Double M/S – a Surround recording technique put to test

This text is essentially based on a paper presented at the Tonmeistertagung 2006:
"Doppel-MS - eine Surround-Aufnahmetechnik unter der Lupe" by

Helmut Wittek, Christopher Haut, Daniel Keinath.

For comments refer to wittek@schoeps.de.
English translation by Kevin Schädler.

Abstract

Double-M/S is a recording technique for two-channel as well as multichannel stereo. It has considerable practical advantages which justify its growing popularity.

What is really behind Double M/S?

Closer examination shows that it can provide high-quality pickup not only of ambient sound but also of music, provided that certain criteria are considered in the dematrixing/decoding of the three microphone signals. Recordings made simultaneously with the "IRT Cross" and "OCT Surround" reference systems reveal that Double M/S can achieve comparable overall quality, given optimal decoding.

The decisive parameters for the quality of decoding are crosstalk level, localization, and the degree of correlation among the loudspeaker signals. These characteristics can be optimized only if the L/R and LS/RS channel pairs are both obtained from all three of the microphone signals. In other words, the optimal decoding needs a variable M pattern generated through a mix of front- and rear-facing cardioids. This capability also makes Double M/S an interesting option for two-channel recording, similar to Ambisonics.

Experiments have revealed the maximum allowable crosstalk level above which the sound color and sound image begin to be adversely affected.

The Double M/S system turns out to be a system that is suitable for more than just the usual applications in documentary sound and radio drama.

Contents

Introduction

The Double-M/S recording technique is now quite popular as well as being one of the most established recording techniques for certain applications. Nonetheless, relatively little is known about its properties, and thus there is a need for objective descriptions and the sharing of experiences concerning this type of microphone arrangement. As well as that, the various methods of decoding the M/S signals and the existence of new tools for optimized decoding are still relatively unknown.

The author himself has fallen into the common trap of assuming the inferiority of M/S-recording to other techniques in recent years, but has reviewed this opinion. The M/S recording technique is by no means a compromise, just as using M/S for multichannel recording is not a problem as long as the properties of the technique are clear to the user. As with many things, the underlying principle is that familiarity with the strengths and weaknesses of a technique allows optimal results to be obtained; no technique is flawless. Luckily, there is no such thing as a "fool-proof" recording technique; any method or product which claims to be must be regarded critically because the results obtained will be neither good nor bad, and thus hardly satisfying. Similarly, the Double-M/S technique will produce useless or disappointing results if it is decoded wrongly; the engineer must be in full control of the decoding for satisfactory recordings. Like other techniques, it also requires care and attention to be taken to ensure quality, as for example the setting of the offset angle in XY-recording. This seemingly obvious idea shall now be demonstrated by the results of the investigations of this paper.

The paper is divided into a theoretical and a practical analysis of Double-M/S recording. The theoretical analysis investigates important parameters of the microphone arrangement such as channel correlation in diffuse fields, directional imaging characteristics and crosstalk behavior. By analyzing these parameters, important characteristics of the microphone arrangement can be predicted; this objective analysis simplifies the assessment of existing arrangements.

However, the significance of this analysis is also limited as not everything is explainable and predictable by it. This is partly due to the fact that stereophonic and spatial perception have complicated underlying attributes which are not fully researched to this day. Another reason is that various other individual parameters affect recording quality so that it is impossible to make general statements. Examples of these parameters include the type of recording room, acoustic source and musical content and of course the individual taste of the recording engineer and listener.

For this reason, it is dangerous to use purely theoretical analysis without practical trials and experience to postulate general rules about optimal recording techniques. This is especially true of coincident recording (in relation to the Ambisonics playback technique); a lot of theoretical research has been done which claims its superiority (eg. [7],[15]). The problem with this research is that it relies on certain basic presumptions which are in dire need of critical review, for which reason the conclusions reached mostly concern directional depiction and fail to include other important parameters.

This paper shall try to analyze the applications of coincident recording in surround sound in an unbiased, critical objective manner as well as including some subjective analysis on the topic.

1 The M/S principle

The signals of a coincident microphone arrangement can be matrixed with the M/S (Mid/Side) principle by calculation of sum and difference values. The "Mid" signal is the sum of the mono signals, and the "Side" signal is their difference. L and R signals are reobtained by combining sum (M) and difference (S) values.

$$M = L + R$$
$$S = L - R$$

$$L = k_1 * M + k_2 * S)$$
$$R = k_1 * M - k_2 * S)$$

to reobtain the original signals set
k_1 and k_2 = 0.5

A distinct advantage of M/S coding is that it allows the recording angle and stereo width to be varied by simply trimming the factors k_1 and k_2 which control the levels of the M/S signals.

Figure 1:

A typical setup for M/S stereophonic recording: shotgun microphone with an attached figure-8 microphone for use in a windjammer. [20]

In M/S recording, two microphones, one for the M-signal and one for the S-signal, are used to record directly in coded form. The M-microphone is directed forwards, whereas the S-microphone directed perpendicular to the shotgun microphone's axis. Every M/S arrangement has an XY equivalent to which it can be converted. Figure 3 illustrates various combinations (= different factors k_1, k_2) of M/S signals (M=cardioid) in the top row with their equivalent XY arrangements after decoding underneath.
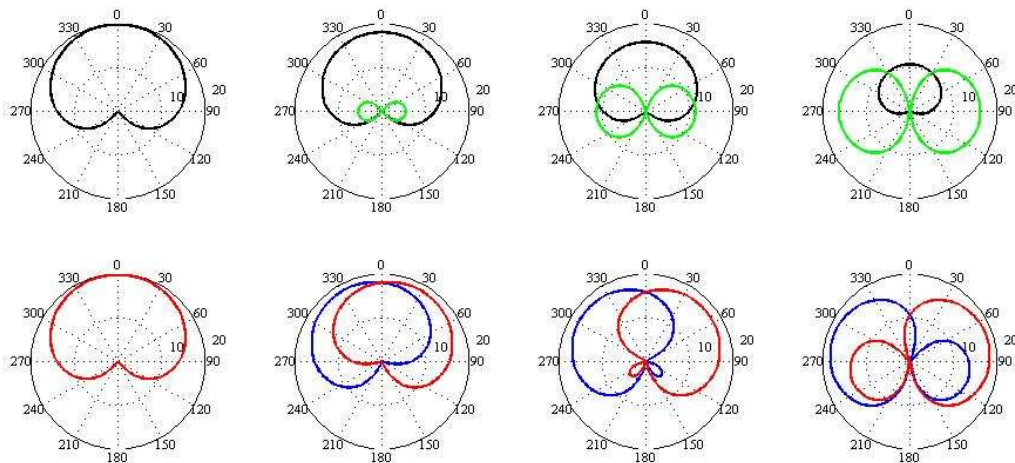

Figure 2: Illustration of the M/S principle: Top Row: M-signal (black) and S-signal (green) Bottom row: the resulting signals L (blue) and R (red)

2 Description of the Double-M/S arrangement

2.1 Microphone arrangement

Double-M/S is a recording technique for two- or multi-channel stereophony which relies solely on signal level differences.

The underlying principle of the Double-M/S arrangement is the grouping of two M/S microphone pairs. With this arrangement, only three microphones are needed as the figure-8 microphone is used for both pairs. Hence a Double-M/S arrangements consists of three microphones. Figure 3 illustrates this principle:
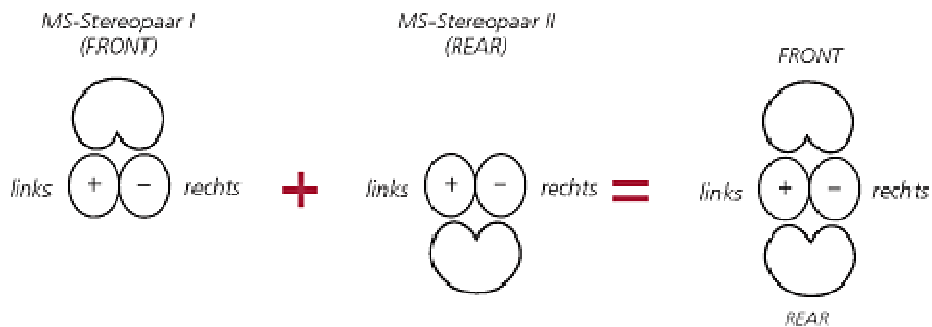
Figure 3: The principle of the Double-M/S arrangement: combination of two M/S pairs [20]
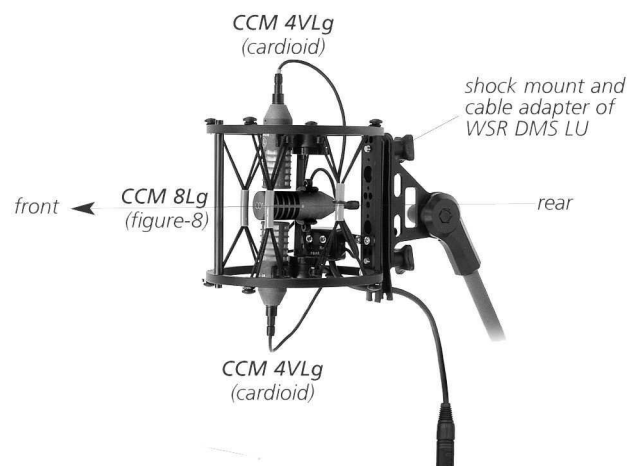
The three microphones/channels are named:

- M_{front}

- S

- M_{rear}

By using three compact small-diaphragm microphones it is possible to arrange them almost perfectly coincidentally, i.e. all at the same point.

Figure 4:

Double-M/S arrangement:

by employing compact microphones (SCHOEPS CCM 4V and CCM 8), the smallest possible spacing between the three diaphragms is achieved

2.2   Generation of 2/0-Stereo and 3/2-Stereo Signals

It is possible to generate signals for two- or multi-channel stereophony from the three Double-M/S microphone signals obtained using the setup. This is conventionally done using an M/S-Matrix to generate the L/R signals from M_{front}/S, and a second matrix to

generate the LS/RS signals from M_{rear}/S. Furthermore, a centre speaker can be powered by signals from the M_{front} microphones (See part 5).

However, using the Double-M/S arrangement allows for much better decoding, as after all, the disadvantage of M/S is that the directional pattern of the virtual microphones depends upon the mixing ratio of the M and S signals. With increasing signal level, the resulting directional pattern develops from cardioid to figure-8 (see fig.5). With a Double M/S arrangement, the signals from the microphones can be mixed to create any directional pattern for the virtual microphone.
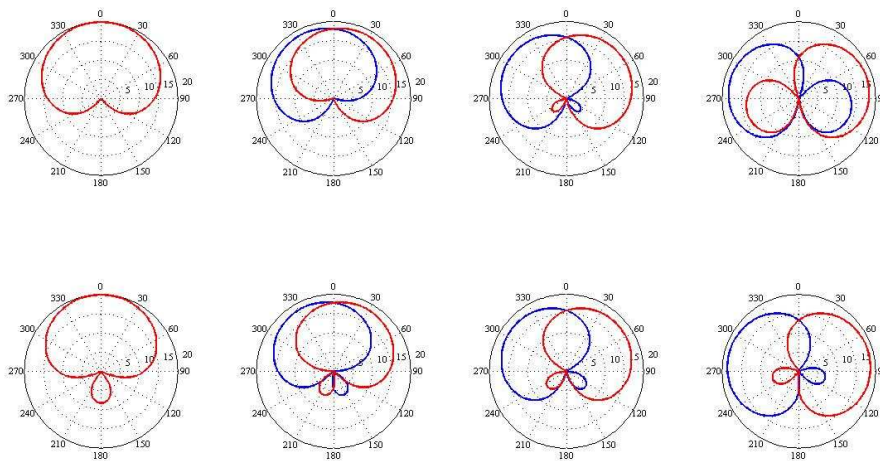

Figure 5: Illustration of the difference between decoding with and without variable directional pattern of the M_{front} signals. The polar pattern of the decoded channels L,R. Top row: fixed directional pattern; M_{front}=cardioid. Bottom row: variable directional pattern, L,R=supercardioid.

This advantage is vital for optimized coincident recording. It enables the variation of the recording angle without changing the directional pattern and correlation of the resulting virtual microphone pair. It also makes maximum decorrelation of the two signals possible- an important aspect for Double-M/S system. The importance of these parameters shall be discussed at a later stage in this paper. Practical tools for Double-M/S decoding are described in more detail in section 5.

2.3   Similarities of the Double-M/S system to the Ambisonics system

Ambisonics is a recording and playback technique invented by Michael Gerzon [7]. This technique relies on coincident recording. The theory on which the technique is based is the splitting of the sound field into so-called "spherical harmonics"; functions which

describe the motion of the incoming sound waves. The higher the order of these spherical harmonics, the greater their descriptive precision.

Figure 6 shows spherical harmonics up to and including order three. To this day, it is only possible to record first order spherical harmonics. Recording such first order Ambisonics signals produces so-called "first order B-format" signals:
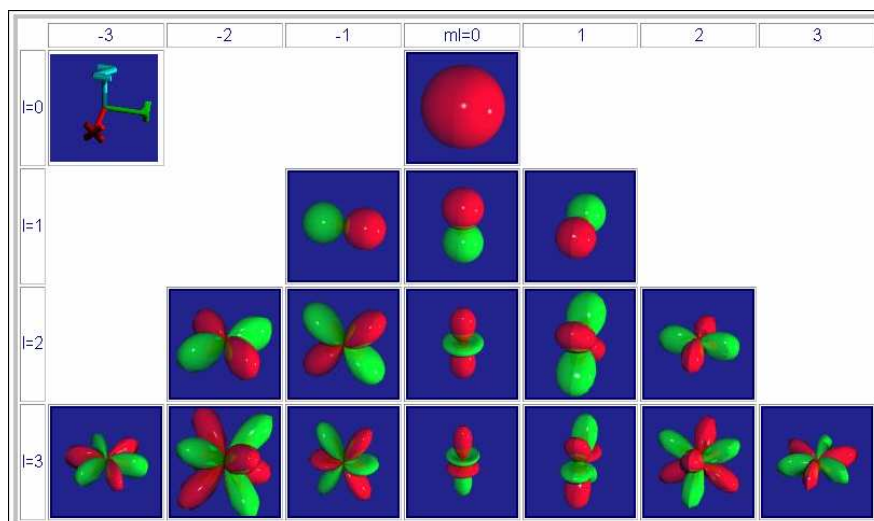

Figure 6: Visualization of spherical harmonics to order three;
l defines the order, and ml the dimension

First order B-format*:*

0th order:

$$W = 1;$$

1st order:

$$X = \cos(\Theta) * \cos(\phi);$$

$$Y = \sin(\Theta) * \cos(\phi);$$

$$Z = \sin(\phi);$$

where Θ is an angle in the x/y plane (z=0) and ϕ is an angle in the z-plane.

These four signals can be obtained in different ways using microphone recordings:

According to Gerzon, the four B-format signals can be obtained from four microphones arranged tetrahedrally. These microphones already exist, for example the "Soundfield" microphone produced by Soundfield UK (see [24]). The advantages of this method are the regular spacing due to the tetrahedral shape, and good coincidence in all spatial directions. However this method has the disadvantage of the need to convert form the terahedral "A-format" signal to B-format. A different method is to use a special microphone setup to obtain the B-format signal directly. This arrangement consists of one spherical microphone (W) and three orthogonal figure-8 microphones (X,Y,Z) and is referred to as "native B-format recording" (see [3] for an illustration). If three-dimensional playback is not required, the third figure-8 microphone can be omitted, which leaves an arrangement that is easily setup. This format with only three microphones is termed "horizontal B-format" by Benjamin [1]. With sufficiently small microphones, it is possible to achieve perfect horizontal coincidence. In principle, Double-M/S signals can also be converted to "horizontal B-format" by addition and subtraction:

$W = M_{front} + M_{rear}$;

$X = M_{front} - M_{rear}$;

$Y = S$;

Hence the Double-M/S signals are identical to first order Ambisonics signals except for the missing Z-component (height) of the Double-M/S signal. This is makes no difference if playback of the signals is on a conventional speaker system without a Z-component (eg. 2.0, 5.1, etc.).

Benjamin [1] compared the two different recording methods. This comparison showed that a native array of single capsules (Benjamin used Schoeps MK2 and MK8) led to better polar patterns for B-format, yet sound from any other direction than horizontal resulted in rather less ideal polar patterns. The tetrahedral setup provided good, consistent polar patterns independent of sound direction, but irregularities occurred above frequencies of about 6kHz. See Flock [6] for more details.
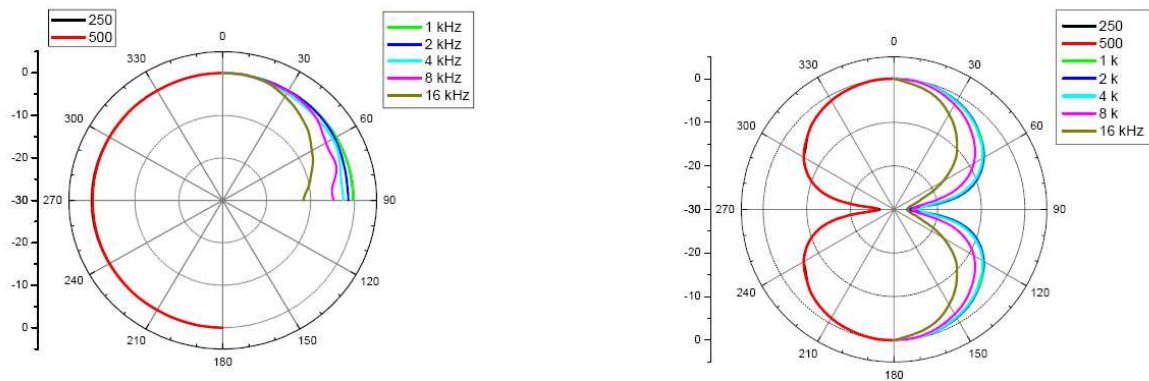
Figure 7: Polar patterns produced by a "horizontal B-format" setup by Benjamin [1]. Left: SCHOEPS MK2 (=W-signal; with a 90° angle for the omnidirectional microphone the pattern would be perfect in the horizontal plane). Right: SCHOEPS MK8 (=X-signal)
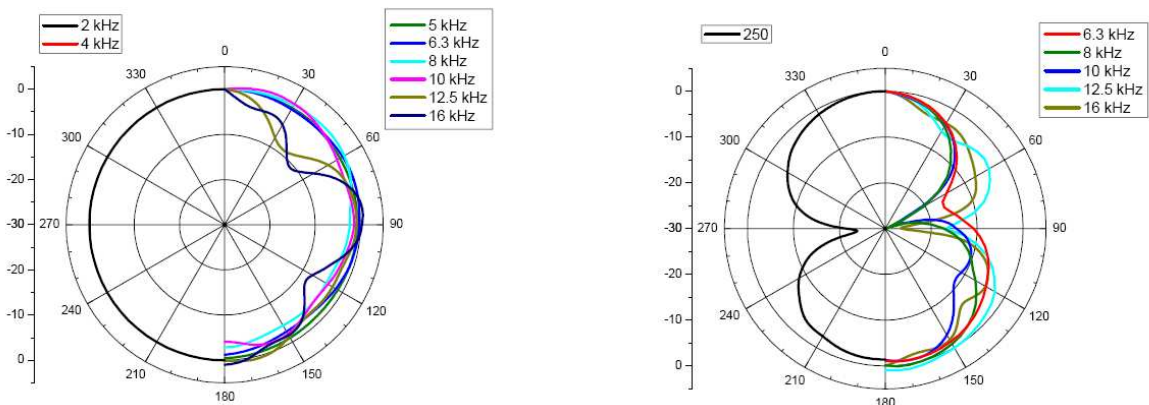

Figure 8: Polar pattern produced by a tetrahedral array (Soundfield MkV microphone system) from Benjamin [1]. Left: W-signal. Right: X-signal. Note: the frequencies used to produce these diagrams are not the same as those in Figure 7, thus direct comparison is not possible!

As mentioned above, the underlying principle of the Ambisonics theory is the analysis the sound field by splitting it into different directional components. During playback, the sound field is reproduced by the mixing of all the speaker signals. Due to this, it is not unlikely that two or more speakers will have correlated signals. The splitting of the sound field does not follow the rule that a phantom source is created using level and time differences between two neighboring speakers, but rather aims to create physical summing in the sweet spot. This leads to different properties of the system, especially concerning the parameters discussed already. It is a completely different approach to playback of a surround signals compared to the phantom source theory. Consequently, the two theories cannot be compared directly. If a first order Ambisonics signal was to be

evaluated with regard to the parameters this paper is concerned with, the result would be negative. In fact, many engineers reject mixes that were recorded in this way. However, it is important to note that this is not necessarily due to the coincidence of the recording, as is often stated. There are many types of coincident recording and also many ways of judging their quality and optimizing them accordingly.

3 Parameters for the theoretical analysis of the Double-M/S technique

In this chapter, various important parameters for the objective evaluation of the Double-M/S technique - whether for two-channel stereo or surround sound - shall be discussed. They were chosen specifically for the discussion of Double-M/S, but can also be used for other techniques.

These parameters are:

- level and time differences for directional imaging

- correlation

- crosstalk.

These parameters influence various attributes of perception such as localization, sound color and spatial perception. The relevance of these different parameters depends on the application in question. An important point to appreciate is that no parameter, no matter how important, should be considered by itself.

3.1 Directional imaging (localization)

3.1.1 General description

This parameter describes the ability of a microphone arrangement to recreate the sound field between the speakers according to the engineer's wishes. It is often desired that the sonic image captured during recording is proportionally reproduced in playback. In this case, the recording angle plays an important part. The recording angle is the angle in the recording room which is reproduced between the front speakers (L/C/R) during playback. For a more detailed description of localization and recording angle please refer to [30] and [31]; the following description shall not go to any great depth on the subject.

The shift of the phantom source is achieved by differences in time and level between microphone signals; these cause the source to be shifted right or left of the centre between two speakers. Theile explains how these two types of signal difference add; the total phantom source shift is the sum of the source shifts due to time differences and level differences between the signals. This can be represented as:

$$\Phi_{total} = \Phi_L + \Phi_t$$

see Theile: [25], [26]

This linear addition is only valid for phantom source shifts up to 50% of the maximum deflection. After this point, there is a gradual saturation up to the point where only one speaker depicts the signal. The author of this paper describes this behavior as a mathematical function ([30], [31], [32], [34], [35]). This approximation is illustrated in fig. 9 below. It must be noted that the deflection of the phantom source is proportional to speaker separation. For this reason, the source deflection is expressed in % to be universally valid. In a normal stereo triangle a deflection of +/- 100% corresponds to a deflection of +/- 30°. A 100% deflection would correspond to only a single speaker creating the sound field.
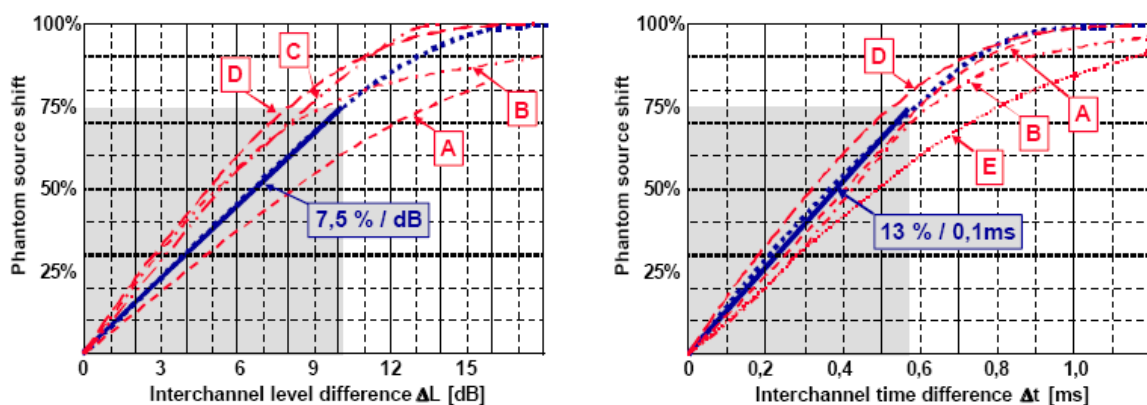

Figure 9: Relationship between level difference (left) and time difference (right) and phantom source deflection. From:

Bold curve: Wittek [34], 7.5%/dB and 13%/0.1ms

A: Leakey [16]
B: Mertens [18]
C: Brittain and Leakey [2]
D: Simonson [23], Basis of the „Williams curves" [28]
E: Sengpiel [22]

With the aid of this approximation, it is possible to calculate the stereophonic image of two microphones. This concept has been realized in the form of the "Image Assistant"; a JAVA applet to simulate the situation. The applet can be used to calculate localization curves, and is available for use online (see www.hauptmikrofon.de and fig.10, also Wittek [33]).

The localization curve describes the deflection of the phantom source depending on the angle of the sound source in the recording room. The main page in the Image Assistant shows the sound source angle in degrees on the abscissa and the deflection of the phantom source between the front speakers, in %, on the ordinate. The recording angle can be found by looking for the area on the graph in which the source is depicted between the speakers. This area is shaded light blue, and the recording angle (100%/ 75%) is displayed in a window in the top left corner of the main page.
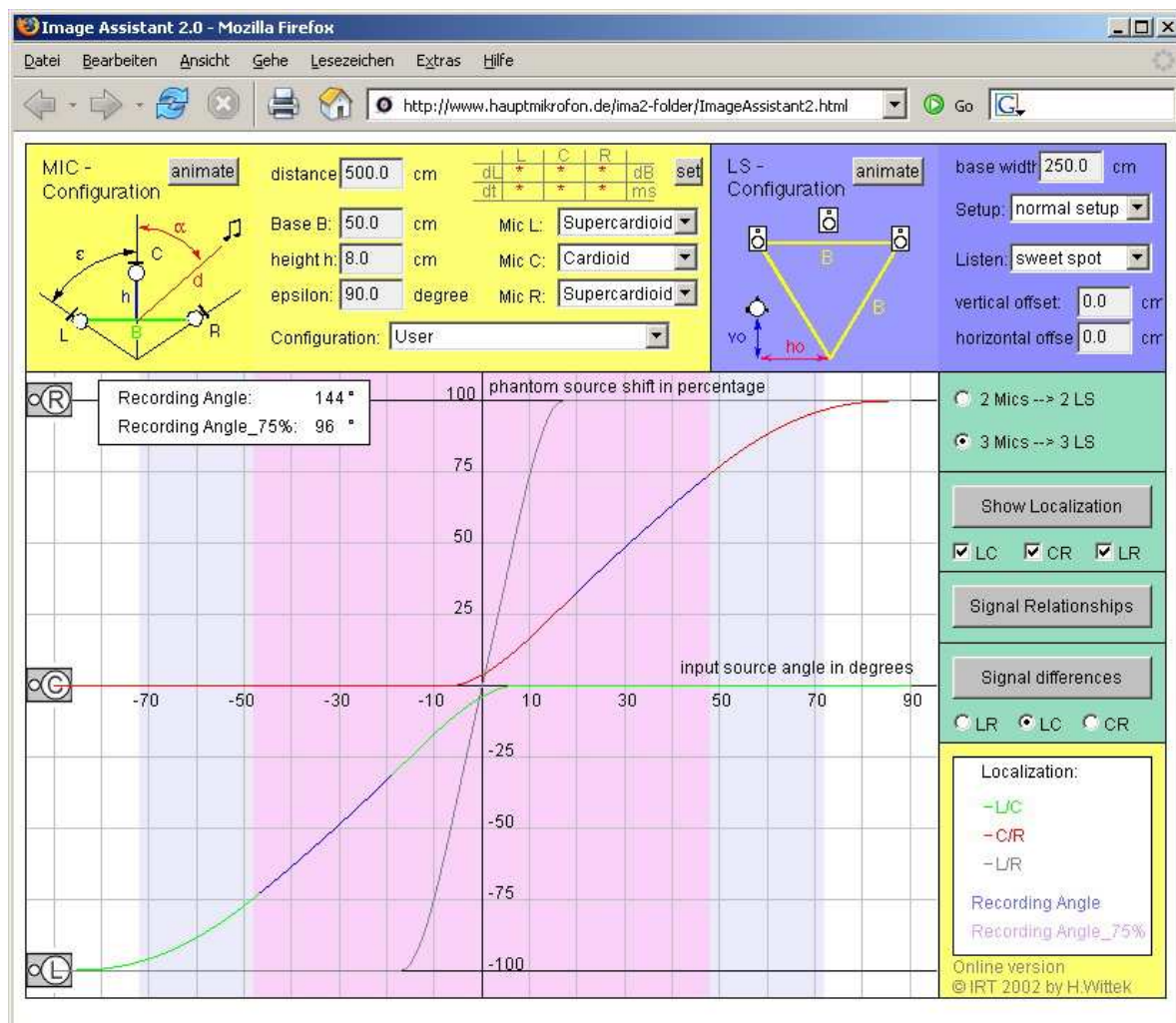

Figure 10: Simulation using the "Image Assistant" [33]:
localization curve of the OCT setup.

3.1.2 Comparison of the imaging characteristics of stereophonic setups

Many authors have designed their microphone setups such that the parameter of imaging characteristics dominates and determines the type of setup. This is not only due to the fact that this parameter is regarded as vital, but also because a setup optimized for this parameter can have good other attributes regarding the other parameters mentioned above as well. Nonetheless, it is important to keep an overview over all parameters when developing new combinations and techniques. The setups proposed by Williams (MMA, see [27] and [29]) are purposely designed for optimal 360-degree imaging of the sound field. According to Williams, other parameters such as spatial imaging are influenced largely by the setup's directional imaging capabilities.

Theile argues similarly to this but strongly recommends certain arrangements to ensure that other parameters apart from directional imaging are well provided for. An example of this type of microphone setup is the OCT setup (see fig.11, [25] and [20]) which has various advantages regarding the parameters already discussed.
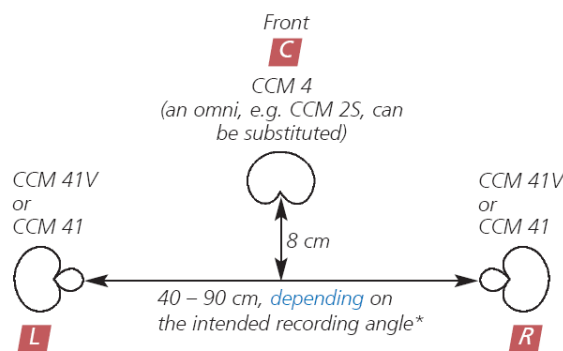

Figure 11: The OCT arrangement proposed by Theile [25], from [20]

For one, the localization curve of this setup, shown in fig. 10, is very linear which entails very natural directional imaging without geometrical distortion. As well as this, crosstalk between the imaging areas of the microphones in an OCT setup is minimized. This has advantages concerning spatial imaging, timbre and the robustness of the image. This aspect of the OCT technique is discussed in detail in part 3.3 of this paper.

The imaging properties of a Double-M/S setup are dependent on the decoding of the signals. This is clear as the decoding makes any variation of a coincident arrangement with 5 or even more microphones possible. Five first-order microphones could not even

produce sufficiently different signals to create optimal tonal and directional imaging. For this reason, care must be taken; frequently, a 4-channel decoding produces better results than a 5-channel recording when it comes to 360-degree imaging.

Figure 12 shows the resulting frontal directional image of the 4-channel setup. A very linear and thus proportional image is produced between the speakers which is typical of an MS/XY recording.

The frontal directional imaging characteristics of the 5-channel variation are shown in fig.13. The simulation shows clearly that the theoretical ideal of a regular, balanced image between the speakers is not possible with the 5-channel setup. The reason for this is reduced channel separation (caused by crosstalk, see 3.3). The *Image Assistant* shows that the central area of playback is produced by three speaker pairs. This theoretical multiple playback results in reduced image sharpness, and decreased locatability[1]. Even so, this does not necessarily result in worse characteristics of the recording as crosstalk also occurs with other recording techniques. Other parts of this paper shall discuss the negative effects due to crosstalk.

A paradox arises between the two theories of localization: Ambisonics allows the use of any amount of speakers, yet only focuses on signal summation in the "sweet spot". On the other hand, a correlated signal on more than two speakers for the creation of a phantom source (see Theile [25]) has negative effects for sound color and localization. If localizationwas to be calculated using Ambisonics theory, the signals of all speakers would have to be accounted for, not just as pairs as with the *Image Assistant*. According to the Ambisonics theory, crosstalk is not a negative parameter.

---

[1] The locatability (or locatedness) of a source describes the degree to which a source can be said to be in a particular direction [38].
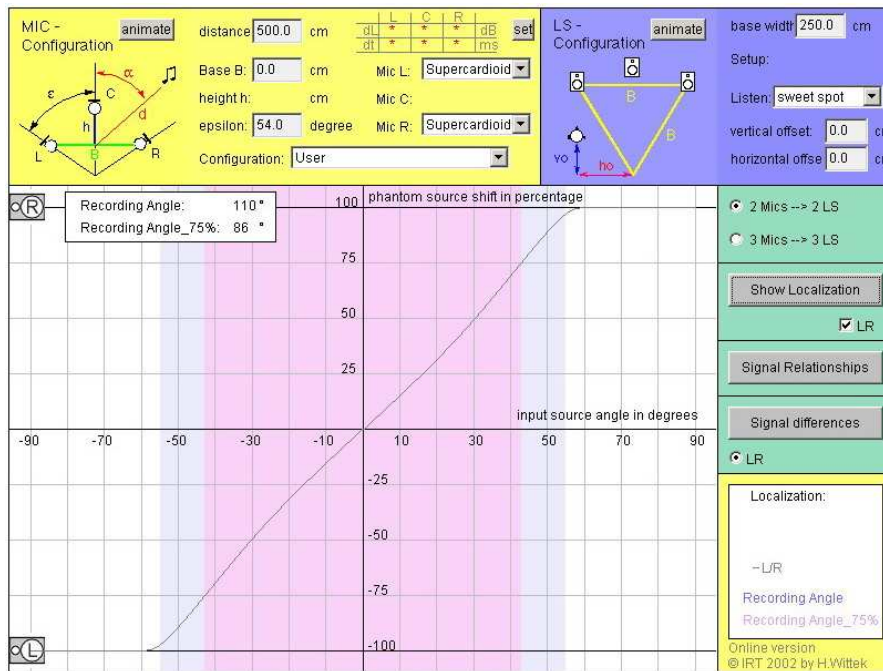
Figure 12: Directional image between L/C/R speakers using 4-channel DMS decoding (= 4-channel setting of the Schoeps Hardware Matrix, see part 5), simulated using "Image Assistant" [33].
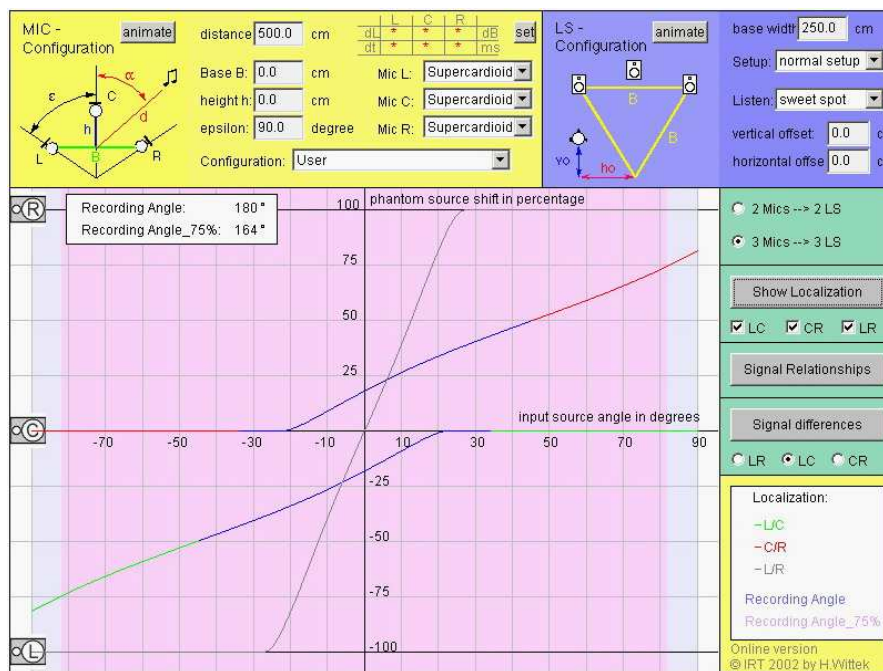

Figure 13: Directional Image between L/C/R speakers using 5-channel DMS decoding, simulated using "Image Assistant".

3.2 Coherence / Correlation

3.2.1 Correlation and its significance in M/S-stereophony

The Coherence (or correlation[2]) in the diffuse sound field between channels of a stereophonic recording is often regarded as a parameter which influences spatial perception and sound color significantly (see [5], [8], [17], [19] and [25]). Diffuse field correlation is seen as a deciding parameter for the differences in perception between various types of stereophonic recording (see figures 30-33). For example, arrangements with increased microphone spacing are known to have better spatial imaging qualities (see, for example, [36])

The parameter "interchannel correlation" in the diffuse sound field should be given more consideration with respect to Double-M/S recording as it is essential for directional imaging and tonal differences between decoding methods. These parameters are not independent though in a coincident recording, but very closely related. Correlation increases with an larger recording angles. Furthermore, correlation and imaging properties are influenced by the type of microphone used. In figure 14 the correlation coefficient of a coincident microphone setup in the diffuse field is shown as a function of the microphones' offset angle[3]. The correlation coefficient is independent of the directional pattern of the two microphones. One can see that the diffuse-field correlation coefficient between two coincident cardioids never falls below 0.5. If a diffuse-field correlation coefficient of zero is called for, this can only by achieved by directional patterns that have approximately the maximal omnidirectional component of a supercardioid (see Griesinger [11]).

Figure 14 shows that diffuse-field correlation decreases with larger offset angles. However, it must be noted that the recording angle decreases accordingly. The table below shows the correlation and recording angle of three arrangements with similar offset angels, but different directional patterns.

---

[2] Coherence of two channels is a measure of similarity of signals in the frequency range, regardless of phase. Correlation is a measure of the similarity of two signals in the time domain.

[3] The offset angle of an arrangement is the angle between the two microphones.

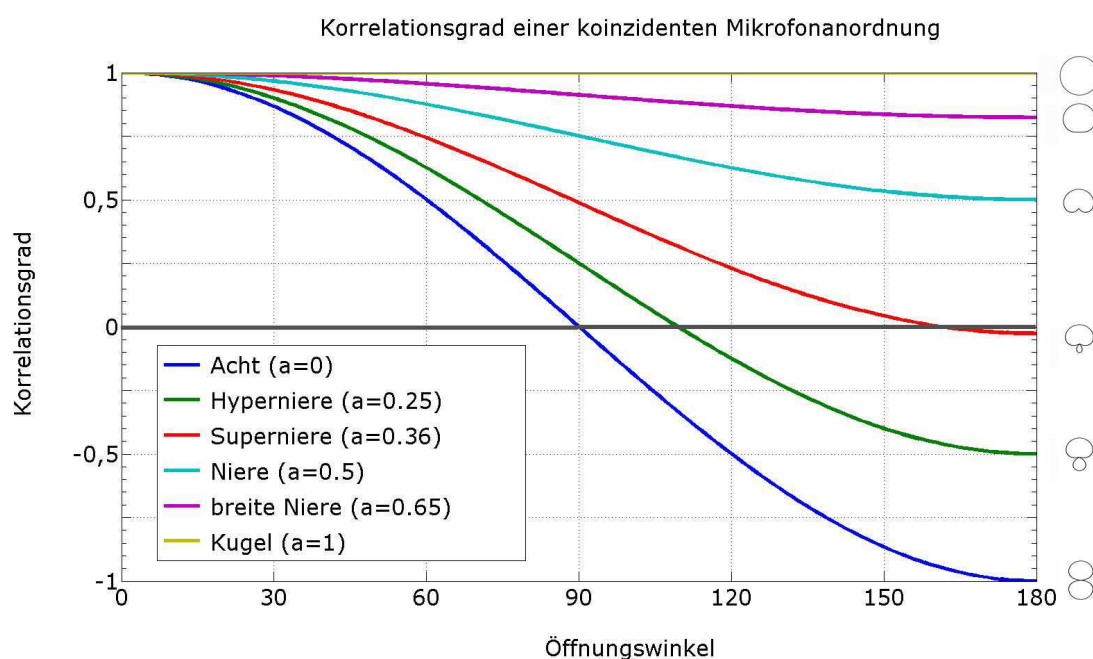| Offset angle=90°	**Cardioids**	**Supercardioids**	**Figure-8s**
Correlation coefficient	0.75	0.49	0.00
Recording angle, (*recording angle 75%* [31])	180° (*142°*)	130° (*104°*)	72° (*58°*)

Figure 14:

Graphs of the correlation coefficient in the diffuse field against offset angle between the microphones for several different directional patterns. The omnidirectional part of the microphone characteristic is also given according to the formula:

sens = a + (1-a) * cos (φ)

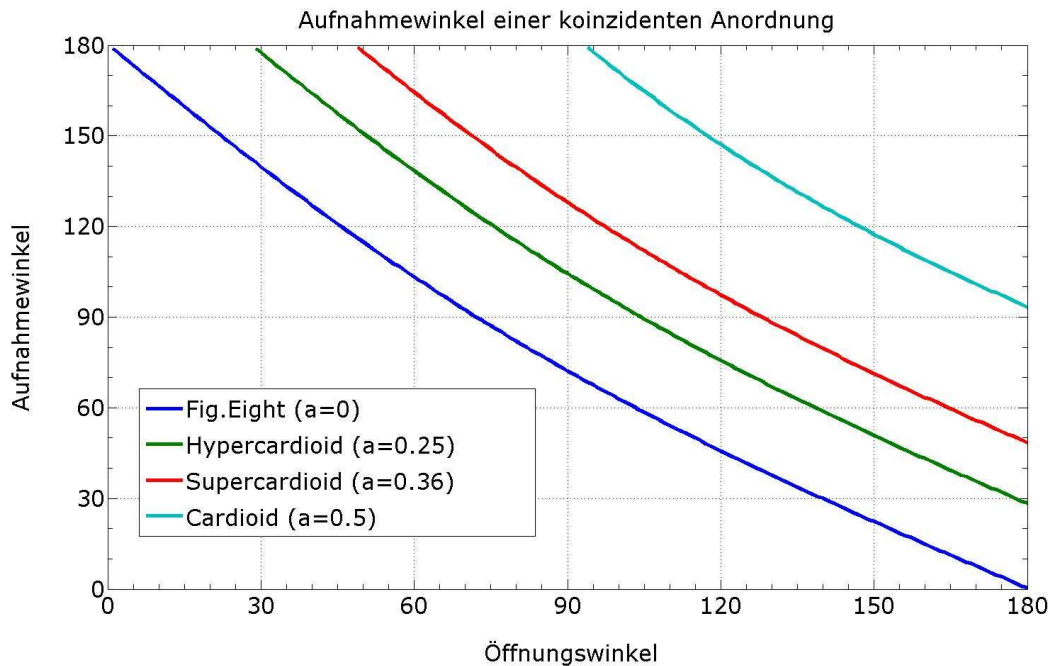**Aufnahmewinkel einer koinzidenten Anordnung**

Figure 15:

Graphs of recording angle against offset angle for a coincident microphone arrangement for several different directional patterns. There is no recording angle for a coincident arrangement consisting only of wide cardioid or omnidirectional microphones.

It would be interesting to investigate the correlation of the setup with the resulting recording angle. With that, it would be possible to assess, for a given recording angle, which arrangement has optimal decorrelation in the diffuse field. This can be done by calculating the recording angle in dependency of offset angle and directional pattern. The results of this calculation can be seen in figure 15.

The calculation was performed in the following way: the recording angle was defined as twice the smallest offset angle for which there was a level difference of at least 16dB between the two microphone signals.

With these values, the correlation coefficient for a coincident arrangement of two microphones, with any directional pattern, dependent of the resulting recording angle, can be obtained:

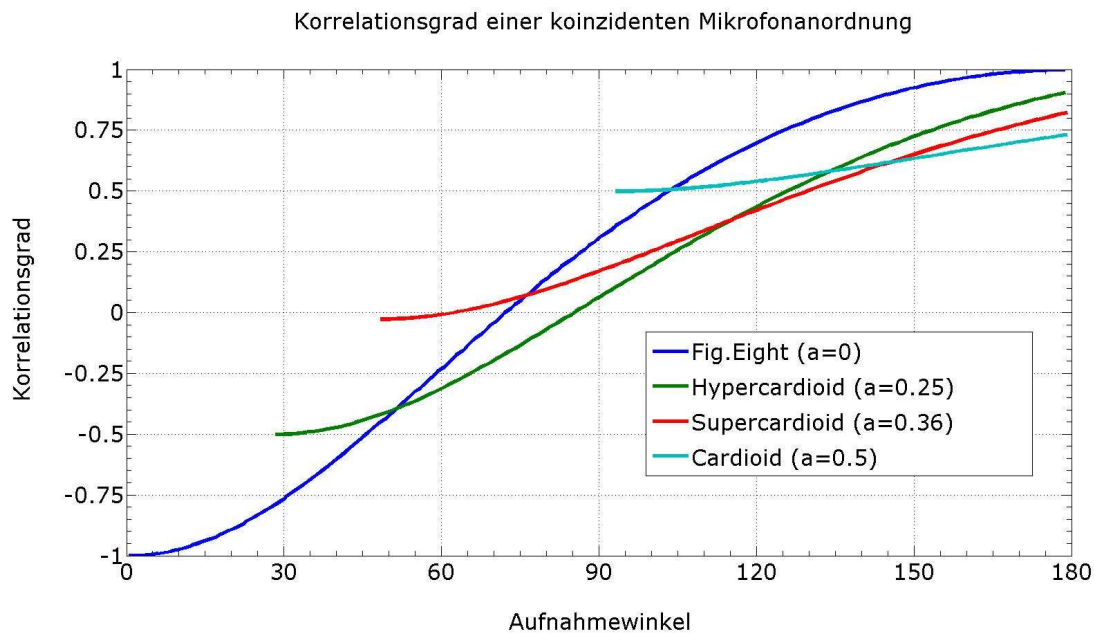Korrelationsgrad einer koinzidenten Mikrofonanordnung

Figure 16: Correlation coefficient for a coincident arrangement of two microphones in a diffuse field against the recording angle of the arrangement for different directional patterns. There is no recording angle for a coincident arrangement consisting only of wide cardioid or omnidirectional microphones.

Figure 16 shows that the diffuse field correlation coefficient of a coincident setup with a fixed recording angle is quite strongly dependent of the microphones' directional patterns. There are of course restrictions on the choice of directional patterns; not all values shown in figure 16 correspond to arrangements that are realistic or useful in practice and differences in correlation are often minimal. Whether existing differences have any audible effect cannot be judged with certainty at this point. The rationale is that the differences in direct sound imaging are too large and mask the diffuse field differences.

In summary, at this point it can be said that the diffuse field correlation between the channels of a stereophonic recording is an important parameter for spatial perception, which is however difficult to control independently of the recording angle. With that, the diffuse field correlation in coincident recordings is mostly dependent on the desired recording angle.

3.2.2 Optimization of Double-M/S systems with respect to diffuse-field correlation

The Double-M/S system corresponds to a coincident recording with 4 or 5 first-order microphones. Maximum signal separation, homogenous directional imaging and minimal diffuse-field correlation between channels is achieved when the following conditions are fulfilled:

- the angle between the virtual microphones is maximum

- the directivity of the virtual microphones is maximum

The first of these requirements is easily fulfilled if the decoding results are monitored. The second demands a choice of a directional pattern which is the best possible compromise between strong directionality and a disturbing back lobe – the supercardioid. Hence an ideal setup would consist of 4 or 5 supercardioids set up at a maximum angle to each other.

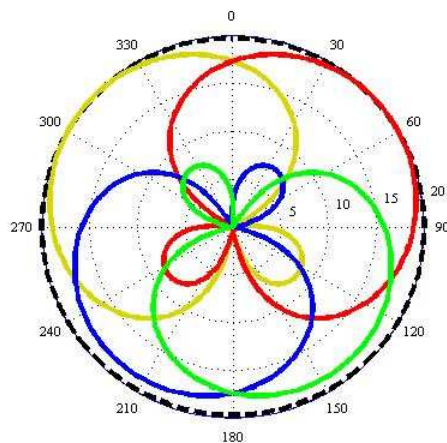These requirements lead to the following 4-channel or 5-channel decodings:

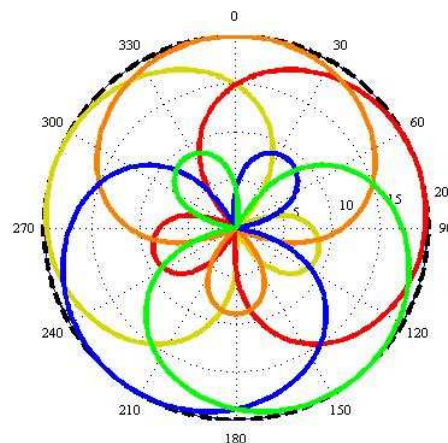*Figure 17a: an optimized 4-channel decoding of the Double-M/S system*

Figure 17b: an optimized 5-channel decoding of the Double-M/S system

The resulting degrees of correlation of these optimized Double-M/S setups are:

5-ch setup:

	L-C	L-R	L-LS	LS-RS	LS-C	L-RS
Offset angle	72°	144°	72°	72°	144°	144°
Correlation coefficient	**0.66**	0.11	0.66	0.66	0.11	0.11

4-ch setup:

	L-C	L-R	L-LS	LS-RS	LS-C	L-RS
Offset angle	-	108°	72°	72°	-	198°
Correlation coefficient	-	**0.36**	0.51	0.66	-	0.04

From the above table, one can see that minimal diffuse-field correlation is achieved with the 4-channel setup. Areas of higher correlation are only found further back in the sound field. The 5-channel setup shows higher correlation due to smaller offset angles of the virtual microphones.

It can be supposed that a Double-M/S setup with a high correlation coefficient is most suited for recording situations where good directional imaging is required. For sufficient spaciousness and enveloping, high degrees of decorrelation are needed according to Griesinger ([8], [10], [11]).

With a two-channel M/S or XY recording, diffuse field correlation coefficients of zero are possible. This property cannot be achieved with two cardioids as their correlation coefficient never falls beneath 0.5. Partly for this reason, M/S and XY recording has a worse reputation than it deserves. XY recordings are often described as narrow, centered and not suitable for the imaging of rooms. This is true only for XY recordings with high correlation and hence does not apply to optimized XY setups.

In figure 14 the offset angle for an ideal XY setup can be determined by reading off those x-values which correspond to a correlation of 0. The following values were found:

- Figure-8 (a = 0): 90° (Blumlein setup, recording angle)

- Hypercardioid (a = 0.25): 110° (recording angle 85°)

- Supercardioid (a = 0.36): 160° (recording angle 64°)

This is also described in [11]. With the Double-M/S setup, these and other coincident arrangements can be reproduced and optimized concerning to correlation.

In multichannel coincident recording, a diffuse field correlation coefficient of zero cannot be achieved for all microphone pairs. With increasing numbers of channels, the danger of correlated microphone pairs increases.

The practical consequences for recording are described in part 4. It turns out that the extraction of 4 channels is easily achieved by good decoding. Five-channel extraction is more complicated and only produces reasonable results if the engineer decodes with a critical eye and takes appropriate measures such as the inclusion of delays, level control etc. where necessary (see part 4). Attempting to extract more than 5 channels is not possible without strong interchannel similarities, which is why proposals of (near-) coincident arrangements for new formats such as 7.1 or even 10.2 must by regarded critically. The number of channels extractable from a microphone system is by no means also a measure of its quality. On the other hand, there are also microphone systems which provide insufficient channel separation, which is the other extreme. The aim of this paper is to describe a simple and yet efficient system that delivers high quality sound.

3.3 Crosstalk

3.3.1 Theoretical analysis

Phantom sources are created by the reproduction of a coherent signal on two speakers. If a third speaker is added which also emits the coherent signal, unwanted, and often audible, comb-filter effects appear.

This third signal is termed the crosstalk signal. If the crosstalk signal is out-of-phase, it is less distorting than an in-phase signal. Literature in this field is provided by Theile [25] who seeks to avoid multiple imaging due to crosstalk by the use of suitable microphone arrangements. Lee and Rumsey [14] investigated different multichannel microphone setups and found negative effects on spaciousness and locatability due to crosstalk.

Crosstalk also decreases the listening area as even small movements towards the speaker can result in localization problems caused by the precedence effect. The effect on localization for listeners not in the sweet spot can be visualized using the *Image Assistant* [33].

When designing a microphone arrangement, it is important to make the crosstalk level as low as possible. To optimize the Double-M/S arrangement with respect to crosstalk the

same two requirements discussed in part 3.2.2 can be applied: as before, the optimum arrangement consists of virtual supercardioids at maximum offset angles to each other.

The optimized decodings shown in figure 17 result in the crosstalk behavior shown in figure 18:
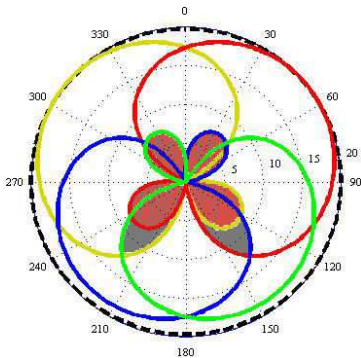

Figure 18a: Crosstalk level of the optimized 4-channel Double-M/S decoding. (black areas are in-phase, red areas are out-of-phase)
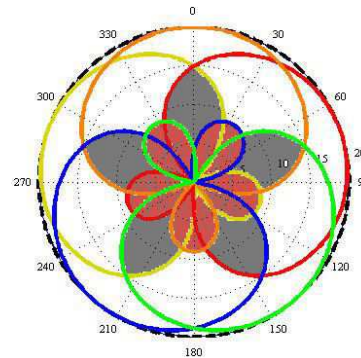
Figure 18b: Crosstalk levels of the optimized 5-channel Double-M/S decoding (black areas are in-phase, red areas out-of-phase)

The 4-channel decoding shown in figure 17a has a maximum crosstalk level of approximately −7.5dB. This level is reached at two angles; during simultaneous playback of the speakers L, LS and RS as well as R, RS and LS. The crosstalk in the front area consists of an out-of-phase signal and has a relatively low level. The 5-channel decoding has maximum crosstalk levels of −5dB approximately. This value is reached in multiple positions in the sound field.

The comparison of the 4-channel and 5-channel decodings results in a theoretically better quality using 4-channel decoding rather than 5-channel decoding. The extent of practical disadvantages implied by this theoretical disadvantage shall by discussed in the following chapters.

Another important part of a good decoding of the Double-M/S system is the homogenous level of the phantom sources. The dotted line shown in figures 17 and 18 illustrates a uniform spread of energy in all directions. The crosstalk levels quoted are with respect to this total energy.

3.3.2 Practical Analysis

To investigate the impact of crosstalk on different aspects of perception, an experiment was performed (see [11] for details). The aim of this investigation was to find the perceptive threshold of a crosstalk signal with respect to the following attributes of the phantom source:

- Width

- Direction

- Locatability

- Sound Color

The phantom source was created between the Center and Right speakers, the level difference being 3.7dB. The crosstalk signal was created by the Left speaker with levels ranging from

-20dB to -5dB.

The participants of the experiment, positioned in the "sweet spot", heard a series of test samples consisting of two groups of four arranged in A-B-A-B and A-C-A-C fashion. A was the reference signal without any crosstalk, and so was either signal B or C. The remaining signal (B or C) contained crosstalk. The addition of crosstalk to signal B or C was randomized and unknown to the participants. This was done to ensure that any prejudices on the candidate's part could be overcome.

The candidates were to record their judgments on the following scale. If the crosstalk stimulus was perceived in the wrong group, it was counted as a negative result.

Ausdehnung
Ist die Phantomschallquelle bei Beispiel 2 ausgedehnter als bei Beispiel 1?

Keine
Änderung   viel
                              ausgedehnter

```
├──┼──┼──┼──┼──┼──┼──┼──┤
1   2   3   4   5   6   7   8   9   10
```
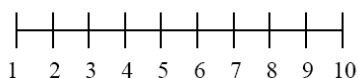
[Source Width: is the phantom source broader in Sample 2 than in Sample 1? Scale from 1 (no change) to 10 (much broader)]

A total of 15 candidates took part in the experiment. The test involved different sound stimuli such as dry speech, speech recorded in a room, dry castanets and castanets recorded in a room.

The results of the experiment are shown in figures 19-23. Figures 18-21 exemplify the results for speech recorded in a room. The diagrams illustrate the how audible the changes of the different sample attributes were perceived by the candidates. The scale above was reduced by one to define zero as "no change perceived". The perceived change is illustrated relative to the change in crosstalk level of the third speaker.
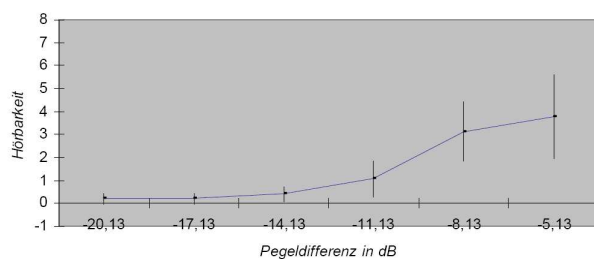

Figure 19: from [11]: Perception of phantom source width. Signal: dry speech. Arithmetic mean, including 95% confidence interval.
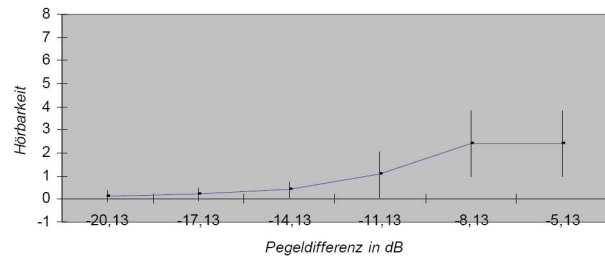

Figure 20: from [11]: Perception of change in phantom source locatability. Signal: dry speech. Arithmetic mean, including 95% confidence interval.
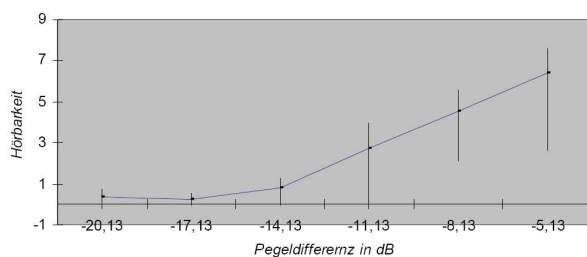

Figure 21: from [11]: Perception of change in direction. Signal: dry speech. Artithmetic mean, including 95% confidence interval.
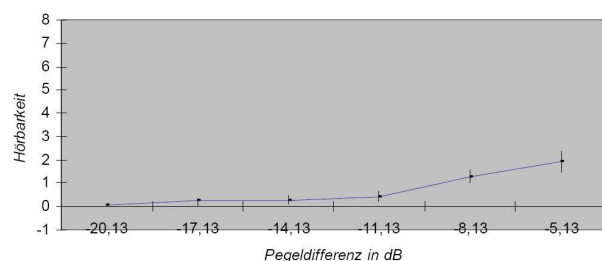

Figure 22: from [11]: Perception of sound color change. Signal: dry speech. Arithmetic mean, including 95% confidence interval.

The results showed that changes in direction and width of the phantom source were more readily detected than changes in locatability and sound color. The threshold (relative to the sum level of the phantom source C/R) is approximately -12dB for changes in direction, and approximately -9 dB for changes in width. Locatability decreases from -9dB and the sound color changes from -6dB.

Figure 23 shows the results for all test signals for the change in phantom source width.
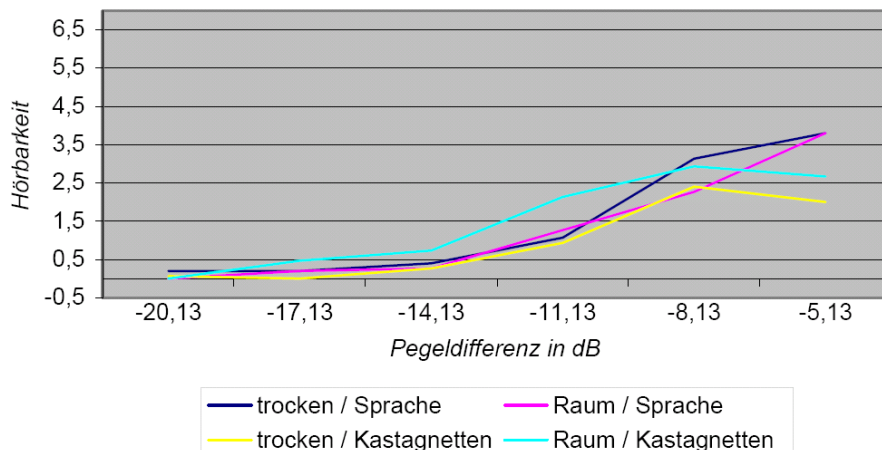
Figure 23: Perception of change in source width. Results for all four test signals. Arithmetic mean.

These results show clearly that crosstalk should be avoided where possible to avoid negative effects. This can be done if the Double M/S signal decoding is optimum. The crosstalk in this optimized decoding is however still above the audibility threshold for individual attributes, depending on the setup used.

4 Practical investigation by means of different test recordings

Despite the fact that "nothing is more practical than a good theory" (Gerhard Steinke), a good theory has its limits and cannot explain everything we perceive. For this reason, a practical investigation must also be included in the discussion of Double-M/S setups.

The aims of this practical investigation are:

- To test different decoding methods

- Examination of the advantages and disadvantages of Double-M/S recording techniques established in the previous chapters.

- Examination of the usability of Double-M/S in different situations such as music, atmo, theater, radio drama, documentation/film and television studios.

- Comparison of Double-M/S to other referenced recording setups.

The quality of a recording is not easily evaluated by scientific means, partly due to differences between individual expectations and priorities of different listeners. For this reason, no general results shall be postulated here; the focus shall rather be on describing experience and the collected comments of others.

4.1 Different methods of Double-M/S decoding

The following variations of Double-M/S recording, denoted A-F, were investigated practically (see [13]). Apart from polar patterns and the description of offset angles, the examination of each method includes a level matrix with which the decoding was performed.

The delay used for the rear channels is also stated.

A) 4 supercardioids => decoding similar to MDMS U (4-channel); see part 5 b)

The supercardioid directional pattern has established itself as a good compromise as the directionality of the supercardioid with its strong rear signal suppression entails low crosstalk levels. In this decoding, the omission of the center channel allows for larger offset angles.

Directional imaging is balanced and precise in all areas. The recording angle for the front L/R basis is 110°, and the suppression of direct sound in the rear channels works well.

These decoding characteristics predestine the setup for use in music and radio drama recording. Whether the varianti with the Center channel is preferred depends upon individual taste in the author's experience.
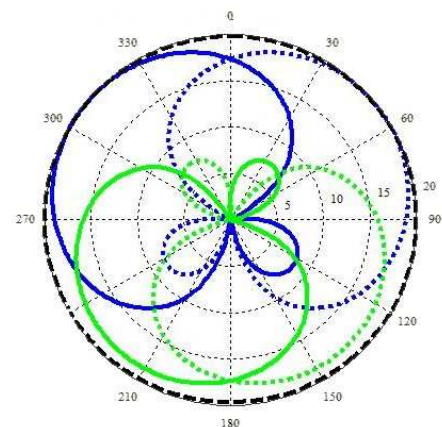
Loudspeaker Feeds

	L	R	C	LS	RS
Front: Cardioid	-0.0	-0.0	-inf	-14.9	-14.9
Fig.8 ☑ MK 8	-0.9	-0.9	-inf	-5.1	-5.1
Rear: Cardioid	-inf	-inf	-inf	0.0	0.0

☑ normalize   ☑ dB   yellow Color marks inverted phase

Virtual Microphones:

	L,R	LS,RS	
Polar Pattern	0.369	0.36	0=Fig.of 8 1=Omni
Angle (+/-)	54.26	144	°
Level	1.5	0	dB

B) 4 supercardioids, broader imaging

To decrease the recording angle of a stereo recording, either the microphone offset angle or the directionality must be increased. This example differs only slightly from the previous one as any attempt to significantly decrease the L/R recording angle and yet to maintain the LS/RS stereophonic imaging would result in uneven energy spreading and thus channel overload. With this setting, the L/R recording angle is 90°. (as compared to 110° in variation A).
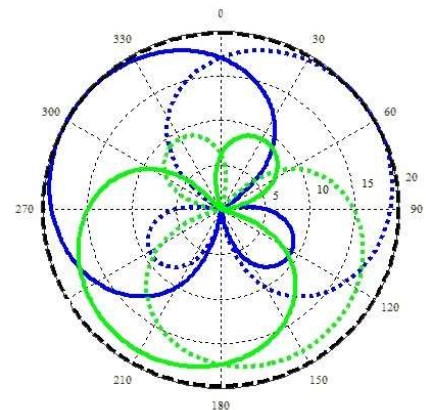
Loudspeaker Feeds

	L	R	C	LS	RS
Front: Cardioid	-0.3	-0.3	-inf	-11.2	-11.2
Fig.8 ☑ MK 8	-0.5	-0.5	-inf	-5.7	-5.7
Rear: Cardioid	-inf	-inf	-inf	0.0	0.0

☑ normalize ☑ dB yellow Color marks inverted phase

Virtual Microphones:

	L,R	LS,RS	
Polar Pattern	0.355	0.325	0=Fig.of 8
Angle (+/-)	56.55	148	1=Omni
Level	0.472	-1.24	dB

C) "Conventional" MS-decoding (4-channel)

This example was included as a negative example. The large back lobes of the microphones result in prominent crosstalk and hence strong acoustic irregularities for listeners outside the "sweet spot". Furthermore, the energy distribution is not ideal. The listening results show that this variant is rated worse in terms of sound and spaciousness.
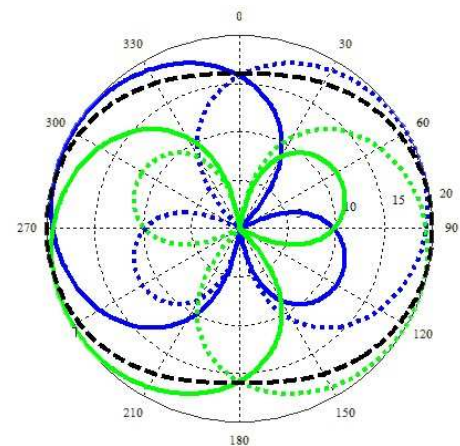
Loudspeaker Feeds

	L	R	C	LS	RS
Front: Cardioid	-2.3	-2.3	-inf	-inf	-inf
Fig.8 ☑ MK 8	0.0	0.0	-inf	0.0	0.0
Rear: Cardioid	-inf	-inf	-inf	-2.3	-2.3

☑ normalize ☑ dB yellow Color marks inverted phase

Virtual Microphones:

|	L,R	LS,RS	
Polar Pattern	0.309	0.309	0=Fig.of 8
Angle (+/-)	63.43	116.6	1=Omni
Level	-1.84	-1.84	dB

D) 5 supercardioids => decoding similar to MDMS (5-channel); see part 5 b)

To obtain a balanced localization and energy distribution, the supercardioids are rotated further apart compared to the 4-channel version. A balanced image is obtained, with added stability due to the center channel. The use of this variant or variant A is also a matter of personal preference.

The Center level can be reduced to avoid crosstalk in the front area and to improve directional imaging. Acoustically, the 5-channel variation is inferior to the 4-channel one if further measures are not taken. However, the Center channel can play an important part in the imaging of solo instruments or in radio drama productions, as well as providing stability.

Loudspeaker Feeds

	L	R	C	LS	RS
Front: Cardioid	-5.1	-5.1	0.0	-16.0	-16.0
Fig.8 ☑ MK 8	-2.0	-2.0	-inf	-6.2	-6.2
Rear: Cardioid	-15.8	-15.8	-11.1	-1.1	-1.1

☑ normalize ☑ dB yellow Color marks inverted phase
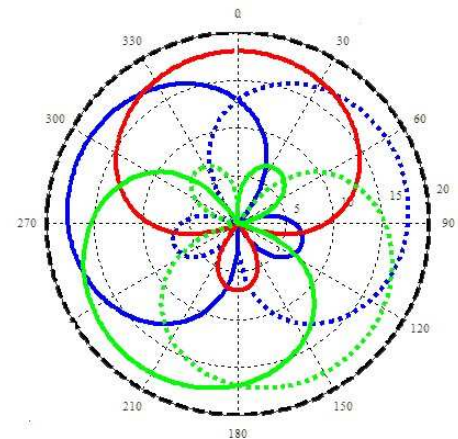
Virtual Microphones:

	L,R	C	LS,RS	
Polar Pattern	0.36	0.36	0.36	0=Fig.of 8 1=Omni
Angle (+/-)	71.99	0	144	°
Level	-1.97	-1.97	-1.97	dB

E) "Conventional" MS-decoding (5-Channel)

This decoding variation is also included here as a negative example as channel overlap is strong with this setup.

The setup shows that bad decoding without control produces bad results; displeasing acoustic imaging and missing transparency to name but few. Added to this, the acoustic color changes with small head movements.
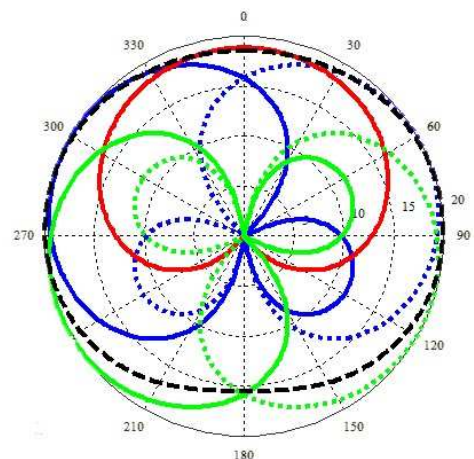
Loudspeaker Feeds

	L	R	C	LS	RS
Front: Cardioid	-3.0	-3.0	0.0	-inf	-inf
Fig.8 ☑ MK 8	-0.7	-0.7	-inf	-0.7	-0.7
Rear: Cardioid	-inf	-inf	-inf	-3.0	-3.0

☑ normalize   ☑ dB   yellow Color marks inverted phase

Virtual Microphones:

	L,R	C	LS,RS	
Polar Pattern	0.309	0.5	0.309	0=Fig.of 8 1=Omni
Angle (+/-)	63.43	0	116.6	°
Level	-1.84	-3	-1.84	dB

F) Delayed surround channels

This setup is an improvement of setup D) as it increases channel independence in the front area by increased offset angels and in the rear by the delay (Δt=10ms).

This setup is ideal for many applications where the Center channel is needed. However, a front-emphasized sound must be used as the delay prevents stable rear-localization.
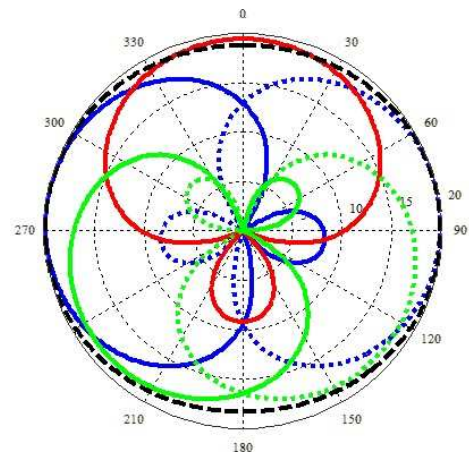
Loudspeaker Feeds

	L	R	C	LS	RS
Front: Cardioid	-6.5	-6.5	0.0	-27.1	-27.1
Fig.8 ☑ MK 8	-1.2	-1.2	-inf	-5.0	-5.0
Rear: Cardioid	-10.1	-10.1	-10.2	-3.0	-3.0

☑ normalize   ☑ dB   yellow Color marks inverted phase

Virtual Microphones:

	L,R	C	LS,RS	
Polar Pattern	0.369	0.346	0.368	0=Fig.of 8 1=Omni
Angle (+/-)	83.16	0	131.3	°
Level	-0.593	-1.15	-2.01	dB

4.2   Subjective experiences with Double-M/S recording

Experience in different situations shows that Double-M/S recording can certainly yield good results. These are our subjective experiences, which were also discussed with others', and found to be quite similar.

For example, an atmo with direct sound incident from all directions (town square) was depicted very well and showed similar quality to the simultaneous recording by an IRT cross with 4-channel decoding. The flexibility of the Double-M/S system proved to be especially useful when a tram, traveling at a skew angle to the setup, was recorded.

For an atmo with the aim of more effective spaciousness (fireworks at a festival), the IRT cross proved to be the better solution. The IRT cross setup achieves good 360° imaging as well as better envelopment and spaciousness.

Some music recordings (piano concert, chamber ensembles and large orchestras) worked surprisingly well with the Double-M/S system, however, the parallel recording using an OCT setup provided even better results that were more effective and spacious. The

decision between these methods is left to the engineer. In our experience, Double-M/S is more suited for small spaces whereas larger spaces require OCT or other setups. For music recordings, the addition of a low-passed omnidirectional microphone for low frequencies is recommended. The mix of the Double-M/S signals with a large AB configuration of omnis results in the spacious sound desired. This option also provides decorrelated low-frequency signals.

An a-capella ensemble with the need for good imaging between speakers was recorded well using a Double-M/S setup; both variations with and without the center signal were possible.

A jazz ensemble with audience in a jazz club was recorded using a Double-M/S setup and individual microphones for the instruments. The atmosphere and spaciousness of the Double-M/S setup mixed well with the individual microphone signals. A parallel ORTF recording provided similar but less flexible results. This flexibility was important as different stereo widths were desired for musical passages and applause.

Figure 24: Test recordings using the Double-M/S and reference setups. Left: Double-M/S and OCT surround system in a live theater performance. Right: Double-M/S and an IRT cross setup for an atmo recording at Durlach town square.

The live theater recording worked better with Double-M/S than with OCT. The reason for this appears to have been the bad room sound due to the location of the microphones above the stage area, and hence a strong presence of the stage acoustics. The Double-M/S system put less emphasis on the room itself.

A live TV broadcast with an audience also worked well using either the IRT or Double-M/S setups.

The use of Double-M/S for radio drama is favorable due to need for coincident recording in these productions, and because it provides more flexibility and more multichannel support. Especially with radio drama, there is often a need for downward or even mono compatibility, which is easily achieved with Double-M/S.

The use of Double-M/S in film sound recording has long established itself. Its use in documentary filming inspired new ideas for the microphone selection (see [37]): it is often useful to replace the front-facing cardioid microphone with a more directional one, as this is used to record the center dialogue discretely. For this reason, a supercardioid or even shotgun microphone can be used. This latter setup requires a special microphone setup to ensure optimum coincidence (see figure 25). The Double-M/S method works well with this setup; the advantages of surround sound on the boom are apparent especially in documentary fimling where authenticity is important and thus a subjective sound perspective can be used.

Figure 25: A Double-M/S setup with a shotgun microphone. Left: Implementation using the Schoeps Double-M/S set "CMIT" (microphones used: CMIT 5U, CCM 4 and CCM 8). Right: the setup in action ([37], photograph courtesy of André Zacher)

5 Tools for Double-M/S decoding

As the application range for Double-M/S setups is very broad, there are various decoding possibilities. As for the decoding, it makes no difference whether it is done during or after recording.

The following three basic principles of decoding shall be discussed:

a) Two M/S matrices in a mixer or editing software

b) Hardware Double-M/S matrix MDMS U

c) Software VST/RTAS plug-in "Double-M/S Tool"

a) In principle, Double-M/S recordings can be decoded similarly to M/S recordings by simply using two instead of just one M/S matrices. Furthermore, the front-facing cardioid can be used as the Center signal. This method produces good results in most cases, but these results should be controlled by an engineer to ensure that unfavorable decodings are avoided (see part 4). If a shotgun microphone is used in the setup for the M-signal, this double-matrix setup is recommended, as a mixing of the shotgun and rear cardioid signals makes no sense.
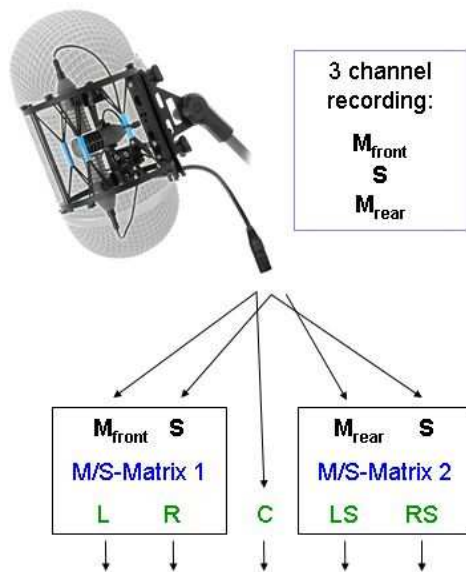

Figure 26:

Double-M/S decoding using method a)

As discussed in previous chapters, the properties of a Double-M/S setup can be improved using a suitable decoding. These decodings involve the combination of all the three microphone signals for the synthesis of the L/R/LS/RS channels and the two cardioids for the Center channel. but troublesome using level matrices or mixers. Decoding in this way is possible, but it is simpler to use of special tools for optimum decoding:

b) Hardware Matrix Schoeps MDMS U

This analog, passive matrix produces the 4 or 5 channels L/R/C/LS/RS from the three Double-M/S signals. It can be used during recording or in post-production, and offers the decoding options A and D described in part 4. The matrix equalizes the different sensitivities of the Schoeps capsules.
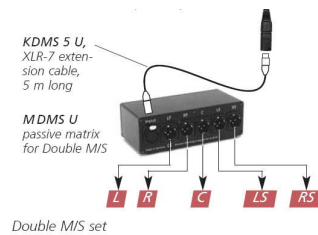
Figure 27:

Hardware Matrix
Schoeps MDMS U

c)   Software VST/RTAS plug-in „Double M/S Tool"

This plug-in is used in a sequencer program, the so-called "host". It is excellently
suited for flexible and intuitive decoding of the Double-M/S signals. The operation
of the plug-in is self-explanatory as all changes are immediately shown on the
polar pattern, and the audio signals are modified in real-time to variable
parameters. Similar to the hardware, the plug-in has three inputs (from the
Double-M/S setup) and five outputs (L/R/C/LS/RS). It is adjusted to the capsule
sensitivities and equalizes the CCM/MK 8.

The plug-in is available at the SCHOEPS website free of charge. In addition, a
couple of Double-M/S audio samples are available for download to try out the
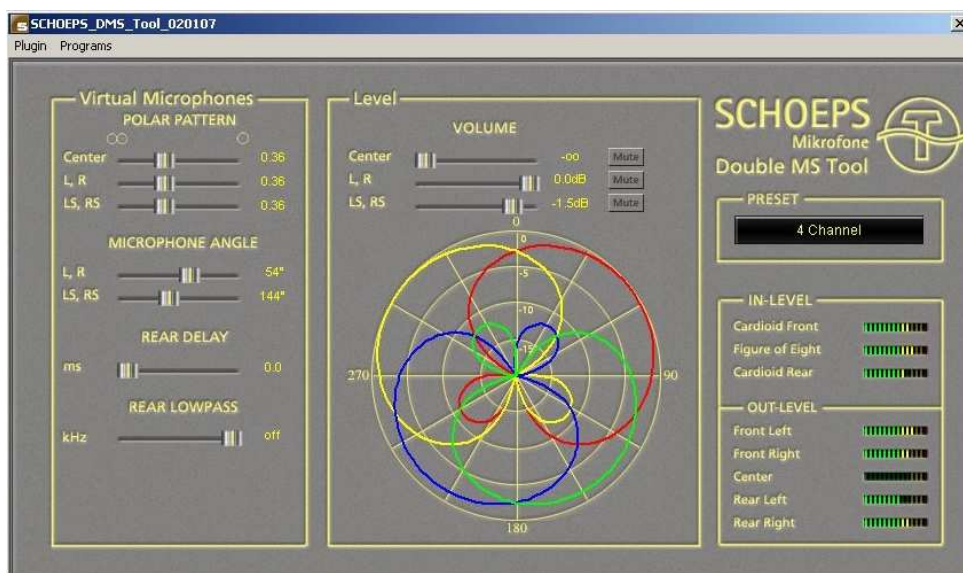plug-in. The plug-in is available for Windows and Mac.

Figure 28:

Plug-in "Double-
M/S Tool"

6 Acknowledgements

7 References

[1]   Benjamin E., Chen T.: "The native B-format Microphone, Part I", 119th AES Convention, New York, 2005, Preprint No.6621

[2]   Brittain, F. H., Leakey, D. M.: "Two-channel stereophonic sound systems. Wireless World 206-210, 1956

[3]   Camerer, F.: „Practical Surround-Sound-Production - Part 2: TV-Documentaries", AES 19th International Conference, Ellmau, May 2001

[4]   Cremer: „Zur Verwendung der Worte Korrelationsgrad und Kohärenzgrad", ACUSTICA, Vol.35, No.3, p.215-218, Juni 1976

[5]   Damaske, P.: „Subjektive Untersuchung von Schallfeldern", ACUSTICA, Vol.19, No.4, p.199-213, 1967/1968

[6]   Flock, S.: "Theoretische sowie objektive und subjektive Messungen am Soundfield-Mikrofon", Diplomarbeit im Studiengang Ton und Bild, FH Düsseldorf

[7]   Gerzon, M.: "Periphony: With-Height Sound Reproduction", Journal of the Audio Engineering Society, 21(1):2-10, 1973.

[8]   Griesinger, D.: „Spaciousness and Localization in Listening rooms – how to make coincident recordings sound as spacious as spaced microphone arrays", 79th AES Convention, New York, 1985, Preprint No.2294

[9]   Griesinger, D.: „New Perspectives on Coincident and Semi Coincident Microphone Arrays", 82nd AES Convention, London, 1987, Preprint No.2464

[10] Griesinger, D.: „Physik, Psychoakustik und Surround-Technik", Production Partner Spezial 21.Tonmeistertagung, Musik-Media-Verlag Ulm, 2000, 2.

[11] Griesinger, D.: „Räumliches Hören in Theorie und Praxis: Wie ergänzt man Tiefe und Halligkeit mit künstlichem Nachhall ohne Beeinträchtigung der Deutlichkeit - The Theory and Practice of Perceptual Modeling - How to use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity", www.world.std.com/~griesngr, Stand Oktober 2006

[12] Haut, C.: „Programmierung und Evaluierung einer parametrisierbaren Doppel-MS Dekodierung für eine 5.1 Surround-Abmischung", Diplomarbeit des Studienganges Hörtechnik & Audiologie an der FH Oldenburg, 2006.

[13] Keinath, D.: „Testaufnahmen mit dem Doppel-MS Verfahren", interner Praktikumsbericht der Firma Schoeps, Karlsruhe, 2005.

[14] Lee, H. K. und Rumsey F.: " Investigation into the effect of interchannel Crosstalk in Multichannel Microphone Technique", 119th AES Convention, New York 2005, Preprint No.6374.

[15] Lipshitz, S.P.: "Stereo Microphone Techniques: Are the Purists Wrong?", AES Convention, Anaheim, May 1985, Preprint No.2261

[16] Leakey, D. M.: "Further thoughts on stereophonic sound systems". Wireless World, 154-160, 1960

[17] Martin, G.: „The Significance of Interchannel Correlation, Phase and Amplitude Differences on Multichannel Microphone Techniques", 113th AES Convention, 2002, Los Angeles, Preprint No.5671

[18] Mertens, H.: "Directional hearing in stereophony theory and experimental verification", Europ. Broadcasting Union Rev. Part A, 92, 1-14, 1965

[19] Pfanzagl, E.: „Über die Wichtigkeit ausreichender Dekorrelation bei 5.1 Surround-Mikrofonsignalen zur Erzielung besserer Räumlichkeit", Bericht zur Tonmeistertagung 2002, Hannover

[20] Schoeps-website: www.schoeps.de, Stand 2006

[21] Schoeps-website: Download des VST plug-in „Double M/S Tool" unter www.schoeps.de/dmsplugin.htm, Stand 2006

[22] Sengpiel, E., Vorlesungsunterlagen: „Theoriegrundlagen:"Intensitäts"-Stereofonie", „Theoriegrundlagen: Laufzeitstereofonie", http://www.sengpielaudio.com/TheorieGrundlaIntensitaet.pdf, http://www.sengpielaudio.com/TheorieGrundlaLaufzeit.pdf, Stand 12. Mai 2006

[23] Simonson, G.: Master´s Thesis. Lyngby, Denmark, 1984

[24] Soundfield-website: www.soundfield.com, Stand 2006

[25] Theile, G.: "Multichannel natural recording based on psychoacoustics principles", 108th AES convention, 2000, Preprint No. 5156.

[26] Theile, G.: "On the performance of two-channel and multi-channel stereophony", 88th AES Convention, 1990, Preprint No.2932

[27] Williams, M.: "Multichannel sound recording practice using microphone arrays", 24th AES International Conference.

[28] Williams, M.: "The Stereophonic Zoom: A Practical Approach to Determining the Characteristics of a Spaced Pair of Directional Microphones", 75th AES Convention, 1984, Preprint No.2072

[29] Williams, M.: "Microphone Arrays for Stereo and Multichannel Sound recording", Il Rostro, Milan, 2000

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

[31] Wittek, H., Theile, G.: „The recording angle – based on localization curves", 112th AES Convention, Munich 2002, Preprint No.5568

[32] Wittek, H.: „Untersuchungen zur Richtungsabbildung mit L-C-R *Hauptmikrofonen*", Diplomarbeit des Studienganges Ton- und Bildtechnik an der FH Düsseldorf, 2000.

[33] Wittek, H., „Image Assistant", Java Applet auf www.hauptmikrofon.de, Stand 2006

[34] Wittek, H., Neumann, O., Schaeffler, M, Millet, C.: „Studies on Main and Room Microphone Optimization", AES 19th Int.Conference, Ellmau, 2001

[35] Wolfram-Homepage, Lösungen zu Kugelflächenfunktionen, http://mathworld.wolfram.coMSphericalHarmonic.html, Stand: Juni 2006.

[36] Wuttke, J.: Principles of microphones and stereo recordings, in: Mikrofonaufsätze, auf www.schoeps.de, Stand 2006

[37] Zacher, A.: „Dragonsongs – Lang Lang in China, Dokumentarischer Dreh", Bericht auf http://www.proaudio.de/reports/praxis/langlang/index.html, Stand: Oktober 2006

[38] Blauert, J.: "Spatial Hearing", Cambridge, Massachusetts: MIT Press, 1997