Audio Engineering Society

Convention Paper

Presented at the 130th Convention
2011 May 13–16 London, UK

Prediction of perceived width of stereo microphone setups

Hans Riekehof-Boehmer[1], Helmut Wittek[2]

[1] *HAW-Hamburg*

[2] *SCHOEPS Mikrofone*

Correspondence should be addressed to Hans Riekehof-Boehmer (`h.riekehof@gmail.com`)

ABSTRACT

The diffuse-field correlation of the two signals generated by a stereophonic microphone setup has an effect on the perception of spatial width. A correlation meter is often used to measure the correlation coefficient. However, due to the frequency dependence of the correlation function, the correlation coefficient is not an appropriate value for predicting the perceived width when it comes to time-delay stereophony.

By using the newly defined "Diffuse-Field Image Predictor" (DFI Predictor) presented in this paper, an attempt is made to reliably predict perceived width. Listening tests show that the DFI Predictor is fairly suitable for this task. The aim of the study is to compare the spatial properties of different stereophonic microphone techniques by one calculated value.

1. INTRODUCTION

The resulting directional image of a stereo microphone setup can be predicted by calculating the so-called localization curve. This can easily be done by means of the Image Assistant [1] or the Williams Curves [2]. It is, however, not as easy to predict the spatial quality of a stereo microphone setup.

The "DFI Predictor" introduced in this paper is designed to support this type of prediction. It is known that the interchannel correlation of the diffuse sound field in a recording is important for the perception of spatial width and envelopment. The diffuse sound should be reproduced as being decorrelated, so as to get a wide image in two-channel stereo or a sensation of envelopment in surround [3]. However, other sources suggest that the diffuse sound should have a certain level of correlation, and that complete decorrelation leads to undefined imaging [4], [5].

For now, the DFI (Diffuse-Field Image) Predictor is defined only for two-channel microphone setups.

The DFI Predictor is calculated by using the coherence function for differential microphones in diffuse fields [6]. The correlation of the diffuse sound has also an effect on sound color. The greater the correlation in the reproduction of a diffuse sound field, the more coloration will be perceived.

Differences in diffuse-field correlation between the two channels are likely to significantly influence the preference for a specific recording technique. The diffuse-field correlation of spaced microphones is dependent on frequency. Thus, the correlation coefficient as well as the degree of coherence of the microphone setup in a diffuse sound field can vary depending on the power spectrum of the excitation signal. The DFI Predictor can be calculated for every two-channel microphone setup and includes a frequency-dependent weighting function.

The weighting function is applied to the coherence function [6]. This frequency-dependent weighting function is the main difference between the DFI Predictor and the correlation coefficient or degree of coherence. The results of the listening test show that the DFI Predictor is able to give a rather good prediction of the perceived width.

The DFI Predictor, however, is meant only as a first step in the search for a reliable prediction of perceived width and envelopment.

2. SIGNAL CORRELATION MEASUREMENT

In many psychoacoustic studies the correlation coefficient or the degree of coherence is used to determine signal “similarity”. The degree of coherence and the correlation coefficient are both calculated in the time domain. When it comes to spaced microphones in a diffuse field, the signal correlation differs for each frequency and has a $\frac{\sin(x)}{x}$ characteristic (see figure 2). This frequency dependence is not included in the correlation coefficient or the degree of coherence. The DFI Predictor, however, is based on the coherence function [6] and therefore includes the frequency-dependent correlation of spaced microphones [7].

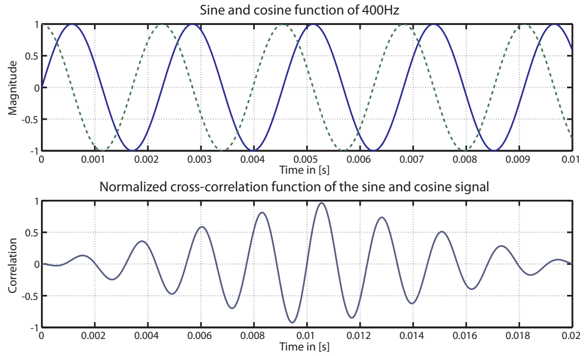


Fig. 1: Top: Sine and cosine signals. Bottom: normalized cross-correlation function of the two signals.

2.1. The correlation coefficient

The correlation coefficient is certainly the most common measurement for signal similarity. It is defined at $m = 0$ of the normalized cross-correlation function $p^E_{sg}(m)$. The discrete type of the normalized cross-correlation function is defined:

$$p^E_{sg}(m) = \frac{\sum_{n=-\infty}^{n=+\infty} s^*(n) \cdot g(n+m)}{\sqrt{E_s \cdot E_g}} \tag{1}$$

with $s(n)$, $g(n)$ being the input signals and E_s, E_g being the corresponding signal energy. The correlation coefficient of the following example is zero because the input signals are orthogonal (see figure 1). The correlation coefficient describes the phase shift between two signals. The real “similarity” of the signals is described inaccurately.

Due to the scaling on the overall energy (see equation 1), the spectral power density is important for the value of the correlation coefficient. Two input signals with the same phase shift but different spectral power densities result in a different correlation coefficient (see figure 7).

2.2. The degree of coherence

The degree of coherence is also based on the normalized cross-correlation function. It is defined [8]:

$$k = \max \left| p^E_{sg}(m) \right| \tag{2}$$

Because the maximum absolute value is used, the phase shift between the two signals cannot be described any more. The degree of coherence for a sine and cosine signal of the same frequency would always be 1 (see figure 1) regardless of the phase shift. The degree of coherence also is dependent on the spectral power density of the input signals.

2.3. The coherence function

The coherence function is able to give a more precise description of the “similarity” of two signals.

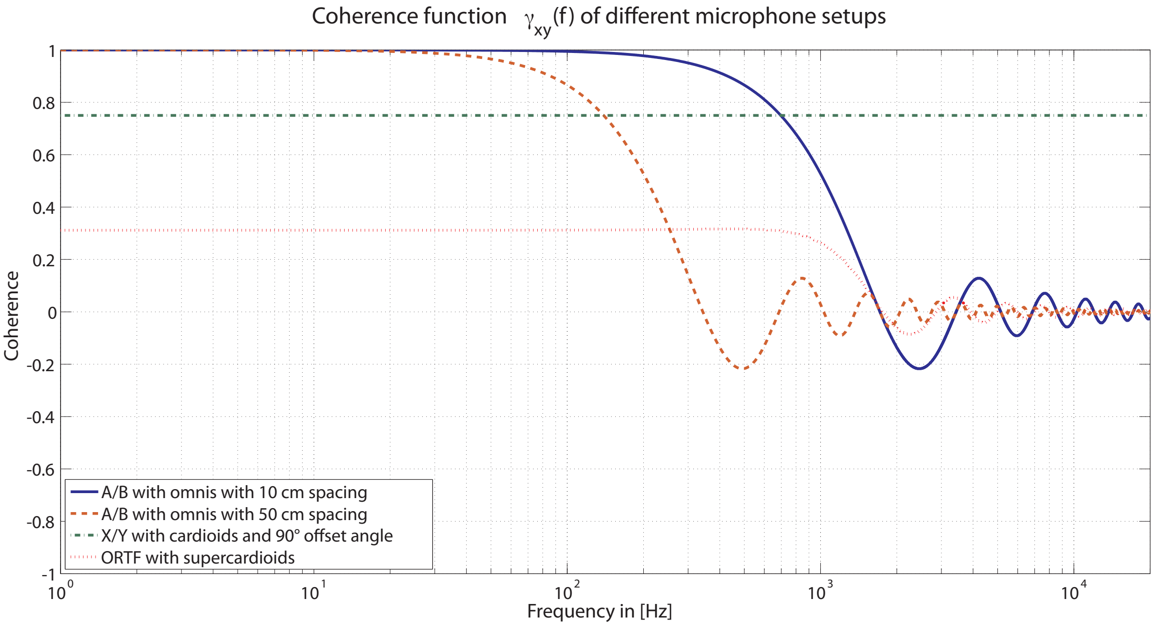


Fig. 2: Coherence functions of different stereo microphone setups in a diffuse sound field.

The advantage of the coherence function is that you have the "similarity" of the input signals at every frequency. In [9] the coherence function is defined as:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f) \cdot P_{yy}(f)} \tag{3}$$

with $P_{xy}(f)$ being the cross power density spectrum and $P_{xx}(f)$ and $P_{yy}(f)$ being the power density spectra. This type of coherence function is called the magnitude squared coherence. In [6] another version of the coherence function is defined (see figure 2):

$$\gamma_{xy}(f) = \frac{P_{xy}(f)}{\sqrt{P_{xx}(f) \cdot P_{yy}(f)}} \tag{4}$$

again with $P_{xy}(f)$ being the cross power density spectrum and $P_{xx}(f)$ and $P_{yy}(f)$ being the power density spectra.

The complex coherence function $\gamma_{xy}(f)$ is able to describe the phase shift between the two input signals. In a nutshell it can be said that the complex coherence function $\gamma_{xy}(f)$ denotes a correlation coefficient for every frequency. Thus, the "similarity" of two broadband input signals can be described more precisely than with the correlation coefficient or the degree of coherence. Furthermore, the coherence function is not dependent on the power spectrum density of the input signals. In [10] it is shown that low frequencies have a strong effect on the spatial impression. In figure 2 the coherence function for different stereo microphone setups in a diffuse field is shown.

An A/B setup with omnidirectional microphones with a spacing of 10 cm is almost completely mono below 400 Hz. For coincident setups the signal correlation is the same for every frequency. In this case, the correlation coefficient leads to a better conclusion on signal correlation.

3. THE DFI PREDICTOR

The DFI Predictor is based on the complex coherence function for microphones in the diffuse sound field. A weighting function $\chi(f)$ is applied which describes a 3 dB per octave attenuation of the coherence function (see figure 3). By summing up the weighted coherence function, the DFI Predictor represents the frequency-dependent correlation of microphones in a diffuse field with a single value.

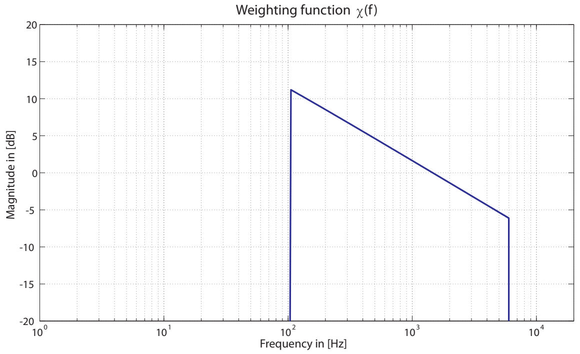


Fig. 3: Weighting of the coherence function.

It is defined as:

$$DFI = \frac{1}{n} \cdot \sum_{f=100Hz}^{f=6000Hz} [\gamma_{xy}(f) \cdot \chi(f)]^2 \qquad (5)$$

with n being the FFT length, $\gamma_{xy}(f)$ being the complex coherence function and $\chi(f)$ being the weighting function.

In [11], [12] a frequency range from 40 Hz to 1.5 kHz was used to calculate the DFI Predictor. However, the frequency range most likely has to match the stimulus used in the listening test.

In the listening test described below, the stimulus was female speech. The effect on the results of the listening test are shown in chapter 4. The DFI Predictor can be considered as a correlation coefficient. A small DFI Predictor defines a low correlation in the frequency range 100 Hz to 6 kHz. A high DFI Predictor value will indicate a high correlation in that frequency range.

The definition of the DFI Predictor is a first approach until further studies can be performed to check and refine this definition.

4. LISTENING TEST

The stimuli for the listening test were created in Matlab. The aim was to create different stimuli with a defined coherence. The basis is an arbitrary stereo microphone setup and its theoretical coherence function in the diffuse sound field [6].

Based on this coherence function two noise signals are created having approximately the same coherence as predicted by this calculation. These two noise signals were separately convolved with the diffuse part of a mono room impulse response. To isolate the diffuse part of the impulse response, the first 100 ms were cut off.

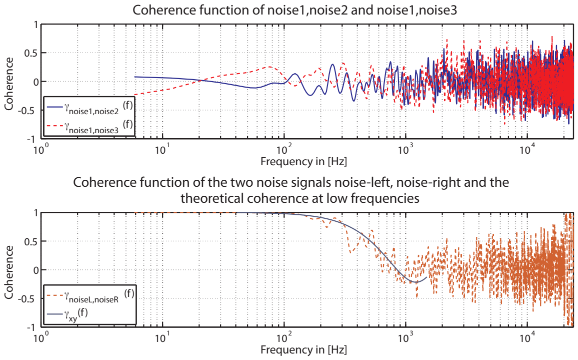
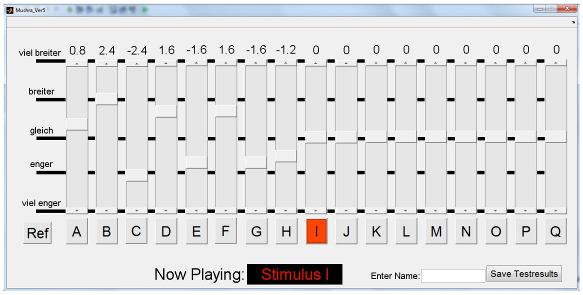


Fig. 4: Top diagram: Coherence of the initial noise signals. Bottom diagram: Result for an A/B setup with omnis at 0.2 m distance. DFI Predictor = 0.2957

After the convolution a stereo diffuse room impulse response is created which has a defined coherence. This impulse response is then convolved with a dry mono recording. The result is the stimulus. Figure 4 shows an example of a noise signal with a defined coherence. The subjects were listening to the stimuli via headphones. The test software used for the listening test is based on a MUSHRA test (see figure 5). After a short introduction the subjects were able to run the listening test with the test software by themselves.



Now Playing: Stimulus I

Enter Name:

Fig. 5: The test software was specially designed for the listening test

In this MUSHRA-like test design, the subjects com-

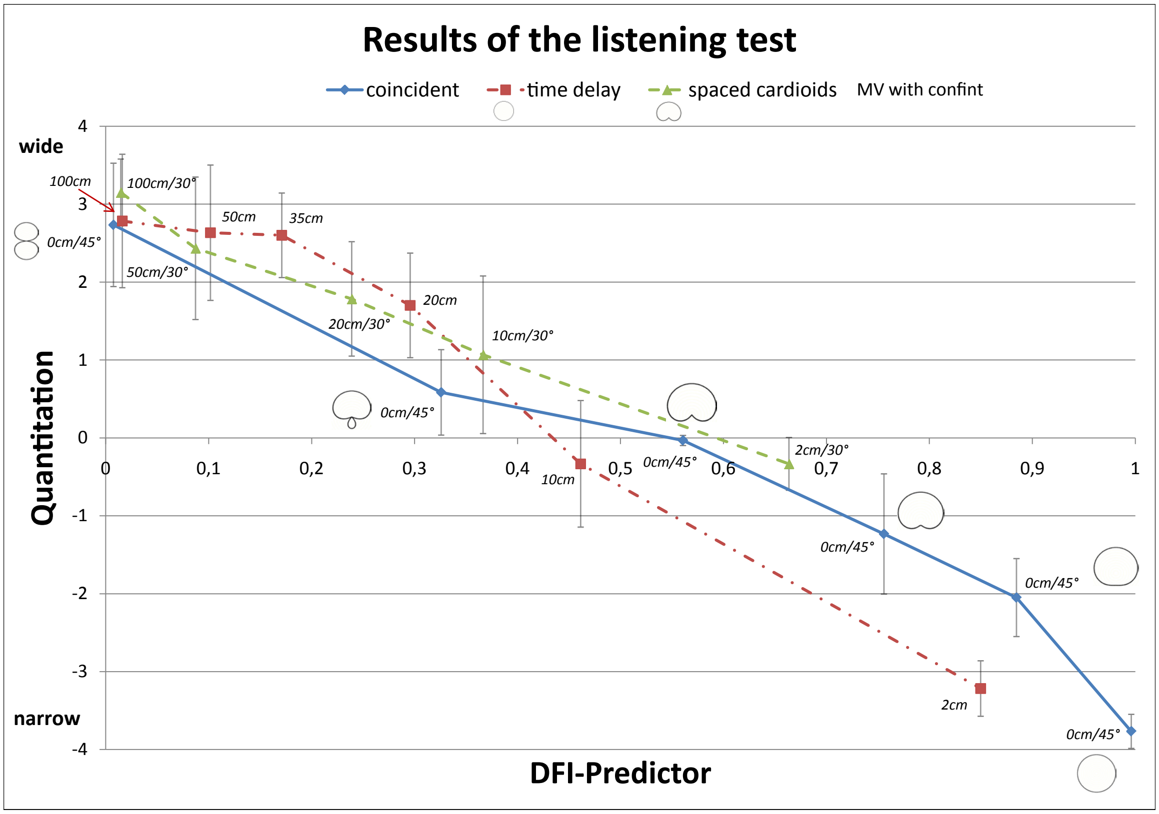


Fig. 6: Relationship between the DFI Predictor and the perceived spatial width.

pared the stimuli to the reference and all other stimuli. A stimulus that is perceived as wider gets a higher quantitative rating on a scale from -4 to 4. If no difference is perceived between a stimulus and the reference, the rating is zero.

In the perfect case there should be a gradual characteristic from narrow to wide after quantifying all stimuli. The reference in the listening test was an X/Y setup with cardioids and a 90-degree offset angle (DFI Predictor = 0.5607). The microphone setups simulated for the listening test are shown in figure 6. A shuffle function was integrated in the test software. As a result, every subject listened to a different order of stimuli. The dry recording used for all stimuli was female speech (SQAM-CD). Eight subjects participated in the listening test.

The results of the listening test are also shown in figure 6. All curves show a similar behavior. All stimuli with a smaller DFI Predictor value than the reference were perceived as narrower and vice versa. The hidden reference was well recognized by all subjects and as expected was positioned in the middle of the scale. These results already show that the DFI Predictor can be used to predict perceived width. At the moment the definition of the DFI Predictor is fairly rough. In further studies the stimuli for the listening test should be expanded with regard to frequency range to improve the significance of the DFI Predictor.

Figure 7 shows the results of the listening test plotted against the correlation coefficient of the two noise signals used for the decorrelation of the impulse response. Depending on whether the diffuse sound field was simulated with pink noise or white noise, the results vary. The graphs show that the correlation coefficient is less able to lead to conclusions about perceived spatial width.

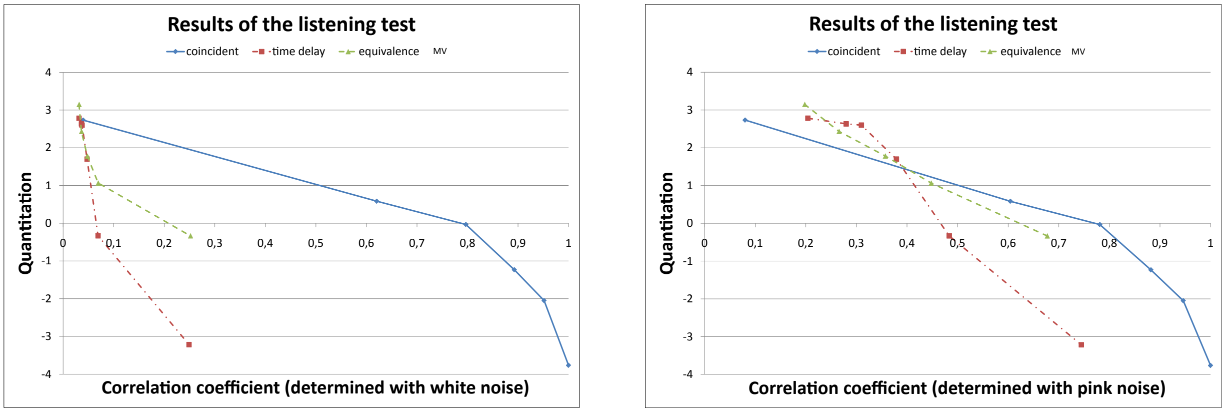


Fig. 7: The results of the correlation coefficient vary, depending on whether the diffuse sound field was simulated with pink or white noise.

5. CONCLUSIONS

By the method presented here, it is possible to predict the perceived spatial width of an arbitrary stereo microphone setup. The results of a listening test show that the DFI Predictor is already able to give a fairly good prediction of perceived width.

The frequency dependency of the correlation is important and should be considered for studies on spatial perception of spaced microphone setups.

The results show that the DFI Predictor offers a direct comparison of coincident and spaced microphone techniques with regard to their spatial imaging properties. This can be particularly important for the design of multichannel microphone setups.

6. REFERENCES

[1] Wittek, H.: *The "'Image Assistant 2.1"' is an interactive JAVA Applet that calculates the localization curves of arbitrary 2- and 3-channel stereo microphone configurations.* www.hauptmikrofon.de, 2008

[2] Williams, M.: *Microphone Arrays for Stereo and Multichannel Sound Recordings: a variable two channel microphone array; The training of the ear: experimental recording and listening tests; More on microphones: a short tutorial.* Editrice Il Rostro, 2004. – ISBN 8873650732

[3] Griesinger, D.: The psychoacoustics of listening area, depth, and envelopment in surround recordings, and their relationship to microphone technique. In: *AES 19th Intern. Conference*, 2001, S. 182–200

[4] Theile, G.: Multichannel natural music recording based on psychoacoustic principles. In: *AES 19th International Conference*, 2001

[5] Damaske, P.: Subjektive Untersuchung von Schallfeldern. In: *Acustica* 19 (1967), Nr. 199, S. 68

[6] Elko, G.W.: 4 Spatial Coherence Functions for Differential Microphones in Isotropic Noise Fields. In: *Microphone arrays: signal processing techniques and applications* (2001), S. 61

[7] Drews, M.: Mikrofonarrays und mehrkanalige Signalverarbeitung zur Verbesserung gestoerter Sprache. In: *Technische Universität Berlin, Germany, Diss* (1999)

[8] Blauert, J.: *Spatial hearing.* MIT press Cambridge, Mass., 1997

[9] Bendat, JS ; Piersol, AG: Engineering applications of correlation and spectral analysis. (1980)

[10] Griesinger, D.: General overview of spatial impression, envelopment, localization, and externalization. In: *Proceedings of the 15th International Audio Engineering Society Confer-*

ence, Copenhagen, Denmark Citeseer, 1998, S. 136–149

[11] Riekehof-Boehmer, H. ; Wittek, H. ; R., Mores: Voraussage der wahrgenommenen Breite einer beliebigen stereofonen Mirkfonanordnung. In: *26. Tonmeistertagung* (2010)

[12] Riekehof-Boehmer, H.: *Auswirkungen der Diffusfeldkorrelation auf die raeumliche Wahrnehmung*. BSc, HAW-Hamburg, 2010