Audio Engineering Society

Convention Paper

Presented at the 14th Regional Convention
2009 July 23–25 Tokyo, Japan

M/S Techniques for Stereo and Surround

Helmut Wittek

SCHOEPS Mikrofone GmbH, Karlsruhe, Germany
wittek@SCHOEPS.de

ABSTRACT

Coincident microphone setups are known for their unique flexibility in terms of stereophonic imaging, but their reputation for spatial reproduction has been less positive. The latter opinion was caused by setups having insufficient signal separation and by the use of less-than-optimal microphones in the 1960s and ’70s–shortcomings that can well be avoided nowadays. Coincident setups and, in particular, M/S setups for stereo and surround exist that are outstandingly practical. When care is taken with parameters such as directional imaging and diffuse-field correlation, a coincident setup can compete with spaced setups even in regard to spatial reproduction. A particular look is taken at the Double M/S technique for stereo and surround.

1. INTRODUCTION

The M/S recording technique is now quite popular, and is one of the best established recording techniques for certain applications. The Double M/S technique increases still further the capabilities of M/S for stereo and surround recording. Various methods of decoding the Double M/S signals exist, and new tools for optimized decoding are available. Nonetheless, relatively little is known about its properties. Thus there is a need for objective description and the sharing of experiences concerning this type of microphone arrangement.

Coincident microphone setups have a negative reputation for spatial reproduction. The author himself has fallen into the common trap of assuming the inferiority of M/S recording to other techniques, but has since reviewed this opinion. The M/S recording technique is by no means a compromise, just as using Double M/S for multichannel recording is not, as long as the properties, i.e. the advantages and constrictions of the technique are clear to the user. As with many things, the underlying principle is that no technique is flawless, and that familiarity with the strengths and weaknesses of a technique allows optimal results to be obtained. Any method or product which claims to offer a "fool-proof" recording technique must be regarded very critically.

This paper analyzes the M/S technique for stereo and surround according to certain decisive parameters. It is divided into a theoretical and a practical analysis. The theoretical analysis investigates the 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. Finally a practical analysis of different Double M/S decoding variants supplements the view of this technique.

2. THE M/S TECHNIQUE

2.1. M/S Encoding and Decoding of X/Y

The signals of a normal (X/Y) coincident microphone arrangement can be matrixed according to the M/S (Mid/Side) principle by calculating the sum and difference values. The “Mid” signal is the sum of the two signals, and the “Side” signal is their difference.
To decode or dematrix these signals, a combination of sum and difference values is obtained as described in the formulas below. The parameters k_1 and k_2 determine the newly derived stereo image. When k_1, $k_2 = 0.5$, the original signals are reobtained. A distinct advantage of M/S coding is that it allows the recording angle and stereo width to be varied by simply trimming the parameters k_1 and k_2.

$$\mathrm{M} = \mathrm{L} + \mathrm{R} \qquad \mathrm{L} = k_1 \times \mathrm{M} + k_2 \times \mathrm{S}$$

$$\mathrm{S} = \mathrm{L} - \mathrm{R} \qquad \mathrm{R} = k_1 \times \mathrm{M} - k_2 \times \mathrm{S}$$

2.2. M/S Recording for two-channel stereo

In M/S recording, two microphones–one for the M signal and one for the S signal–are used to record directly in “encoded” form. The M microphone is directed forward while the S microphone is directed perpendicular to the M microphone’s axis.

Every M/S arrangement has a theoretical X/Y equivalent to which its signals can be converted. Figure 2 illustrates various combinations of M/S signals (M = cardioid) in the top row, with their equivalent X/Y arrangements (after decoding) underneath. The X/Y patterns which would theoretically produce the equivalent of the dematrixed M/S signals can also be called “virtual microphones”.

Figure 1: A typical setup for M/S stereophonic recording: shotgun microphone with an attached figure-8 microphone for use in a windscreen. [21]

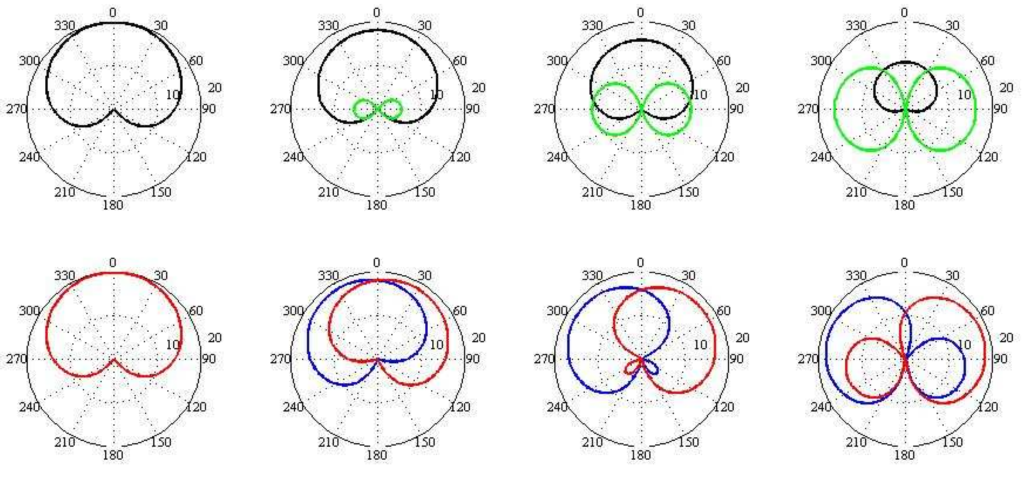


Figure 2: Illustration of the M/S principle:
Top Row: M signal (black) and S signal (green); Bottom row: the resulting signals L (blue) and R (red)

2.3. Double M/S arrangement for two-channel and multichannel stereo

2.3.1. Double M/S: microphone arrangement

Double M/S is a recording technique for two- or multi-channel stereophony which relies solely on signal level differences, not on arrival-time differences.
The underlying principle of the Double M/S arrangement is the grouping of three microphones into two M/S microphone pairs which share one figure-8 microphone. Figure 3 illustrates this principle:

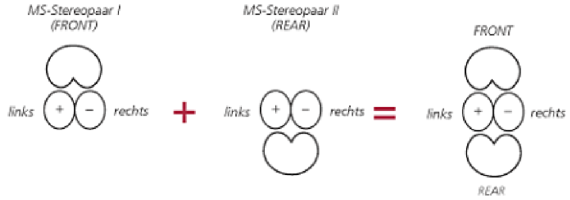


Figure 3: The principle of the Double M/S arrangement: combination of two M/S pairs [21]

The three microphones/channels are named:

- M_{front}
- S
- M_{rear}

By using three compact, small-diaphragm microphones it is possible to arrange them in almost perfect coincidence, *i.e.* all at the same point, see Figure 4.

2.3.2. Double M/S: generation of 2/0-Stereo and 3/2-Stereo Signals

It is possible to generate signals for two- or multichannel stereophony from the three Double M/S microphone signals. This can be done by using one conventional M/S matrix to generate the L/R signals from M_{front}/S and a second such matrix to generate the LS/RS signals from M_{rear}/S. Furthermore, a center speaker can be driven by the signals from the M_{front} microphone. However, the Double M/S arrangement also allows for a much better form of decoding in which the directional patterns of the "virtual microphones" are independent of the mixing ratio chosen for M and S (see section 1.1 for a how-to).
This overcomes the basic disadvantage of conventional M/S, which is that the directional pattern of the virtual microphones depends on the mixing ratio chosen for the M and S signals. This means that with increasing "S" signal level, the resulting directional pattern develops from cardioid to figure-8 (see Figure 2). With a Double M/S arrangement, the signals from the three microphones can be mixed to create any directional pattern for the virtual microphones. As an example, Figure 5 shows both conventional decoding (top row) and Double M/S decoding to constant supercardioid patterns (bottom row).

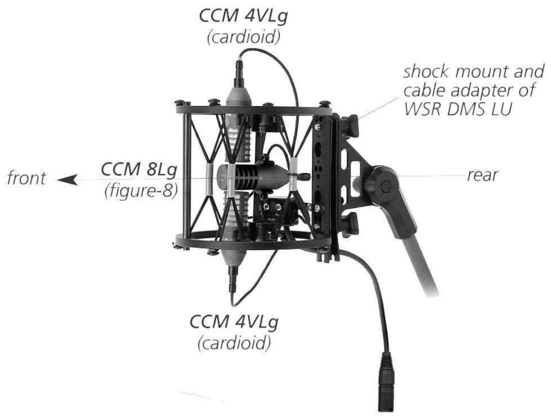


Figure 4: Double M/S arrangement: by employing compact microphones (SCHOEPS CCM 4V and CCM 8), the smallest possible spacing between the three diaphragms is achieved

This advantage is vital for optimized coincident recording. It enables the user to vary the recording angle without changing the directional pattern. It makes it possible to adjust the correlation of the resulting virtual microphone pair without changing the recording angle. Hence, it enables maximum decorrelation of the two signals–an important consideration for the M/S technique. The importance of these parameters will be discussed in section 3.2. Practical tools for Double M/S decoding are described in more detail in section 1.1.

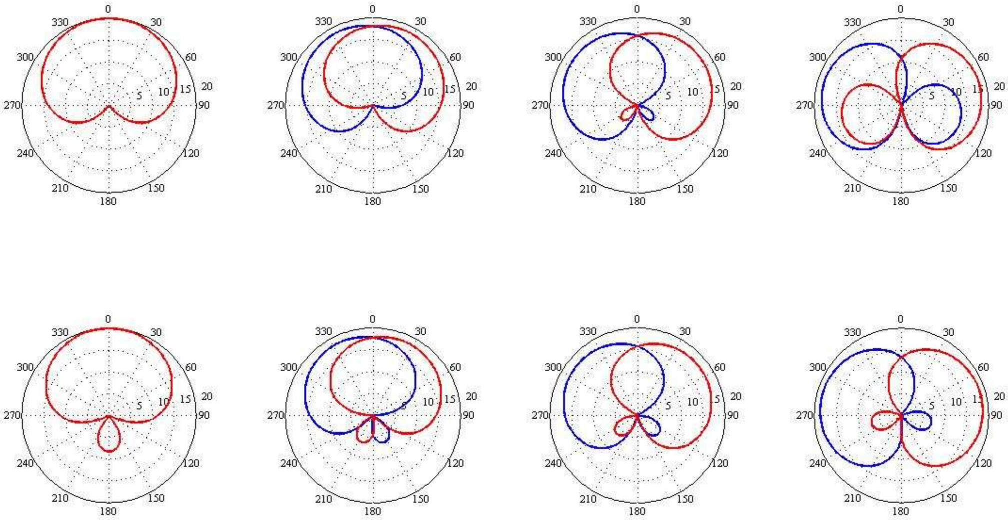


Figure 5: Illustration of the difference between decoding with and without variable directional pattern of the M_{front} signals. Indicated is the polar pattern of the decoded channels L, R.

Top row:	**fixed directional pattern of M_{front};**	**M_{front}=cardioid (compare Figure 2)**
Bottom row:	**variable directional pattern of M_{front};**	**L, R=supercardioid**

2.3.3. Similarities of the Double M/S system to the Ambisonics system

Ambisonics is a recording and playback technique invented by Michael Gerzon [8]. 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 is their descriptive precision.

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

First order B-format*:*

0^{th} order: $W = 1;$

1^{st} 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.

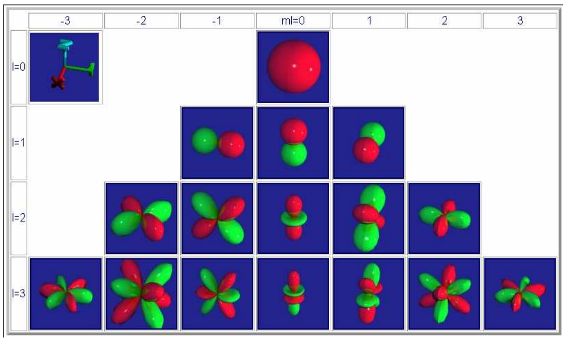


Figure 6: Visualization of spherical harmonics to order three; l (y-axis) defines the order, and ml (x-axis) the dimension, from [35]

These four signals can be obtained from microphones in different ways. 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 [25]). 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 needing to convert from the tetrahedral "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 omnidirectional microphone (W) and three orthogonal figure-8 microphones (X, Y, Z) and is referred to as "native B-format recording" (see [4] 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 set up. 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). This makes no difference if the playback of the signals is on a conventional speaker system without a Z-component (e.g. 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 MK 2 and MK 8) led to better polar patterns for B-format, yet sound from any other direction than horizontal resulted in rather less ideal response. The tetrahedral setup provided good, consistent polar patterns independent of sound direction, but irregularities occurred above frequencies of about 6 kHz. See also Flock [7] for more details.
As mentioned above, the underlying principle of the Ambisonics theory is the analysis of 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 localization from the phantom source theory. Consequently the two theories cannot be compared directly. If a first order Ambisonics signal were to be evaluated with regard to the parameters this paper is concerned with, the result might be negative. In fact, many engineers reject mixes that have been recorded in this way. However, it is important to note that this is not necessarily due to the coincident nature of the recording, as is often stated. There are many types of coincident recording and many ways of judging their quality and optimizing them [16].

3. PARAMETERS FOR THE THEORETICAL ANALYSIS OF THE M/S TECHNIQUE

In this section, various important parameters for the objective evaluation of the M/S technique—whether for two-channel stereo or surround sound—will be discussed. 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 environment 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 [31] and [32]; the following description will not go into any great depth on the subject.
The apparent positioning 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: [26], [27]

This linear addition is valid only for phantom source shifts up to 50% of the maximum shift. After this point, there is a gradual saturation up to the point the source is localized in the direction of one speaker. The author of this paper describes this behavior as a mathematical function ([31], [32], [33], [34]). This approximation is illustrated in Figure 7 below. It must be noted that the shift of the phantom source is proportional to speaker separation. For this reason, the source shift is expressed in per cent so as to be valid for any playback system geometry. In a normal stereo triangle a shift of ±100% would correspond to a shift of ±30°. A 100% shift means that the source is localized in the direction of one speaker. The graphs in Figure 7 show that different data on the phantom source localization exist. This is also due to the fact that the source shift depends on the type of stimulus (static or impuls) and on the spectral composition.
With the help 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 Figure 8; [33], [34]).

The localization curve describes the shift of the phantom source as a function of 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 shift of the phantom source between the front speakers, in percent, on the ordinate. The recording angle can be found by looking for the area on the graph in which the source is localized 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.

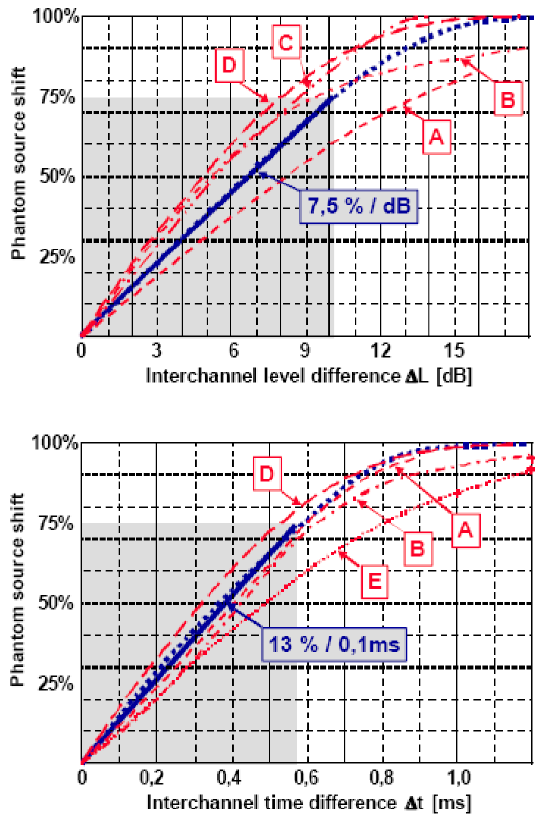


Figure 7: Relationship between level difference (top picture) and time difference (bottom picture) and phantom source shift. From:

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

A: Leakey [17]
B: Mertens [19]
C: Brittain and Leakey [3]
D: Simonson [24], Basis of the "Williams curves" [29]
E: Sengpiel [23]

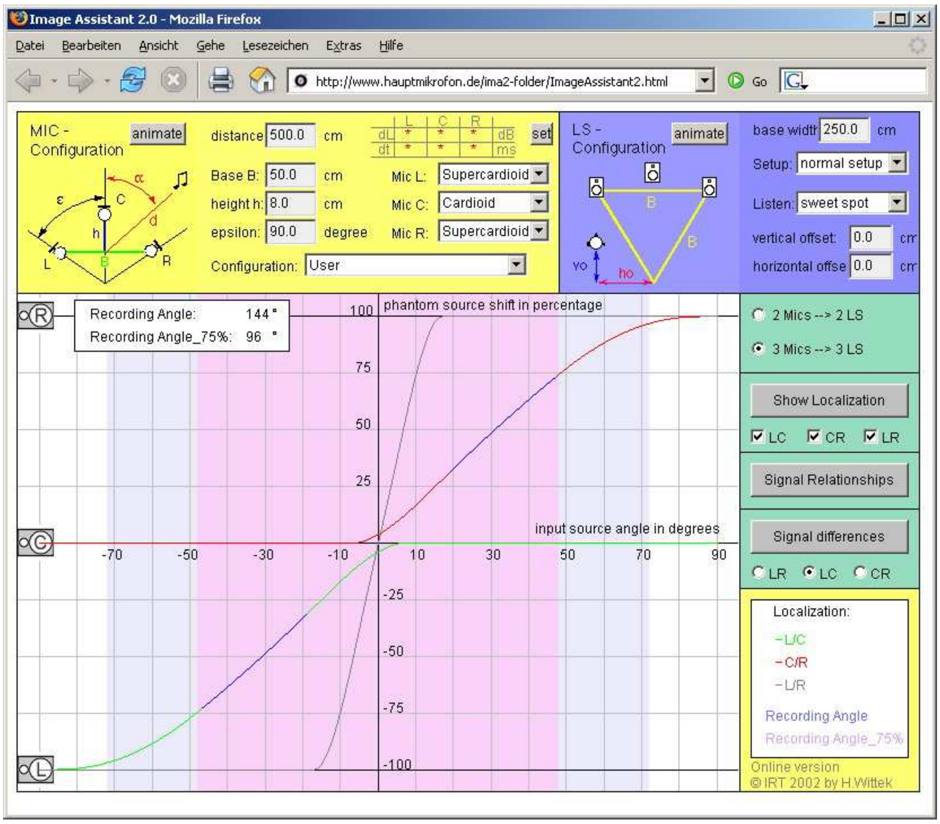


Figure 8: Simulation by the "*Image Assistant*": localization curve of the OCT setup.

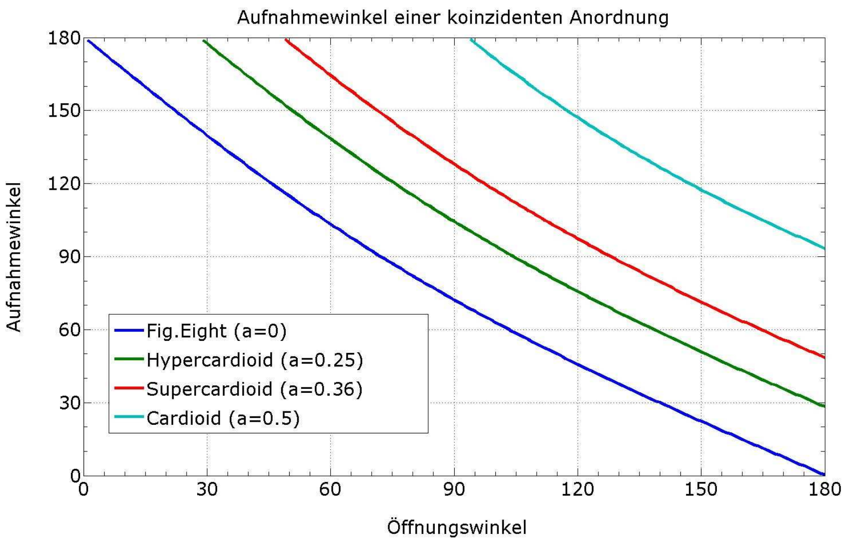


Figure 9: Graphs of recording angle against offset angle for a XY microphone arrangement applying several different directional patterns. (There is no meaningful recording angle for a coincident arrangement consisting only of wide cardioids or omnidirectional microphones).

3.1.2. Comparison of the imaging characteristics of stereophonic setups

Many authors have designed their microphone setups such that the directional imaging 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 automatically have other good attributes regarding the other parameters mentioned above as well. Nonetheless, it is important to observe all parameters when developing new combinations and techniques. The setups proposed by Williams (MMA, see [28] and [30]) are purposely designed for optimal (360-degree) directional 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, but recommends certain arrangements to ensure that parameters other than directional imaging are well provided for. An example of this type of microphone setup is the OCT setup (see Figure 10, [26]) which has various other advantages. First, the localization curve of this setup, shown in Figure 8, is very linear which entails very natural directional imaging without geometrical distortion. In addition, crosstalk between the imaging areas of the microphones is minimized in an OCT setup. In other words, a signal is never reproduced coherently in all of the three channels. This has advantages for spatial imaging, timbre and the robustness of the image.

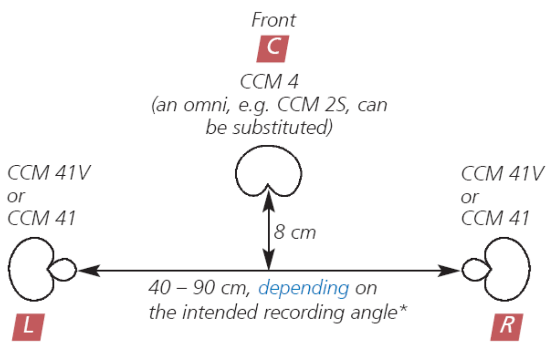


Figure 10:The OCT arrangement proposed by Theile [26], from [21]

The directional imaging properties of an M/S setup depend on the decoding of the signals. This is clear as the decoding makes any variation of a coincident arrangement with 2, 4, 5 or even more virtual microphones possible. Figure 9 shows the recording angle vs. the offset angle and the chosen pattern for a two-channel coincident X/Y setup. Note that not every combination is useful, since there may be an uneven energy balance (for example, a combination of two cardioids at 180° has a 3 dB loss of energy at 0°).

For Double M/S surround, the imaging is critical as five coincident first-order microphones simply cannot produce different enough signals to create optimal tonal and directional imaging. For this reason, care must be taken; frequently a 4-channel decoding (without a center channel) produces better results than a 5-channel decoding when it comes to 360-degree imaging.
The frontal directional imaging of a setup without a center channel is not critical (see Figure 11). The frontal directional imaging of the 5-channel variation is shown in Figure 12. The simulation clearly shows 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 section 3.3). The *Image Assistant* shows that the central area of playback is produced by three speaker pairs. This multiple reproduction results in reduced image focus and decreased locatedness[1]. Even so, it does not necessarily result in worse characteristics for the recording, since crosstalk also occurs with other recording techniques (see section 3.3).

A paradox arises between the two theories of localization: Ambisonics allows the use of any number of speakers, yet focuses only 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 [26]) has negative effects for sound color and localization. If localization would 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] Locatedness is defined as the "spatial distinction of a source" [2] or "the degree to which an auditory event can be said to be clearly perceived in a particular location"

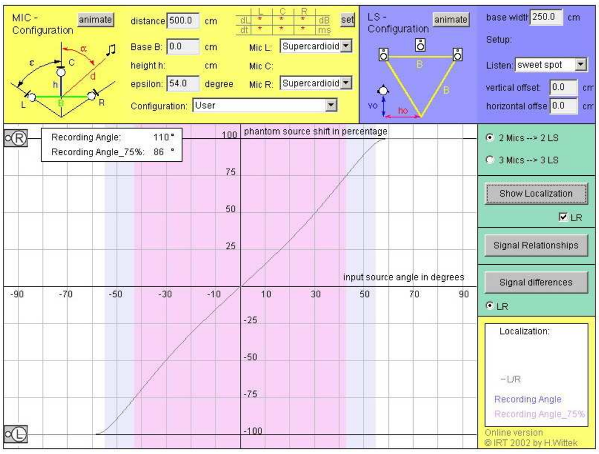


Figure 11: Directional image between L/C/R speakers using 4-channel DMS decoding (= 4-channel setting of the SCHOEPS Hardware Matrix, see section 1.1), simulated using "*Image Assistant*" [33]

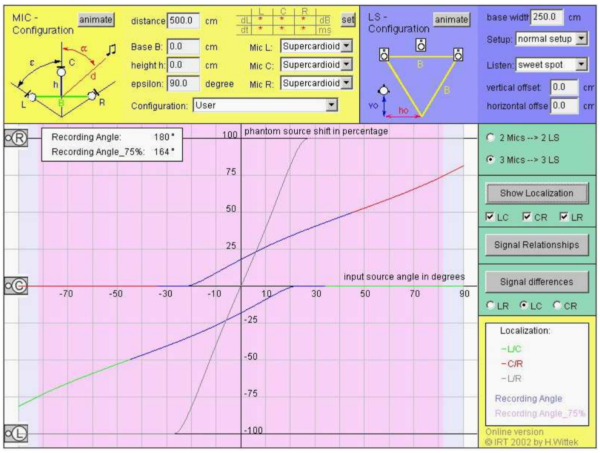


Figure 12: Directional Image between L/C/R speakers using 5-channel DMS decoding, simulated using "*Image Assistant*" [33].

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 the channels of a stereophonic recording is often regarded as a parameter which influences spatial perception and sound color significantly (see [6], [9], [18], [20], [26]). Diffuse-field correlation is considered as a decisive parameter for the differences in perception between various types of stereophonic setups. For example, arrangements with increased microphone spacing are known to have better spatial imaging qualities (see, for example, [36]); this is due in large part to their low diffuse-field correlation.

Care has to be taken when analyzing the correlation of a stereophonic signal. Naturally, both the direct sound and the early reflections should be reproduced coherently in adjacent channels. This is because they are to be localized as phantom sources within the loudspeaker base. In particular the lateral reflections are important for the spatial perception. However, for the perception of envelopment the reverberation is responsible, i.e. the diffuse sound after 50-100 ms. Therefore the diffuse sound has to reproduced diffusely.

It is difficult to create a diffuse sound field with a limited number of loudspeaker channels. Hence, it is vital that the reverb is reproduced by largely incoherent loudspeaker signals. If this postulation is not fulfilled, which means if the reverb is reproduced with coherent loudspeaker signals, the consequence will be a degradation of both sound color and spatial impression. The reverb then will be localized and will sound unnatural. Furthermore, reverberation can be considered similar to a noise signal: correlated contributions on two loudspeakers can cause severe comb filtering while head movements.

Hence, the parameter “diffuse field correlation” should be given more consideration with respect to M/S recording since it is essential for the timbral and spatial differences between decoding methods.

---

[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 [5].

The parameters directional imaging and diffuse field correlation are not independent in a coincident recording, but are very closely related. Correlation increases with larger recording angles. Furthermore, correlation and imaging properties are influenced by the type of microphone used.
In Figure 13 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 depends on 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. We demand a correlation coefficient of zero, hence, this can only by achieved in a coincident setup with directional patterns that have an omnidirectional component not larger than that of a supercardioid microphone (see also Griesinger [12]).
Figure 13 shows that the 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 angles, but different directional patterns:

Offset Angle = 90°	**Cardioids**	**Super-cardioids**	**Figure-8s (Blumlein)**
Correlation coefficient	0.75	0.49	0.00
Recording angle, (Recording angle 75% [32])	180° (142°)	130° (104°)	72° (58°)

It is interesting to investigate the correlation vs. the resulting recording angle. It then becomes possible to assess, for a given recording angle, which arrangement has optimal decorrelation in the diffuse field. This can be done by first calculating the recording angle as a function of the offset angle and directional pattern. The results of this calculation can be seen in
Figure 9. 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 16 dB between the two microphone signals.

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

With these values, the correlation coefficient for a coincident arrangement of two microphones, with any directional pattern, can be obtained as a function of the recording angle. Figure 14 shows that the diffuse field correlation coefficient of a coincident setup with a fixed recording angle is quite strongly dependent on the microphones' directional patterns. For a given recording situation this figure can help to choose the appropriate (=least correlated) microphone pair or (Double) M/S decoding. There are of course restrictions on the choice of directional patterns; not all values shown in Figure 14 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, it can be said that the reduction of the diffuse-field correlation between the channels of a stereophonic recording is vital for spatial perception and timbre. However, it is difficult to control this parameter independently of the recording angle.

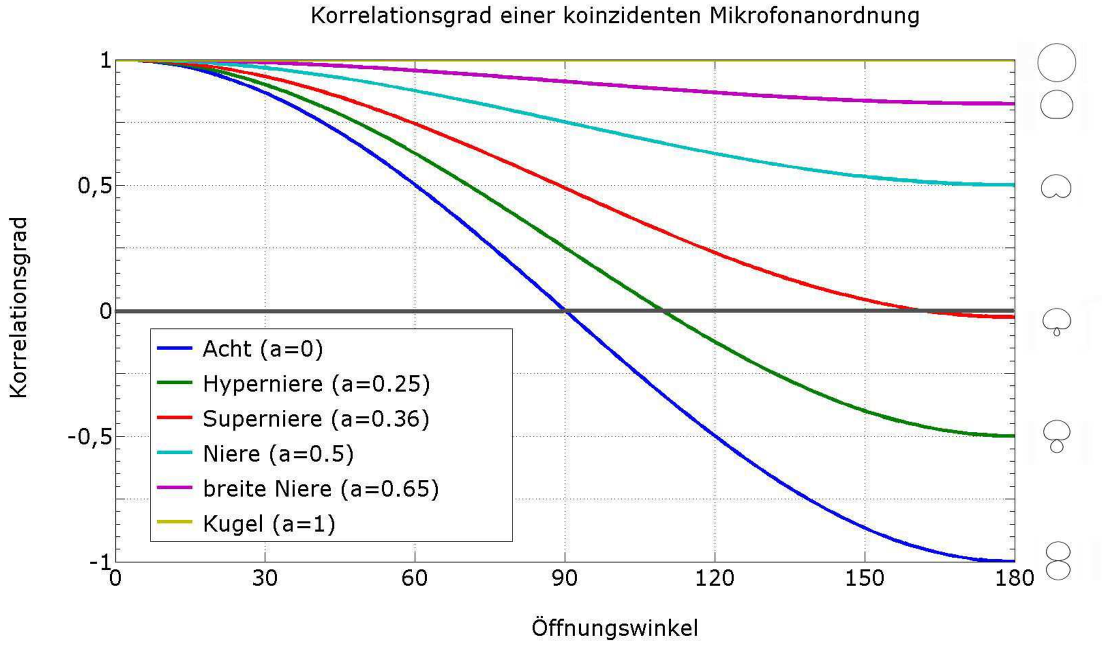


Figure 13: The correlation coefficient in the diffuse field vs. the offset angle between the microphones for several different directional patterns. The omnidirectional portion "a" of the microphone pattern is also given according to the formula: Output level = a + (1-a) * cos (ϕ); with ϕ being the source angle.

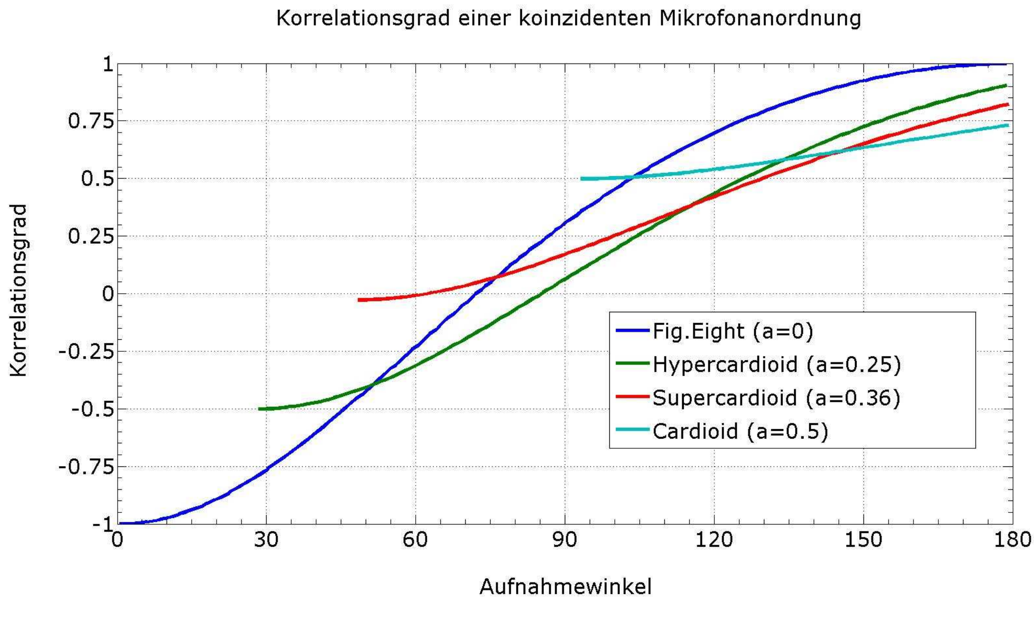


Figure 14: Correlation coefficient of a coincident arrangement of two microphones in the diffuse field vs. the recording angle of the arrangement, for different directional patterns. (There is no meaningful recording angle for a coincident arrangement consisting only of wide cardioid or omnidirectional microphones.)

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, homogeneous directional imaging and minimal diffuse-field correlation between channels is achieved under the following conditions:

- 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 analyzed. Meeting the second condition requires a choice of a directional pattern which is the best possible compromise between strong directionality and the potentially disturbing effects of the back lobe: the supercardioid. Hence an ideal setup would consist of four or five supercardioids set up at a maximum angle to each other.
These requirements lead to the 4-channel or 5-channel decodings in Figure 15 and Figure 16. The resulting values of diffuse-field correlation of these optimized Double M/S setups are given in the tables below.

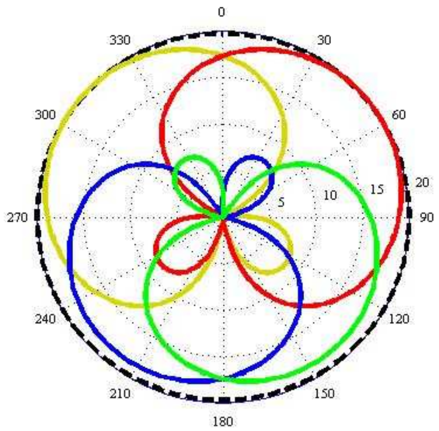


Figure 15: An optimized 4-channel decoding of the Double M/S system

	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

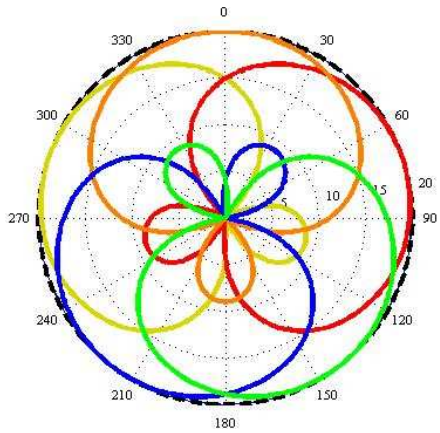


Figure 16: An optimized 5-channel decoding of the Double M/S system

	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

From the above tables one can see that lower diffuse-field correlation is achieved with the 4-channel setup. Areas of higher correlation are found only in the rear. 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 best suited for recording situations where robust directional imaging is required. For sufficient spaciousness and enveloping, high degrees of decorrelation are mandatory (e.g. Griesinger [9], [10], [11], [12]).
With a two-channel M/S or X/Y recording, diffuse-field correlation coefficients of zero are possible. This cannot be achieved with two cardioids since their correlation coefficient never falls beneath 0.5. Partly for this reason, M/S and X/Y recording has a worse reputation

than it deserves. X/Y recordings are often described as narrow, overly "centered" and unsuitable for the imaging of rooms. But this is true only for X/Y recordings with high correlation, and hence does not apply to optimized X/Y setups. Furthermore, X/Y were in the past often done with double-membrane microphones which inherently have a loss of directionality at low frequencies. The low frequency decorrelation and thus the spatial impression then is even poorer.

In Figure 13 and Figure 14, the offset angle for an ideal X/Y 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 72°)
- Hypercardioid (a = 0.25): 110°
  (recording angle 85°)
- Supercardioid (a = 0.36): 160°
  (recording angle 64°)

With the Double M/S setup, these and other coincident arrangements can be reproduced and thus optimized in regard 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 section 4. It turns out that the generation of four channels is easily achieved by good decoding. A five-channel mix 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, filters, level control etc. where necessary. Attempting to generate more than 5 channels is not possible without strong interchannel crosstalk and correlation, which is why proposals for (near-)coincident arrangements for new formats such as 7.1 or even 10.2 must by regarded critically. The number of channels generated from a microphone system is by no means a measure of its quality.

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 potentially audible comb filtering appears.
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 [26] who seeks to avoid multiple imaging due to crosstalk by the use of suitable microphone arrangements. Lee and Rumsey [15] investigated different multichannel microphone setups and found negative effects on image width and source focus due to crosstalk. Crosstalk also decreases the listening area, since even small movements towards one speaker can result in localization problems caused by the precedence effect. The effect on localization for listeners aside the sweet spot can be approximated 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 the previous section can be applied. As before, the optimum arrangement consists of virtual supercardioids at maximum offset angles to each other.
The optimized decoding variants shown in Figure 15 and Figure 16 result in the crosstalk behavior shown in Figure 17 and Figure 18.
The 4-channel decoding shown in Figure 17 has a maximum crosstalk level of approximately -7.5 dB. This level is reached at two angles; through 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 approximately -5 dB. This value is reached in multiple positions in the sound field.

The comparison of the 4-channel and 5-channel decodings shows that better quality can be achieved using 4-channel decoding rather than 5-channel decoding. The extent of practical disadvantages implied by this theoretical disadvantage will be discussed in the following section.

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

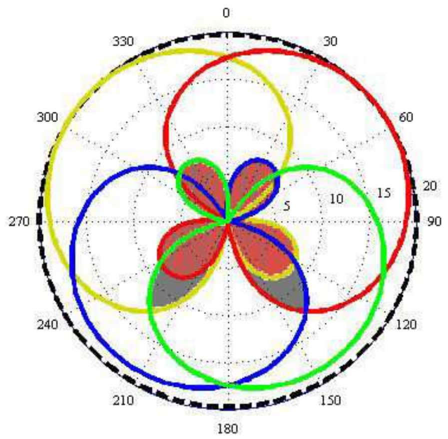


Figure 17: Crosstalk levels of the optimized 4-channel Double M/S decoding. (black areas are in-phase, red areas are out-of-phase)

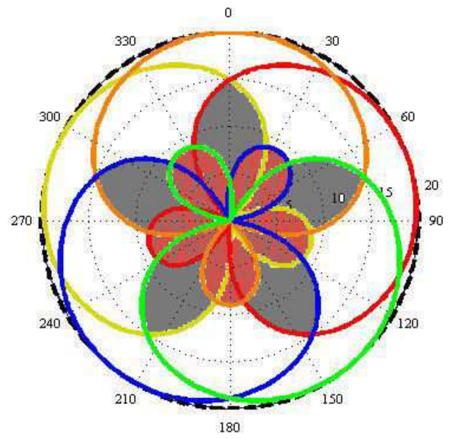


Figure 18: Crosstalk levels of the optimized 5-channel Double M/S decoding (black areas are in-phase, red areas out-of-phase)

3.3.2. Practical Analysis

To investigate the impact of crosstalk on different aspects of perception, an experiment was performed (see [13] for details). The aim of this investigation was to find the perceptual threshold of a crosstalk signal with respect to the following attributes of the phantom source:

- Width
- Direction
- Locatedness[1]
- Sound Color

The phantom source was created between the center and right speakers, the level difference being 3.7 dB. The crosstalk signal was created by the left speaker with levels ranging from -20 dB to -5 dB.
The participants, 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 a scale from 1 (no change in the respective parameter) to 10 (big change). If the crosstalk stimulus was perceived when it was not actually present, this was counted as a negative result.
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.
Some of the results of the experiment are shown in figures 19–22 (the results for dry speech). The diagrams show how audible the changes of the different sample attributes were perceived to be by the candidates. The scale above was reduced by one to define zero as "no change perceived". The perceived change is given vs. the change in crosstalk level of the third speaker.
The results showed that changes in direction and width of the phantom source were detected more readily than changes in locatedness or sound color. The threshold (relative to the sum level of the phantom source C/R) is approximately -12 dB for changes in direction, and approximately -9 dB for changes in width. Locatedness decreases from -9 dB, and the sound color changes from -6 dB.
These results show clearly that crosstalk has negative effects and should be avoided where possible. This concern can definitely be taken into account when determining the optimal decoding of Double M/S signals. But even with optimized decoding, the level of crosstalk may still exceed the audibility threshold for some attributes, depending on the setup that is used.

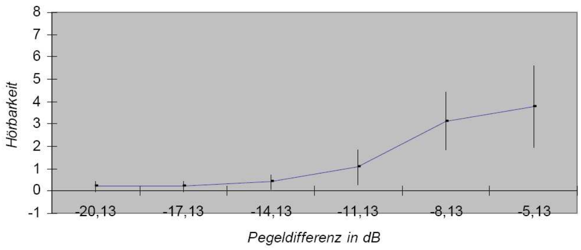


Figure 19: from [13]: Perception of phantom source width. Signal: dry speech. Arithmetic mean, including 95% confidence interval.

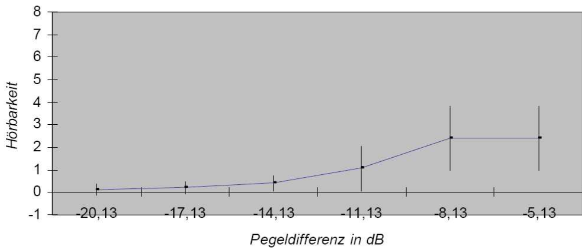


Figure 20: from [13]: Perception of change in phantom source locatedness. Signal: dry speech. Arithmetic mean, including 95% confidence interval.

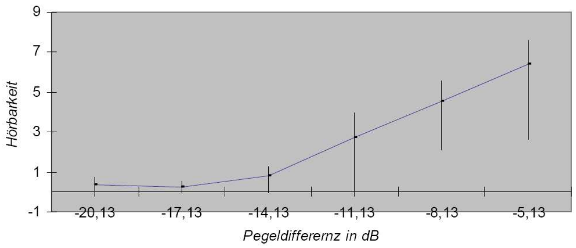


Figure 21: from [13]: Perception of change in direction. Signal: dry speech. Arithmetic mean, including 95% confidence interval.

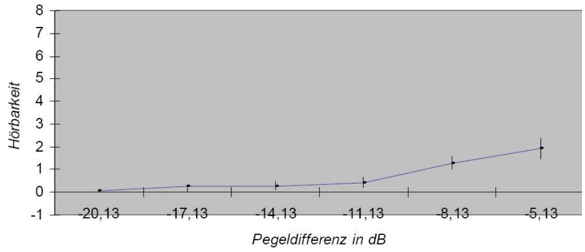


Figure 22: from [13]: Perception of sound color change. Signal: dry speech. Arithmetic mean, including 95% confidence interval.

4. PRACTICAL INVESTIGATION BY MEANS OF DIFFERENT TEST RECORDINGS

Despite the fact that “nothing is more practical than a good theory” (Gerhard Steinke), even a good theory has its limits and cannot explain everything we perceive. For this reason, 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
- To examine the advantages and disadvantages of Double M/S recording techniques as established in the previous sections
- To examine of the usability of Double M/S in different situations such as music, “atmo,” theater, radio drama, documentation/film and television studios
- To compare Double M/S with 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 will be postulated here; the focus will rather be on describing experience and the collected comments of others.

4.1. Different methods of Double M/S decoding

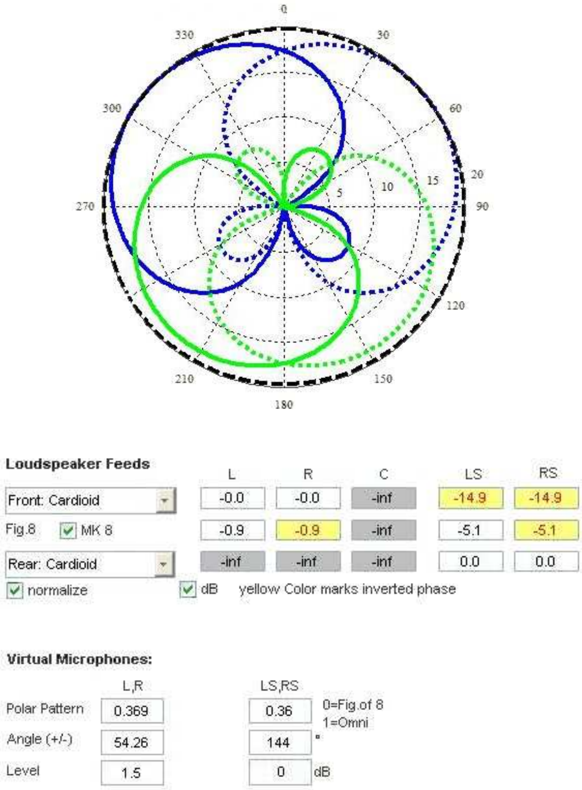
The following variations of Double M/S recording, denoted A–F, were investigated practically (see also [14]). The discussion describes the subjectively perceived impressions on the sound quality.
The polar patterns and offset angles of the resulting virtual microphones together with the overall energy (=loudness) level are shown. Furthermore, the examination of each method includes the level matrix with which the decoding was performed. Note that the level matrix includes the correction of the SCHOEPS MK 8 sensitivity, i.e. the Fig-8 level is 2.3 dB lower than stated.

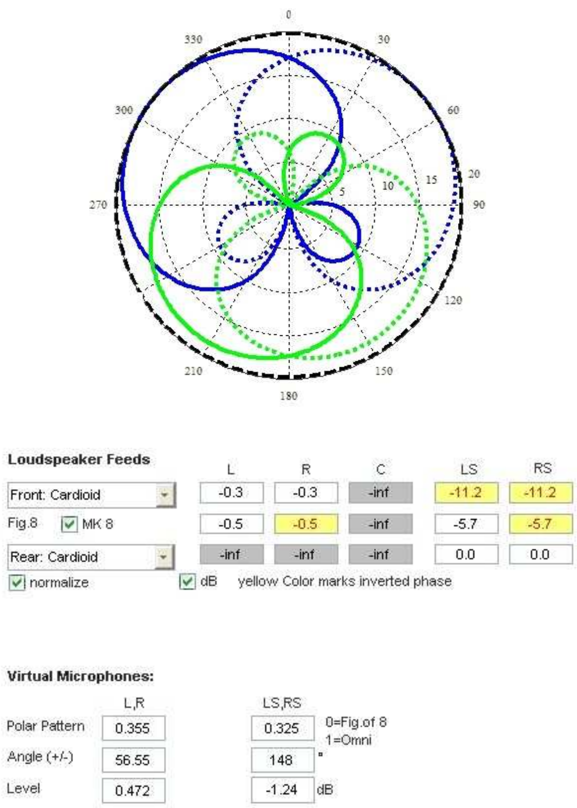
A) Optimized 4-ch decoding: 4 supercardioids

The supercardioid directional pattern has established itself as a good compromise, since the directionality of the supercardioid along with its strong rear signal suppression leads to low crosstalk levels. In this decoding, the omission of the center channel allows for smaller offset angles.
Directional imaging is well-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 make the setup suitable for music and radio drama recording. Whether the variants with a center channel are preferred depends upon individual taste in the author's experience.
This decoding is realized in the MDMS U device (4-channel), see section 1.1.

B) 4 supercardioids, broader imaging

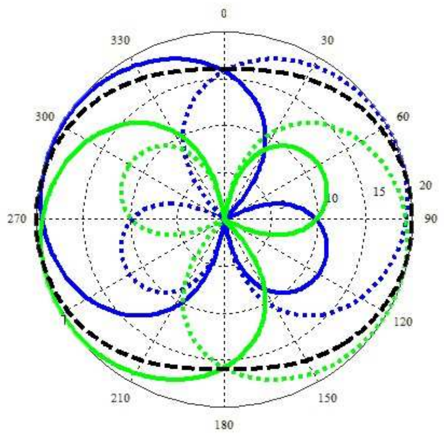
To decrease the recording angle of a stereo recording, either the microphone offset angle or the directionality must be increased. This example differs only slightly from the previous one, since any attempt to significantly decrease the L/R recording angle while maintaining the LS/RS stereophonic imaging would result in an uneven distribution of energy.
With this setting, the L/R recording angle is 90°. (as compared to 110° in variation A).





C) "Conventional" M/S-decoding (4-channel)

These pattern result if the decoding was done with k_1, $k_2 = 0.5$ (see section 2.1). 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

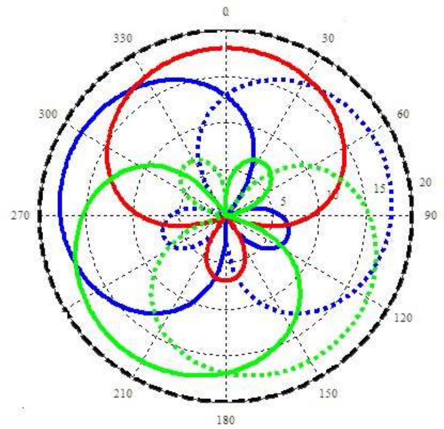
☑ normalize ☑ dB yellow Color marks inverted phase

Virtual Microphones:

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

D) Optimized 5-ch decoding: 5 supercardioids

To obtain balanced localization and energy distribution, the supercardioids for L, R are rotated further apart than in the 4-channel version. A balanced image is obtained, with added stability due to the center channel. The choice between this variant and 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 unless further measures are 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.
This decoding is realized in the MDMS U device (5-channel); see section 1.1.



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:

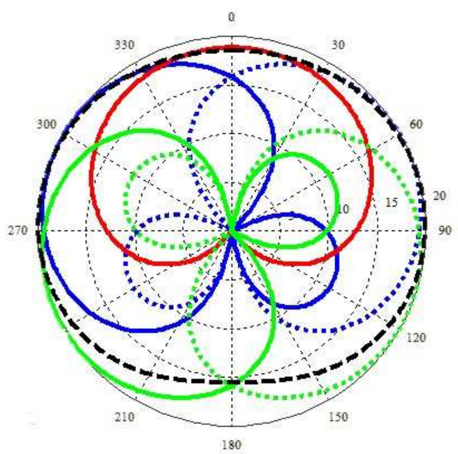
	L,R	C	LS,RS	
Polar Pattern	0.36	0.36	0.36	0=Fig.of 8 1=Omni
Angle (+/-)	71.99	0	144	°
Level	-1.97	-1.97	-1.97	dB

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

These pattern result if the decoding was done with k_1, k_2 = 0.5 (see section 2.1) and the center channel was provided with the frontal cardioid only. In this decoding variation the channel overlap is very strong, which is a disadvantage. The setup shows that bad decoding without control produces bad results; displeasing acoustic imaging and missing transparency to name but a few issues. Furthermore, the sound color changes with small head movements.

F) Delayed surround channels

This setup is an improvement over setup D) since it increases channel independence in the front area by means of increased offset angles, and in the rear by means of a delay (Δt = 10 ms).
This setup is ideal for many applications in which the center channel is needed. However, a front-emphasized recording scenario is mandatory, as the delay prevents stable rear-localization.

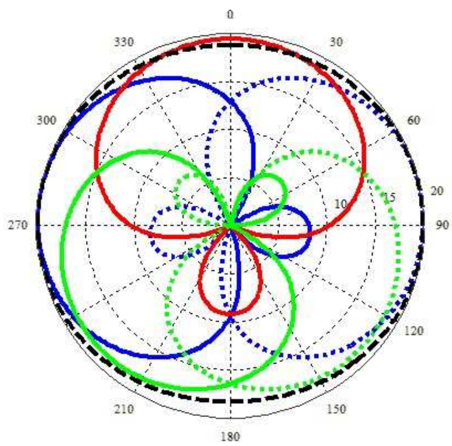


Loudspeaker Feeds	L	R	C	LS	RS
Front: Cardioid	-3.0	-3.0	0.0	-inf	-inf
Fig.8 MK 8	-0.7	-0.7	-inf	-0.7	-0.7
Rear: Cardioid	-inf	-inf	-inf	-3.0	-3.0

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



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 experience with Double M/S recording

Experience shows that the Double M/S technique can yield good results in a variety of recording situations. These are our subjective experiences, which were also compared with others' and found to be quite similar. For example, an ambience recording with direct sound incident from all directions (town square) was depicted very well and showed similar quality to a simultaneous recording with an IRT cross. The flexibility of the Double M/S system proved 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), on the other hand, 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, but a parallel recording using an OCT setup provided results that were even more effective and spacious. The decision between these methods is left to the engineer. In our experience, Double M/S is more suitable for small spaces whereas larger spaces require OCT or other setups. For music recordings, the addition of a low-passed omnidirectional microphone for the low frequencies is recommended. The mix of the Double M/S signals with a large A/B configuration of omnis results in the spacious sound that is often desired. This option also provides decorrelated low-frequency signals.
An *a capella* choir with the need for good imaging between speakers was recorded well using a Double M/S setup; variations both with and without a center-channel 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.

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

Figure 23: Test recordings using the Double M/S and reference setups: Top picture: Double M/S and OCT surround system in a live theater performance. Bottom picture: Double M/S and an IRT cross setup for an ambience recording at Durlach town square.

A live TV show with audience also worked well using either the IRT cross 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 downmix or even mono compatibility, which is easily achieved with Double M/S.

The use of Double M/S in film sound recording is long established. Its use in documentary film has 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 a shotgun microphone can be used. This latter setup requires a special microphone setup to ensure optimum coincidence (see Figure 24). The Double M/S method works well with this setup; the advantages of surround sound on the boom are apparent especially in documentary filming where authenticity is important, thus allowing a “subjective” sound perspective to be used.

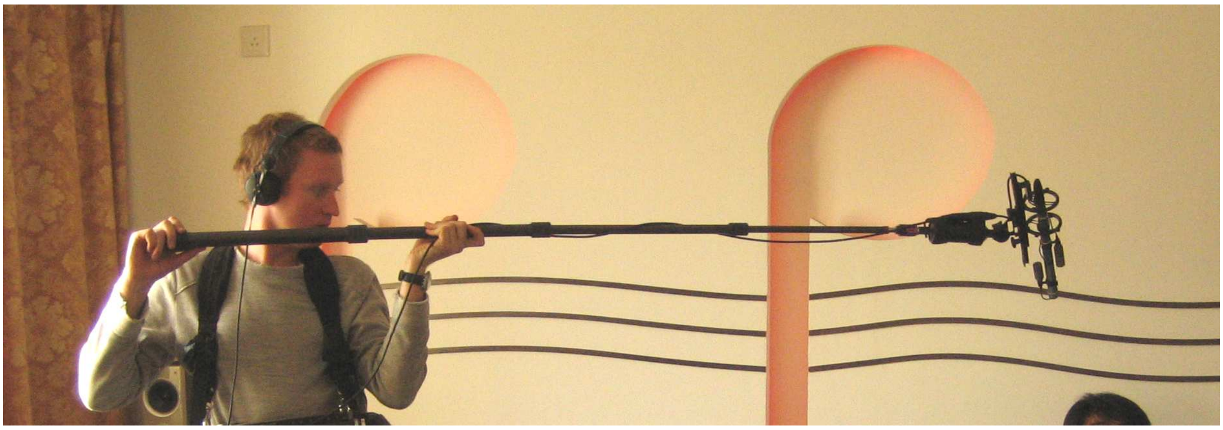


Figure 24: A Double M/S setup with a shotgun microphone.
Top: Implementation using the SCHOEPS "CMIT" Double M/S set (microphones: CMIT 5, CCM 4 and CCM 8) [21].
Bottom: the setup in action ([37], photograph courtesy of André Zacher).

4.3. 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 recording or after recording during post-production.
The following three basic principles of decoding will be discussed:

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

b) MDMS U Double M/S hardware matrix

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

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

In principle, Double M/S recordings can be decoded like M/S recordings simply by using two M/S matrices instead of one (see Figure 25). 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. If a shotgun microphone is used for the M signal, this double-matrix setup is recommended, since a mixing of the shotgun and rear cardioid signals makes no sense.

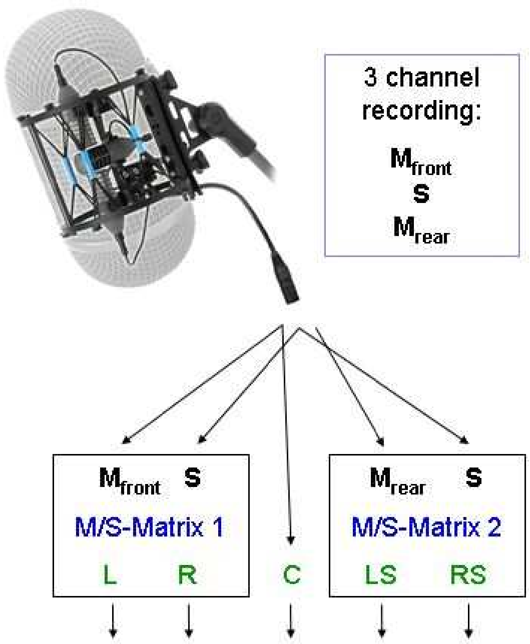


Figure 25: Double M/S decoding using method a)

As discussed in previous chapters, the properties of a Double M/S setup can be improved by a suitable decoding. These decodings involve the combination of all three microphone signals to synthesize the L/R/LS/RS channels and the two cardioids for the center channel. This can be difficult to achieve by only using level matrices or mixers. It is simpler to use these special tools for optimum decoding:

b) SCHOEPS MDMS U hardware matrix.

This analog, passive matrix directly produces the 4 or 5 channels L/R/(C)/LS/RS from the three Double M/S signals (see Figure 26). It can be used during recording or in post-production, and offers the decoding options A and D as described in section 4.1. The matrix equalizes the different sensitivities of the SCHOEPS capsules.

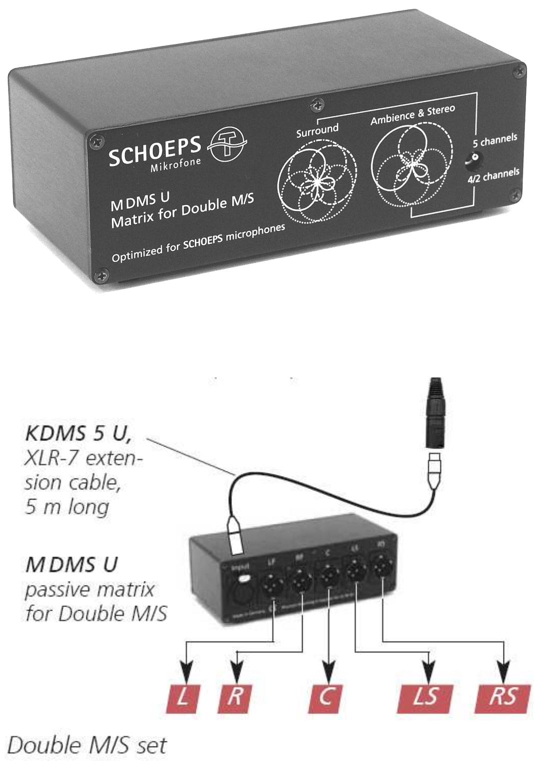


Figure 26: SCHOEPS MDMS U hardware matrix and signal flow.

c) Software VST/RTAS Plug-in „Double M/S Tool“, see Figure 27.

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 the variable parameters. Like 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. A number of presets offer the optimal decoding variants.
The plug-in is available at the SCHOEPS website (www.schoeps.de/dmsplugin.htm [22]) free of charge. In addition, short Double M/S audio samples are available to try out the plug-in. The plug-in is available for Windows and Mac.

ACKNOWLEDGEMENTS

I would like to thank Christopher Haut and Daniel Keinath for their excellent contribution to this work during their time spent at SCHOEPS for their diploma thesis and work experience respectively. Many thanks also to Prof.Dr. Jörg Bitzer of the IHA at the Fachhochschule Oldenburg and Rainer Lorenz of the Institut für Musikwissenschaft/Musikinformatik at the HfM Hochschule für Musik Karlsruhe for their support.

REFERENCES

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

[2] Blauert, J.: “Spatial Hearing”, Cambridge, Massachusetts: MIT Press, 1997

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

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

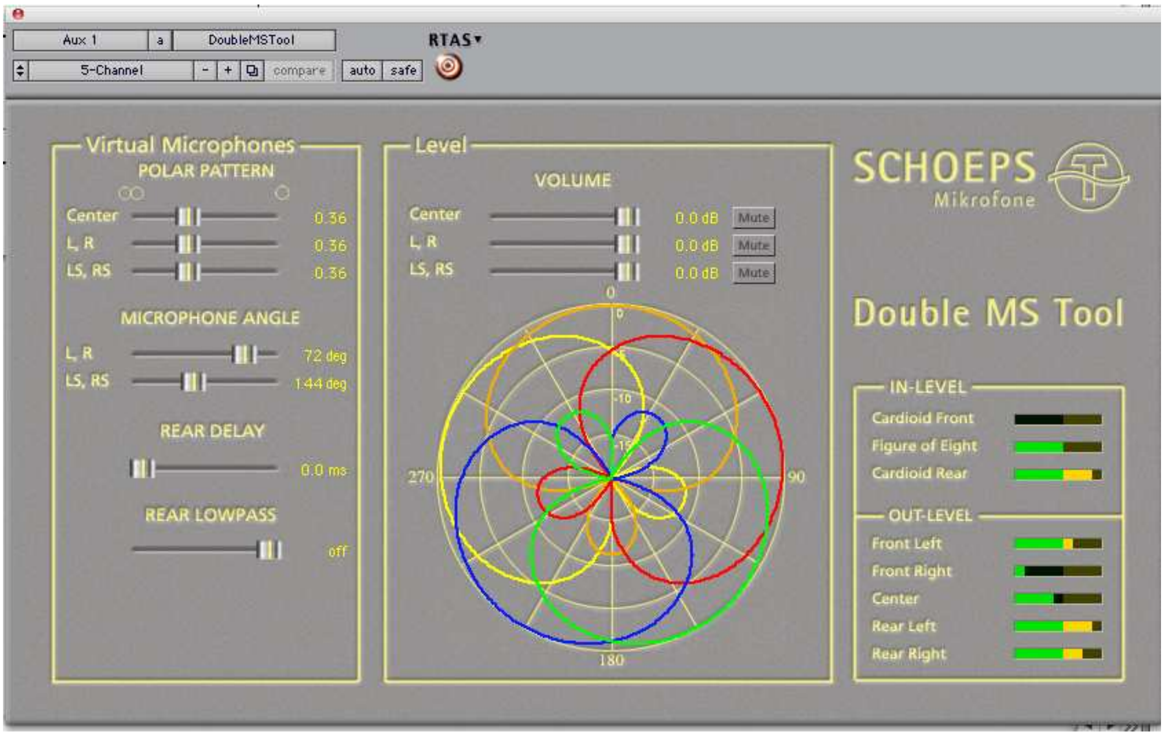


Figure 27: “Double M/S Tool” software plug-in [22]

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

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

[7] Flock, S.: “Theoretische sowie objektive und subjektive Messungen am Soundfield-Mikrofon”, Master thesis, FH Düsseldorf

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

[9] 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

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

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

[12] 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, October 2006

[13] Haut, C.: „Programmierung und Evaluierung einer parametrisierbaren Doppel-MS Dekodierung für eine 5.1 Surround-Abmischung“, Master thesis Hörtechnik & Audiologie at the FH Oldenburg, 2006.

[14] Keinath, D.: „Testaufnahmen mit dem Doppel-MS Verfahren“, internal bulletin for SCHOEPS, Karlsruhe, 2005.

[15] 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.

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

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

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

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

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

[21] SCHOEPS-website: www.schoeps.de, 2009

[22] SCHOEPS-website: Download of the plug-in „Double M/S Tool“ at www.schoeps.de/dmsplugin.htm, 2009

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

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

[25] Soundfield-website: www.soundfield.com, 2009

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

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

[28] Williams, M.: “Multichannel sound recording practice using microphone arrays”, 24th AES International Conference.

[29] 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

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

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

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

[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, on www.SCHOEPS.de, 2009

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