Multichannel Natural Recording based on Psychoacoustic Principles

5156 (R - 3)

Günther Theile
Institut für Rundfunktechnik, München, Germany

Presented at
the 108th Convention
2000 February 19-22
Paris, France

MULTICHANNEL NATURAL RECORDING BASED ON PSYCHOACOUSTIC PRINCIPLES

Günther Theile, IRT, Germany
theile@irt.de

ABSTRACT

The new 3/2-stereo format can offer an improved stereophonic presentation for a number of listeners with regard to imaging parameters such as spatial impression, enveloping atmosphere, direction, depth. 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, inter-channel correlation and level balancing of direct/indirect sound. A correspondingly designed recording concept is suggested to be the adequate basis to take maximum advantage of five stereophonic channels. Practical concepts for natural music recording and surround sound reportage are discussed.

1. INTRODUCTION

Since many years it has been widely recognised that the 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 to satisfy the "purist" music-lover when, for example, he is listening to DVD-Audio or Super-Audio-CD of a classical music concert, without accompanying picture.

This contribution concentrates on the latter goal. Psychoacoustic principles are considered in order to outline the possibilities but also the limits of the new stereo format. A kind of overview is given in **TABLE 1**. It illustrates that the possibilities of stereophonic imaging are still limited with respect to a number of parameters. On the other hand, the 3/2 stereo format enables the sound engineer to create a new dimension of enveloping atmosphere and spatial impression. This will be succeeded the better, the more accurately the psychoacoustic principles are understood and taken into account from the technical and artistic points of view. 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 principal limits are given with respect to presentation of direction and distance, when for instance compared with dummy-head stereophony.

What does optimum naturalness mean? The simplest answer would be: the reproduced sound image must be as identical as possible with 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 perhaps for the reproduction of a speaker's voice through loudspeakers, but it is probably appropriate to a limited extent only for the reproduction of the sound of a large orchestra through loudspeakers. 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) - in other words, the essential problems of loudspeaker stereophony actually still force a deviation from identity. The desired natural stereophonic image should therefore meet two requirements: it should satisfy aesthetically and it should match the tonal and spatial properties of the original sound at the same time.

2. SURROUND SOUND

The term "surround sound" originates from the movie industry where it is used mainly to represent the acoustic environment 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 surround channels are presentation of space, atmosphere, and effects, completing or supporting the frontal stereophonic image.

A principal remark is related to the nature of 3/2-stereophonic sound. It seems to be useful for recording and mixing to distinguish three different types of sound (**TABLE 2**). Localization of (phantom) sources, perception of spatial impression and perception of enveloping atmosphere should be understood as phenomena of spatial hearing, each of them governed by specific psychoacoustic laws and thus needing specific microphone set ups and mixing technologies.

Precedent (direct) sound

Located sound sources ⟶ directional / distance perception

Non-precedent (indirect) sound

Reflections / reverberation ⟶ spatial impression

Environmental (non-reflected) sound

Ambient non-located sound sources ⟶ enveloping atmosphere

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

Localization of (phantom) sources, perception of spatial impression and perception of enveloping atmosphere should be understood as phenomena of spatial hearing, each of them following specific laws and thus requiring suitable types and locations of microphones, as well as distinct sound balancing and handling of delay.

For example, it is well-known already from two-channel stereophony that in the concert hall there are two different basic tasks, besides directional and loudness balancing of the orchestra:

- Generation of the adequate and natural spatial impression
- Imaging of natural enveloping atmosphere (e.g. applause)

Each of the two tasks could require suitable types and locations of the room microphone, and distinct handling of level and delay. Details are presented in the next chapters.

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 [3]. The first is "reverberance" described as a temporal slurring of auditory events caused by late reflections and reverberation. The second is "auditory spaciousness", which denotes a characteristic spatial spreading of the auditory events, caused by the early lateral sound which reaches the listener's ears from lateral directions about 10 to 80 ms after the direct sound (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. Also, the overall level of the direct sound and of the reflections has been found to be of central importance [4]. In [5] 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 there is a clear dependence of the extend of spaciousness on the angle of sound incidence of lateral reflection. The results are of practical importance because they demonstrate that reflections from the side is the most effective way of achieving spaciousness. In contrast, early reflections in the media plane are disadvantageous.

2.2 Three-dimensional space imaging

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 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 localization 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.

3. NATURAL MUSIC RECORDING

3.1 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.

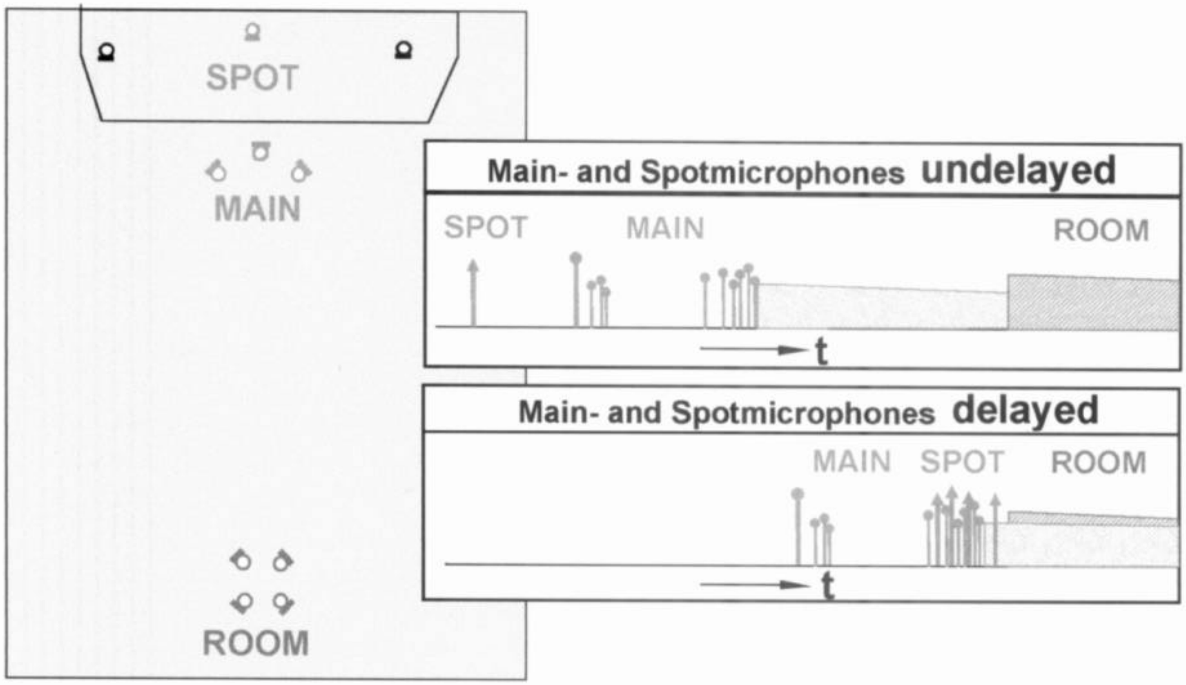

FIG. 1: Delay design according to the natural room response

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 is possible with any kind of microphony.

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.

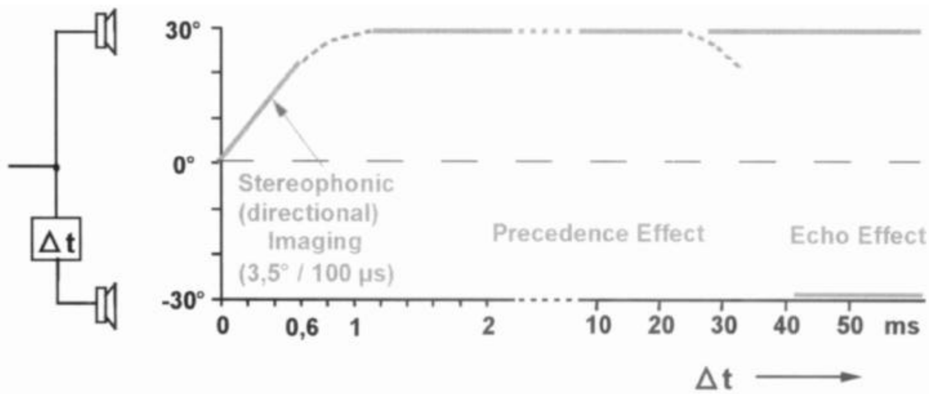

FIG. 2: Inter-channel delay inducing phantom source, Precedence Effect, or echo

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).

3.2 Time-balancing

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**.

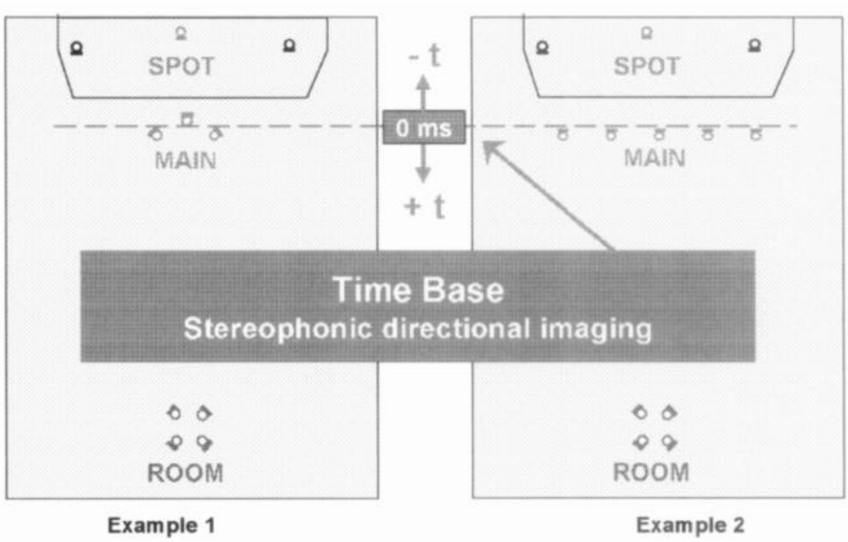

FIG. 3: Room-related time balancing: Setting the time base
A common time base should be set in any microphone configuration. On this basis the delay of each of the microphone can be designed according to the natural pattern (see e.g. **FIG. 1** and **TABLE 3**).

The basis of a delay plan is an exact conception of the pattern of reflections to be reproduced (see **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 [1]. 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: 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.

[1] 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**)

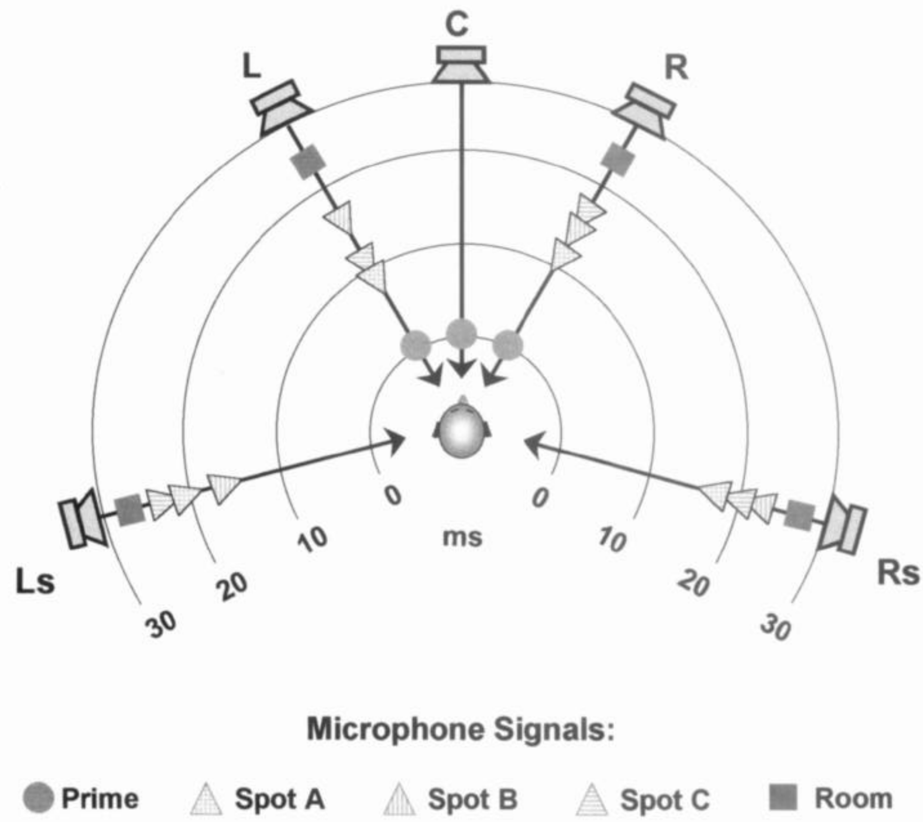

FIG. 4: Delay design according to the natural room response (Example 1)

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 section 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. 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) [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.

3.3 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.

$L_0 = L + 0{,}7C + k\ Ls$	***The choice:***
$R_0 = R + 0{,}7C + k\ Rs$	**k = 1, 0.7, 0.5, 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].

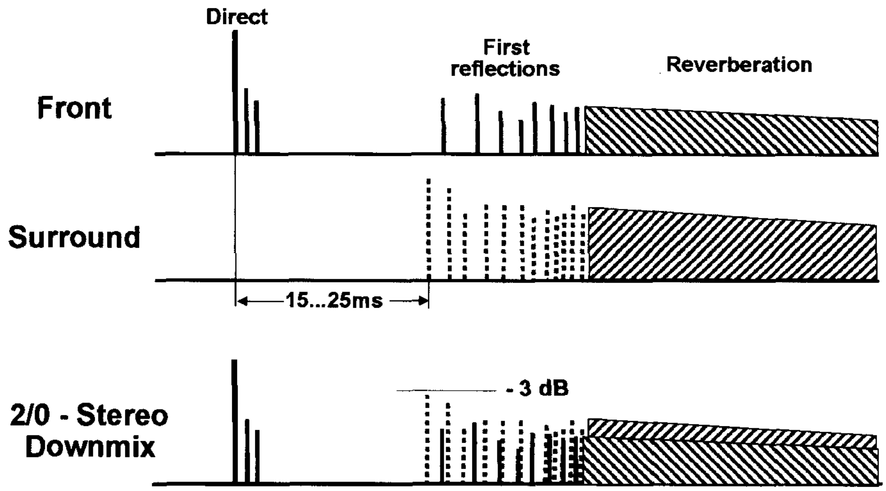

FIG 5: Pattern of reflections in the 3/2-mix and in the 2/0-downmix
In the 2/0-downmix the room information of the 3/2-stereo mix is completely preserved. Two-channel reproduction allows corresponding spatial impression in the simulation plane. However, optimum stereophonic quality is not ensured in 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 localization" 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.

3.4 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 psycho-acoustic 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: 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 psycho-acoustic 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 more than 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.

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 (enhanced 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 requires greater attention.

An optimum 3/2-stereo main microphone should exploit these principal benefits. The L-C-R stereophonic representation should ensure maximum localization focus and avoid coloration due to combing effects. When we look at typical main microphone methods used in practical situations it appears that none of them works perfectly with this respect. So far no method is known, which is sufficient for all request.

A substantial problem is the (probably unavoidable) "triple phantom source" (see the sketch **FIG. 6**). In all cases where three microphone capsules are used there are more or less correlated signals resulting in 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 three-channel microphone is in principle characterized by a decrease of the localization focus and clarity, and by coloration effects.

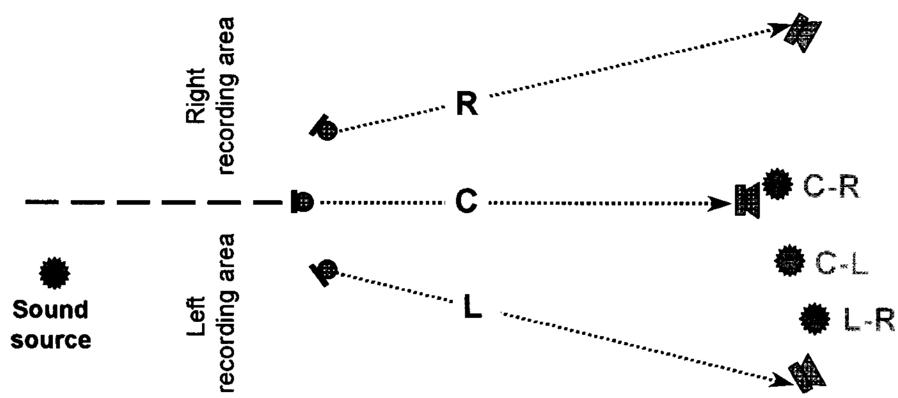

FIG 6:

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 localization 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 minimized 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 such as cardioid, or super-cardioid do not prevent sufficiently from interfering L-C-R-crosstalk. However, a number of different practical solutions have been used in practical situations, as shown in **FIGS. 7, 8, 9, 12, 13**.

The multi-channel stereophonic microphones 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 inter-channel temporal relations (time-of-arrival, phase, correlation, see e.g. [6], [18], [19]).

2. Cardioid or super-cardioid microphones are applied not only to minimize the triple phantom source effect 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 omnidirectional 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 emphasized 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" /[2]. [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 optimization regarding the attachment of adjacent recording areas does not imply a minimization regarding artifacts due to the triple phantom source effect. Minimum impairments of localization focus, clarity, and timbre may not be achievable with this configuration, because the acoustical channel separation is not sufficient.

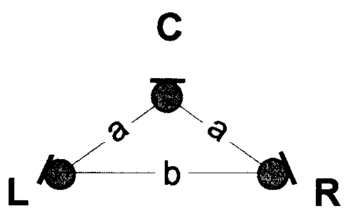

FIG 7: Configuration "*INA 3*" on the basis of [21]

The triangle arrangement can be desired 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

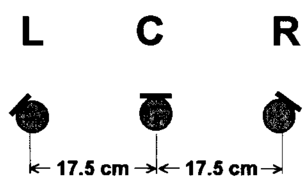

FIG 8: Near-coincident configuration [24]

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.

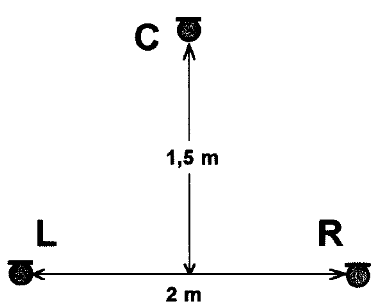

FIG 9: Widely spaced omnis ("Decca-Tree", [19])

Three omnidirectional microphones are widely-spaced in a triangle configuration. Traditionally the middle micro-phone was mixed into L and R, ensuring "a solid center for the stereo sound stage" [19].
A similar triangle configuration is applied in the Fukada-Tree [12], however, cardioit microphones are used. The dimensions of this triangle are: L-R spacing 2 m, distance from C to the L-R basis 1 m (a = 1.3 m, b = 2 m).

/[2] 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	130°
A / B	Omni	0°	35 cm	160°

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 localization 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 localization curves are useful. **FIG 10** shows the typical localization curve of a usual narrow-spaced two-channel stereophonic microphone.

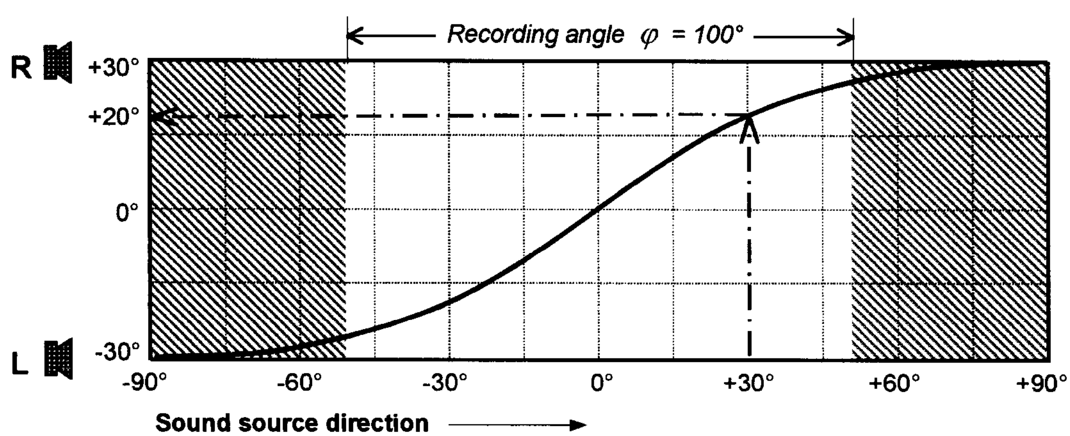

FIG 10: Localization curve of a narrow-spaced two-channel microphone
The curve displays the directional translation of the stereophonic microphone. For example, a sound source located 30° off-center right of the microphone will be perceived approximately 20° off-center right in the standard two-channel loudspeaker arrangement, due to the channel signal difference delivered from the microphone. The useful recording angel of this microphone is ± 50°. The curve ensures within this sector 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 localization curves of each pair, R-C, L-C, L-R are plotted. We can see that the triple image effect 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.

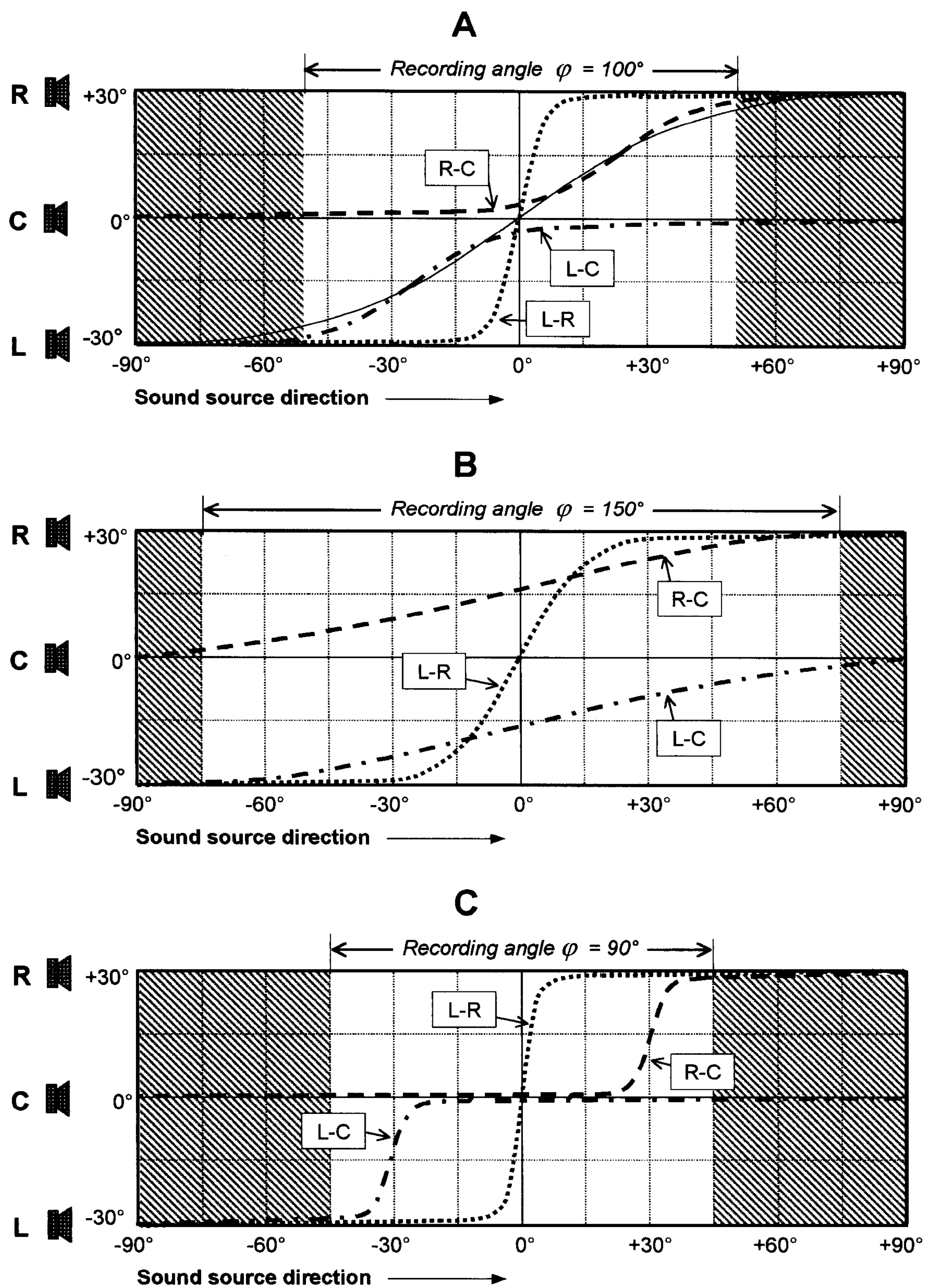

FIG 11: Localization curves and recording angles of three-channel microphones

A: *INA 3* (FIG 7) **B: Near-coincident (FIG 8)** **C: *Decca-Tree***

Considering the *INA 3* configuration (**A**), the pairs L-C and R-C provide the desired recording angle and the desired localization 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 even 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 delay is 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"* is a triangle configuration similar to **FIG 7**, however, omni-directional microphones are applied and widely spaced, see **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 localization 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 localization. Only the curves L-C and R-C are relevant. They demonstrate that all sources in the recording sector ±25° are reproduced in the center or very close to it. All sources outside the sector ±35° 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 to 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 cardioit triangle configuration is applied in the *Fukada-Tree* reported in [12]. The dimensions are: L-R spacing 2 m, distance from C to the L-R basis 1 m (a = 1.3 m, b = 2 m). In the *Fukada-Tree* the triangle is flanked by additional omni-directional microphones on the sides (1 m distance from L or R) "to present a sense of the orchestra width and to smooth the sound connection between the front and the rear"[19].

In some situations it may be suitable to move the microphone towards the orchestra 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 orchestra, the differential distance between them and the front and back of the stage will be minimized. This helps reduce acoustical imbalance between the nearer and more distant elements of the orchestra.

This is valid also for another widely-spaced configuration shown in **FIG 12**. Five microphones are distributed 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 localizable 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 benefits of widely-spaced microphones with regard to spatial imaging.

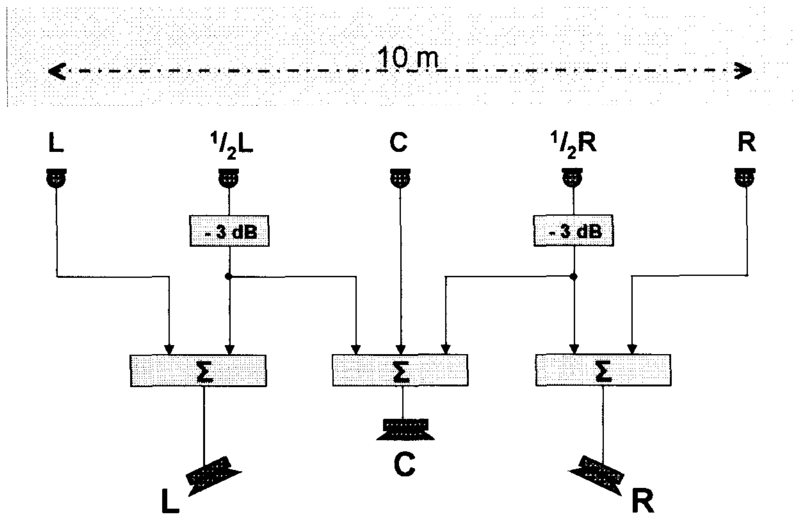

FIG 12: Five microphones in line and widely spaced

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 is 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 localization 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 localization curve)

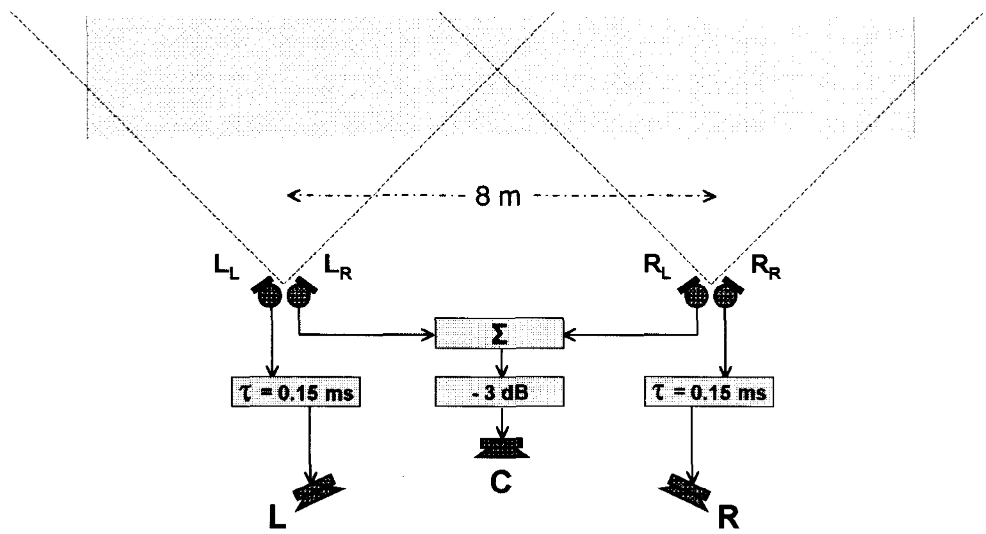

FIG 13: Separate 2-channel main microphones [25]

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 solo piano? Here we need another concept to achieve directional stability from the center loudspeaker and – at the same time – ensure optimum stereophonic quality.

Many practical questions remain to be answered.

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 Localization. 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.: Theory and design of a digital audio signal processor for home use. J. Audio Eng. Soc. 37, 1989, pp. 40-50.

[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 optimizing 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