NORDIC SOUND SYMPOSIUM XX, BOLKESJØ, OCTOBER 4.-7. 2001

# Multichannel Natural Music Recording Based on Psychoacoustic Principles /[1]

Günther Theile
IRT
D-80939 München, Germany

theile@irt.de



/[1] *Extended version of the paper presented at the AES 19th International Conference, May 2001*

# Multichannel Natural Music Recording Based on Psychoacoustic Principles

Natural 5.1 recording of orchestra elements, soloists, room acoustics, audience requires adequate stereophonic representation of direct sound, early and late reflections, reverberation. Particular concepts are proposed to create directional clarity, spatial depth, spatial impression, envelopment, ambient atmosphere, as well as directional stability within the sweet area. Corresponding microphone configurations, mixing methods and signal processing tools are offering realization of a wide range of spatial sound design ideas.

## 1 INTRODUCTION

The new 3/2-stereo format according to Recommendation ITU-R BS 775-1 [1] provides an additional center channel and two surround channels (**Fig. 1**), completing the left and right stereo channels, thereby offering enhanced quality of the stereophonic presentation for a number of listeners who are not positioned at the ideal reference point of the loudspeaker arrangement.

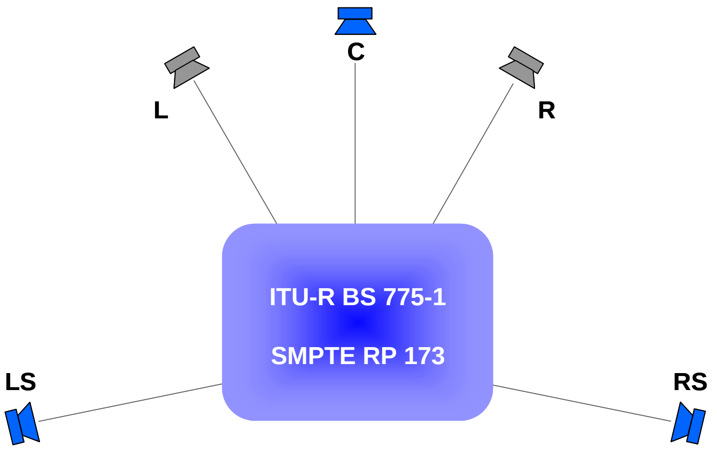



**Fig. 1:**

**The 3/2-stereo format**

It provides easy programme exchange with film sound and overcomes some of the weakness of conventional two-channel loudspeaker stereophony when music is reproduced. The center channel C is intended to avoid imperfections of the phantom sound source in the center area, in particular to improve directional stability (enlarged listening area) as well as localisation focus and sound colour. – The surround channels LS / RS offer improved stereophonic presentation with regard to imaging parameters such as perception of distance and spatial depth, spatial impression, envelopment.

The format is destined for universal use in the home - for audio-only applications as well as for applications with accompanying picture (see SMPTE RP 173 [2]). It provides easy programme exchange with film sound and - at the same time – overcomes some of the weakness of conventional two-channel loudspeaker stereophony when music is reproduced.

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 spatial depth, spatial impression, and enveloping atmosphere. 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 | WFS | Dummy-head |
|---|---|---|---|---|
| Horizontal direction | *+30°...-30°* | *+30°...-30°. surround effects* | *surround* | *surround (instable front)* |
| Elevation | *not possible* | *constraints* | *constraints ?* | *possible* |
| Near-head distance | *not possible* | *no?* | *possible ?* | *possible* |
| Distance, depth | *simulated* | *constraints?* | *possible* | *possible* |
| Spatial impression | *simulated* | *possible ?* | *possible* | *possible* |
| Envelopment | *not possible* | *constraints?* | *possible* | *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 (WFS, see e.g. [3], [4] ), 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 loud-speakers. Artistic intentions of the sound engineer, aesthetic irregularities in the orchestra, poor re-cording 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, down-ward 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. He should be able to apply usefully specific imaging parameters for creating a natural stereophonic image of the orchestra, the soloists, the room acoustics, the ambient acoustical atmosphere (**Table 2**).

| | Orchestra elements | Soloists | Room | Audience |
|---|---|---|---|---|
| **Horizontal direction** | ●●● | ●●● | ● | ●● |
| Elevation | ○ | ○ | | ○ |
| Near-head distance | | | | ○○○ |
| **Distance, spatial depth** | ●●● | ●●● | | ●● |
| **Spatial impression** | | | ●●● | ● |
| **Envelopment** | | | ●● | ●●● |
| **Sound colour** | ●●● | ●●● | ●●● | ●● |

**Table 2:**

**Consideration of specific imaging parameters for natural sound design**

Imaging of orchestra elements, soloists, room acoustics, or ambient atmosphere (audience) requires the application of specific phenomena of spatial hearing, each of them governed by particular laws and needing adequate microphone configurations and mixing meth-ods. However, in the case of 3/2-stereo there are constraints regard-ing perception of elevation and near-head distance.

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 “localisation focus”, “clarity” and “sound colour” of the image in the center area. Consequences regarding microphone and recording methods are discussed in the next chapter.

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, supporting or completing the frontal stereophonic image.

In the case of natural music recording the supplemental surround channels offer improved stereophonic presentation with regard to imaging parameters such as perception of distance and spatial depth, spatial impression, envelopment (due to reverberation and / or ambient non-located and non-reflected sound, e.g. applause), see **Table 3**. 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 is worth to say that the pros and cons of this concept on the one hand and poly-microphony on the other are not discussed here. Rather, the psychoacoustic principles can also be applied to poly-microphony. Modern mixing consoles include tools for the synthesis of arbitrary spatial creations on the same basis (see e.g. [5], [6] ).

| | Direct sound | Early reflections | Reverberation | Environmental non-reflected sound |
|---|---|---|---|---|
| **Horizontal direction** | ●● | ● | | |
| Elevation | ○○ | ○ | | |
| Near-head distance | ○○ | | | |
| **Distance, spatial depth** | | ●● | ● | |
| **Spatial impression** | | ●● | ●● | |
| **Envelopment** | | | ●● | ●● |
| **Sound colour** | ●● | ● | ●● | |

**Table 3: Four types of 3/2-stereophonic sound to be designed for natural music recording**

Adequate microphone configurations and mixing methods are desirable in order to apply usefully specific imaging parameters for creating a natural stereophonic presentation of the orchestra, the soloists, the room acoustics, the ambient acoustical atmosphere. Characteristics of direct sound, early reflections and reverberation related to spatial hearing can be artificially designed by means of modern signal processing tools.

A further principle remark is related to the aesthetical goal “natural music recording”. As stated above, there are many reasons causing loudspeaker stereophony to deviate from identity. It is clear that the 3/2-stereo music production is governed by a number of parameters (see **Fig. 2,** left side), in particular the artistic intention of the conductor, sound director, producer may differ from the natural recording idea fundamentally.

A wonderful artistic idea of sound picture may require aversion from a recording techniques offering precise directional distribution of orchestra elements, image of spatial depth and spatial impression. Rather, for example, three stably localisable orchestra elements may be created ("L-C-R thrice mono stereophony", e.g. by means of the "Decca Tree"), completed by panned elements in the foreground of the stage (between the loudspeakers), and enveloping non-localisable indirect sound, resulting in a dense and "open" sound picture perceptible over a huge listening area.

This kind of sound picture could be compared with an aquarelle, obviously requiring other painting tools than a painting showing more realistically objects in the foreground and background. Natural recording is aiming a precisely dispersed, deeply staged, stable and wide image of the orchestra in the room, where the foreground distance is not fixed at the loudspeaker distance and the spatial impression and acoustical environment allow a convincing illusion of "being in the hall" over a reasonable listening area (see **Fig. 30**).

The reader of this paper should consider that basically natural music recording is intended here. In particular, the paper is striving for getting knowledge about tools suitable for natural recording. This does not mean that natural music recording is proposed as the only aesthetical goal. Rather, it is suggested that this knowledge about tools for creating depth, spatial impression, envelopment etc. can be useful for the realization of any artistic idea.

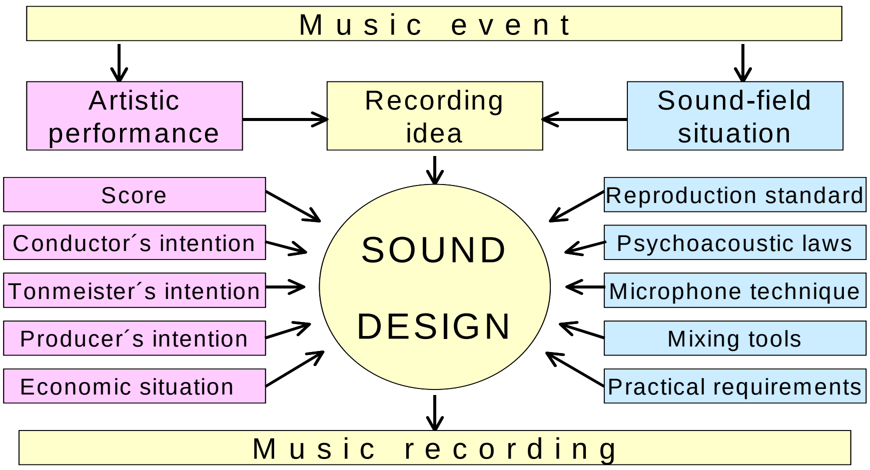



**Fig 2: Parameters of music production**

Music sound design as a result of artistic intentions (left) and application of technical tools (right). Psychoacoustic phenomena as well as the principle characteristics of reproduction standard, microphone techniques, mixing device must be known and exploited to create the desired product. This paper is dealing with the technical tools, namely in case where natural imaging is intended.

## 2 L-C-R STEREOPHONIC IMAGING

It is the aim to achieve imaging characteristics equivalent to those of an optimum two-channel main microphone (**Fig. 3**). 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 [7]. 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. [8], [9]) 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 [9] and requires "greater attention".

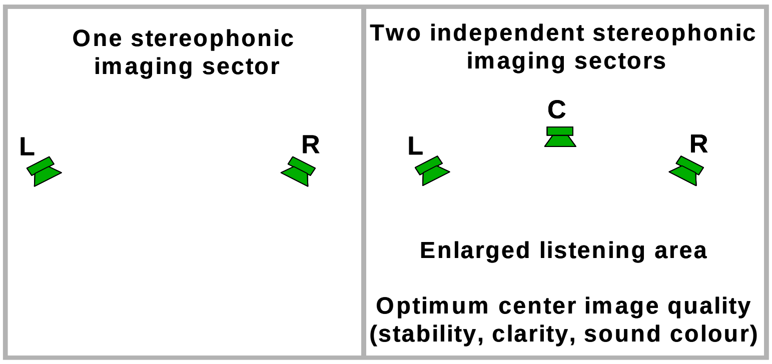



**Fig. 3:**

**Two vs. three front channels**

The primary purpose of the center channel C is to divide the stereophonic loudspeaker basis into two stereophonic "sub-areas" [7], using the two 30° imaging sectors L-C and C-R instead of one 60° imaging sector L-R, in order to increase directional stability (to broaden the listening area) as well as to improve "clarity" and "sound colour" of the image in the center area.

Two recording techniques are principally suitable for use with the L-C-R loudspeaker arrangement, see **Fig. 4**. First, it is evident that the psychoacoustic principles of loudspeaker stereophony could be applied to achieve stereophonic presentations within each section of the three-channel imaging plane, thereby offering an overall presentation as similar as possible to that in the two-channel imaging plane but with the advantage of an enlarged listening area. In order to preserve the imaging capabilities of conventional two-loudspeaker stereo as far as possible, this "stereophonic sub-area concept" [7] appears to be the appropriate approach.

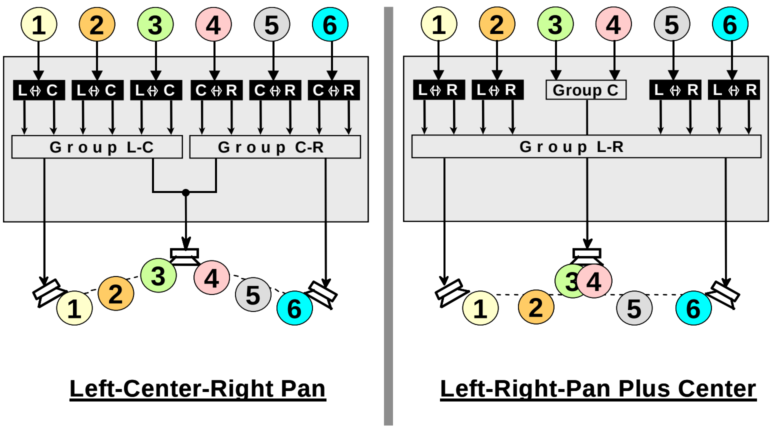



**Fig. 4:**

**True L-C-R stereophony vs. L-R stereophony plus C**

Left side: The center channel C offers increased directional stability of the complete L-C-R stereophonic image, which is divided into two stereophonic sub-areas. Right side: The center channel C is used for a stable center image (e.g. soloist), in addition to the usual two-channel stereophonic representation of the orchestra.

The second possibility is drawn from the well-tried mixing techniques for film/television sound applications. In this domain phantom sound sources are only rarely used for important (picture-related) sources such as dialogue; it is assigned to the centre channel. Stereophonic images are generated across L-R without providing stable localization between the loudspeakers. In contrast to the sub-area stereophony concept, the signals of adjacent loudspeakers are barely correlated. - In practical music recording applications (e.g. orchestra plus soloist) both techniques may be used in combination.

## 2.1 L-C-R microphone configurations

A consequent stereophonic sub-area approach for large orchestra situations is known from [10], see **Fig. 5**. 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 sound 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 [10] 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)

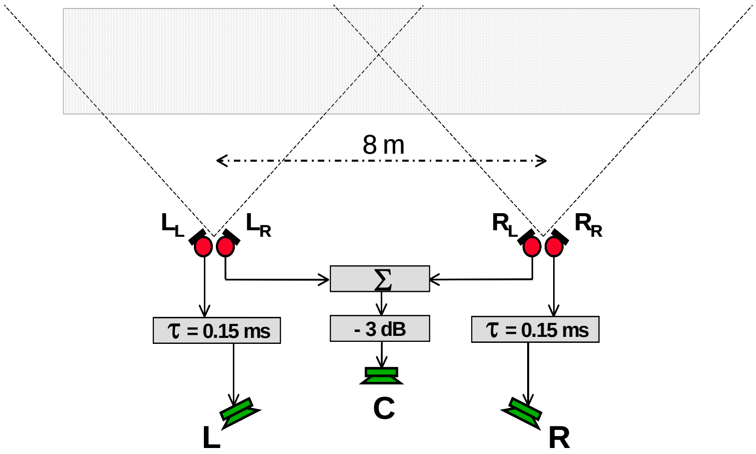



**Fig. 5:**

**Separate 2-channel main microphones [10]**

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 is available without compromises
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.

Another configuration shown in **Fig. 6** is based on widely spaced microphones (“*Multiple-A/B*”). 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, the exploitation of the precedence effect to reduce the multiple phantom sound source problem. Secondly, the provision of a “stable” phantom sound source half left between L and C and half right between C and R.

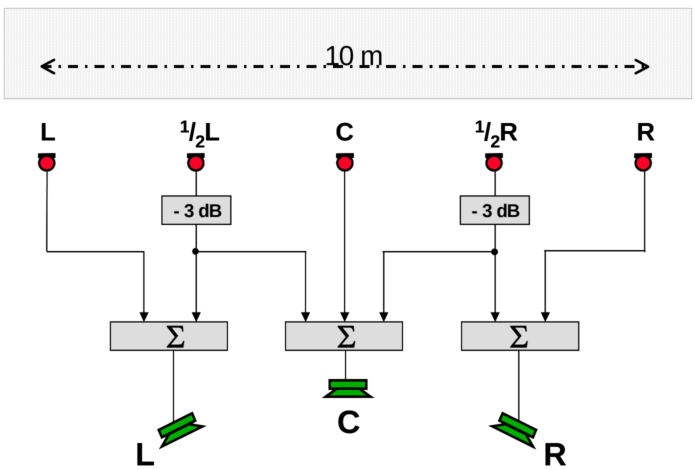



**Fig. 6:**
**Five microphones in line and widely spaced (“Multiple-A/B”)**

The microphones are distributed across the stage width, providing five negligibly correlated signals to produce three stable sources plus two phantom sound 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.

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.

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). The “double-main” method or widely spaced configurations could perhaps satisfy only in a limited range of applications. For example, how to record 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, i.e. maximum localisation focus and minimum coloration due to combing effects..

## 2.2 The interchannel crosstalk problem

An optimum three-channel L-C-R stereophonic microphone should exploit the principle benefits of three front channels. However, a substantial problem is the “crosstalk phantom sound source" (see the sketch **Fig. 7**). In all cases where three microphone capsules are used there are results interchannel crosstalk.

More or less correlated signals generate three phantom sound 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.

The triple phantom sound source and associated comb filter effects particularly in the two-channel downmix (see e.g. [11]) 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.

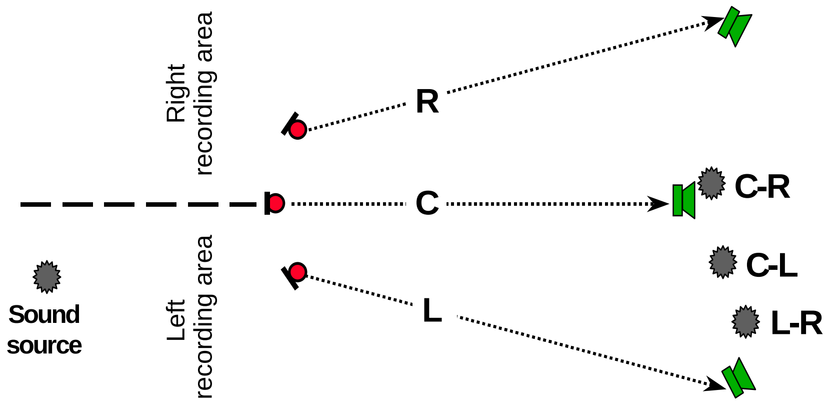



**Fig. 7: The “crosstalk phantom source” problem arising with a 3-channel stereophonic microphone**

In principle each 2-channel stereophonic basis C-L, C-R, L-R produces its own phantom sound 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. In this case the phantom sources C-R and L-R are undesired.

### 2.3 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. [12], [13], [11]).

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 [48] 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. 8** and proposed in [14] ("*INA 3*"). The geometrical arrangement has been designed under consideration of the "recording angles" /2 [15], [16] [62]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 artefacts 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.

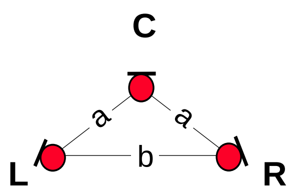



**Fig. 8: Configuration "*INA 3*" on the basis of [15]**

The triangle arrangement has been designed in line with the so-called "*Williams-Curves*" [15] aiming optimum attachment of the recording areas for L-C and C-R. In [14] 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 ε = ½ φ

/2 The recording angle (also known as "useful acceptance angle" [52]) 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 (TABLE 4):

| Configu-ration | 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°** |

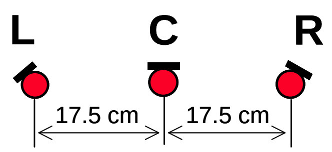



**Fig. 9: Near-coincident configuration [17]**

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

A near-coincident in-line configuration shown in **Fig. 9** 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 sound 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 ε is much smaller and a more directional polar characteristic is used at one site. Therefore, in contrast to the triangle arrangement according to **Fig. 8**, 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.

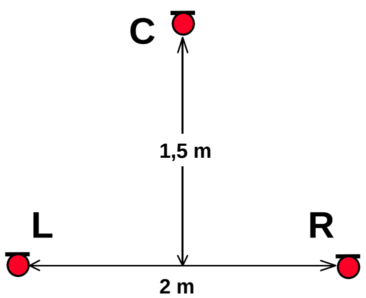



**Fig. 10: Widely spaced omnis ("*Decca-Tree*" [11])**

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* [18], however, cardioid microphones are used.

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. 11** shows the typical localisation curve of a usual narrow-spaced two-channel stereophonic microphone. The curve displays the directional translation of the two channel L-R or three-channel L-C-R stereophonic microphone.

The useful **recording angle** can be defined on this basis. In the case shown in **Fig. 11** the recording angle is φ = 100° /[3]. Within this sector the localisation curve ensures a well-balanced directional image. Sound source directions outside ± 50° will produce too big stereophonic signal differences in the channels L-C or C-R.

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). The L-C-R microphone channels should produce a linear localisation curve in the center area. If this is achieved, the center channel enables a natural and well-balanced distribution of sources across the stereophonic stage L-C-R (the "**unobtrusive center channel**").

/[3] The recording angle φ of two- or three-channel stereophonic microphones based on the localisation curves can easily determined by means of the calculation tool "***Image Assistant***" [23], see **Chapter 2.4**.

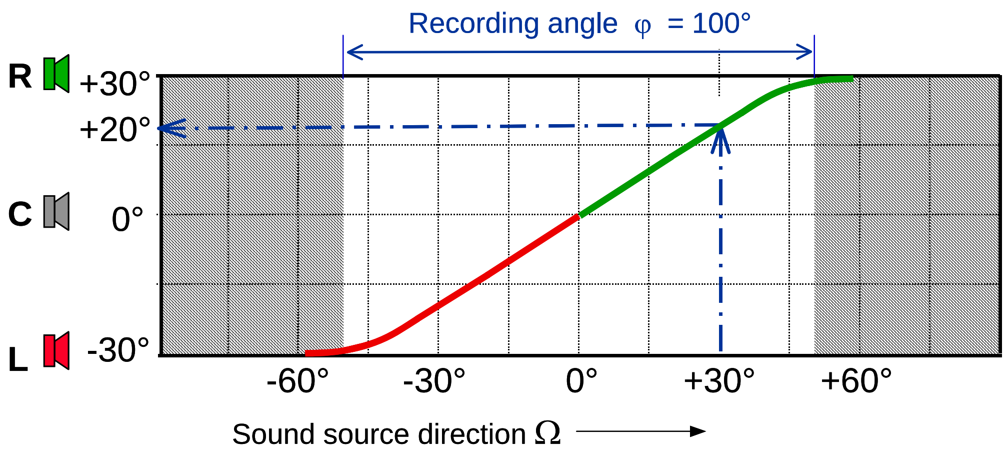



**Fig. 11: Desirable localisation curve**

The curve displays the directional translation of the two channel L-R 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 same shape is desired for three front channels. The L-C-R microphone channels should translate directions accordingly, in particular show a linear curve in the center area. If this is achieved in the case of L-C-R stereophony, the desirable so called "**unobtrusive center channel**" will be effective.

Three examples are outlined in **Fig. 12**. The first (A) corresponds with the "*INA 3*" configuration (**Fig. 8**), the second (B) with the near-coincident microphone according to Fig. 9, and the third one (C) indicates directional characteristics of the "*Decca-Tree*" (**Fig. 10**). 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.

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. 9**, **12 B**). 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. [11]) is a triangle configuration similar to **Fig. 8**, however, omni-directional microphones are applied and widely spaced (**Fig. 10**). 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 sound 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. 12 C** 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 with respect to localisation. They demonstrate that the precedence effect is effective, and that therefore all sources in the recording sector ±45° are reproduced in the center or very close to it. This center effect may be the reason of reports that the center is disturbing without level reduction. Sources outside the sector ±60° are reproduced in L or R

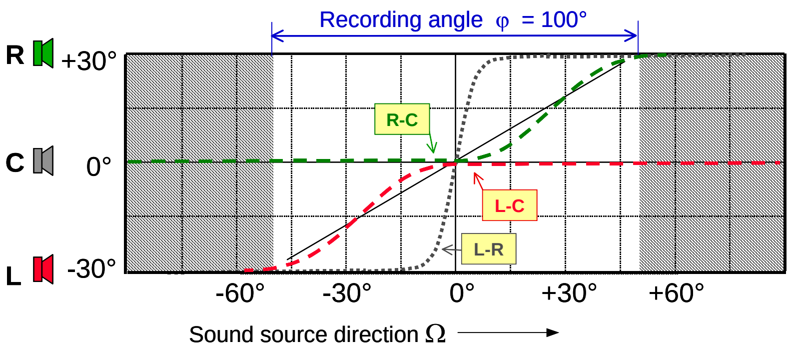



**Fig. 12 A:**

**Principal localisation curves of three-channel microphone *INA 3* (Fig. 8, φ = 100°)**

In the centre of the recording sector (Ω = 0°) the undesired "crosstalk phantom sound source L-R" has the level of about -3 dB and is about 1 ms delayed

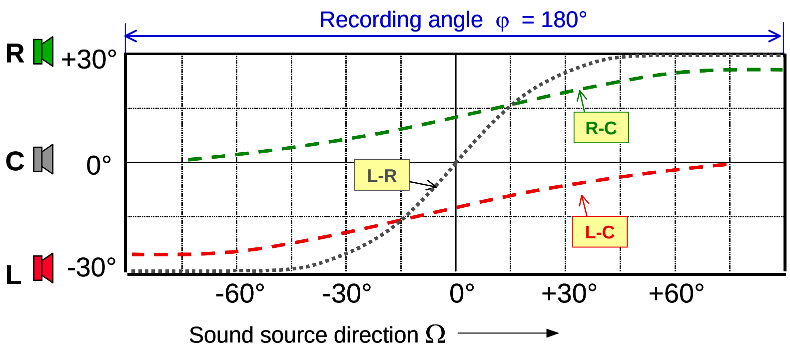



**Fig. 12 B:**

**Principal localisation curves of near-coincident three-channel microphone (Fig. 9)**

In the centre of the recording sector (Ω = 0°) the undesired "crosstalk phantom sound source L-R" has the level of about -1 dB and is about 0 ms delayed

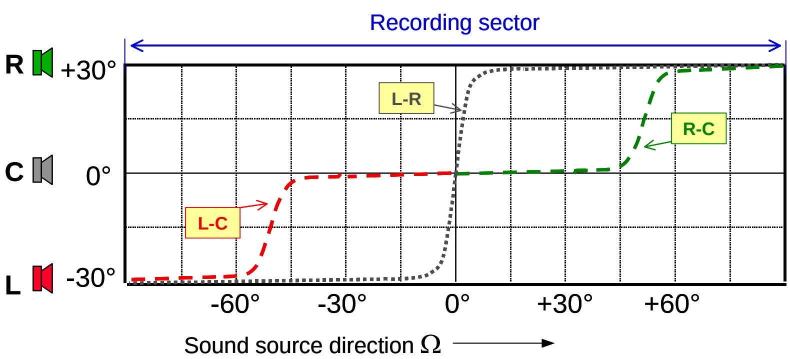



**Fig. 12 C:**

**Principal localisation curves of the *Decca-Tree* (Fig. 10)**

In the centre of the recording sector (Ω = 0°) undesired "crosstalk phantom sound source L-R" has the level of about 0 dB and is about 4,5 ms delayed

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 ε = 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 [18].

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. - It is possible for any widely-spaced configuration, since stereophonic directional imaging according to localisation curves does not happen anyway.

When we look at the L-C-R microphone concepts discussed above, it appears that none of them works completely sufficient with respect to the resulting localisation curve, sound coloration, and localisation focus. It is necessary to reduce the interchannel crosstalk as much as possible. A corresponding approach is called “Optimised Cardioid Triangle” (*OCT*), which has been introduced recently in [19].

## 2.4 Optimised Cardioid Triangle (*OCT*)

Can the center channel increase directional stability without decreasing the stereophonic quality? There are a number of psychoacoustic principles and basic requirements to be considered carefully if we try to optimise the main microphone configurations shown in **Figs. 8, 9, 10.** 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.

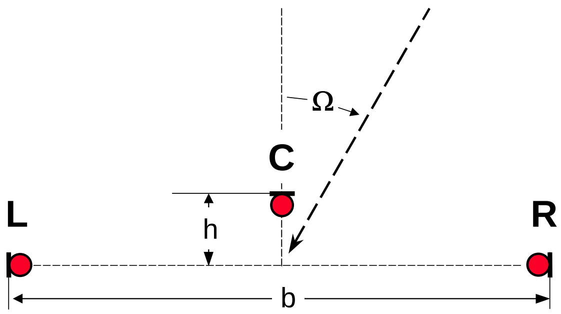



**Fig. 13:**

***OCT* - an optimised triangle configuration**

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

Distance b depends on the recording angle φ (see Fig. 15). Distance h = 8 cm. - If cardioids are used instead of super-cardioids, it is h = 12 cm.

The principle configuration is shown in **Fig. 13.** 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 (off-center angle ε = 90°). 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.

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

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

The level differences ΔL are based on the directivity pattern of unidirectional microphones (see **Table 5**):

The time differences Δt between L and C or R and C are calculated according to the geometrical dimensions. The equation is:

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

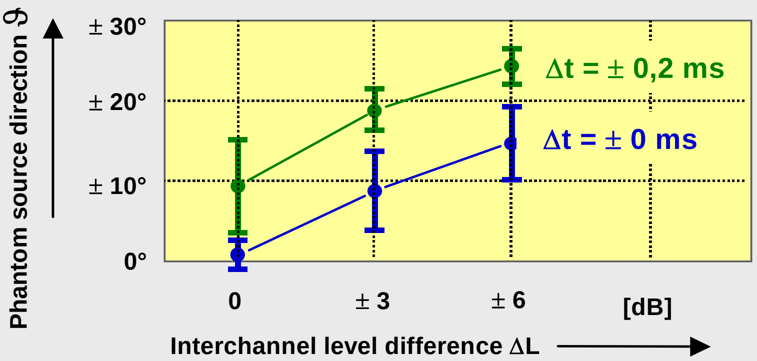
The determination of distance b depends on the desired recording angle φ. Distance h determines the time difference Δt. In the case $\Omega \approx 0°$ it is h = 8 cm, which implies Δt = 0,24 ms. As a result we get the level difference ΔL = 9 dB and the time differences Δt = 0,24 ms between L and C or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

| **Direction** (rel. to mico axis) | **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 |
| **Figure of eight [dB]** | 0 | -0,3 | -1 | -3 | -6 | -12 | < -18 | -12 | -6 | -3 | -1 | 0 |

**Table 5: Directivity pattern of unidirectional microphones**

It has been shown in former papers [7], [20] that the phantom sound source perceived in the middle between two loudspeakers can be shifted in certain limits according to the rules as presented in **Fig. 14.** There are two important facts. Firstly, if the phantom sound 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

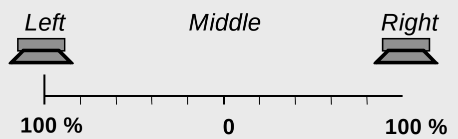
$$\vartheta\,(\Delta L, \Delta t) = \vartheta\,(\Delta L) + \vartheta\,(\Delta t).$$




| | Two-channel stereo<br>Loudspeakers L – R<br>Stereo base angle Ψ = 60° | Three-channel stereo<br>Loudspeakers L – C – R<br>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

**Fig. 14: Intensity and time shift factor**

Secondly, the phantom sound 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 $\Psi$ = 60° to $\Psi$ = 30° will correspondingly halve the shift: it is $\Delta\vartheta$ = 6,6°. (By the way, the "constant relative shift phenomenon" of the phantom sound source is contradictory to summing localisation theories, however, it is explained by the Association Model, see for example [7], [20]).

We understand now the principle function: If a stereo base angle $\Psi$ = 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.

Based on this knowledge the perceived source angles $\vartheta$ as well as the interchannel crosstalk for three configuration examples has been calculated, see **Table 6**. It is assumed here that the source is located right side (according to **Fig. 13**). 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 configurations A and C. However, when cardioids are applied for L and R instead of super-cardioids the crosstalk is not neglectable, and imaging quality degradation (colouration, focus) occurs. With this respect the pure cardioid configuration B has drawbacks. As regards the directional translation (perceived angle $\vartheta$) the three configurations demonstrate similar characteristics. The recording angles are approximately $\varphi$ = 105° - 110°.

Configuration C is a practical compromise solution intended in order to improve the low frequency end (see also **Fig. 18**). The center omni capsule will provide satisfying results with respect to low frequencies without significant decrease of the good directional imaging performance, however, it could be possible in cases where surround channels are used for spatial imaging that the omni center capsule pick up to much indirect sound energy (see also [65] and **Chapter 3**).

| | Configuration A | | Configuration B | | Configuration C | |
|---|---|---|---|---|---|---|
| | **S-CARD. – CARD. – S-CARD.** h = 8 cm / b = 70 cm | | **CARD. – CARD. – CARD.** h = 12 cm / b = 70 cm | | **S-CARD. – OMNI – S-CARD.** h = 8 cm / b = 80 cm | |
| **Source angle $\Omega$** | **Crosstalk →L** | **Perceived angle $\vartheta$** | **Crosstalk →L** | **Perceived angle $\vartheta$** | **Crosstalk →L** | **Perceived angle $\vartheta$** |
| + 90° | -11 dB | ( 30°) | <-18 dB | ( 30°) | -11 dB | ( 30°) |
| + 60° | <-18 dB | 30° | <-18 dB | 30° | <-18 dB | 29° |
| + 45° | <-18 dB | 25° | -15 dB | 24° | <-18 dB | 25° |
| + 30° | <-18 dB | 18° | -11 dB | 16° | <-18 dB | 18° |
| + 15° | -14 dB | 9° | -8 dB | 8° | -14 dB | 9° |
| 0° | -9 dB | 0° | -5 dB | 0° | -9 dB | 0° |

**Table 6: Imaging characteristics of three *OCT* configurations (right sector)**

The corresponding localisation curves and the resulting recording angle of configuration A are plotted in **Fig. 15**. They show that the desired linear directional translation in the center area is achieved. The center channel enables a natural and well-balanced distribution of sources across the stereophonic stage L-C-R ("unobtrusive center channel", **Fig. 11**). Due to the minimized crosstalk the localisation curve L-C is effective only in the allocated left section, and L-R only in the right section. For the same reason there is no significant interfering correlated signal energy in channels L and R.

In the next **Chapter 2.5** results of corresponding listening test are presented, showing imaging characteristics of configuration A with respect to the localisation curve, localisation focus, sound colouration.

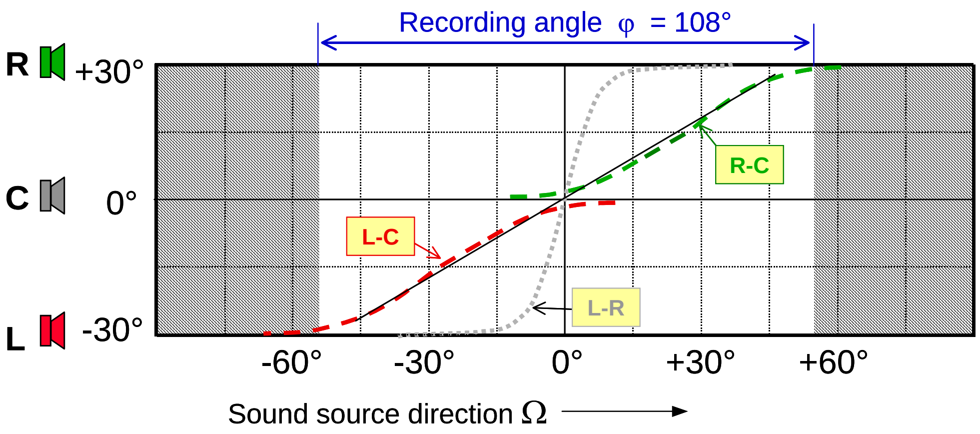



**Fig. 15:**

**Calculated localisation curves of *OCT 70* (Table 6, Configuration A)**

In the centre of the recording sector (Ω = 0°) the undesired “crosstalk phantom sound source L-R” has the level of about -10 dB and thus can be neglected. The *OCT 70* curves may be compared with the *INA 3* curves shown in **Fig. 12A**.

### Calculation tool “*Image Assistant*”

In [21] listening tests have shown that the calculation basis used above is valid. The psychoacoustic laws for lateral phantom sound source shifts **ϑ(ΔL, Δt)** shown in **Fig. 14** have been confirmed for the two-channel loudspeaker arrangement (L-R imaging sector 60°) as well as for the L-C-R arrangement (imaging sectors 30°), for details see [22], [66]. On this basis a calculation tool (“***Image Assistant***") has been developed [23].

It offers the calculation of localisation curves (and corresponding recording angle), interchannel level und time differences, overall sound level, for any two-channel or three-channel microphone configuration, including omni, cardioid, super cardioid, broad cardioid, figure of eight characteristics. For example, the recording angle of the *OCT* configuration as a function of distance b (h = 8 cm) is determined in this way and plotted in **Fig. 16.**

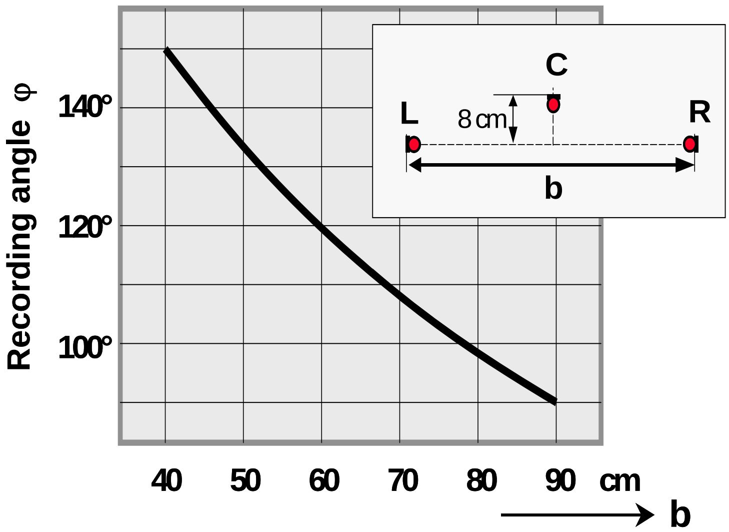



**Fig. 16:**

***OCT* recording angle φ**

as a function of distance b of capsules L and R.
L, R: Super-cardioid
C: Cardioid
The determination is based on the localization curve, as shown in **Fig. 11**.

The *OCT* recording angle φ can give orientation for practical situations. Experience has shown however that in some cases the main microphone is located too distant if even the string instruments in the foreground of a deeply scaled orchestra are covered inside the recording sector.

### Adjustment of the recording angle

In practical situations the recording angle adjustment can be important in order to get freedom of microphone location and resulting direct/indirect sound-balance as well as sound colour (see also **Chapter 3.2**). **Fig. 17** 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).

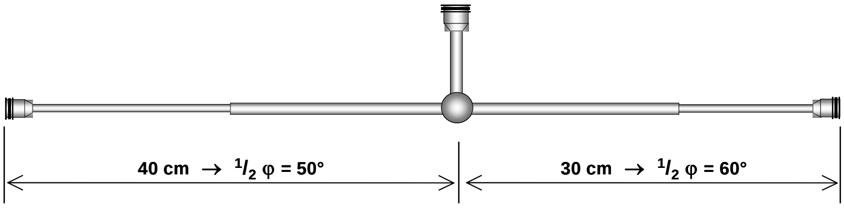



**Fig. 17: Individual adjustment of the recording angle for each side**

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.

### *OCT* low frequency response

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. [24], [52]).

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 [25] ). Applied to *OCT* according to **Fig. 18 A** or **Fig. 18 B**, above 100 Hz there results directional imaging performance as described above. Configuration **Fig. 18 A**: Below 100 Hz the center microphone C does not contribute, which means that here the characteristics of a usual narrow spaced AB microphone are effective. Configuration **Fig. 18 B / 18 C**: Below 100 Hz the center microphone C contributes.

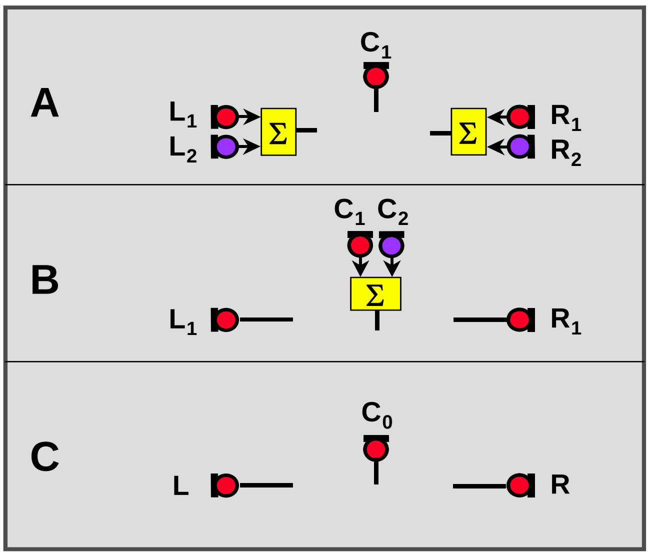



**Fig. 18:**

**Alternative *OCT* configurations for improved bass response**

L1 and R1: S-Cardioid 100 Hz highpass filtered
L2 and R2 : Omni, 100 Hz lowpass filtered
L and R : S-Cardioid
C1: Cardioid, 100 Hz highpass filtered
C2: Omni, 100 Hz lowpass filtered
C0: Omni

**Configuration B**: The bass information C2 can alternatively be added to L1 and R1

**Configuration C**: This is a simple *OCT* low cost version, see also **Table 6**, right column.

### Practical applications

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, clarity, and last not least sound colour and bass response. The *OCT* system is a design initially based on theoretical and psychoacoustic principles.

Recently a number of sets are available [26], and sound engineers have got the possibility to test them in practical situations. First results are available, e.g. [22], [70], [75]. It may be interesting to see that the frequency response of super-cardioid microphones can be realized astonishingly linear in the range 100 Hz – 18 kHz for frontal and lateral directions. The two frequency responses for 0° and 90° of the SCHOEPS capsule MK 41V are plotted in **Fig. 19**, as an example for technical feasibility.

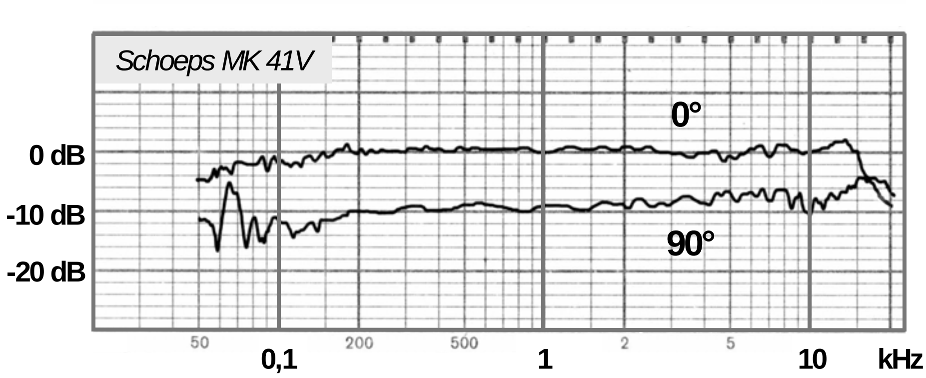


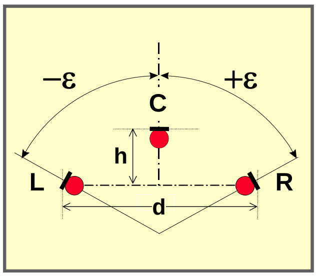

**Fig. 19:**

**0° and 90° frequency responses of a super-cardioid microphone used for *OCT* channels L and R**

## 2.5 Listening tests [22]

In a recent study three-channel L-C-R stereo microphone systems for the front channels have been investigated in theory and practice, particularly *OCT* and INA3. Careful listening tests were performed to examine location, extension, and character of the perceived phantom sound sources. Special interest was related to the resulting characteristics of the frontal stereophonic image, in comparison with the result from a corresponding two-channel stereo microphone: Details can be found in [21] and [22].

The test signals, derived from recordings through the participating stereo microphone configurations, were reproduced through the front loudspeakers of a 3/2-Stereo-configuration. Three different microphone configurations were investigated, see **Fig. 20**.




| | Quasi-ORTF | INA 3 | OCT 70 |
|---|---|---|---|
| Microphone L | Cardioid, ε = -30° | Cardioid, ε = -60° | S-Cardioid, ε = -90° |
| Microphone C | --- | Cardioid, ε = 0° | Cardioid, ε = 0° |
| Microphone R | Cardioid, ε = +30° | Cardioid, ε = +60° | S-Cardioid, ε = +90° |
| Distance d | 20 cm | 92 cm | 70 cm |
| Distance h | --- | 27 cm | 8 cm |

**Fig. 20: Three microphone configurations under test**

During listening the subjects used two seats, the first one at the sweat spot (“center seat”), the second one 50 cm beside to the left (“off-center left seat”). This was useful in order to demonstrate the expected imaging improvement with respect to stability, focus, and sound colour due to the center channel.

**Fig. 21** shows a first result for *OCT 70* and the center seat. The plotted experimental data and the calculated localisation curve are matching very well. This confirms firstly that the psychoacoustic laws used for the calculation are correct. Secondly, it demonstrates that a localisation curve can be achieved which is very similar to the curve known from two-channel stereo microphones (see **Fig. 11**): The L-C-R configuration enables a natural and well-balanced distribution of sources across the stereophonic stage in the front.

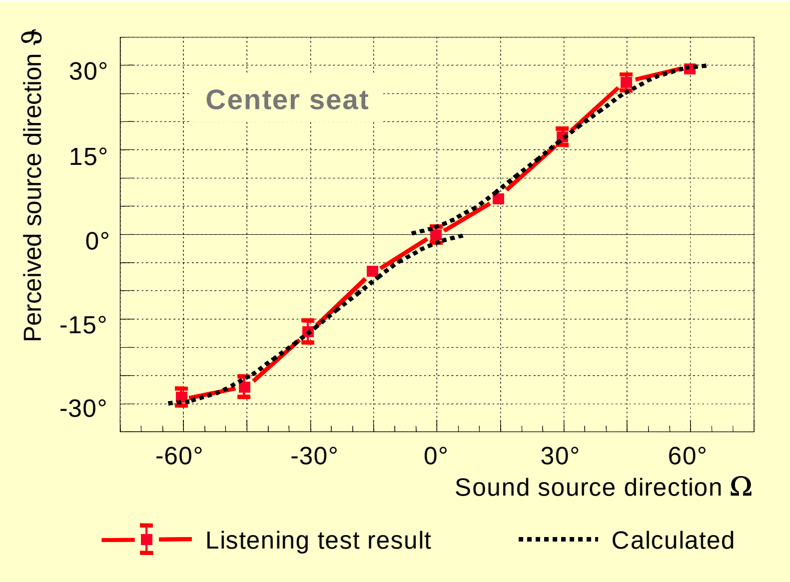



**Fig. 21:**

***OCT 70* localization curve**

The experimental data show good match with the computed prediction (dotted curve, see also **Fig. 15** and **Fig. 11**). The desired linear directional translation in the center area is achieved with high accuracy. Low standard deviations indicate good localisation focus and stability, see **Fig. 23**.

In **Fig. 22** for all three microphones under test the perceived directions at the non-optimal listening position 50 cm left beside the sweet spot are plotted. It is evident that the two-channel microphone “**Quasi-ORTF**” cannot provide a satisfying distribution of phantom sources between the loudspeakers. Sound source directions in the center area Ω = 0°..30° are perceived close to the left speaker (φ = -30°..25°).

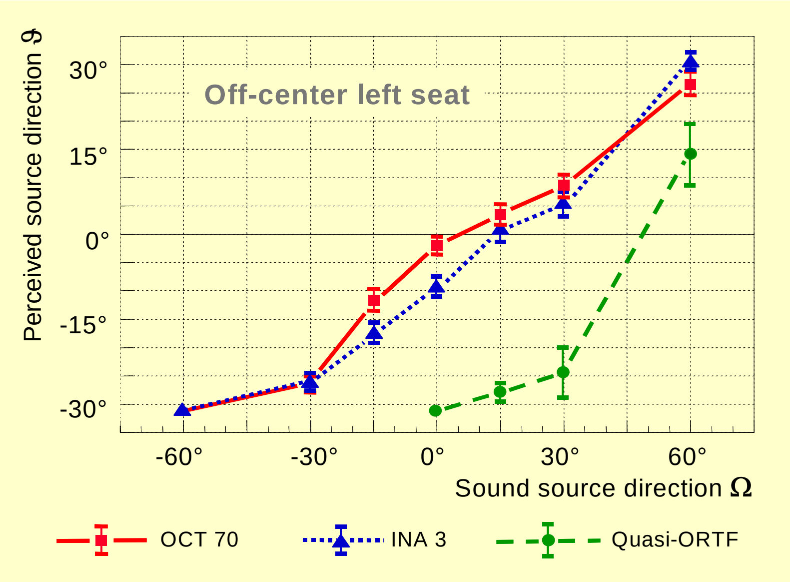



**Fig. 22:**

**Localization curves for non-optimal listening position**

The center channel provides directional stability, however, this depends on the microphone configuration. In the case of INA 3 interchannel crosstalk between channels L and R produces a 10° image shift in the center area.

The image of the L-C-R configuration **INA 3** looks better. However, sound source directions in the center area are shifted about 10° towards the left loudspeaker. This result is due to interchannel crosstalk between channels L and R, inducing too high level in L.

The ***OCT 70*** result is fairly matching the desired linear curve. There is only a very small shift of the center image. Of course it remains a left shift of phantom sound sources in the stereophonic sub-areas. This corresponds with the psychoacoustic laws and could only be avoided by increasing the number of speakers [7], [73].

A further study concentrates on the imaging quality. Each phantom source was compared with a (reference) single loudspeaker positioned in the same direction as the phantom source. It was fed with one of the microphone signals and adjusted to the same level. A five-grade scale was used for comparison tests.

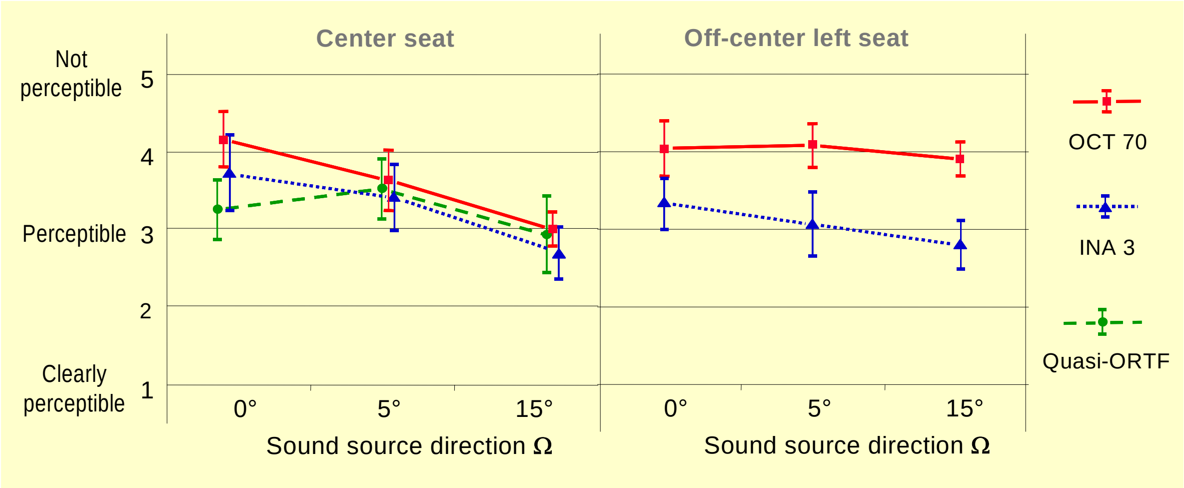



**Fig. 23: Focus of phantom sound source**

The focus of the resulting phantom sound source was compared with the focus of the corresponding real sound source. The differences were assessed in a five-grade scale.

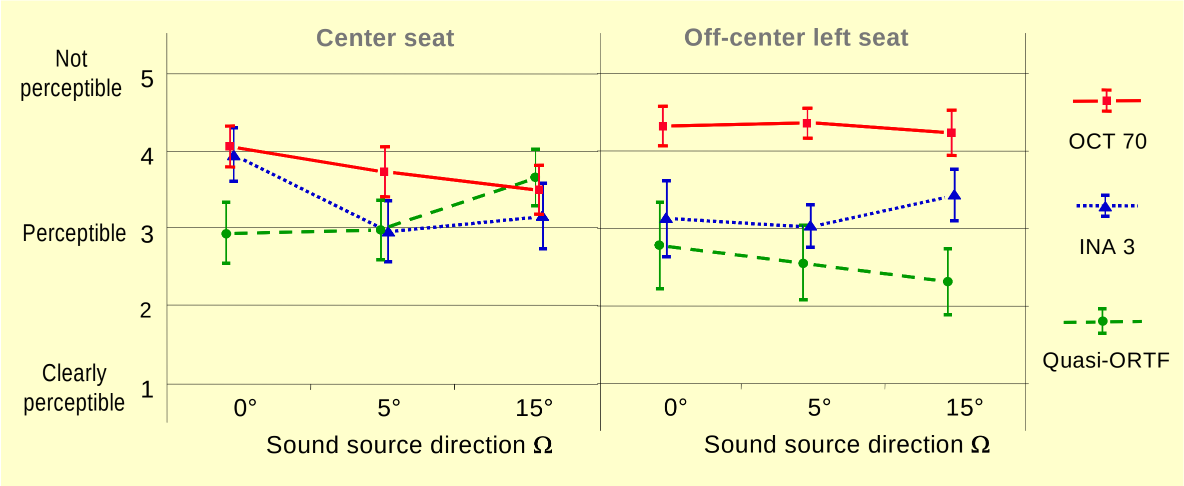



**Fig. 24: Sound colour of phantom sound source**

The sound colour of the resulting phantom sound source was compared with the sound colour of the corresponding real sound source. The differences were assessed in a five-grade scale.

In **Fig. 23** and **Fig. 24** results are plotted. The grade 5 (differences “not perceptible”) is taken as “excellent performance”, the grade 1 (differences “clearly perceptible”) is taken as “poor performance” of the stereophonic microphone. The data confirm first of all that the imaging quality of the phantom sound source is always worse than the corresponding real source. For example, if we look at the **Quasi-ORTF** results (in particular sound source direction $\Omega = 0°$), the center phantom sound source image was assessed about one grade lower than the corresponding center loudspeaker image with respect to both attributes focus as well as sound colour. The L-C-R stereophonic microphone is clearly superior compared with two-channel microphones, not only regarding directional stability (see **Fig. 22**) but also regarding imaging quality.

When comparing the data of the two L-C-R microphones we find superiority of ***OCT 70.*** As expected from the theoretical considerations in **Chapter 2.3 and 2.4** the interchannel crosstalk between L and R must be reduced as much as possible. Due to the crosstalk **INA 3** shows considerable degradations of the imaging quality with respect to focus and sound colour, particularly when the listener is seating off-center left or right.

## 3 THREE-DIMENSIONAL SPATIAL IMAGING

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 roll off due to air absorption. In a recording the directional cues in the direct sound have even less weight than in natural hearing [27]. 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.

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 principle result is a “perspective picture in the simulation plane” between the loudspeakers [12]. It is comparable with spatial visual imaging (**Fig. 25**).: The distance of the 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.

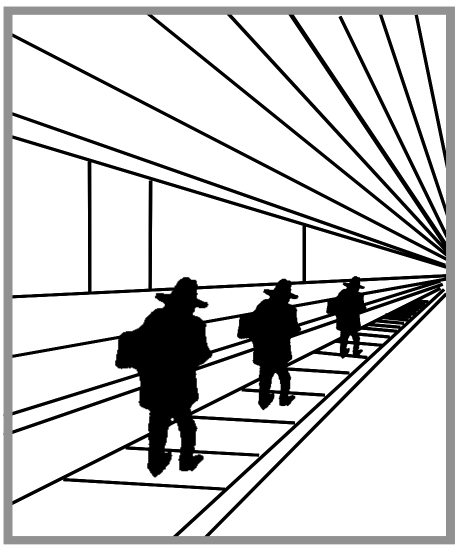


**Fig. 25:**

**Simulation of spatial depth**

The distance of this picture can be compared with the distance of (two-channel) stereo loudspeakers in front of the listener. The visual perspective, which is simulated by applying phenomena of spatial vision, can be compared with the stereophonic perspective, which can be simulated by applying corresponding phenomena of spatial hearing. The dominant cue is a pattern of early reflections in the region 15…50 ms, see next **Fig. 26**.

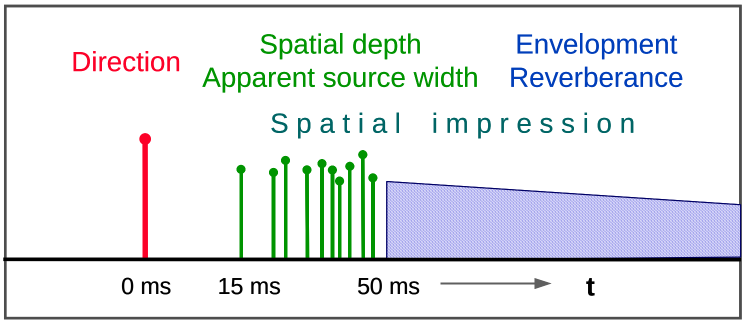



**Fig. 26: Generation of spatial attributes with reflected sound**

The dominant cue to create distance, spatial depth and apparent source width is the natural pattern of early reflections in the region 15…50 ms. In addition with reverberation the spatial impression can be perceived. Furthermore, reverberation can generate perception of reverberance and envelopment.

In the case of the 3/2-stereo format, the listener's acoustic environment can be tailored by the use of additional surround loudspeakers. The reproduction of early lateral reflections (in the region 15 ms –50 ms) leads to a realistic perception of **distance** and **spatial depth**, see **Fig. 27 B.**

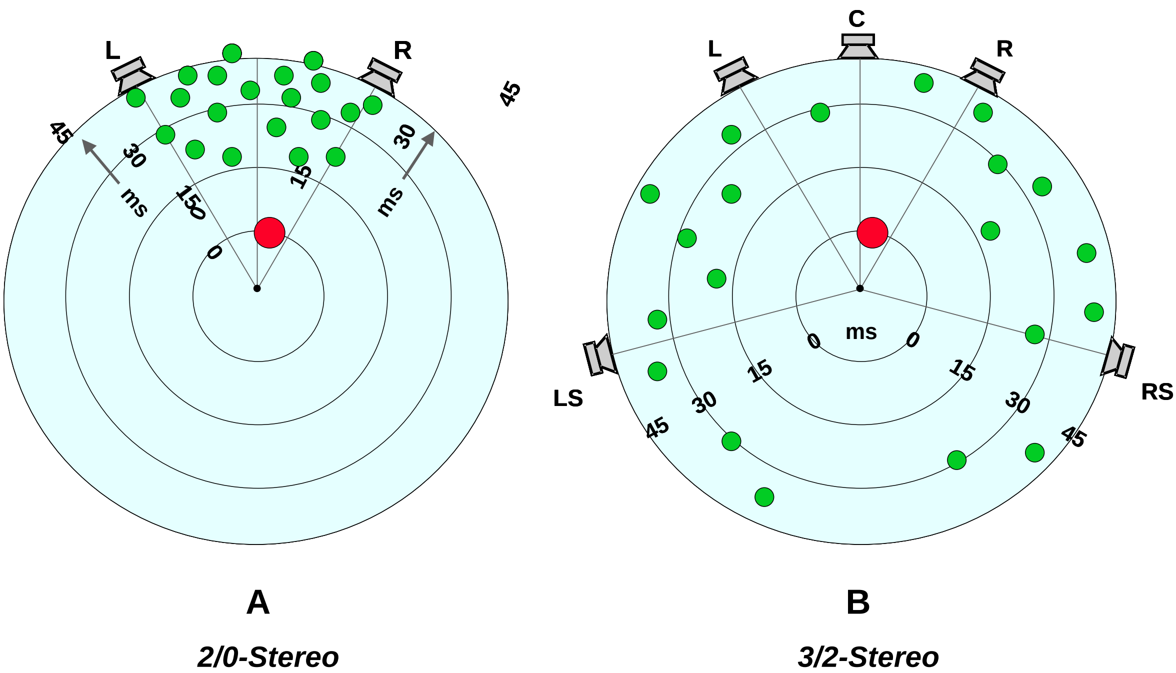



**Fig. 27: Spatial distribution of reflection pattern**

The change from two-cannel stereophony to 3/2-multicannel stereophony implies the change from a simulated to a real impression of space. If the lateral reflection pattern is delivered with the help of additional surround loudspeakers and actually arrives at the listener from lateral directions, the stereophonic quality can be improved dramatically. It is worth to point out that the lateral distribution of reflections allows perception of distance. When we switch from 2/0-stereo to 3/2-stereo we get the realistic impression of a distant source behind the front loudspeakers. The surround sound LS and RS will neither be perceived as lateral or surround sound nor as envelopment. In this application the additional surround speakers do not enlarge the scope of artistic intentions, rather they improve the quality of the stereophonic image in the front by creating depth – an old wish in two-channel stereophony! - However, in combination with reverberant sound, reflected surround sound can create additionally perception of spatial impression, envelopment and reverberance. Overall, if desired, loudspeakers disappear, leaving the listener in the impression that he is in the hall.

It was found earlier in the context of the so-called room-related balancing technique [12], [28] that the presentation of distances in the simulation plane is successful even if only one left and right lateral reflection can be imitated. However a reflection pattern of about ten or twenty early reflections mainly laterally distributed are advantageous, also to avoid colouration effects due to combing (see also **Fig. 43**). It is reported in [29] that early reflections that do not come from the same direction of the sound source can create a perception of distance and depth without affecting the intelligibility or clarity of the sound, and without added muddiness.

Moreover, in combination with reverberation, a largely natural, real **spatial impression** /[4] 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 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. - Experimental results reported in [30] 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 apparent source width. 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 [20]), and to the precedence effect [31]. 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. [32], [33], [34], [35], [58], [65], [71], [72]).

Methods for describing the spatial quality of reproduced sound and the applicability of suitable attributes has been investigated in recent studies [67, [68].

### 3.1 Lateral stereophonic sound

Three dimensional spatial imaging and envelopment requires lateral sound. It has been shown above that the indirect lateral reflection pattern has a certain psychoacoustic function. Furthermore, reproduction of enveloping reverberation as well as the stereophonic representation of environmental (non-reflected) sound (enveloping atmo, applause) are special applications of the lateral stereophonic imaging areas L-LS and R-RS (see **Fig. 28**).

---

/[4] 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. 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. The second is auditory "apparent source width" (ASW), 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 15 to 50 ms after the direct sound. 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.

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 ASW. The dependence of the ASW 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 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.

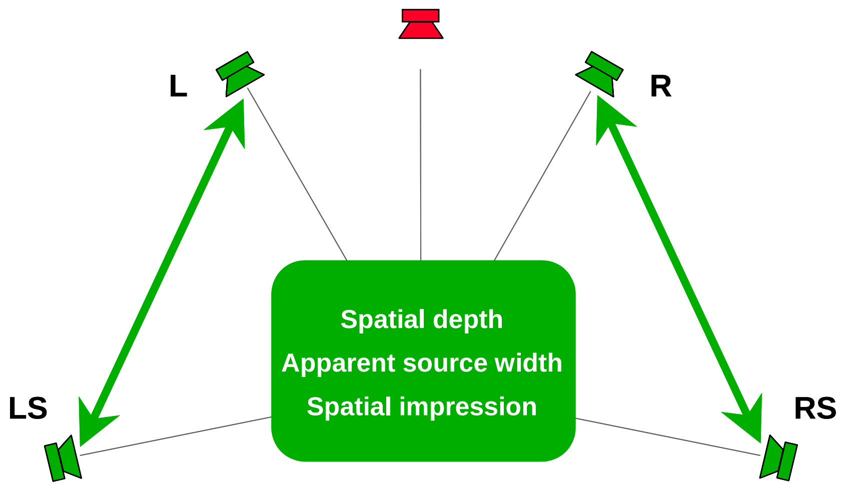



**Fig. 28:**

**Use of the lateral stereophonic areas for spatial imaging**

The stereophonic representation of L-LS and R-RS is aiming to create important spatial sensations, according to **Fig. 26** and **Fig. 27**.

It is not evident that the stereophonic lateral areas can suitably be applied for this purpose, since the loudspeaker arrangement [1] is not intended for lateral placement of stable sources and for presentation of completely surrounding and localisable auditory events. It is well-known in theory [20] and from a number of experiments (e.g. [36], [37], [38], [53]) that lateral phantom sound sources are extremely unstable and sensitive regarding signal spectrum and listener´s location. Corresponding test results are shown in **Fig. 29**.

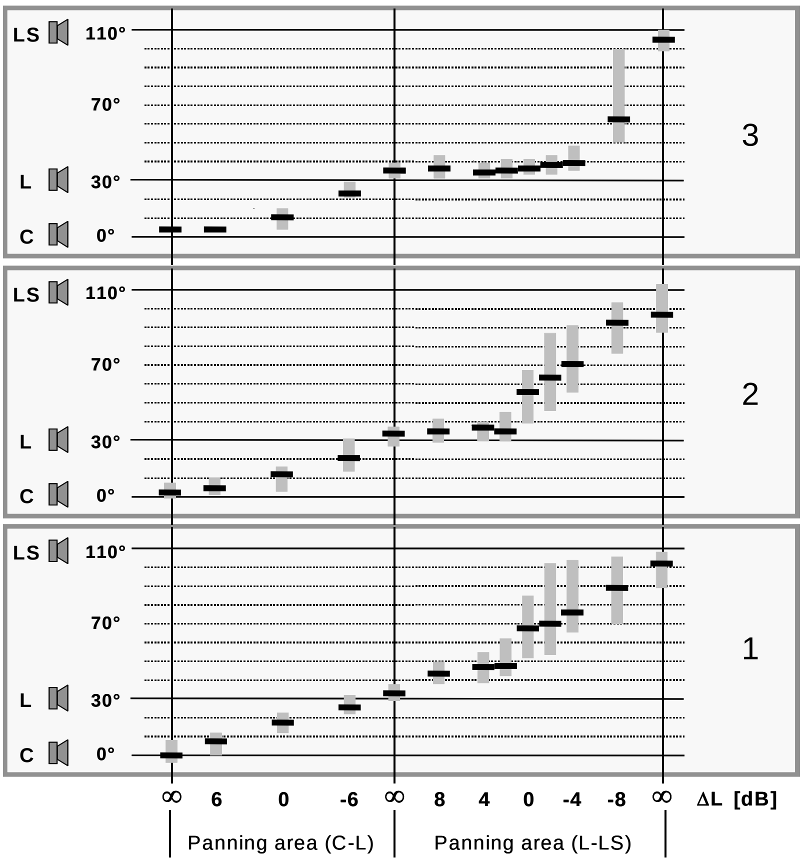



**Fig. 29:**

**Perception of lateral phantom sound sources due to level panning [38]**

Perceived directions are plotted as a function of interchannel level difference in channels C-L and L-LS. Subjects were seated at the sweet spot (plot 1), one seat left hand beside the sweet spot (plot 2), and one seat in front of the sweet spot (plot 3). Standard 5.1 loudspeaker arrangement [1] in the studio, loudspeaker distance 2,5 m. For details see [38].

We can see that the perceived directions of phantom sound sources in the frontal loudspeaker basis C-L are much more resistant against deviations of the seat location from the sweet spot. In contrast, the stability of lateral phantom sound sources (loudspeaker basis L-LS) is rather poor even when the listener is seating within the so called “sweet area”, which is the listening area where the stereophonic quality of the surround sound recording should be sufficiently stable, see **FIG. 30.** It is suggested to be one principle requirement that sufficient stereophonic quality inside the sweet area is ensured. This is not achievable if the mix contains important auditory elements to be localized on basis of lateral phantom sound sources.

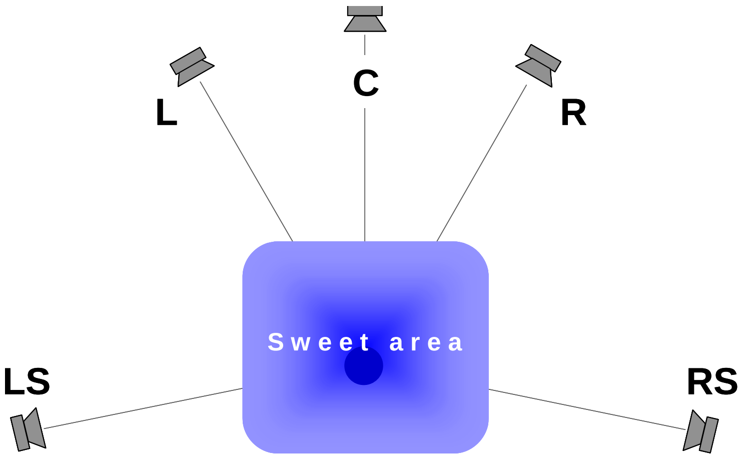



**Fig. 30**

**Sweet area**

It is assumed that the “normal” home sweet area is approximately a 1,5 m square. Inside this area the stereophonic quality of the surround sound recording should be sufficiently stable.

Fortunately, in the case of three-dimensional spatial imaging, directional stability of lateral phantom sound sources is not an indispensable requirement. It can easily be shown that the quality of the stereophonic presentation in terms of spatial depth, apparent source width, spatial impression, and envelopment is surprisingly resistant against deviations of the seat location from the sweet spot. This phenomenon has not yet been investigated carefully. It seems to depend on the characteristics of the reproduced indirect sound.

It is common experience and no surprise, for example, that fully decorrelated signals (interchannel incoherence factor k = 0 ) /[5] feeding two loudspeakers will generate extremely stable images, because phantom sound sources obviously do not contribute. If these decorrelated signals have a different “Gestalt” [20] we will perceive two separated auditory events, for example the first is L, the other is LS. If however the loudspeakers are reproducing decorrelated reverberation, the two sources are fusing more or less, resulting in a broad auditory “cloud” between the loudspeakers. This “stereophonic” image is very stable, but correspondingly it contains no spatial information.

The situation does not change, of course, if the four speakers L, LS, RS, R are used instead of two to reproduce decorrelated reverberation. The result is enveloping reverberance, giving the feeling to be included in the acoustic event, however, producing no spatial impression (see also **Chapter 3.3, FIG. 39**). 100% interchannel incoherence of front signals and rear signals implies that the lateral stereophonic areas are not exploited for reproduction of reflection patterns. Each lateral reflection principally contributes with a certain portion of coherent sound.

It is clear that the lateral stereophonic signals must contain spatial information. Reproduction of natural reflection patterns in the lateral imaging areas L-LS and R-RS is even much more important than in the frontal or rear area. Since the indirect sound does not change directions of auditory events, the principle weakness of lateral stereophonic presentation as described above is kept less annoying if a succession of reflections is reproduced.

---

/[5] The degree of coherence is usually identical with the maximum absolute value of the normalized cross-correlation function (see [31] and also [50]). “Interchannel coherence“ is a particularly relevant measure for interchannel and interaural relationship in spatial hearing. The term “correlation” used in audio practice is not really specified. For example, if we delay one channel of a stereophonic pair of signals, this will not change the physical correlation, rather, the peaks of the normalized cross-correlation function have a corresponding distance. If this distance exceeds a certain limit, stereophonic non-impulsive signals are perceived as “decorrelated”. Usual correlation measurement units for studio applications indicate neither the ”degree of subjective correlation” nor the physical degree of coherence.

If the interchannel time difference between channels L-LS or R-RS becomes too large due to deviations from the sweet spot inside the sweet area, there could still remain a perceptible interchannel relationship giving spatial information. However, a number of questions remain to be answered here.

Obviously additional lateral channels between L-LS and R-RS could improve the performance. A practical solution could perhaps be the layout shown in **FIG. 31**, which is also according to Recommendation ITU-R BS 775-1 [1], see also [7]. The optional loudspeakers LL and RR are fed with supporting information generated from L-LS and R-RS by means of suitable 2 → 3 upwards conversion processing, resulting in an enlarged sweet area and more freedom of creative sound design.

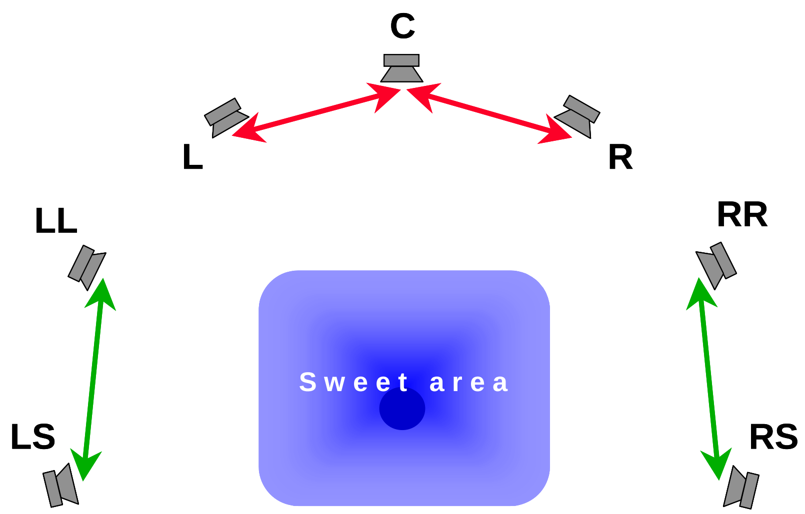



**Fig. 31:**

**Optional speaker arrangement according to ITU-R 775**

Optional channels LL and RR are applied offering improved stereophonic quality of lateral indirect sound inside an enlarged sweet area.

Interesting imaging areas are L-LL and R-RR, for example, to extend the orchestra width and/or to bridge the frontal and lateral image.

## 3.2 Surround sound main microphones

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 principle, 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" [31] during reproduction (direct sound).
>
> **Spatial imaging:**
> Picking up 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 [39] ). For example, under these conditions the so-called sphere microphone [12] 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. We must consider frontal directional imaging by means of the three loudspeakers L, C, R, forming the two stereophonic sub-areas [7] L-C and C-R, and thus requiring a suitable three-channel pick-up of the frontal direct sound. At the same time 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.

The realisation and even the application of a multichannel main microphone appears to be particularly difficult. The microphone has to pick up the direct sound, early reflections, reverberation, and possibly enveloping sources (e.g. applause), and to deliver the complete five-channel mix which must satisfy with respect to many parameters, such as sound colour, directional imaging, spatial imaging, and envelopment (see again **TABLE 2** and **3**). The parameters are governed by psychoacoustic principles and practical constraints leaving not much room to get everything well.

It is always 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. It is suggested that a 3/2-stereo main microphone will be preferred for a small ensemble or a soloist, e.g. piano concert.

Initially three solutions are discussed below which are based on the L-C-R configurations presented in **Chapter 2.3.** Further interesting approaches are introduced for example in [17], [40], [41], [60], [61], [62].

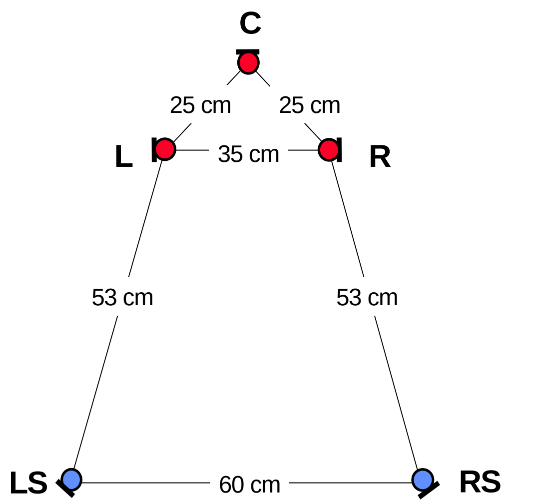



**Fig. 32:**

**Configuration "*INA 5*" [14]**

The L-C-R triangle arrangement "*INA 3*" (see **Fig. 8**, recording angle φ = 180°) is supplemented by the surround microphones LS and RS. The distances are calculated again according to [15], aiming optimum attachment of three additional recording areas (left, right, in the rear), each covering the recording angle φ = 60°. Therefore the surround cardioids are not facing 180° backwards. The off-center angles are ε = 150°.

***INA 5***

The first example shows **Fig. 32**. 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 [15], aiming an attachment of five recording sub-areas:

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

However, regarding the subdivision of recording sectors there are two concerns. The first one has a more principal nature and is related to the psychoacoustic view. It has been shown in **Chapter 3.1** already that lateral phantom sound 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 **Fig. 14**) and accordingly the "*Williams Curves*" [15] are not useful here. With the exception of moving source effects directional imaging of localisable lateral sources can not be recommended, because even at the sweet spot directional stability is critical (see **Chapter 3.1, Fig. 29**). For this reason the exact arrangement of surround microphones on this theoretical basis is not reasonable. In particular, the theory implies that the surround cardioids of ***INA 5*** are not facing backwards, rather, the off-center angles are ε = 150°. In practical situations there results needlessly not the optimum suppression of direct sound, mainly for instruments located at the side of the orchestra (see also **Fig. 35** and **Fig. 40**).

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 2.2**, 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 φ = 180°. This choice is 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, they will pick up too much direct sound from the instruments at the side of the orchestra. 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, as well as the corresponding layout of the direct sound. In the case of ***INA 5*** the wide recording angle φ = 180° limits considerably the possibilities of microphone placement and the optimisation of indirect sound and direct sound at the same time.

Therefore the 3/2-stereo main microphone should provide a suitable compromise: It is suggested to provide a relatively narrow recording angle (φ = 90°...110°) in order to allow the microphone placement in the area of "critical distance" (see **Chapter 3.4**) 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.

### *Fukada-Tree*

An interesting 3/2-stereo main microphone approach is the "*Fukada-Tree*" [18] (**Fig. 33**). The soundstage imaging triangle L-C-R is a modified version of the *Decca Tree*, (see **Fig. 10**), 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 2.3**.

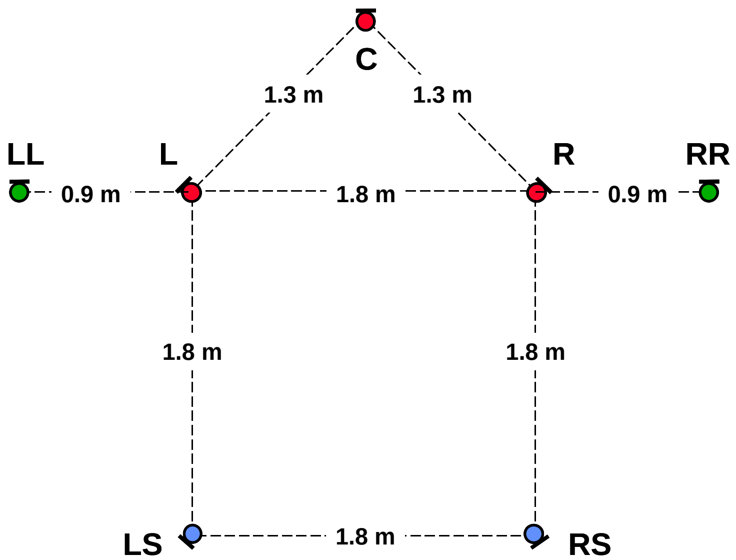



**Fig. 33:**

**"*Fukada-Tree*" [18]**

The triangle configuration is a modified "Decca Tree", The Microphones L, C, R, RS, LS are cardioids. Supplementing 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" [18]. Spatial imaging is realised by means of the 1,8 m spaced square L, R, RS, LS. - It is also stated in [18]: "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."

Because of the wide microphone spacing a recording angle does not exist. Due to the precedence effect all sources in the recording sector around ±45° are reproduced in the center or very close to it. Sources outside the sector ±70° 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 [11]. 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 [42].

### Proposed *OCT-Surround*

As stated above there are more than two parameters to be considered in the optimisation process: frontal direct sound, early lateral reflections, reverberation, and possibly the audience in the hall (see also **Table 3**). How can we control the level balance of lateral reflections and reverberation without affecting the direct sound? How can we control the R/D-ratio in the front channels on the one hand and in the surround channels LS and RS on the other hand?

Assuming 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, we can state:

- The five channel main microphone should reproduce direct sound as well as the indirect sound not only with respect to the original level and time balance but also with respect to the directions of direct sound and early reflections.
- Consequently the surround microphones LS and RS have to suppress the direct sound as much as possible, and the microphone pairs L-LS, LS-RS, RS-R must pickup early reflections and - ideally - to provide a stereophonic representation for the lateral and back areas.

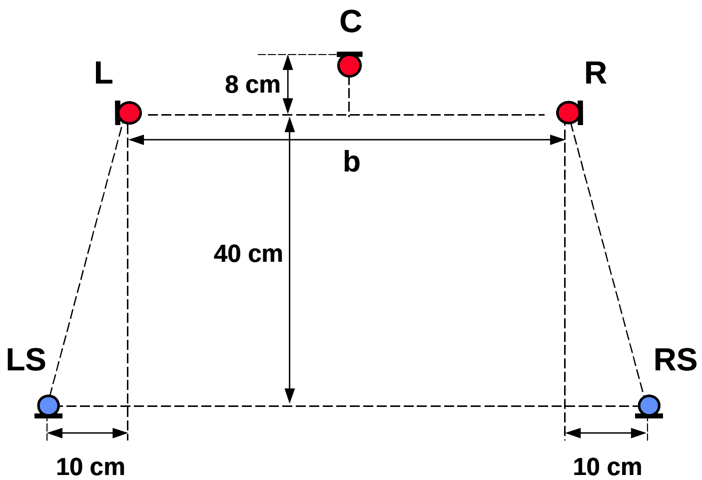
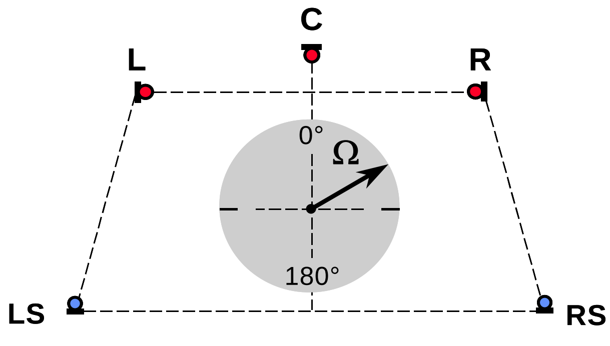



**Fig. 34:**
**"*OCT-Surround*"**

*OCT* is supplemented by LS and RS cardioids facing backwards in order to suppress the direct frontal sound (see **Fig. 35**). In this case the sound colour results from the diffus-field response of the capsules. Good experience with backwards facing cardioids for surround has been made in a number of situations (see e.g. [71], part 2).

Considering this, the so-called "***OCT-Surround***" configuration is proposed to be tested in 3/2-stereo microphone applications. The *OCT* L-C-R microphone is supplemented by a pair of surround cardioid capsules facing exactly backwards (**Fig. 34**). This has the effect that the direct sound from frontal directions will be suppressed according to the directional characteristics of the cardioids. The resulting level balances are plotted in **Fig. 35**.

The table demonstrates satisfying separation of stereophonic imaging areas with respect to any sound source direction. In particular, suppression of direct sound (Orchestra, $\Omega = 0° \dots 45°$) in LS and RS is in the range < - 25 dB … - 13 dB. In contrast, early reflections from any direction are reproduced with adequate level in the corresponding stereophonic pair of channels. This will result in a balanced image of the actual R/D-ratio and of the directional characteristics of early reflections.




| Ω | L | C | R | RS | LS |
|---|---|---|---|---|---|
| 0° | - 4 | + 5 | - 4 | < -20 | < -20 |
| 45° | < -20 | + 3 | + 3 | - 10 | - 10 |
| 90° | - 6 | - 1 | + 5 | - 1 | - 1 |
| 135° | < -20 | - 10 | + 3 | + 3 | + 3 |
| 180° | - 4 | < -20 | - 4 | + 5 | + 5 |

**Fig. 35:**

***OCT-Surround***
**at the critical distance**

The table displays the level of direct sound for a number of sound source directions Ω in relation to the level of indirect sound in the individual microphone channels. In each channel the diffuse sound level is 0 dB.

Regarding the microphone spacing we should consider that the relatively close distance (40 cm) of LS and RS to the front microphones L-C-R is provided in order to ensure optimum stereophonic representation of the lateral sound. Obviously the distance between front and surround microphones is not intended to introduce sufficient delay of direct sound in the surround channels. Rather, the direct sound is sufficiently suppressed, delay is not required for this reason in principal.

Looking at the lateral stereophonic pair L-LS or R-RS and assuming a lateral reflection (Ω = 90°) we gather from the table in **Fig. 35** that the resulting level difference is ΔL = 6 dB. However, due to the geometrical configuration the lateral reflection will lead about 0,3 ms in RS or LS and thus compensate the directional “level bias”.

Looking at the rear stereophonic pair LS-RS and assuming rear reflections we gather from the table in **Fig. 35** that the resulting level difference is always ΔL = 0 dB. For the rear sound a pure A/B-microphone (60…110 cm spaced) is effective.

The behaviour of the *OCT-Surround* system with respect to reverberation and corresponding creation of natural spatial impression and envelopment should be excellent for suitable recording situations. The level of the diffuse sound is equal in each channel, and the adequate degree of correlation is ensured (see also **Chapter 3.3**). Furthermore, the natural level balance of early reflections (from all sides) and reverberation will be preserved sufficiently.

In practical recording situations a successful application of *OCT-Surround* will depend strongly on the adequate choice of the distance and height from the orchestra. Finding the optimum location is a difficult and time consuming task, because many parameters are to be considered, such as loudness and directional balance of instruments, R/D-ratio, sound colour, orchestra width, applause, noise of the audience. It is clear that convincing results can be achieved better with small ensembles, and that sound engineer´s experience is very important.

For *OCT-Surround* applications it may be useful to consider a kind of check list for natural recording (see **Table 7**), showing the suggested order of decisions and practical steps. It is important to determine the optimum microphone location at first dependent on R/D-ratio optimisation and sound colour (check point 2). When the location is found the recording angle can be adjusted according to the resulting directional situation of the instruments. The last two check points will become more clear in the following chapters.

Order of steps:

1. Is natural recording preferable?
2. Where is the optimum R/D-ratio?
3. Suitable recording angle adjustment
4. No separation of main and room microphone?
5. Spot microphones desirable?

**Table 7:**

**Check list for natural recording using a main microphone**

If natural recording is possible and desired, a main microphone can successfully be applied. Optimisation of its location with respect to the balance of direct / indirect sound should be the first preparation step, followed by steps 3, 4, 5. For each decision Tonmeister´s ear is the only relevant touchstone!

## 3.3 Four channel room configurations

Suitable stereophonic representation of early lateral reflections, reverberation and environmental non-reflected sound is required for imaging of distance, spatial depth, spatial impression, and envelopment (see **Table 3**). The *OCT-Surround* microphone is a consequent solution if a five-channel main microphone is desired. It should be emphasized here that there is always the basic requirement of four channels, forming four stereophonic microphone pairs, each covering one side of the room. This is obviously true also for room microphones.

It has been shown in the chapters before that four channels are necessary to pick up lateral sound for imaging of spatial depth and impression. Another important parameter is envelopment. Sensation of envelopment includes two different types of originating sound. The first is reverberation, reflected diffuse sound in the room. The second is environmental, non-reflected diffuse sound, such as applause. Natural sensation of envelopment requires exploitation of four channels as well (**Fig. 36**).

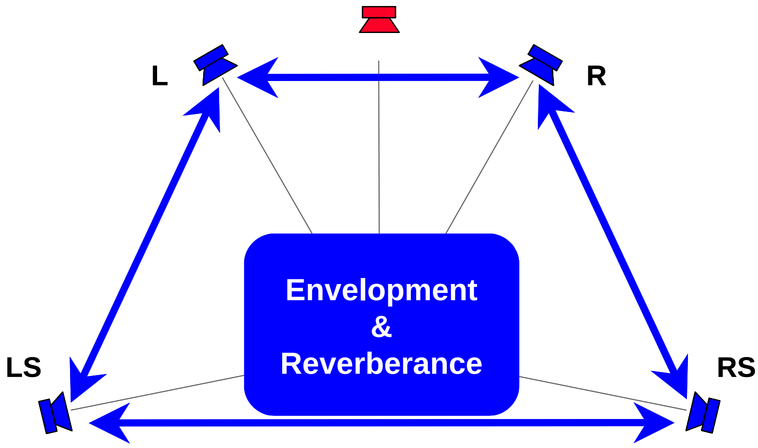



**Fig. 36:**

**Use of four stereophonic areas for imaging of envelopment and reverberance**

Each of the four stereophonic areas should be exploited in order to create realistic sensation of surrounding sound in both situations reverberant sound as well as environmental non-reflected sound (atmo, e.g. applause).

Which kinds of four channel room microphone configurations are suitable? The answer is difficult, and theoretical considerations can only help in practical recording situations. It is not evident that the same four-channel microphone ensures optimum results in any case. For example, experience has shown that the so called “*Atmo-cross*” [44] (**Fig. 37**) which has been optimised for recording the atmosphere (ambient non-reflected sound) is not principally the best solution for room microphone applications in the concert hall. In most situations the direct sound in the hall can cause problems even if the Atmo-cross is delayed correctly according to the concept presented in **Chapter 4**.

In order to catch the lateral sound and to avoid direct sound in the room microphone channels as much as possible the figure of eight microphone arrangement according to **Fig. 38** is the better solution. This so called “*Hamasaki-Square*” has been proposed in [54], [63] and proved very successful in a number of situations, see e.g. [70]. It has become already the standard room microphone in the world-famous Vienna Musikvereinssaal.

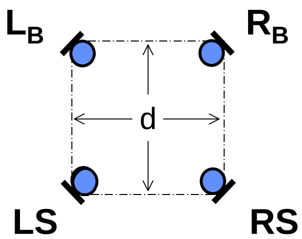



**Fig. 37:**

**"*IRT-Cross*" for atmo recording**

A square arrangement of cardioids (d = 20…25 cm) is known as "Atmo-cross" and offers realistic stereophonic images of ambient sound such as for example applause. More details can be found in [44]. However, this configuration can have disadvantages in room microphone applications according to **Fig. 41** if the R/D-ratio level picked with LB and RB is not large enough (distance from the L-C-R main microphone to short (**Fig. 40**).

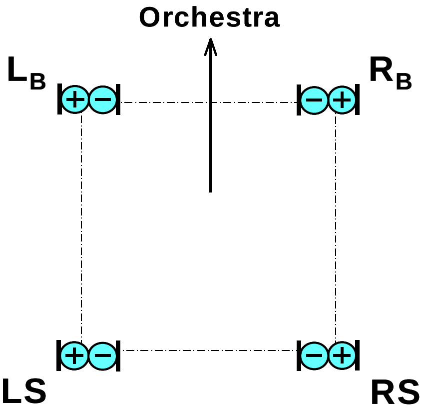



**Fig. 38:**

**"*Hamasaki-Square*" for spatial imaging**

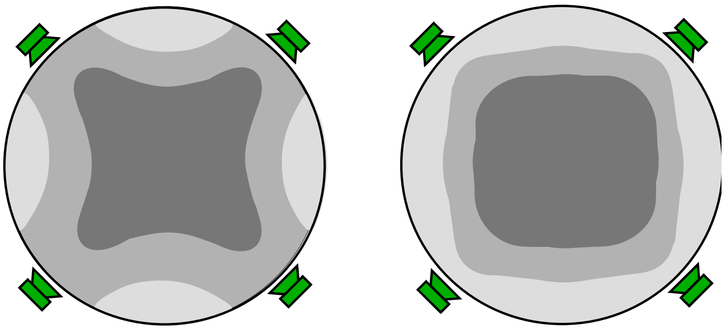
In [54], [63] the square arrangement is proposed to be configured with figure of eight microphones (spacing d = 1 m). The null point of the directional characteristic of each microphone is facing the stage in order to reduce the energy of direct sound as much as possible. For lateral sound a pure A/B-microphone is effective. The array has been placed "at a very high position in the concert hall where the sound is very diffused and energy from direct sound is diminished" [54]. The microphones $L_R$ and $R_R$ are routed to channels L and R or panned between L-LS and R-RS, the microphones LS and RS are routed to channels LS and RS. See also **Fig. 41**.

A general recommendation for optimum spacing and microphone characteristic is not reasonable. Experiments with omnis, cardioids and figure of eights have shown that preference is given dependent on the situation (orchestra, hall, production) and Tonmeister´s intention. Further experiences and listening tests would be useful. Also, a suitable remote control of microphone characteristics would offer flexibility and optimisation according to the actual sound field constellation (for example figure-of-eight for music, and omni for applause).

For optimisation of microphone spacing a relevant aspect is 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 [31]. The resulting sound field can be called "subjectively diffuse" if a well-balanced distribution of auditory events is perceived around the listener.

In [43] the perception of subjectively diffuse sound fields has been investigated, in particular the impact of interchannel coherence 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. 39**.

The degree of subjective diffuseness or subjective envelopment depends on the spacing d 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 nearly coincident, a phantom sound source becomes audible above listener´s head, and subjective envelopment disappears accordingly.



d = 60 cm → k = 0,2 d = 25 cm → k = 0,35

**Fig. 39:**

**Effect of coherence on subjective diffuseness [43]**

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 effects could be studied in practical recording situations, for example, if we compare reverberation or applause in the concert hall recorded with different microphone types in square arrangement and different spacing, e.g. d = 20 cm, 50 cm, 300 cm. Optimum microphone distances should be found in the region of about d = 40 cm…1 m if directional microphones are applied. This is still subject to further experiences. Corresponding investigations are reported in [21], [45], [64]. The results confirm principally the effect shown in **Fig. 39**, however, this is not significant if the room microphone is used in association with the main L-C-R system. The perception of envelopment and homogeneity depends to a large extent on the microphone characteristic and sound field constellation.

Principally at least four equivalent stereophonic channels are desirable for realistic imaging of spatial impression and enveloping atmosphere. The microphone square shall provide suitable lateral sound and interchannel correlation. Usually the placement of a separate square should be far from the critical distance in order to provide a high R/D-ratio, adequate density and spectrum (see next **Chapter 3.4, Fig. 40**). Placement and choice of microphone characteristics will depend on the intended stereophonic perspective and actual recording situation.

It is clear that a number of practical aspects will have an influence and that compromises must be found. For example, if the sweet area 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.

### 3.4 Separation of main and room microphone

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 for useful lateral sound, adequate R/D-ratio in each channel, proper density and spectrum of reverberant sound, suitable interchannel correlation and time-of-arrival relations.

However, a compact 3/2-stereo main microphone would require additional consideration of directional imaging parameters for capsules L, C, R as presented in **Chapter 2.4.** Obviously there result particular disadvantages with regard to the arrangement and directional characteristics of capsules L and R. The front microphone configuration is primarily designed with regard to stereophonic imaging of direct sound, and thus are not optimised with regard to spatial imaging at the same time.

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. Thus in many recording situations a compact main microphone causes an unfavourable compromise. This is true even if we consider the variable recording angle of the *OCT* system, offering some margin in the microphone distance to the orchestra and thus in balancing the R/D-ratio.

It is useful to imagine on the level relation of direct and diffuse sound in the hall, see **Fig. 40**. It is well known experience that the main microphone in most cases is placed approximately near the critical distance, which means that the energy of the direct sound in the surround channels LS and RS of a compact main microphone must be reduced to a great extend.

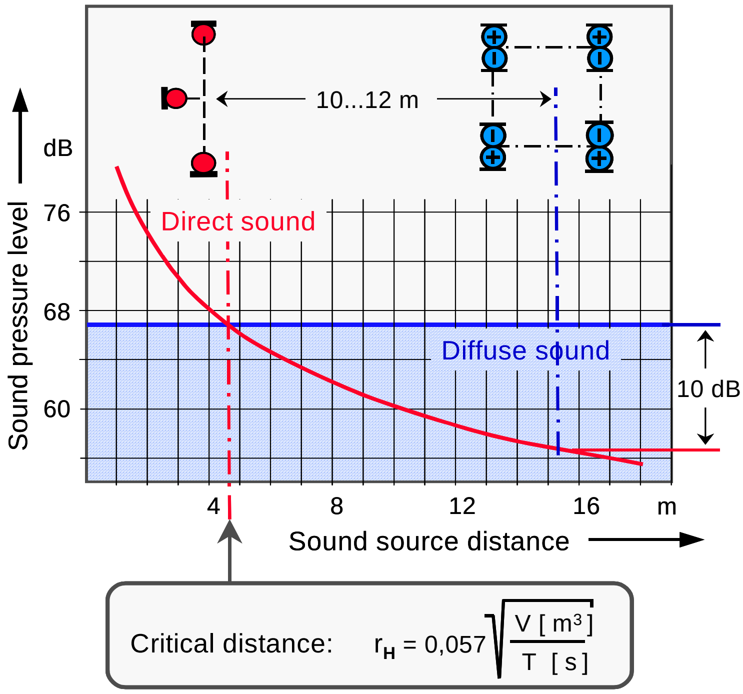



**Fig. 40:**

**R/D-ratio and critical distance in the hall /[6]**

In normal situations the main microphone is located near the critical distance. Suppression of direct sound in the surround channels should be at least 10 dB.
More suppression is necessary if the level of lateral sound in the hall is low. Separation of main and room microphone is beneficial with respect to R/D optimisation, because it does not restrict the margin for the location of the L-C-R part of the main microphone.

**Table 8** shows that the compact main microphone configurations discussed above provide quite different diffuse sound levels in relation to the direct front sound incidence Ω = 0° ("effective directivity indexes").

| | L | R | LS | RS |
|---|---|---|---|---|
| ***INA 5*** | +1 dB | +1 dB | +6 dB | +6 dB |
| ***Fukada Tree*** | -3 dB | -3 dB | +10dB | +10 dB |
| ***OCT-Surround*** | +4 dB | +4 dB | > + 15 dB | > + 15 dB |

**Table 8:**

**R/D-ratios of the main microphone capsules resulting at the critical distance**

**Fig. 40** illustrates the advantage of room microphone placement far behind the L-C-R main microphone: reduction of direct sound energy in the surround channels is less difficult. On the other hand, it illustrates that the distance of about 10…12 m (which would be the right choice regarding delay of direct sound) may not be large enough with respect to the R/D-ratio if omnis are used, we can expect only less than about 10 dB here.

---

/[6] **Critical distances of music halls [59]**

| | V [m³] | T [s] | $r_H$ [m] |
|---|---|---|---|
| Musikvereinssaal, Wien | 14600 | 2,2 | 4,6 |
| Herkulessaal, München | 13400 | 2,0 | 4,7 |
| Théatre National de l´Opéra, Paris | 9960 | 1,1 | 5,4 |
| Teatro alla Scala, Mailand | 11250 | 1,2 | 5,5 |
| Neues Gewandhaus, Leipzig | 21560 | 2,0 | 5,9 |
| Philharmonie, Berlin | 24500 | 2,0 | 6,4 |
| Cornegie Hall, New York | 24300 | 1,7 | 6,8 |
| Royal Albert Hall, London | 86600 | 2,5 | 10,6 |

If we are using a single compact five-channel microphone we should not touch the front/back level balance in order to keep the stereophonic representation of lateral sound (see **Chapters 3.1** and **3.2**). In this case front/back level balancing 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.

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. 40** and **Fig. 41**. Regarding directional imaging, a suitable front channel microphone configuration (e.g. *OCT* according to **Chapter 2.4**) 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.

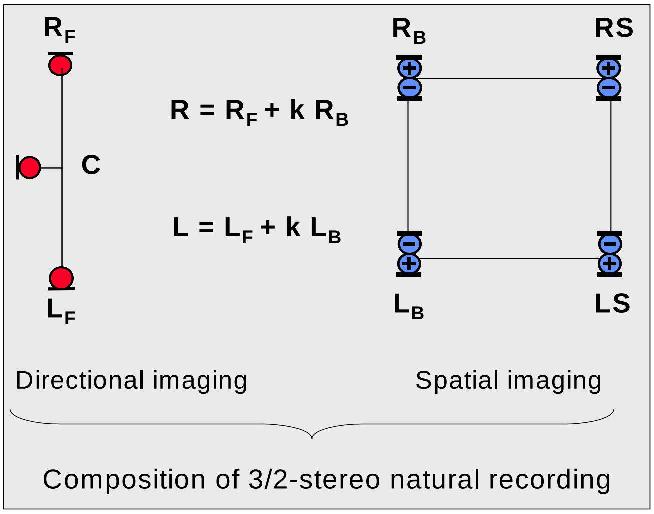



**Fig. 41:**

**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. The L-C-R main microphone as well as the room microphone configuration can be optimised and placed with respect to its purpose and related psychoacoustic principals, in particular 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 can be created.

In this way the microphone square can be applied providing suitable information 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 (**Fig. 40**).

Separation of L-C-R main microphone and room microphone allows level balancing of direct /indirect sound without affecting the stereophonic representation of lateral sound.

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 the next **Chapter 4**, without touching the desirable stereophonic four-channel information picked up with the room microphone square.

### *Listening tests*

Various multichannel microphone set-ups have been investigated recently in practice (classical music recording at the ORF-Radio station in Vienna [70]). Among other configurations the figure of eight ***"Hamasaki-Square"*** was used in combination with ***OCT***, ***INA 3***, and ***Decca-Tree***, moreover the 3/2 stereo systems ***OCT-Surround*** and ***INA 5***. Careful listening tests were performed to examine the performances with respect to the spatial presentation of the orchestra, spatial imaging of the concert hall, and sound colour. A detailed report will be available soon [75].

## 4 SPATIAL DESIGN USING DELAY

Natural imaging of depth and spatial impression requires careful layout of the delay situation, that is to say according to principles of the room-related balancing technique (RRB) [12], [28], [51]. The basic concept is indicated in **Fig. 42**, showing the case "orchestra, main microphone plus room microphone and spot microphones", as an example. It should be emphasised here that the concept is not restricted to the main / spot microphone method but can be applied accordingly to poly-microphony [5], [6], [29].

In the case of undelayed spot microphone signals (**Fig. 42**, upper situation) they are reproduced earlier than the corresponding main-microphone signal. Thus the ear interprets the spot-microphone signal as the direct sound. Due to the precedence effect (e.g. [31], see **Fig. 43**) the localisation of an auditory event is in principle determined by the sound arriving first at the ear. Accordingly the panned direction and the "close" character of the added spot microphone signal is dominant, and favourable characteristics of the time information of the indirect sound are lost. Such recordings sound unnatural, without spatial depth and spatial impression.

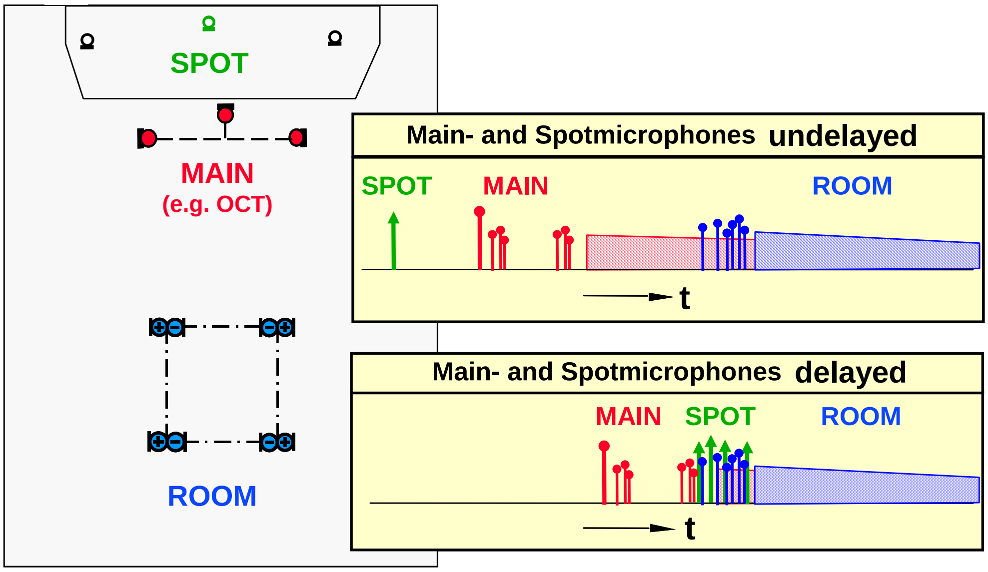



**Fig. 42: Delay design according to the natural room response**

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. In cases where an L-C-R main microphone is used in combination with a (square) room microphone it is important to adjust the arrival-time gap due to the natural delay of early reflections. In special recording situations (particularly small ensembles, e.g. chamber music) the resulting stereophonic image can be excellent with respect to the directional balance and spatial attributes. If corrections with respect to the loudness balance, sound colour etc. are desired, the spot microphones should be delayed according to the **Room Related Balancing (RRB)** concept [12], [28].

It is not possible to moderate this space-disturbing effect by compensating the delay of the main-microphone signal. Those technique is 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 depth, apparent source width and spatial impression, the spot-microphone signal should be converted to a number of reflections. This reflection pattern is delayed much more than necessary for the compensation, so as to fall within the region of the early reflections ("arrival-time gap" [31], [9] of about 15...25 ms, see **Table 9** and **Fig. 45**).

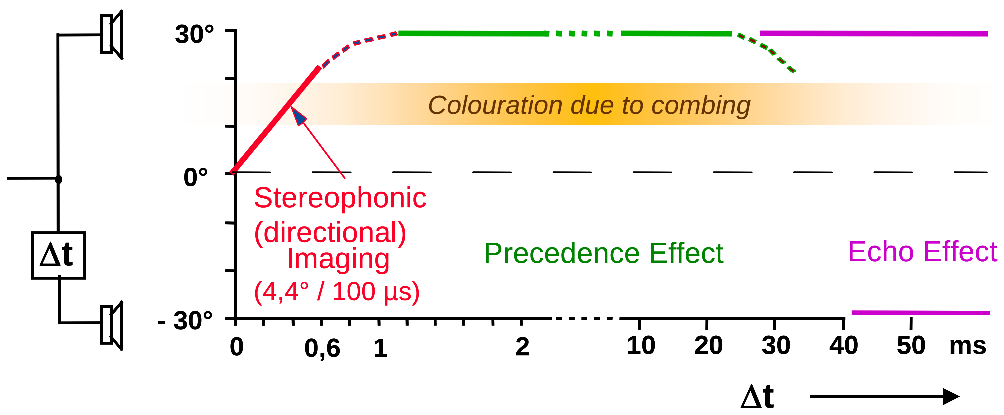



**Fig. 43: Delay effects**

Interchannel delay inducing phantom sound source shift, precedence effect, or echo. In the range Δt = 0,6…40 ms the lagging source can cause colouration resulting from combing effects at listener´s ears. Colouration depends strongly on the source nature and on the number of reflections.

The RRB technique causes that the temporal reflection pattern of direct sound and early reflections, which is given by a surround sound main microphone concept such as shown in **Fig. 43**, remains authentic (**Fig. 42,** 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).

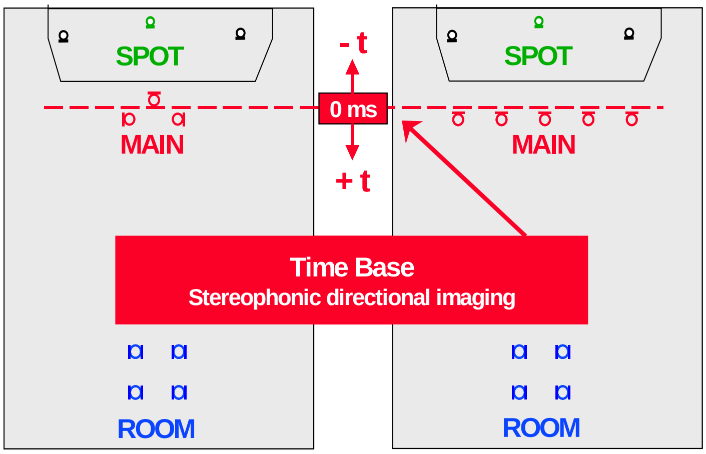
Since the spot microphone signal is converted to a number of delayed reflections, it will not impact the directional information (direction of phantom sound source) produced by the main microphone. Consequently the RRB technique is not applicable if the main microphone does not produce a satisfying directional image (directional balance, see e.g. **Fig. 12 C**). For example, if a *Decca Tree* (**Fig. 10**) or widely spaced microphones ("*Multiple-A/B*", e.g. **Fig. 6**) are used, the result would be unsatisfying because delayed spot microphones do not provide a desirable directional stable images. Sound engineers who give preference to widely spaced main microphone configurations obviously reject from spot microphone delay techniques. The two principal concepts are illustrated in the following table:

| Concept | L-C-R Main Microphone | Spot Microphones |
|---|---|---|
| A | *OCT* | delayed (RRB) |
| B | *Decca Tree* | undelayed |

Moreover, as regards the complete temporal pattern, **Fig. 42** 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, see **Fig. 43**).

## 4.1 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 9** (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. 44**.




Example1 Example 2

**Fig. 44:**

**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. 42** and **Table 9**).

The basis of a delay plan is an exact conception of the reflection pattern to be reproduced. This is resting on the natural situation as presented in **Fig. 27 B**. The intended result to be created is shown in **Fig. 45**. The delay plan determines the temporal order and the spatial allocation of direct sound and early reflections for the sweet spot of the listening area. (see **Fig. 42,** 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. 42** and **Table 9** this is the main microphone for the complete orchestra. All further microphones involved supply either leading or lagging signals (column 2 of **Table 9**).

| Microphone | Lead / Lag ( + / - ms) | Compensa-tion (ms) | | Arrival-time Gap (ms) | Comp+Gap (ms) | Resulting Delay (ms) | Initial Direc-tion |
|---|---|---|---|---|---|---|---|
| Main L | Time Base | | 0 | 0 | 0 | - 45 | - |
| Main C | Time Base | | 0 | 0 | 0 | - 45 | - |
| Main R | Time Base | | 0 | 0 | 0 | - 45 | - |
| Spot A | + 25 | Refl. 1: | - 25 | - 12 | - 37 | - 82 | - 30° |
| | | Refl. 2: | - 25 | - 9 | - 34 | - 79 | + 30° |
| | | Refl. 3: | - 25 | - 17 | - 42 | - 87 | - 110° |
| | | Refl. 4: | - 25 | - 20 | - 45 | - 90 | + 110° |
| Spot B | + 35 | Refl. 1: | - 35 | - 19 | - 54 | - 99 | - 30° |
| | | Refl. 2: | - 35 | - 21 | - 56 | - 101 | + 30° |
| | | Refl. 3: | - 35 | - 22 | - 57 | - 102 | - 110° |
| | | Refl. 4: | - 35 | - 25 | - 60 | - 105 | + 110° |
| Spot C | + 45 | Refl. 1: | - 45 | - 17 | - 62 | - 107 | - 30° |
| | | Refl. 2: | - 45 | - 11 | - 56 | - 101 | + 30° |
| | | Refl. 3: | - 45 | - 19 | - 64 | - 109 | - 110° |
| | | Refl. 4: | - 45 | - 23 | - 68 | - 113 | + 110° |
| Room L | - 60 | | + 60 | - 15 | + 45 | 0 | - 30° |
| Room R | - 60 | | + 60 | - 15 | + 45 | 0 | + 30° |
| Room LS | - 60 | | + 60 | - 15 | + 45 | 0 | - 110° |
| Room RS | - 60 | | + 60 | - 15 | + 45 | 0 | + 110° |

1 m ↔ 3 ms / 1 ms ↔ 0,33 m

**Table 9: Proposed delay plan for practical recording application (Example 1)**

The delay plan corresponds to the situation shown in **Fig. 42**. In this example an L-C-R microphone (e.g. *OCT*) and the square room microphone are applied, as indicated in **Fig. 41**. It is assumed that the room microphone is located 20 m behind the main. In order to achieve a suitable arrival-time gap between main (direct sound) and room microphone (indirect sound), the main needs 35 ms delay.- Furthermore, it is assumed that three spot microphones A, B, and C are required to balance the loudness and/or sound colour of orchestra elements. In order to preserve the quality of directional and spatial impression the "RRB" concept is applied: 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. 43**.

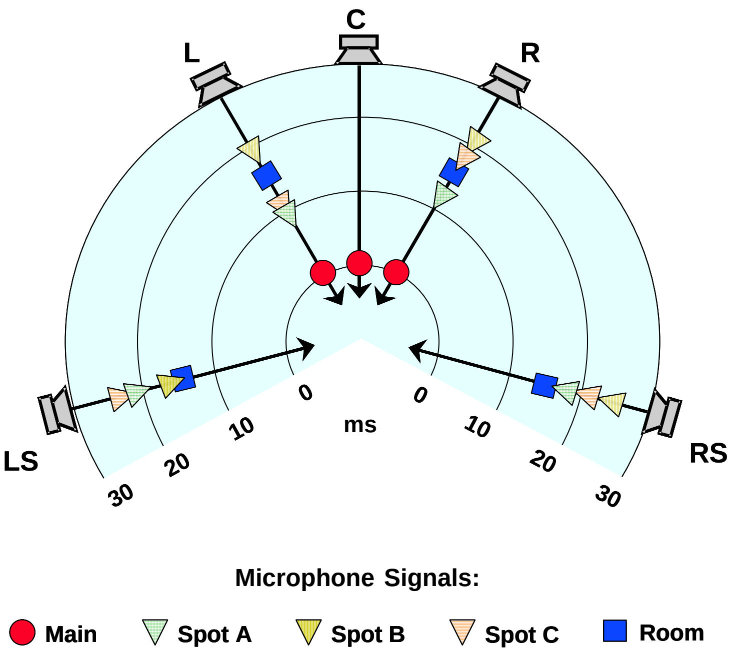



**Fig. 43:**

**Delay design according to the natural room response (Example 1)**

The arrival-time gap values plotted in column 5 of **Table 8** 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 inside the sweat area 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 3**. 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. 10 …. 20) 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 [12], [27], [29].

---

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" [31] during reproduction.

The individual time-arrival gaps should be 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 spot microphone levels within a wide range, without changing the perception of direction and spatial depth.

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 8**, 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 [5], [6]. Digital processing allows to realise further optimisation such as panning across the lateral channels (according to **Fig. 27 B**), 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) [5], [12], [28].

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 RRB concept can also be applied to polymicrophony. Modern mixing consoles could comprise room-related balancing tools for the synthesis of arbitrary natural spatial impressions [5], [6].

## 4.2. 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 10**. 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 spatial depth.

As regards the spatial information, the downmix result concerning temporal order of direct sound, early reflections and reverberation is correct, see **Fig. 46**. The resulting pattern of reflections corresponds with the desirable pattern for natural two-channel recordings as described in [12], see also **Fig. 27**: 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.

| | |
|---|---|
| $L_0$ = L + 0,7C + k LS | ***The choice:*** |
| $R_0$ = R + 0,7C + k RS | **k = 1, 0.71, 0.50, 0.36, 0** |

**Table 10: 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” [46] or “Meta-data” [47] 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. The coefficient k = 0 is important in cases where the 3/2-stereo mix has been generated from an existing two-channel mix (upwards conversion).

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 mainly 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 but even for two-channel reproduction [12]. This is true also with respect to quality improvements due to “natural panning” suggested in [5], [6], [28].

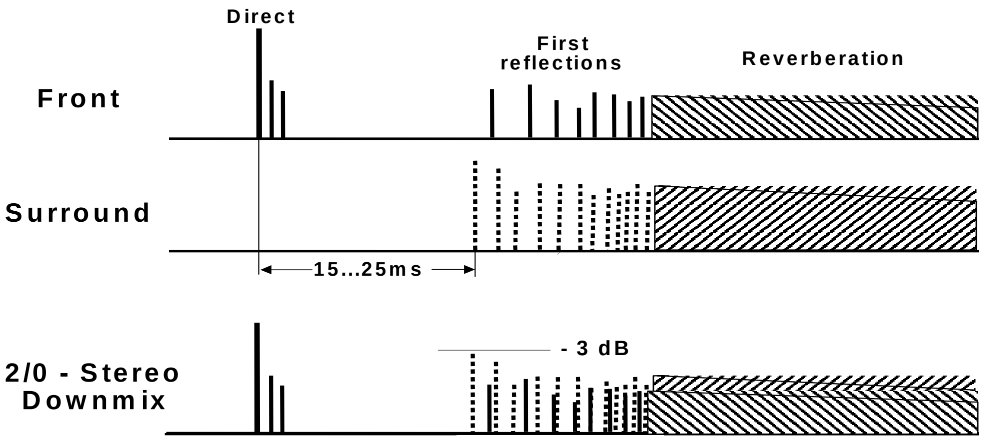



**Fig. 46: Pattern of reflections in the 3/2-mix and in the 2/0-downmix**

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 [31], 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 programme material.

Different experience is reported in [18], [48]: “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.

Considering **headphone reproduction** it must be stated that the simple downmix according to ITU-R BS 775-1 (**Table 10**) does not represent the optimal solution. The well-known “in-head localisation” effect [31] 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 [49], [69] 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.

## REFERENCES


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

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

[03] Boone, M., de Vries, D.: Wave Field Synthesis Using Array Technology.
20th Tonmeistertagung, 1998, Proceedings (ISBN 3-598-20361-6), pp. 265-279

[04] de Bruijn, W., Boone, M., de Vries, D.: Sound Localisation in a Videoconferencing System based on Wave Field Synthesis. 108th AES convention, 2000, Preprint 5144

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

[06] Horbach, U., Karamustafaoglu, A.; Corteel, E.; Pellegrini, R.: Implementation of an Auralization Scheme in a Digital Mixing Console using Perceptual Parameters.
108th AES Convention, 2000, Preprint 5099

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

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

[09] 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

[10] 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

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

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

[13] 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

[14] 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

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

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

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

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

[19] Theile, G.: Multichannel natural music recording based on psychoacoustic principles. AES Convention, 2000, Preprint 5156, supplementing handout February 2000

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

[21] Wittek, H.; Neumann, O.; Schäffler, M.; Millet, C.: Studies on Main and Room Microphone Optimization. AES 19th Intern. Conference, 2001, pp 448-455

[22] Wittek, H., Theile, G.: Investigations on directional imaging using L-C-R stereo microphones. (German) 21. Tonmeistertagung 2000, Proceedings pp. 432-455

[23] Wittek, H.: "Image Assistant". www.hauptmikrofon.de

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

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

[26] Wuttke, J.: General Considerations for Multi-Channel Audio Recording, AES 19th Intern. Conference, 2001, pp 167-175. See also: SCHOEPS: Catalogue 4, 2001

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

[28] 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.

[29] Griesinger, D.: The theory and practice of perceptual modeling – how to use electronic reverberation to add depth and envelopment without reducing clarity. 21. Tonmeistertagung 2000, Proceedings pp. 766-796

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

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

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

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

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

[35] Steinke, G.: How many channels/signals are needed? Pleading for the 3/2 Stereo-Hierarchy and its optimization in the Home (German). 21th Tonmeistertagung 2000.

[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. J. Audio Eng. Soc. (25), 1976, 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] Keibs, L.: Stereo-Ambiophone Schallübertragung (Stereo-ambiophonic sound transmission). Acustica (12), 1962, pp. 118-124

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

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

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

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

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

[45] Neumann, O., Schäffler, M.: Optimizing the room microphone configuration for 3/2-stereo multichannel music recordings (German). Diplomarbeit, FH Stuttgart – Hochschule für Druck und Medien, 2001

[46] 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)

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

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

[49] 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

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

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

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

[53] Martin, G.; Woszcyk, W.; Corey, J.; Quesnel, R.: Sound source localization in a five-channel surround sound reproduction system. 107th AES Convention, 1999, Preprint 4994

[54] Hamasaki, K.; Fukada, A.; Kamekawa, T., Umeda, Y.: A concept of multichannel sound production at NHK. 21. Tonmeistertagung, 2000, Proceedings Paper only on CD-ROM

[55] Griesinger, D.: Speaker placement, externalization, and envelopment in home listening rooms. 105th AES Convention, 1998, Preprint 4860

[56] Rumsey, F.: Surround sound: placing European research in an international operational context. 21. Tonmeistertagung, Proceedings pp. 44-62

[57] Mason, R.; Rumsey, F.: An investigation of microphone techniques for ambient sound in surround sound systems. AES 106th Convention, 1999, Preprint 4912

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

[59] Forsyth, M.: Buildings for Music. MIT Press; 1985. ISBN: 0521268621

[60] Segar, P.; Rumsey, F.: Optimisation and Subjective Assessment of Surround Sound Microphone Arrays. 110th AES Convention, 2001, Preprint 5368

[61] Williams, M.; Le Du, G.: The Quick Reference Guide to Multichannel Microphone Arrays - Part I : using Cardioid Microphones. 110th AES Convention, 2001, Preprint 5336

[62] Williams, M.; Le Du, G.: Multichannel Microphone Array Design. 108th AES Convention, 2000, Preprint 5157

[63] Hamasaki, K.; Shinmura, T.; Akita, S.; Hiyama, K.: Approach and Mixing Technique for Natural Sound Recording of Multichannel Audio. AES 19th Intern. Conference, 2001, pp 176-181, p. 447

[64] Millet, C.: „Studies on the performance of the BRS system concerning soundfield assessment". Diploma work, Conservatoire de Paris, Cité de la Musique, France, 2001

[65] Griesinger, D.: The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings, and their relationship to Microphone Technique. AES 19th Intern. Conference, 2001, pp 182-200

[66] Gernemann, A.: Stereo+C – An All Purpose Arrangement of Microphones Using Three Frontal Channels. 110th AES Convention, 2001, Preprint 5367

[67] Berg, J.; Rumsey, F.: Verification and correlation of attributes used for describing the spatial quality of reproduced sound. AES 19th Intern. Conference, 2001, pp 233-251

[68] Zacharov, N.; Koivuniemi, K.: Unravelling the perception of spatial sound reproduction: Techniques and experimental design. AES 19th Intern. Conference, 2001, pp 272-286

[69] Spikofski, G.; Fruhmann, M.: Optimization of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics. AES 19th Intern. Conference, 2001, pp 124-135

[70] Camerer, F.; Sodl, C.: Classical Music in Radio and TV – a multichannel challenge. InterBEE Tokyo, Nov. 2001

[71] Holman, T.: Capturing the sound. Part 1-5. Surround Professional, Sept. 2000 pp. 46-48, Oct. 2000 pp. 46-49, Nov. 2000 pp. 41-44 and 54, Dec. 2000 pp. 39-42 and 56, Jan./Feb. 2001 pp. 37-40

[72] Dabringhaus, W.: 2+2+2 – the 5.1 reproduction chain (DVD-A) gives us the chance to be used as a true three-dimensional reproduction system for classical music with three stereophony channels (German). 21. Tonmeistertagung, 2000, Proceedings pp. 265-272

[73] Rebscher, R., Theile, G.: Enlarging the listening area by increasing the number of loudspeakers. 88th AES Convention, 1990, Preprint No. 2932

[74] Rumsey, F.: Spatial Audio. Music Technology Series, Focal Press 2001. ISBN 0240 51623 0

[75] Camerer, F.: Listening Comparison of 7 Surround-Sound Microphone Setups for Classical Music Recording. 112th AES Convention, May 2002, in preparation