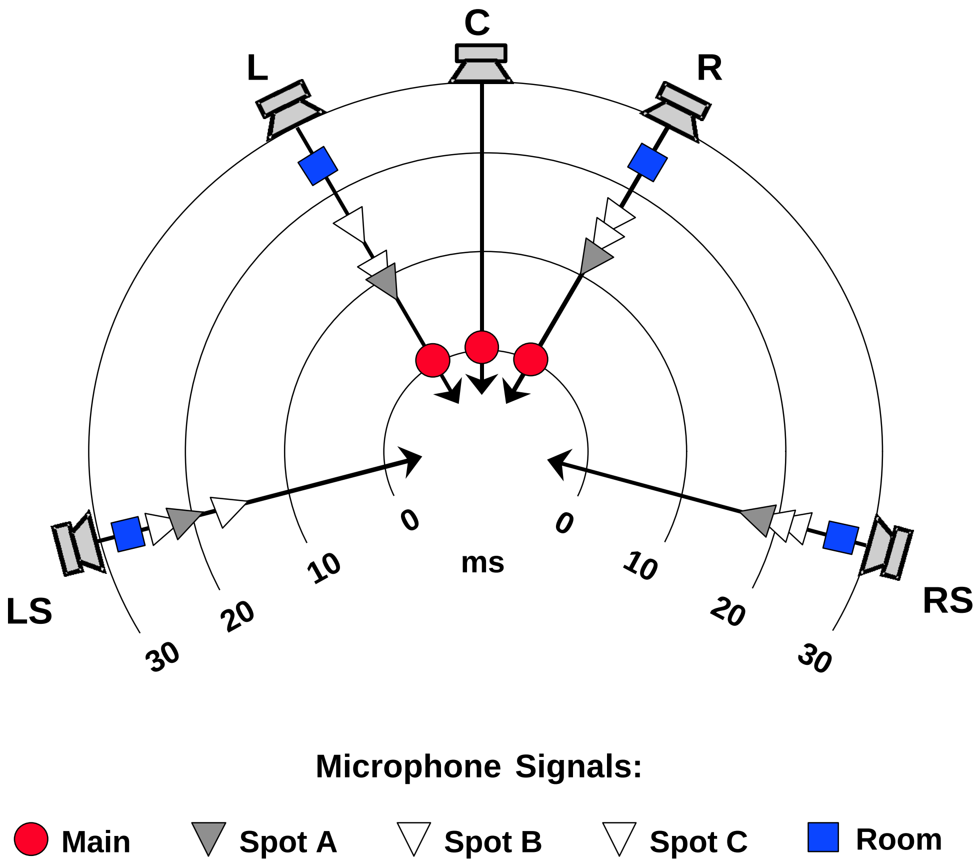


FIG. 4: Delay design according to the natural room response (Example 1)

The “Arrival-time Gap” values plotted in column 5 of **TABLE 3** are displayed here graphically in order to illustrate the intended mixing result (the time scale [ms] defines the arrival time of sound for the listener). The propagation times of the reproduced channels from the speakers to the listener are principally identical at the sweet spot; deviations due to extreme listening positions or asymmetric surround loudspeaker arrangements can be ignored with a certain extent of tolerance (e.g. ± 3 ms or ± 1m). - The time pattern of the indirect sound is designed in order to create the desired spatial impression and perception of depth according to **Chapter 2.2**. It is proposed to allocate the indirect sound derived from the spot and room microphones into the surrounding channels L, LS, RS, R. It is advantageous to generate at least four reflections (the more the better, e.g. 8 or 12) from each spot microphone signal and to use a four-channel room microphone (see **Chapter 3.3**). Reflections in the center channel (median plane of the listener) are unfavourable [5], [6].

Note: In this example the “prime” sound is the direct sound picked up from the main microphone. Generally it is defined as that sound fraction of a source or group of sources which forms the “first wave front” [3] during reproduction.

In principal 3/2-stereo-music recordings enable a convincing reproduction of the spatial impression. Optimum results are attainable even in difficult situations, assuming that each microphone signal (main and spot microphones) can be delayed individually. It is a practicable approach in cases where analogue mixing desks are applied and only a few delay lines are available, to group a number of spot microphones which have approximately the same distance to the main microphone (similar lag, ± 5 ms). These spot groups can be treated as shown in **TABLE 3**, i.e. there are spot groups (A, B, C) instead of single spot microphones. It should be mentioned here that obviously the generation of more than four reflections is useful. Furthermore, the quality of the spatial impression may be improved by generating adequate reverberation, in order to match the reverberance of the supported instrument(s) according to the actual loudness balance.

In principle the room-related balancing algorithm could be implemented into digital mixing desks so that it could be used alternatively to conventional panpot balancing. A corresponding mixing desk has been introduced in [8]. Digital processing allows to realise further optimisation such as distance equalisation (taking into account changes in spectrum, due to absorption effects during sound propagation), additional reflections per spot microphone signal (realistically around 12 ... 24), additional artificial reverberation (generated from the spot microphone signals in line with the artificial reflections), “natural panning” (panning law according to the interaural transfer function of the sphere microphone) [6], [7], [8].

In a further step the early reflections may be synthesised completely by means of signal processing. The main microphone will then become obsolete. Thus the room related balancing concept can also be applied to poly-microphony. Modern mixing consoles could comprise room-related balancing tools for the synthesis of arbitrary natural spatial impressions.

2.3 Aesthetic downward compatibility

An important aspect is downward compatibility of 3/2-stereo music recordings. The downmix equations according to Recommendation ITU-R BS 775-1 are plotted in **TABLE 4**. Although it is possible to determine the downmix coefficient for surround , it is not guaranteed that automatically the resulting two-channel downmix satisfies aesthetically in similar way as the original 3/2-stereo version. Ideally, the downmix should prove just as to sound as an appropriate conventional two-channel recording, which is originally mixed in 2/0-stereo, for example from the same set of microphone signals. It is clear that in practice the downmix will not always perform optimum quality regarding a number of parameters such as reverberance balance, loudness balance, perception of depth.

$$L_0 = L + 0{,}7C + k\,LS$$

$$R_0 = R + 0{,}7C + k\,RS$$

The choice:

k = 1, 0.71, 0.50, 0.36, 0

TABLE 4: Compatibility matrix 3/2 → 2/0

The standard downmix coefficient for the surround is k = 0.7, according to ITU-R BS 775-1. However, it is possible to determine an alternative coefficient at the production side, which can be transmitted via “Ancillary Data” [9] or “Meta-data” [10] in order to ensure an adequate two-channel downmix reproduction for individual program material. The downmix coefficient for the center C is fixed at 0,7, since this value works fairly well for all types of material with respect to directional imaging as well as loudness balance.

As regards the surround information, the downmix result concerning temporal order of direct sound, early reflections and reverberation is correct, see **FIG. 5**. The resulting pattern of reflections corresponds with the desirable pattern for natural two-channel recordings as described in [6]: From this point of view the downmix principally enables perspective imaging of a three-dimensional space in the two-dimensional simulation plane between the two loudspeakers.

It can be stated that the consideration of time relations does principally not differ in both cases, two-channel and multichannel recording. The essential change is related only to the spatial distribution of the indirect sound. Hence careful handling of time-balancing should be considered as an important mixing parameter for natural music recording not only in the case of 3/2-stereo (Section 3.2) but even for two-channel reproduction [6]. This is true also with respect to quality improvements due to “natural panning” suggested in [7], [8].

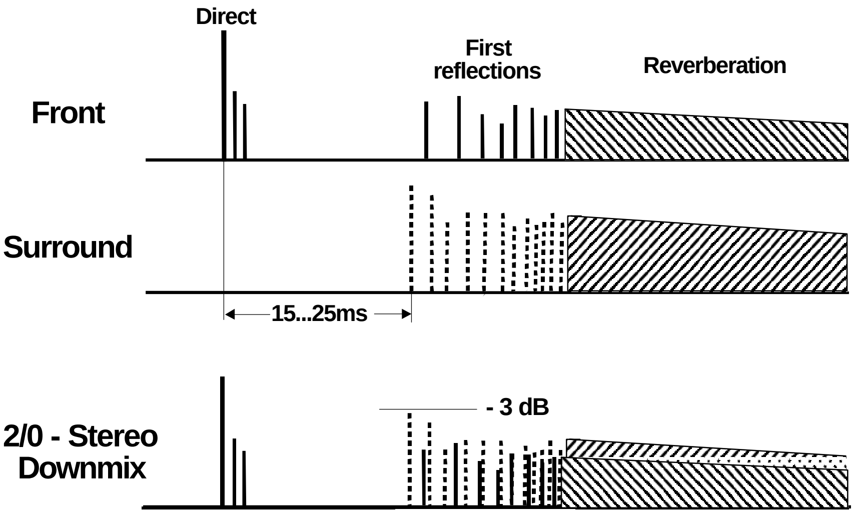


FIG 5:

Pattern of reflections in the 3/2-mix and in the 2/0-downmix

In the 2/0-downmix the room information of the 3/2-stereo mix is completely preserved. Two-channel reproduction allows corresponding spatial impression in the simulation plane. However, optimum stereophonic quality is not ensured in all cases.

On the other hand, it does not seem to be obvious to preserve the originally intended reverberance balance. It is a well-known psychoacoustic effect in binaural hearing that a live room is perceived to be less reverberant than in the case of monaural hearing [3], and a similar phenomenon occurs in the practice of natural music recording when we switch from two-channel stereo to mono or from multichannel to two-channel stereo. This could mean for example that the total energy of the reverberant sound should be smaller in the downmix than in the multichannel presentation, which can be achieved with the surround downmix coefficient k = 0,7 or k = 0,5, depending on the program material.

Different experience is reported in [11], [12]: “For reverberation component energy containing spatial information to be perceived as natural by the audience, the total sound levels in both stereo and surround–recording procedures must be equal.” This would imply k = 1. The reason could be connected with specific density characteristics of the reverberant sound picked up with a particular surround sound microphone concept.

Level-balancing of the indirect sound appears to be a matter of aesthetic feeling rather than recommended practice, particularly with respect to downward compatibility considerations, because perception of auditory perspective and spatial impression is governed by a number of parameters, such as density, temporal and directional distribution, and energy of reflections (the so-called “R/D-ratio [13]).

Considering **headphone reproduction** it must be stated that the simple downmix according to ITU-R BS 775-1 (**TABLE 4**) does not represent the optimal solution. The well-known “in-head localisation” effect [3] is a severe impairment with respect to the perception of space and depth even in comparison with conventional two-channel loudspeaker reproduction. When compared with the real spatial impression achievable by means of a natural 3/2-stereo recording, the lack of aesthetic compatibility appears to be unacceptable. A special “downmix” method for multichannel headphone reproduction is required, which is able to preserve the original three-dimensional spatial impression perceived in a multichannel listening room.

A suitable approach is the application of auralisation concepts in order to achieve virtual loudspeaker reproduction. A corresponding system proposed in [14] is based on binaural data measured in a real multichannel control room. It generates a binaural signal for headphone reproduction, and enables listening in the virtual control room at the sweet spot, thus avoiding any impairment of the spatial impression achievable with loudspeakers.

3 MAIN MICROPHONE

The term “main microphone” is often used in divergent ways, and the weight of characteristic attributes may be different in conventional two-channel or five-channel applications. In principal, the main microphone should combine two basic psychoacoustic functions:

Directional imaging: Picking up the prime sound of a source or group of sources which forms the “first wave front” [3] during reproduction (direct sound).

Spatial imaging /[4]: Picking up the corresponding natural reflections and reverberation (indirect sound).

Realisation of both functions with one stereo (main) microphone appears to be advantageous in the case of conventional two-channel stereophony, provided that suitable recording conditions are given and the correct microphone location is found to ensure the adequate directional image as well as the adequate balance of direct and indirect sound (R/D ratio [13]). For example, under these conditions the so-called sphere microphone [6] has been proven to offer optimum presentation of direction, depth and spatial perspective in the simulation plane between the loudspeakers.

However, in the case of 3/2-stereophony, the psychoacoustic parameters are not identical. Firstly, we must consider frontal directional imaging by means of the three loudspeakers L, C, R, forming the two stereophonic sub-areas [15] L-C and C-R, and thus requiring a suitable three-channel pick-up of the frontal direct sound. Secondly, we must consider that about 50% of the indirect sound energy should be allocated to the surround channels LS and RS, and thus an adequate polar pattern of the microphones is required enabling a sufficient separation of direct and indirect sound.

3.1 L-C-R front channel microphone

It is the aim to achieve imaging characteristics equivalent to those of an optimum two-channel main microphone. Moreover, the stereophonic frontal reproduction is intended to be superior. The first plus point is of course related to the directional stability (enlarged listening area) as shown e.g. in [15]. This is the primary purpose of the center channel. The second advantage concerns sound quality. It is found in a number of studies (e.g. [16], 17]) that the discrete three-channel system is preferable in comparison to the two-channel system on “clarity” “sound colour” of the center image, even when the listener sits precisely on the center line and does not move his head. It is presumed that this preference arises because the center loudspeaker is “easier” to listen to and that the center phantom image principally causes some coloration [17] and requires “greater attention”.

An optimum three-channel main microphone should exploit these principal benefits. The L-C-R stereophonic representation should ensure maximum localisation focus and avoid coloration due to combing effects. When we look at main microphone methods used in practical situations it appears that none of them works perfectly with this respect. So far no method has been used, which is completely sufficient for all request.

/[4] The term “spatial imaging” includes both, imaging of spatial impression as well as imaging of spatial depth. In the concert hall the major cues for perception of depth or distance are delay and level (in relation to the direct sound) of early and late reflections (reflection pattern), and of reverberation. There is only some distance information in the direct sound – namely the relative loudness of different musical sections, and the possible presence of some high frequency rolloff due to air absorption. In a recording the directional cues in the direct sound have even less weight than in natural hearing [5]. So imaging of spatial depth is related to the design of indirect sound rather than direct sound. On the other hand, direct sound provides the relevant localization cues with respect to direction.

3.1.1 The interchannel crosstalk problem

A substantial problem is the "triple phantom source" (see the sketch **FIG. 6**). In all cases where three microphone capsules are used there are results interchannel crosstalk. More or less correlated signals generate three phantom sources whose direction and expansion depend on the resulting level and time differences. It is not possible to find a geometrical arrangement of the microphone capsules which could ensure that the three phantom images are congruent for any source direction. Therefore such a three-channel microphone is in principle characterised by a decrease of the localisation focus and clarity, and by coloration effects.

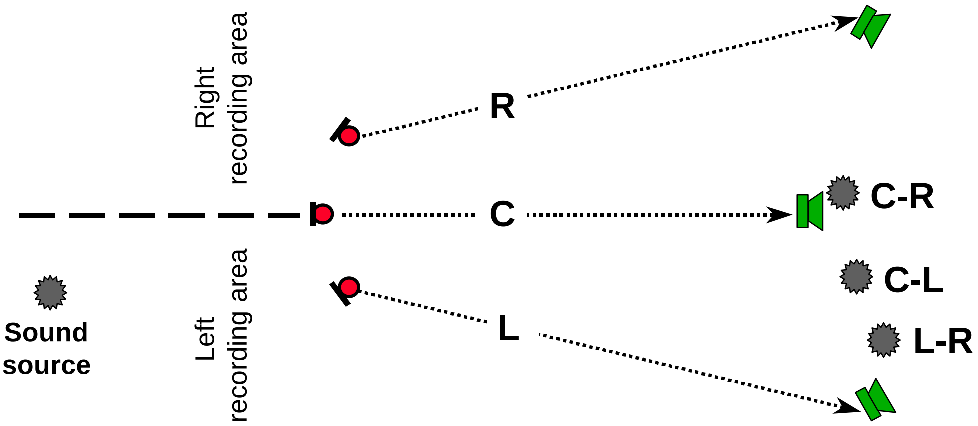


FIG 6:

The "triple phantom source" problem arising with a 3-channel stereophonic microphone

In principal each 2-channel stereophonic basis C-L, C-R, L-R produces its own phantom source, and each of them would be located at divergent places, resulting more or less in a decrease of the localisation focus and clarity, and in coloration effects.

The triple phantom source and associated comb filter effects particularly in the two-channel downmix (see e.g. [19]) should be minimised as perfect as possible. Ideally, a source located in the left recording area must not be picked up by the right capsule, a source located in the right recording area not by the left capsule, and a source located in the center line neither by the left nor by the right capsule. Although of course the channel separation must not exceed about 15 dB, appropriate directional characteristics do not prevent sufficiently from interfering L-C-R-crosstalk. Even hyper- or super-cardioid microphones do not provide enough channel separation in usual arrangements.

3.1.2 Practical approaches

In this passage a number of distinct practical solutions used in standard recording situations or proposed in the literature are described and discussed. The configurations are designed not only for frontal imaging but also to provide additionally the indirect sound for the front channels as frontal portion of the surrounding environment. This has two consequences:

1. Narrow or widely spaced microphone configurations are preferred. It is well-known experience that pure coincidence microphone concepts are not able to produce a satisfying natural spatial impression, due to the lack of adequate interchannel temporal relations (time-of-arrival, phase, correlation, see e.g. [6], [18], [19]).

2. Cardioid or super-cardioid microphones are applied not only to minimise interfering crosstalk but also to attenuate the indirect lateral and rear sound and to ensure a sufficient leeway for allocating a certain portion of the indirect sound energy to the surround channels LS and RS. In [11] it is stated: "An omni-directional microphone has been used as the main microphone for stereo recording in recent years in order to effectively record affluent reverberatory components. However, if a surround microphone were added in such a case, very strong reverberation would be recorded, resulting in an overly emphasised ambience in reproduction."

Cardioid microphones are applied in the triangle configuration as shown in **FIG 7** and proposed in [20] (“*INA 3*”). The geometrical arrangement has been designed under consideration of the “recording angles” /[5]. [21], [22] of the microphone pairs L-C and C-R. The intention is to provide a balanced directional distribution of sources within the left recording area as well as in the right recording area, resulting together in a corresponding complete image across L-C-R. However, as demonstrated later more detailed, the optimisation regarding the attachment of adjacent recording areas does not imply a minimisation regarding artifacts due to interchannel crosstalk. Minimum impairments of localisation focus, clarity, and timbre may not be achievable with this configuration, because the acoustical channel separation is not sufficient.

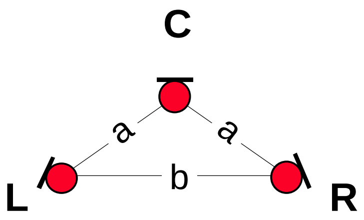


FIG 7: Configuration “*INA 3*” on the basis of [21]

The triangle arrangement has been designed in line with the so-called “Williams-Curves” [21] aiming optimum attachment of the recording areas for L-C and C-R. In [20] the distances a and b are calculated for cardioid capsules dependent on the resulting recording angle φ :
φ = 100°: a = 69 cm b = 126 cm
φ = 120°: a = 53 cm b = 92 cm
φ = 140°: a = 41 cm b = 68 cm
φ = 160°: a = 32 cm b = 49 cm
φ = 180°: a = 25 cm b = 35 cm
The off-center angles of the microphones are always ½ φ

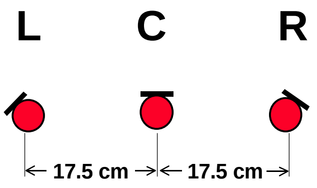


FIG 8: Near-coincident configuration [24]

Three microphones in line. The outside capsules L, R have a super-cardioid polar characteristic (30° off-centre). This avoids producing a strong center phantom image. The center capsule has a cardioid polar characteristic.

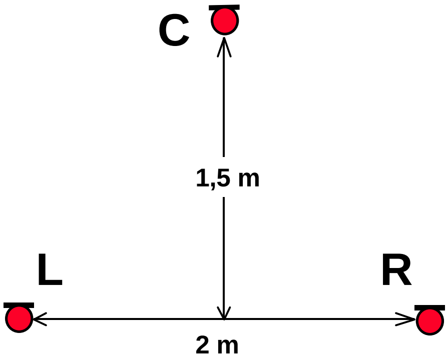


FIG 9: Widely spaced omnis (“*Decca-Tree*”, [19])

Three omni-directional microphones are widely-spaced in a triangle configuration. Due to the wide spacing the configuration is not suitable for accurate stereophonic directional imaging. It is used to produce an open, spacious sound, combined with a solid central image. –
A similar triangle configuration is applied in the Fukada-Tree [12], however, cardioid microphones are used (see **FIG 18**).

/[5] The recording angle (also known as “useful acceptance angle” [23]) indicates that pick-up sector of a stereophonic microphone which results in a balanced directional distribution of sources within the loudspeaker basis. The recording angles according to the “Williams-Curves” [21] of usual two-channel main microphones are for example:

Configuration	Capsules	Off-Center Angle	Spaced	Recording Angle φ
NOS	Cardioid	+/- 45°	30 cm	80°
RAI	Cardioid	+/- 50°	21 cm	90°
ORTF	Cardioid	+/- 55°	17 cm	95°
DIN	Cardioid	+/- 45°	20 cm	100°
A / B	Omni	0°	50 cm	100°
A / B	Omni	0°	40 cm	150°

A near-coincident in-line configuration shown in **FIG 8** provides enhanced channel separation, at least between L and R, because super-cardioid microphones are applied here. On the other hand, a source close at the center line will be picked up by both stereophonic pairs, L-C and C-R, resulting in a double phantom source, one half left and one half right and in correspondingly decreased localisation focus and clarity. Another concern is related to the recording angle φ: Compared with an ORTF pair, the recording angle of each stereophonic microphone pair is wider because the off-center angle is much smaller and a more directional polar characteristic is used at one site. Therefore, in contrast to the triangle arrangement according to **FIG 7**, the central sector of overlapping recording areas is unnecessarily vast (in the range of at least 60°). Furthermore, a cardioid microphone at one site and a super-cardioid microphone at the other causes slight unsymmetrical directional distribution of sources (see later **FIG 11B**).

As mentioned earlier, it is the primary purpose of the center channel to ensure enhanced directional stability. Moreover, it is the aim to achieve directional imaging characteristics equivalent to those of an optimum two-channel main microphone at the same time. Thus the triangle configuration with a certain resulting recording angle φ should theoretically produce the same directional image as a two channel configuration with identical recording angle φ.

For a more detailed consideration the so-called localisation curves are useful. **FIG 10** shows the typical localisation curve of a usual narrow-spaced two-channel stereophonic microphone.

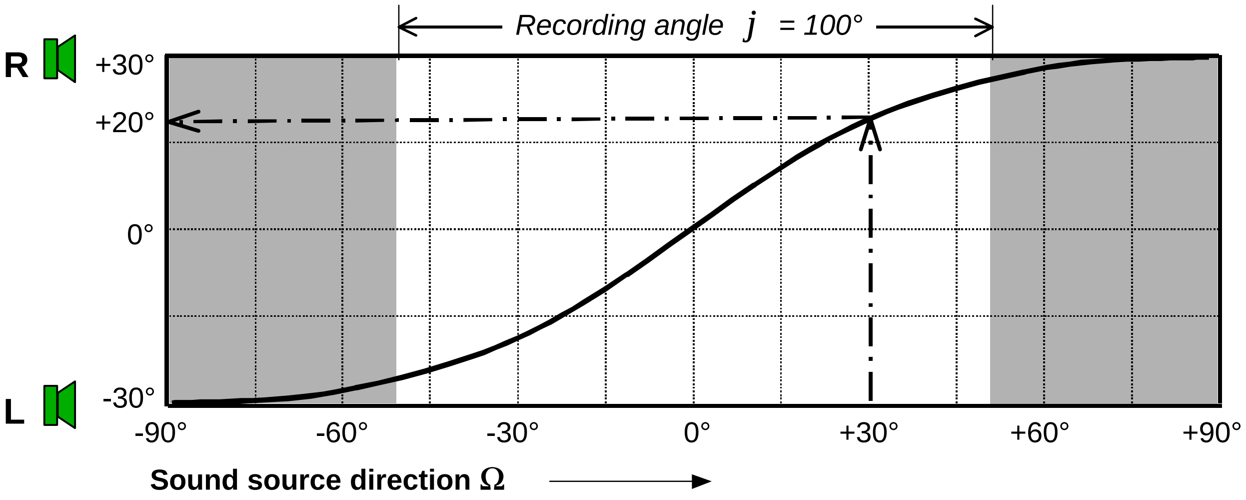


FIG 10: Localisation curve of a narrow-spaced two-channel microphone

The curve displays the directional translation of the stereophonic microphone. For example, a sound source located 30° off-center right of the microphone will be perceived approximately 20° off-center right in the standard two-channel loudspeaker arrangement, due to the channel signal difference delivered from the microphone. The useful recording angle of this microphone is ± 50°. Within this sector the curve ensures a well-balanced directional image.

It is presumed here that the shape of this curve as well as the recording angle is desirable also for a corresponding three-channel L-C-R microphone (reference curve). Three examples are outlined in **FIG 11**. The first (**A**) corresponds with the “*INA 3*” configuration (**FIG 7**), the second (**B**) with the near-coincident microphone according to **FIG 8**, and the third one (**C**) indicates directional characteristics of the “*Decca-Tree*” (**FIG 9**). In all cases the localisation curves of each pair, R-C, L-C, L-R are plotted. We can see that interchannel crosstalk can be more or less problematic. In each configuration the curve L-R is unwanted, as well as the curve L-C in the right sector and the curve R-C in the left sector. However, there are individual differences with respect to level and delay.

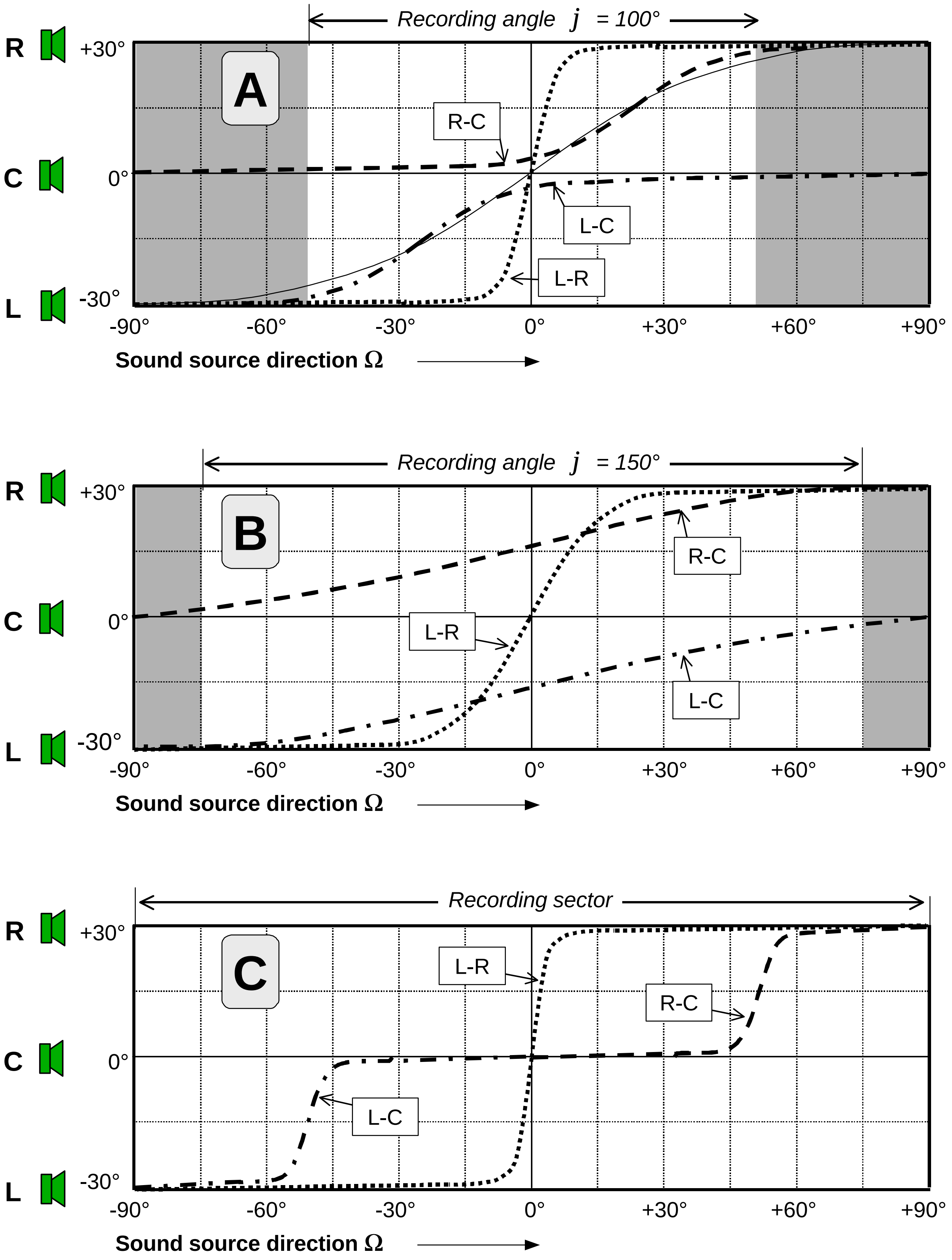


FIG 11: **Localisation curves and recording angles of three-channel microphones**

A: *INA 3,* φ = 100° (**FIG 7**) Localisation curve L-R, Ω = 0°: level -3 dB, delay 1 ms

B: Near-coincident (**FIG 8**) Localisation curve L-R, Ω = 0°: level -1 dB, delay 0 ms

C: *Decca-Tree* (**FIG 9**) Localisation curve L-R, Ω = 0°: level 0 dB, delay 4,5 ms

Considering the *INA 3* configuration (**A**), the pairs L-C and R-C provide the desired recording angle and the desired localisation curve – however either in the left sector or in the right sector. Unfortunately the two other curves are divergent (except for the center area). The related interference effect cannot be neglected, because channel separation is less than 6 dB, and the delay of unwanted acoustical crosstalk is in the range 1...2 ms. In particular, the impact of the curve L-R is considerable, since the associated level is only 3 dB lower than that of the desired phantom sound source L-C or R-C. The delay is up to three times smaller in cases where *INA 3* is configured for broader recording angles.

This situation is not better for the near-coincident in-line configuration (**FIG 8** and **11B**). Channel separation is in the range 1...8 dB, and the interchannel time differences are less than 1 ms. In the central recording sector (about ±30°) the L-R curve is as dominant as the two sub-area curves L-C and R-C. Due to the extreme narrow spacing and the use of super-cardioids for L and R the recording angle is very broad, which may result in a close image around the center in usual main microphone applications.

Since many years another concept has been used successfully. The well-known "***Decca-Tree***" (e.g. [19]) is a triangle configuration similar to **FIG 7**, however, omni-directional microphones are applied and widely spaced (**FIG 9**). This has an advantage, compared with narrow spacing: The delay of acoustical channel crosstalk is in the range 3...5 ms, and the precedence effect is effective: Thus the interfering acoustical crosstalk does not essentially affect the localisation of phantom sources.

On the other hand, the disadvantages of widely spaced (A/B) microphones with respect to directional imaging are well-known. There is no suitable localisation curve which could ensure a balanced distribution of sources between the loudspeakers. In **FIG 11C** we must not consider the curve L-R, because the L-R information arrives 3...5 ms later and is therefore irrelevant regarding localisation. Only the curves L-C and R-C are relevant. They demonstrate that the precedence effect (see **Fig 2**) is effective, and that therefore all sources in the recording sector ±45° are reproduced in the center or very close to it. Sources outside the sector ±60° are reproduced in L or R.

The center microphone in this configuration is certainly an improvement of widely spaced (A/B) microphones, since the "hole in the middle" is filled with solid and clean center information. The spacing provides sufficient time information to produce a dense and "open" sound picture. Comb filter effects, which could arise during two-channel reproduction when the center signal is mixed into L and R, are suppressed because of the spatial separation of L, C and R during reproduction.

As mentioned earlier, in most 3/2-stereo recording situations it is advantageous to use uni-directional microphones to reduce the energy of indirect sound and to provide headroom for allocating the indirect sound energy to the surround channels. For this reason it seems to be useful here to replace the omnis of the *Decca-Tree* by cardioids, each of them facing the front (off-center angles = 0°). This does not change the directional characteristics of the tree, but the indirect sound level is theoretically 4,8 dB lower (hyper-cardioid: 5,7 dB). A similar cardioid triangle configuration is applied in the *Fukada-Tree* reported in [12], see **CHAPTER 3.2**.

In some situations it may be suitable to move the microphone towards the soundstage in order to control the R/D-ratio for the front channels by listening - not only to the complete 3/2-stereo mix inclusive surrounds but also to the two-channel down-mix. By placing the microphones high above the front of the soundstage, the differential distance between them and the front and back of the stage will be minimised. This helps reduce acoustical imbalance between the nearer and more distant elements of the orchestra. - This is possible for any widely-spaced configuration, since stereophonic directional imaging according to localisation curves does not happen anyway.

Another widely-spaced configuration is shown in **FIG 12**. Five microphones are distributed in line across the stage width, the distance between neighbouring microphones is in the range of 2 m or more. Two effects are intended: Firstly, as described above, the exploitation of the precedence effect to reduce the multiple phantom source problem. Secondly, the provision of a "stable" phantom sound source half left between L and C and half right between C and R. As a result, five clearly localisable sources are available for the directional representation of the orchestra. Of course this is again a compromise, however, there results a rather stable and balanced stereophonic image, combined with the typical characteristics of widely-spaced microphones with regard to spatial impression.

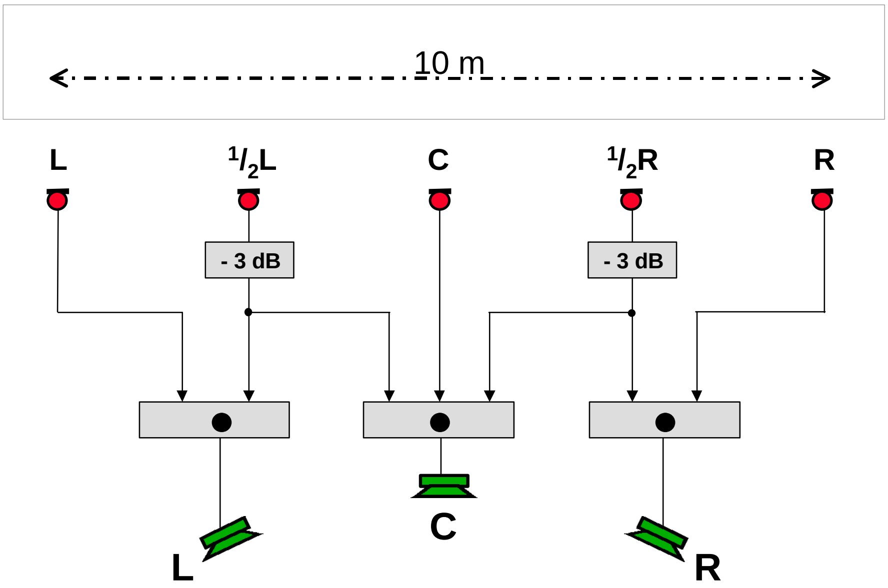


FIG 12: Five microphones in line and widely spaced

The microphones are distributed across the stage width, providing five negligibly correlated signals to produce three stable sources plus two phantom sources for directional imaging. The R/D-ratio and the balance of the orchestra elements can be controlled in certain limits by microphone positioning. Cardioid microphones may be used in order to reduce the indirect sound energy in the front channels.

Obviously this configuration can be useful only for large orchestra situations. It is the wrong way to reduce the microphone distances according to a smaller dimension of the orchestra (e.g. chamber music). In these cases an alternative widely-spaced configuration could be for example the *Decca-Tree*.

An interesting approach for large orchestra situations is known from [25], see **FIG 13**. Two usual two-channel main microphones are widely spaced. Each of both is used in the usual way to pick up the left or right part of the orchestra. The directional shifts of phantom sources due to the attenuation in the center channel should be compensated, for example by means of corresponding delay. In practice it might be sufficient to provide a compensating orientation of the two main microphones axis.

A critical point could still be foreseen in the overlapping area of the two recording sectors. An instrument in the middle of the stage will be picked up equally from both main microphones. It is reported however in [25] that neither a decrease of localisation focus nor coloration (comb filter effect) has been observed. More practical experience with this method should verify this result. Positive factors are:

- The large distance between the two main microphones
- No considerable correlation between L and R (no disturbing L-R localisation curve)

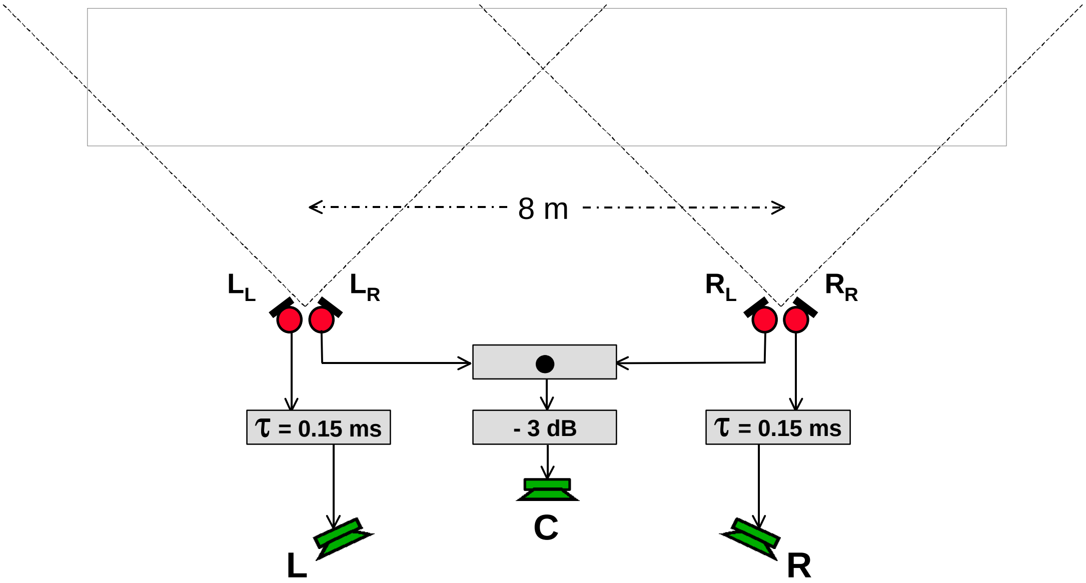


FIG 13: Separate 2-channel main microphones [25]

The two-channel main microphones are widely spaced. Each of both is used to pick up the left or right half of the orchestra in the usual way. It is not necessary that the main microphones are located in line, rather, they can be positioned and adjusted individually according to the recording situation given in the left and right recording area. It is important to avoid overlapping recording angles as perfect as possible.

The most important aspect would be: In recording situations where a main microphone is preferred there is no problem to do it in the same way as for two-channel stereophony. Instead of one stereophonic area there are now two. Spot microphones in the left stage area are added to the left main microphone, and spot microphones in the right stage area are supporting the right stereophonic image produced by the right main microphone. The method would offer two significant benefits:

1. Current two-channel main microphone methods and existing experience could be applied accordingly. The performance of two-channel main microphones (see e.g. **Fig 10**) is available without compromises such as illustrated in **Fig 11**.
2. The location and the recording angle of each two-channel main microphone could be individually optimised according to the situation in the left and right recording area.

It is clear however that the “double-main” method could perhaps satisfy only in a limited range of applications and situations. For example, it seems to be disadvantageous in many occasions to use a pair of sphere microphones, because the indirect sound portion is too high when additionally the surround channels are used for spatial imaging. Furthermore, how to record for instance a string quartet or a solo piano? At least here we need a real three-channel main microphone concept to achieve directional stability from the center loudspeaker and – at the same time – ensure optimum stereophonic quality. A possible approach is introduced in the following chapter.

3.1.3 L-C-R main microphone proposal

There are a number of psychoacoustic principals and basic requirements to be considered carefully if we try to optimise the main microphone configurations shown in **FIGS. 7, 8, 9.** The approach proposed here is basically a conventional narrow-spaced three channel microphone configuration. It is not based on the application of microphone array technology, because further research and development is necessary to achieve suitable directional characteristics, satisfying free-field and diffuse-field frequency response, and optimum audio signal quality at the same time.

The principal configuration is shown in **FIG 14.** It is optimised regarding the interchannel crosstalk and regarding the resulting recording angle. The crosstalk problem is reduced satisfyingly by applying super-cardioid microphones for L and R and facing them towards the sides (90° off-center). Of course these capsules are used quite unfavourable with respect to the frequency response, however, this drawback can be overcome by using capsules with minimum diameter and by equalising the freefield response of L and R for about 60° instead of 0° microphone axis. The center microphone is a cardioid.

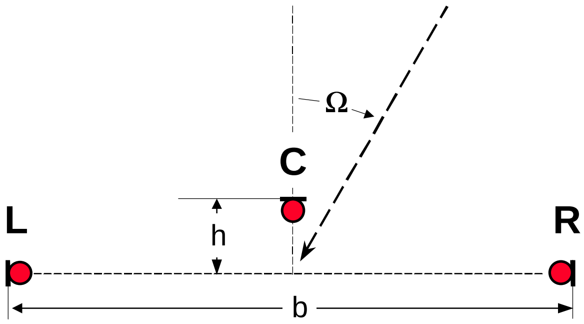


FIG 14: Optimised triangle configuration

The microphone characteristics of capsules L and R are super-cardioids. They are faced sidewards (90° off-center), in order to ensure maximum channel separation. The freefield equalisation of capsules L and R is preferably based on Ω = 30°. The center microphone is a usual cardioid.

Distance b depends on the recording angle φ:.
φ = 90° b = 80 cm
φ = 100° b = 70 cm
φ = 110° b = 60 cm
Distance h = 8 cm. - If cardioids are used instead of super-cardioids, it is h = 12 cm.

In the case of frontal sound directions (Ω ≈ 0°) the following interchannel level / time differences result:

Ω	L	C	R	Δt
0°	-9 dB	0 dB	-9 dB	0,24 ms

The time difference Δt between L and C or R and C is calculated according to the geometrical dimensions /[6]. The determination of distance b depends on the desired recording angle φ. Distance h determines the time difference Δt. In the case Ω ≈ 0° it is h = 8 cm, which implies Δt = 0,24 ms. As a result we get the level difference ΔL = 9 dB and the time differences Δt = 0,24 ms between L and C or R and C, both shifting the phantom sound source into the center (about 10° due to ΔL and additionally about 5° due to Δt).

It has been shown in former papers [15], [29] that the phantom source perceived in the middle between two loudspeakers can be shifted in certain limits according to the rules as presented in **TABLE 5.** There are two important facts. Firstly, if the phantom source is shifted due to ΔL and additionally due to Δt (same direction) the resulting shift is approximately the sum of both single shifts, it is

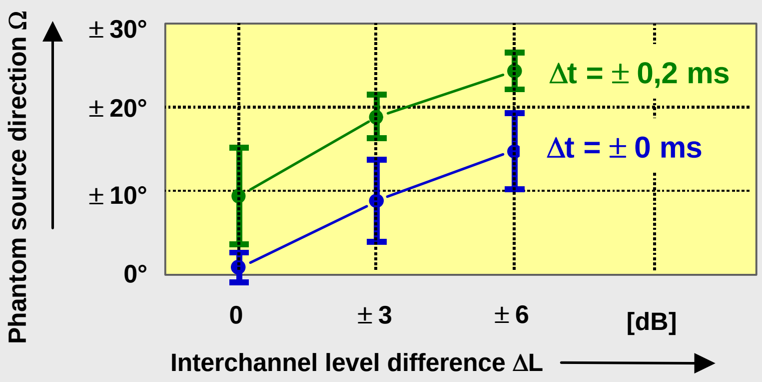
$$\vartheta\,(\Delta L, \Delta t\,) \;=\; \vartheta\,(\Delta L\,) \;+\; \vartheta\,(\Delta t\,).$$

Secondly, the phantom source shift for a certain interchannel level and/or time difference has a constant relation in the loudspeaker basis. For example, if the interchannel level difference is 6 dB, the shift is 45 % , or $\Delta\vartheta_i = 13{,}2°$ in the conventional 60° loudspeaker arrangement. A decrease of the stereo base angle from Ψ = 60° to Ψ = 30° will correspondingly halve the shift: it is $\Delta\vartheta_i = 6{,}6°$. (By the way, the "constant relative shift phenomenon" of the phantom source is contradictory to summing localisation theories, however, it is explained by the Association Model, see e.g. [15], [29]).

We understand now the principal function: If a stereo base angle Ψ = 60° is divided into halves, the same is necessary for the recording angle, and the capsule distance of the narrow spaced microphone has to be increased accordingly. In other words: compared with a two-channel narrow spaced microphone, the distance between capsules L and R is about four times larger when a center capsule C is introduced between L and R, and when the same recording angle and similar perception of time cues (spatial information, "open" sound picture) is desired.

/[6] The equation is:

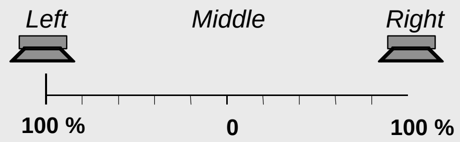
$$\Delta t \;=\; \left[\sqrt{h^2 + (\tfrac{1}{2}b)^2}\;\cdot\;\cos\left(90° - \Omega + \arctan\frac{2\,h}{b}\right)\right]\cdot 0{,}03\,\frac{\text{ms}}{\text{cm}}$$



	Two-channel stereo Loudspeakers L – R Stereo base angle Ψ = 60°	Three-channel stereo Loudspeakers L – C – R Each stereo base angle Ψ = 30°
Intensity shift factor Z_i = 7,5 % / dB	$\Delta\vartheta_i$ = 2,2° / dB	$\Delta\vartheta_i$ = 1,1° / dB
Time shift factor Z_t = 15 % / 0,1 ms	$\Delta\vartheta_t$ = 4,4° / 0,1 ms	$\Delta\vartheta_t$ = 2,2° / 0,1 ms
Effect of intensity plus time shift	$\Delta\vartheta_\Sigma = \Delta\vartheta_i + \Delta\vartheta_t$	

$\Delta\vartheta_i$, $\Delta\vartheta_t$: Shift angle of the phantom source due to the interchannel level or time difference

Shift factor Z :



Z_i = 7,5 % / dB

Z_t = 15 % / 0,1 ms

TABLE 5: Intensity and time shift factor

Based on this knowledge, the perceived source angles ϑ as well as the interchannel crosstalk has been calculated /[7], see **TABLE 6.** It is assumed here that the source is located right side (according to **FIG 14**). In this case no sound should be picked up from the left microphone. We can see that the crosstalk situation is acceptable in the case of the super-cardioid configuration. However, when cardioids are applied for L and R instead of super-cardioids the crosstalk is not neglectable. With this respect the cardioid configuration has drawbacks. As regards the directional translation both configurations demonstrate similar characteristics as plotted in **FIG 10** (reference localisation curve). The recording angle will be approximately φ = 100°.

	SUPER-CADIOID h = 8 cm / b = 70 cm		CARDIOID h = 12 cm / b = 70 cm	
Source angle Ω	Crosstalk →L	Perceived angle ϑ	Crosstalk →L	Perceived angle ϑ
+ 90°	-11 dB	(30°)	<-18 dB	(30°)
+ 60°	<-18 dB	(30°)	<-18 dB	(30°)
+ 45°	<-18 dB	27°	-15 dB	26°
+ 30°	<-18 dB	19°	-11 dB	17°
+ 15°	-14 dB	9°	-8 dB	8°
0°	-9 dB	0°	-5 dB	0°

TABLE 6: Imaging characteristics of two optimised triangle configurations (right sector)

/[7] Directivity pattern of unidirectional microphones used for calculations:

Direction	0°	15°	30°	45°	60°	75°	90°	105°	120°	135°	150°	180°
Cardioid [dB]	0	-0,3	-1	-2	-3	-4	-6	-8	-11	-15	< -18	< -18
Super-Cardioid [dB]	0	-0,3	-1	-2	-3,5	-5	-9	-14	< -18	< -18	< -18	-11
Hyper-Cardioid [dB]	0	-0,3	-1	-2	-4	-7	-12	< -18	< -18	-12	-9	-6

A further optimisation is related to the bass response of the main microphone. It is well-known that the cardioids (in particular super- or hyper-cardioids) principally have more or less bass weakness (see e.g. [34]). A main microphone based on sound pressure capsules (spaced omnis, sphere microphone) is superior with this respect. For this reason it is proposed to combine this advantage of pressure capsules and of the super-cardioid capsules. It seems to be attractive to develop a suitable hybrid (two-way) microphone (known as "Double Transducer Microphone"), offering pure omni characteristics below 100 Hz and super-cardioid characteristics above 100 Hz (see also [35]). Applied in the three-channel main microphone for L and R, above 100 Hz there results directional imaging performance as described above. However, below 100 Hz the center microphone C will not contribute, which means that here a usual 70 cm spaced AB microphone is effective.

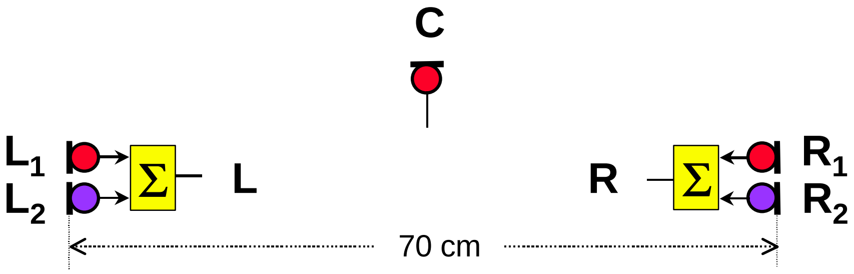


Fig 15: Super-cardioid configuration with improved bass response
Capsules L_1 and L_2 as well as R_1 and R_2 are forming two-way microphones L and R
Capsules L_1 and R_1 : Super-Cardioid, minimum diameter, 100 Hz highpass filtered
Capsules L_2 and R_2 : Omni, 100 Hz lowpass filtered
Capsule C: : Cardioid, 100 Hz highpass filtered

Summarising, it appears to be possible to achieve directional stability from the additional center channel and – at the same time – ensure optimum stereophonic quality in terms of directional translation, localisation focus, spatial imaging, clarity, and last not least sound colour and bass response.

Of course, up to now the proposed three-channel main microphone is an approach designed more or less on theoretical basis, considering psychoacoustic principals as well as practical requirements. It is hoped that sound engineers are interested in testing the tool, and that they will get a chance to do so. Careful listening tests and experienced ears in many practical situations are necessary now.

Fig 16 illustrates an example for realisation. Variable length of the crossbeam between L and R seems to be beneficial, because the recording angle φ can easily be adapted to the actual recording situation.. This offers more flexibility in directional balancing (for example to modify the distribution of orchestra elements to a certain degree).

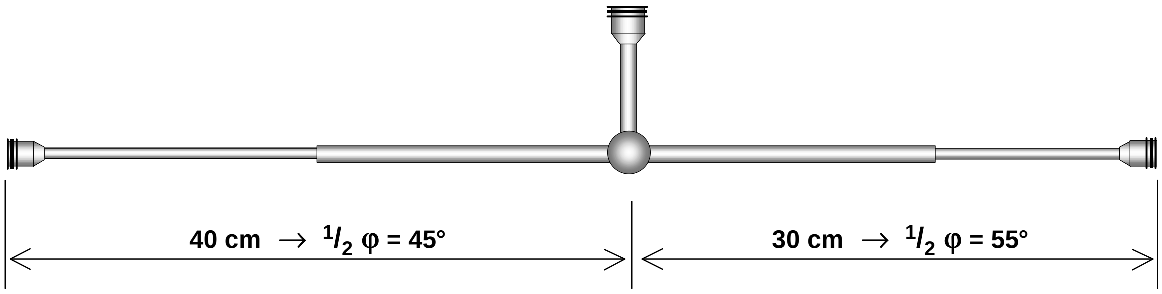


Fig 16: Individual adjustment of the recording angle for each side
The distance between L and C or R and C can be adjusted individually according to the desired recording angle for the left and for the right part of the soundstage. Remote control from the control room is desirable and would be feasible.

3.2 3/2-stereo main microphone

In conventional two-channel stereophony the main microphone should combine the two basic functions directional imaging and spatial imaging. A number of practical solutions are well-known. In the case of multichannel stereophony the realisation appears to be particularly difficult. The microphone has to deliver the complete five-channel mix and to satisfy with respect to many parameters. We know already from **CHAPTER 3.1** that the adequate exploitation of three front channels is a difficult task. The psychoacoustic principals governing localisation of phantom sources leave not much room to pick up the prime sound correctly and to get a corresponding L-C-R stereophonic directional information.

A 3/2-stereo main microphone requires additionally consideration of psychoacoustic laws governing the perception of spatial depth and auditory spaciousness in the concert hall, as described in **CHAPTER 2.** It is the extra task to pick up the indirect sound correctly and to generate a three-dimensional stereophonic representation. Thus suitable time information (see **CHAPTER 2.2.2**) as well as appropriate balance of direct / indirect sound (R/D-ratio) must be recorded by applying the adequate 3/2-stereo microphone at the right location. It is provided here of course that suitable recording conditions for main microphone applications are given. However, the scope of applications is suggested to be even more limited in the case of 3/2-stereo than in the case of two-channel stereo.

What is the purpose of a 3/2-stereo main microphone? Why not using the concept “L-C-R stereo main microphone” plus “room microphone”? Is there any fundamental psychoacoustical reason or just a practical advantage, e.g. easy handling of a compact microphone tree? Initially two solutions are discussed below which are based on the L-C-R configurations presented in **CHAPTER 3.1.2.** Further interesting approaches are introduced for example in [24], [26], [27].

The first example is shown in **FIG 17**. The 3/2-version of the “***INA***” concept is configured in order to provide a recording angle of φ = 360°, to cover the complete surround recording area. The microphone distances are calculated again according to [21], aiming an attachment of five recording subareas:

Sub-area	L - C	C - R	R - RS	RS - LS	LS - L
Recording sector	- 90° - 0°	0° - +90°	+90° - +150°	+150° - -150°	-150° - -90°

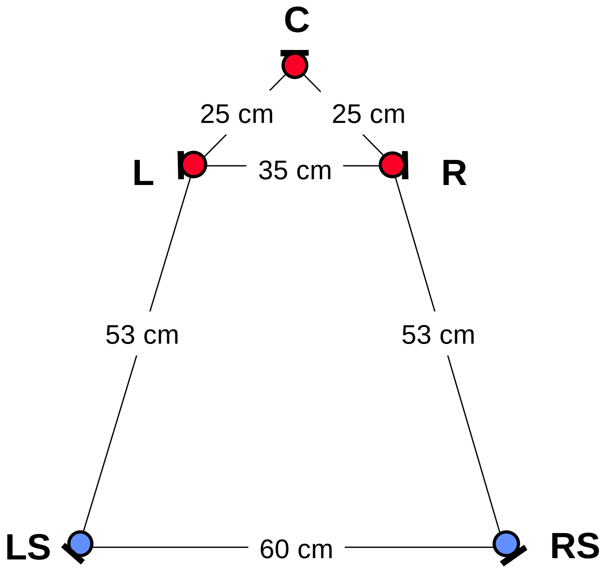


FIG 17: Configuration “*INA 5*” [20]

The L-C-R triangle arrangement “*INA 3*” (see **FIG. 7**, recording angle φ = 180°) is supplemented by the surround microphones LS and RS. The distances are calculated again according to [21], aiming optimum attachment of three additional recording areas (left, right, in the rear), each covering φ = 60°.

However, there are a number of concerns. As regards the psychoacoustic view we must consider principal characteristics of phantom sources. It has been demonstrated in theory [29] and shown in a number of experiments (e.g. [36], [37], [38]) that lateral phantom sources are extremely unstable and sensitive regarding signal spectrum and listener´s location. The intensity and time shift factors measurable for phantom sources in front or behind the listener (see **TABLE 5**) and accordingly the “Williams Curves” [21] are not applicable here. With the exception of moving source effects lateral directional imaging is not possible and correspondingly not an intended application of the 3/2-stereo format (see also **TABLE 1**). For this reason the psychoacoustic basis for the determination of the microphone distances (53 cm – 60 cm – 53 cm) is incorrect. A suitable basis would be for instance the perception of interchannel correlation, because this is a relevant reproduction parameter for the attribute “spatial impression” [3], [5].

Another aspect is related to the recording sector in the front, covered by the triangle arrangement L-C-R. Besides the interchannel crosstalk problem discussed in detail already in **CHAPTER 3.1.1**, there is an additional problem regarding the recording angle. The microphone distances L-C and R-C are small in order to get the recording angle $\varphi = 180°$. This choice may be correct with respect to the intended 360° surround imaging, however, it is abnormal and not useful for main microphone applications. Particularly in this case, where the surround microphones have similar distance to the orchestra, there arises a conflict: Due to the broad front recording angle *INA 5* must be located very close at the orchestra, maybe above the conductor´s head, in order to ensure the adequate distribution of instruments across L-C-R. Obviously this location is unsuitable for the surround microphones LS and RS. On the other hand, if *INA 5* is located optimally with respect to the surrounds, the orchestra will be perceived more or less concentrated around the center loudspeaker (“center effect”).

Obviously a flexible and independent control of the recording angle at the one hand and direct / indirect sound level balance on the other is desirable. This is not possible with fixed 3/2 stereo main microphone configurations. Natural music recording should aim to create a convincing spatial perspective and therefore requires careful design of the R/D-ratio [13], as well as the corresponding layout of the direct sound. In the case of *INA 5* the wide recording angle $\varphi = 180°$ limits considerably the possibilities of microphone placement and the optimisation of indirect sound and direct sound at the same time.

Therefore the main microphone should provide a suitable compromise: It is suggested to provide a relatively narrow recording angle ($\varphi = 90°...100°$) in order to allow the microphone placement in the area of “critical distance” (where the levels of direct and reverberant sound are approximately equal) in the majority of recording situations. Adjustments of the configuration (directional pattern of capsules, geometrical dimensions) should be possible to enable both, optimisation of directional image and of R/D-ratio.

An interesting 3/2-stereo main microphone approach is the “***Fukuda-Tree***” [12] (**FIG 18**). The soundstage imaging triangle L-C-R is a modified version of the *Decca Tree*, (see **FIG 9**), where the omnis are replaced by cardioids to reduce the energy of indirect sound in the front channels. The imaging characteristics of widely spaced front channels are already discussed in **CHAPTER 3.1.2**. Because of the wide microphone spacing a recording angle does not exist. Due to the precedence effect (see **FIG 2**) all sources in the recording sector around ±30° are reproduced in the center or very close to it. Sources outside the sector ±50° are reproduced in or close to loudspeakers L or R.

The widely spaced triangle configurations are not suitable for accurate stereophonic directional imaging. Rather, they are used to produce an open, spacious sound, and due to the center microphone, to maintain a solid central image [19]. They are usually placed a few meters above and behind the conductor, and the distance adjustment is done “without rules”, just by listening to the results [39].

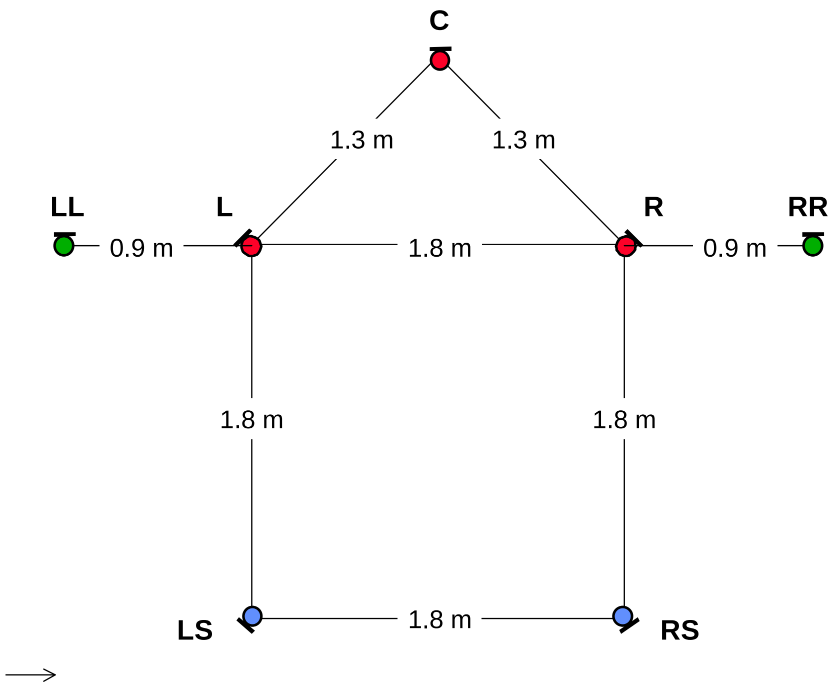


FIG 18: “*Fukuda-Tree*” [12]

The triangle configuration is a modified *“Decca Tree”*, The Microphones L, C, R, RS, LS are cardioids. Additional omni-directional flanking microphones on the sides (LL, RR) are used “to present a sense of the orchestra width and to smooth the sound connection between the front and the rear” [12]. Spatial imaging is realised by means of the 1,8 m spaced square L, R, RS, LS. - It is also stated in [12]:
“The configuration of the tree can vary depending on the hall´s acoustic characteristics, while the microphone intervals may be changed conforming to the orchestra´s size and formation.”

3.3 Proposed separation of L-C-R main and room microphone

As regards recording of the indirect sound, both configurations *INA 5* and *Fukuda Tree* are using supplementing cardioid microphones LS and RS facing backwards left and right. However, the front microphone configurations are primarily designed with regard to stereophonic imaging of direct sound, and thus are not optimised with regard to spatial imaging at the same time. Ideally, suitable balance of direct / indirect sound (R/D-ratio) and proper density and spectrum of reverberant sound in the front channels would require specific microphone characteristics, arrangements and placements. For example, concerning R/D ratio, **TABLE 7** shows that the main microphone configurations discussed above provide quite different reverberant sound levels in relation to the direct front sound incidence Ω = 0° (“effective directivity indexes”).

		L	C	R	LS	RS
INA 5	**FIG 17**	+ 1 dB	- 5 dB	+ 1 dB	+ 6 dB	+ 6 dB
Fukuda Tree	**FIG 18**	- 3 dB	- 5 dB	- 3 dB	+ 10 dB	+ 10 dB
Decca Tree, standard	**FIG 9**	0 dB	0 dB	0 dB	-	-
Decca Tree, cardioids	**FIG 9**	- 5 dB	- 5 dB	- 5 dB	-	-
Near-coincident	**FIG 8**	- 5 dB	- 5 dB	- 5 dB	-	-
Optimised triangle	**FIG 14**	+ 1 dB	- 5 dB	+ 1 dB	-	-

TABLE 7: R/D-ratios of the main microphone capsules resulting at the critical distance

It should be mentioned that the adequate R/D-ratios of the front channels picked up in a certain recording situation can only be set by placing correspondingly the main microphone placement. Suitable level settings are possible at most in the surround channels. However, if we assume that the front / back balance of reverberant sound should approximately equal the natural situation (for a certain spatial perspective), we should provide approximately equal level settings in the four channels L, R, LS, RS. Therefore, as well-known from two-channel main microphones, there is only one parameter to form the stereophonic perspective: that is the adequate microphone placement.

Level balancing of the indirect sound exclusively in the two separate surround channels is understood to be not the right measure for natural stereophonic imaging of a spatial perspective. Rather, natural recording of surrounding indirect and enveloping sound requires (at least) four equivalent channels, two for the front and two for the back, each pair L-R, R-RS, RS-LS, LS-L providing a stereophonic representation of early reflections and reverberation. A corresponding square arrangement of microphones is shown in **FIG 19**.

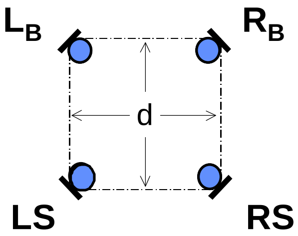


FIG 19: Proposed square arrangement for spatial imaging /[8]

Principally at least four equivalent stereophonic channels are desirable for realistic imaging of spatial impression and enveloping atmosphere. The microphone square should be spaced in order to provide suitable interchannel correlation. Usually the placement of the square should be far from the critical distance in order to provide a high R/D-ratio, adequate density and spectrum. Placement and choice of microphone characteristics will depend on the intended stereophonic perspective and actual recording situation.

The four microphones should be spaced by taking account of psychoacoustic principals regarding the perception of enveloping diffuse sound. In the case of reverberation, the reflected sound can be thought of as being generated by mirror-image sound sources [3]. The resulting sound field can be called “subjectively diffuse” if a well-balanced distribution of auditory events is perceived around the listener. In [41] the perception of subjectively diffuse sound fields has been investigated, in particular the impact of interchannel coherence /[9] in a square loudspeaker arrangement around the listener. The results indicate that a low degree of interchannel coherence (k < 0,2) results in perception of accumulated directions of auditory events located in the loudspeaker regions. The more the contributing loudspeaker signals are incoherent (“dissimilar”), the better the auditory system is able to distinguish the sound sources (loudspeakers). The other way round, an increasing interchannel coherence causes fusion effects. Principal results are outlined in **FIG 20**.

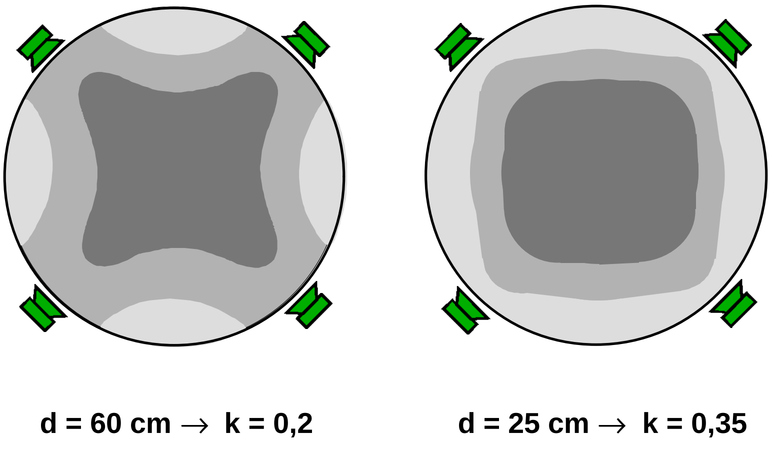


FIG 20: Effect of coherence on subjective diffuseness [41]

Perceived directions in a diffuse field generated by four loudspeakers radiating noise (0,25–2,5 kHz) with interchannel coherence factors k = 0,2 and k = 0,35. The noise was recorded in a reverberation room by means of spaced omni microphones, distances d = 60 cm and d = 25 cm.
Darker shading represents relatively higher statistical frequency of directions of auditory events.

The degree of subjective diffuseness or subjective envelopment depends on the spacing of the square microphone arrangement. If it is too wide there results a loss of subjective envelopment. The balanced distribution of enveloping sources starts to turn into “auditory clouds” around the loudspeakers. If it is too close or even coincident, a phantom sound source becomes audible above listener´s head, and subjective envelopment disappears accordingly.

/[8] A square arrangement of cardioids (d = 20.. 25 cm) is known as “Atmo-cross” and offers realistic stereophonic images of enveloping sources such as for example applause. More details can be found in [40].

/[9] The degree of coherence is usually identical with the maximum absolute value of the normalized cross-correlation function (see [3] and also [42]). “Interchannel coherence“ is a particularly relevant measure for interchannel and interaural relationship in spatial hearing.

The effects could be studied in practical recording situations, for example, if we compare applause in the concert hall recorded with omni microphones in square arrangement and different spacing, e.g. d = 10 cm, 50 cm, 300 cm. Optimum microphone distances should be found in the region of about d = 25 cm (using cardioids, see also [40]) or d = 40 cm (using omnis). However, this is subject to further investigations and experiences.

It is clear that a number of practical aspects will have an influence and that compromises must be found. For example, if the sweet spot should be as large as possible, the interchannel correlation should be minimised and thus the microphone spacing maximised. On the other hand, if natural imaging of space and envelopment are dominating requirements, the preservation of time-of-arrival and diffuse field effects is important and thus the microphone spacing should be determined accordingly.

Summarising, we assume that the proposed square arrangement of microphones L_B, R_B, RS, LS can ensure optimum natural imaging of space and envelopment. The configuration and placement of this four-channel room microphone is aiming adequate R/D-ratio in each channel, proper density and spectrum of reverberant sound, suitable interchannel correlation and time-of-arrival relations.

A compact 3/2-stereo main microphone would require additional consideration of directional imaging parameters for capsules L, C, R as presented in **CHAPTER 3.1.3**. Obviously there result particular disadvantages with regard to the arrangement and directional characteristics of capsules L and R. Furthermore, the placement of the L-C-R main microphone is governed mainly by the recording angle (consideration of direct sound), while the room microphone should be placed due to suitable characteristics of the indirect sound. In many recording situations this causes an unfavourable compromise.

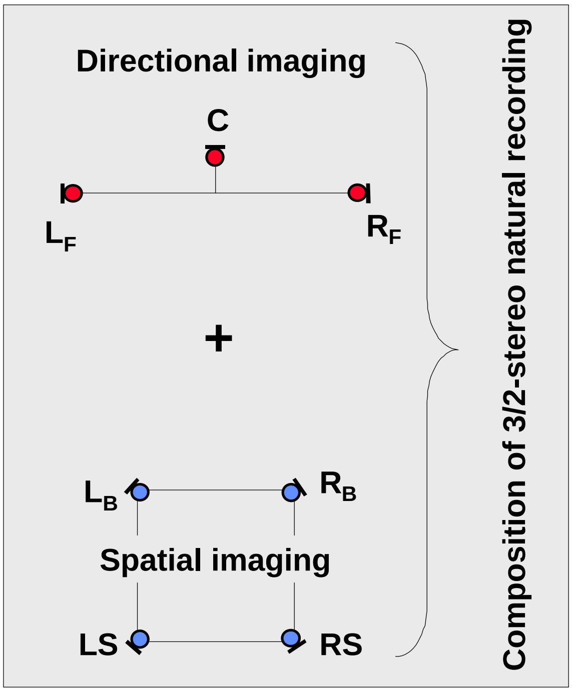


FIG 21:
Distinct psychoacoustic functions are applied separately for natural recording

An L-C-R triangle configuration is used for directional imaging of the orchestra, and a separate L_B, R_B, RS, LS square configuration is applied for spatial imaging. Each microphone configuration can be optimised and placed with respect to the purpose and related psychoacoustic principals, as well as with respect to the actual recording situation and the artistic intention, e.g. stereophonic perspective. Channels L_F and L_B are combined to L, channels R_F and R_B are combined to R The delay between the stereophonic four-channel space information and the directional front information should be designed according to the principals described in **CHAPTER 2.**

It is therefore suggested that sufficient flexibility and adaptability is ensured if the two functional tasks of the main microphone are settled separately with distinct tools, as illustrated in **FIG 21**. Regarding directional imaging, a suitable front channel microphone configuration (e.g. L-C-R triangle according to **CHAPTER 3.1.3**) can be placed with respect to optimum directional and spectral translation of the orchestra, and with respect to a satisfactorily low R/D-ratio providing headroom for adequate four channel spatial imaging.

In this way the microphone square can be applied providing suitable interchannel correlation of four equal channels for realistic imaging of spatial impression and enveloping atmosphere. Placement and choice of microphone characteristics may depend on the intended stereophonic perspective and actual recording situation, since there are no constraints due to directional imaging purposes. The four-channel room microphone can be placed for example far from the critical distance in order to provide appropriate early reflections, a high R/D-ratio, adequate density and spectrum, as well as far from the auditorium in order to avoid disturbing noise.

The initial delay of the spatial imaging channels L_B, R_B, RS, LS in relation to the directional imaging channels L_F, C, R_F can be designed according to **CHAPTER 2,** without touching the interchannel correlation and time-of-arrival relations picked up with the microphone square.

4 CONCLUSION

The 3/2 stereo format enables the sound engineer to overcome some of the constraints associated with conventional two-channel stereophony, and to create a new dimension of spatial impression, enveloping atmosphere, and directional imaging. The more accurately the psychoacoustic principles are understood and taken into account from the technical and artistic points of view, the better the new format will succeed. Specific phenomena of spatial hearing require suitable types, configurations and locations of microphones, as well as distinct design of delay, R/D-ratio, and interchannel correlation. This is particularly true in all cases where optimum naturalness of the stereophonic presentation is desired.

Corresponding recording concepts are suggested as an adequate basis to take maximum advantage of five stereophonic channels. The psychoacoustic principals can be applied not only to the classical recording concept “main microphone / spot microphones / room microphone” but also to poly-microphony. Modern mixing consoles could include an algorithms for the synthesis of arbitrarily realistic spatial impressions based on the same psychoacoustic principles. Psychoacoustic knowledge and digital signal processing technologies are powerful skills and tools for the sound engineer, allowing the creation of convincing and exciting surround sound music recordings.

REFERENCES

[1] ITU-R BS. 775-1: Multichannel Stereophonic Sound System With and Without Accompanying Picture (Geneva, 1992-1994)

[2] SMPTE RP 173: Loudspeaker Arrangements for Audio Monitoring in Television Productions.

[3] Blauert, J.: Spatial Hearing - The Psychophysics of Human Sound Localisation. The MIT Press, Cambridge, Mass, 1996, ISBN 0-262 02413-6 .

[4] Barron, M., Marshall, H.A.: Spatial Impression due to early lateral reflections in concert halls. J. Sound Vibr. 77, 1981, pp. 211-232.

[5] Griesinger, D.: Personal correspondence 15.2.00. See also: Spatial impression and envelopment in small rooms. 103rd AES Convention, 1997, Preprint 4638

[6] Theile, G.: On the naturalness of two-channel stereo sound. J. Audio Eng. Soc.(39), 1991, pp. 761-767.

[7] Wöhr, M.; Theile, G.; Goeres, H.-J.; Persterer, A.: Room-related balancing technique: a method for optimising recording quality. J. Audio Eng. Soc.(39), 1991, pp. 623-631.

[8] Horbach, U:: New techniques for the production of multichannel sound. 103rd AES Convention, 1997, Preprint 4624

[9] ETSI Technical Report. Draft ETR 154 Rev 3 (June 1999): Digital Video Broadcasting (DVB): Implementation guidelines for the use of MPEG-2 systems. (Annex D: Implementation of Ancillary Data for MPEG Audio)

[10] Cossette, S.: Metadata issues for ATSC audio. SMPTE Journal, July 1999, pp. 486-490

[11] Fukada, A.; Tsujimoto, K.; Akita, S.:: Microphone techniques for ambient sound on a music recording. 103rd AES Convention, 1997, Preprint 4540

[12] Sawaguchi, M.; Fukada, A.: Multichannel sound mixing practice for broadcasting. IBC Conference 1999, Proceedings, in print

[13] Keibs, L.: Stereo-Ambiophone Schallübertragung (Stereo-ambiophonic sound transmission). Acustica (12), 1962, pp. 118-124

[14] Karamustafaoglu, A.; Horbach, U.; Pellegrini, R.; Mackensen, P.; Theile, G.: Design and application of a data-based auralisation system for surround sound. 106th AES Convention, 1999, Preprint 4976

[15] Theile, G.: On the performance of two-channel and multi-channel stereophony. 88th AES Convention, 1990, Preprint 2932

[16] Holman, T.: Surround sound systems used with Pictures in cinamas and homes. 8th AES Conference, 1990

[17] Silzle, A., Theile, G.: HDTV-Multichannel sound: Study on imaging quality of the center loudspeaker (German). 16. Tonmeistertagung, 1990, Proceedings (ISBN 3-598-20357-8), pp. 208-218

[18] Wöhr, M.; Nellesen, B., Theile, G.: Studies on the selection of the main microphone method (German). 14. Tonmeistertagung, 1986, Proceedings (ISBN 3-598-20355-1), pp. 106-120

[19] Everest, A.; Streicher, R.: The New Stereo Soundbook.. TAB Books, 1992, ISBN 0-8306-3903-9

[20] Herrmann, U., Henkels, V., Braun, D.: Comparison of 5 surround microphone methods (German). 20· Tonmeistertagung, 1998, Proceedings (ISBN 3-598-20361-6), pp. 508-517

[21] Williams, M.: Unified theory of microphone systems of stereophonic sound recording. 82th AES Convention, 1987, Preprint 2466

[22] Williams, M.; Le Dû, G.: Microphone array analysis for multichannel sound recording. 107th AES Convention, 1999, Preprint 4997

[23] Hugonnet, C.; Walder, P.: Stereophonic Sound Recording. John Wiley & Sons Ltd, 1998, ISBN 0-471-97487-0

[24] Klepko, J.: 5-Channel microphone array with binaural head for multichannel reproduction. 103rd AES Convention, 1997, Preprint 4541

[25] Gernemann, A.: The arrangements of microphones using three front channels, a systematic approach (German). 20. Tonmeistertagung, 1998, Proceedings (ISBN 3-598-20361-6), pp. 518-542

[26] Bruck, J.: Solving the surround dilemma. 19. Tonmeistertagung, 1996, Proceedings (ISBN 3-598-20360-8), pp. 117-124

[27] Steinke, G.: Targets and concepts for multichannel recording technologies. German Surround Sound Forum, Doc. 1-1999

[28] Theile, G.: Main microphone and spot microphones – new aspects on a reliably recording method (German). 13. Tonmeistertagung, 1984, Proceedings pp. 170-184

[29] Theile, G.: On the Theory of Optimum Reproduction of Stereophonic Signals through Loudspeakers and Earphones. (German). Rundfunktechn. Mitt., 1981, pp. 155 – 170

[30] Steinke, G.: Modern sound control technology in large auditoria and in small (reference) listening rooms. 93rd AES Convention, 1992, Preprint Number 3380

[31] Holman, T.: The number of audio channels. 100th AES Convention. 1996. See also: 10.2 Multichannel Sound System. www.tmhlabs.com/home.htm

[32] Woszcyk, W.: Quality assessment of multichannel sound recordings. 12th AES Conference, 1993. Proceedings pp. 197-218

[33] Holman, T.: 5.1 Channel Surround Sound: Up and Running. Focal Press; 1999. ISBN: 0240803833

[34] Wuttke, J.: Principles of microphones and stereo recordings. (German). In: Microphone Papers. Schoeps Brochure. 1998, pp 4-16.

[35] Wuttke, J.: How universal can a microphone be? (German). 19. Tonmeistertagung, 1996, Proceedings (ISBN 3-598-20360-8), pp. 675-683

[36] Radliff, P.A.: Properties of hearing related to quadrophonic reproduction. 1974, BBC Report RD 38,

[37] Theile, G.; Plenge, G.: Localization of lateral phantom-sources. 1976, J. Audio Eng. Soc. (25), S. 196-200

[38] Zieglmeier, W.; Theile, G.: Imaging of lateral sources using the 3/2-stereo format (German). 19. Tonmeistertagung, 1996, Proceedings (ISBN 3-598-20360-8), pp. 159 - 169

[39] Streicher, R.: The Decca Tree in stereo and surround recording. 106th AES Convention, 1999, Workshop Note

[40] Theile, G.: Capabilities and limits of 3/2-stereo recording (German). 19. Tonmeistertagung, 1996, Proceedings (ISBN 3-598-20360-8), pp. 139-158

[41] Damaske, P.: Subjective Investigation of Sound Fields (German). Acustica, Vol.19, 1967/68, pp. 199-213

[42] Cremer, L.: On the use of the terms “degree of correlation” and “degree of coherence”. (German). Acustica, Vol.35, 1976, pp. 215-218