SOUND COLOUR PROPERTIES OF WFS AND STEREO

Helmut Wittek

Schoeps Mikrofone GmbH / Institut für Rundfunktechnik GmbH / University of Surrey, Guildford, UK
Spitalstr.20, 76227 Karlsruhe-Durlach
email: wittek@hauptmikrofon.de
web: www.hauptmikrofon.de

ABSTRACT

The sound colour properties of wavefield synthesis are analysed by listening tests and compared to those of stereophony. A tool for this comparison is the 'OPSI' concept, proposed to avoid spatial aliasing. Colouration was measured for a number of different systems The effect of spatial aliasing on colouration could be shown. Both stereophonic phantom sources as well as OPSI sources were perceived to be less coloured than was predicted by colouration predictors based on the spectral alterations of the ear signals. This leads to the hypothesis that a decolouration process exists for stereophonic reproduction as proposed in the "association model" of Theile. The investigation is completed by experiments on localisation properties described in [15].

0. PERCEPTION MECHANISM FOR STEREO[1]

There are different possible explanations of stereophonic perception. Two loudspeakers can create a phantom source when they reproduce a coherent signal. Level and time differences between the loudspeaker signals determine the perceived direction of the phantom source. This phenomenon is usually called "summing localisation". The phantom source is understood as a substitute sound source, the sum of the two loudspeakers signals at each ear entrance forms the same binaural localisation cues as a correspondingly located real loudspeaker would form.

There are also objections against this explanation of stereophonic perception. Both from the side of researchers (e.g. Theile [1]) and sound engineers it is argued that the properties of the phantom source are much better than one could derive from an analysis of the ear signals. Moreover, Blumlein's basic theory of stereophonic imaging only applies to interchannel level differences. It is well known, however, that a phantom source shift can be achieved also by introducing time differences between the channels [3].

Theile suggested an explanation for these phenomena with his "association model" [1][2]: it is assumed that the function of the auditory spatial system is based on two different processing mechanisms. A current stimulus stemming from a sufficiently broadband sound source gives rise to a location association in the first and to a gestalt association in the second, higher-level processing stage because of auditory experience. The two stages jointly determine in every instance the properties of one or multiple simultaneous auditory events. Figure 0-1 shows the functional principle of the association model.

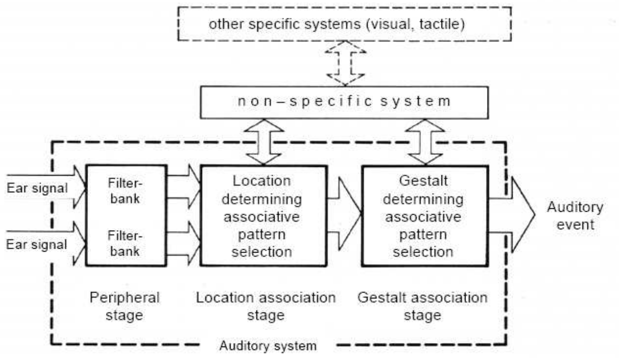


Figure 0-1: Functional principle of Theile's association model, after Theile [1]

The fundamental difference to summing localisation theories is the suggested ability of the auditory system to separately discriminate the two loudspeaker locations in a stereophonic setup. Due to the inverse filtering process of the location association stage postulated in [1] the relations between the left and right loudspeaker signals are recognised basically independent of the binaural crosstalk. Interchannel level and/or time differences determine the lateral displacement of the phantom source in the same way as during headphone listening. This understanding offers new approaches for the explanation of phantom source phenomena, such as colouration, perceived direction, distance, elevation, sound colour, stability.

Apart from Theile and his association model other investigations also address the above described phenomenon. Often the expression "binaural decolouration" is used in this context. It is defined as the "suppression or reduction of colouration through binaural mechanisms" (Brüggen [5][6]; Salomons [9]). Hence, binaural decolouration is

---

[1] The term "stereo" is used for two-channel stereophony. A stereo system produces phantom sources between two loudspeakers through interchannel level and/or time differences

understood as the sound colour improvement when listening with two ears as compared to with one ear. Brüggen presumes that one internal spectrum is responsible for the timbre perception and that this spectrum is built by the mean of the two ear signals. The idea of an internal spectrum or "central spectrum" is utilised by Bilsen [7], Zurek [4], Kates [8], Raatgever and Bilsen [14] and others.

The colouration caused by spatial aliasing in WFS is not produced by discrete signals from different directions rather than from many signals merging to a dense signal. Hence, decolouration consequently does not apply in the same manner to WFS as it could exist in stereo.

1. THE "OPSI" CONCEPT

Single loudspeakers emanate the high frequency part of the WFS source, see Figure 1-1. The WFS array uses the low-passed (<f_{alias}) WFS signals to reproduce the accurate virtual source position. The high frequency source components are generated by the single loudspeakers (solid) which are fed with high-passed (>f_{alias}) signals. These loudspeakers need to be spaced significantly wider than the loudspeakers of the WFS array and thus produce a stereophonic image. The low and the high frequency source are expected to merge unless the difference in their incident angles is not too large. This technique is called *"OPSI" = "Optimised Phantom Source Imaging of high frequency content wavefield synthesis"*. Figure 1-2 illustrates the method of deriving an OPSI signal for the WFS array.

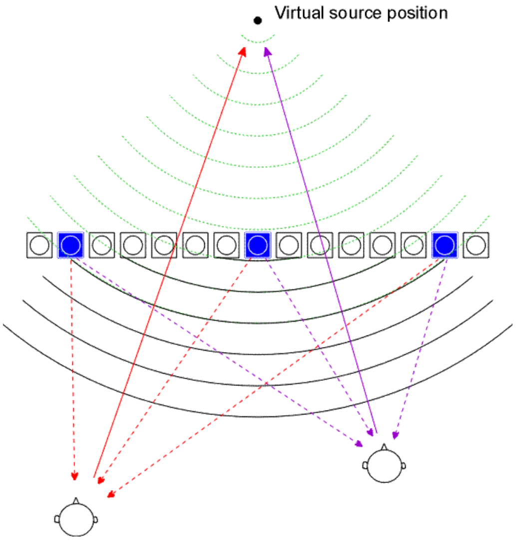


Figure 1-1: Example of an OPSI system: Three loudspeakers replace the WFS array for reproducing the high frequency part

The WFS virtual source and the phantom source are expected to merge and to be perceived as one auditory event. Their directions must not be too different, otherwise they would be separately perceived. This difference shall be called the *OPSI localisation error*. It should be smaller than a maximum allowed OPSI localisation error which was determined in a pilot experiment. A change in the perceived source direction does not exist in cases where the OPSI localisation error does not exceed 5°. The locatedness and the localisation focus of the OPSI source, however, was the subject of another experiment which is not described here [15]. The OPSI system turned out to be roughly similar to a conventional WFS system with the same WFS spacing regarding these attributes.

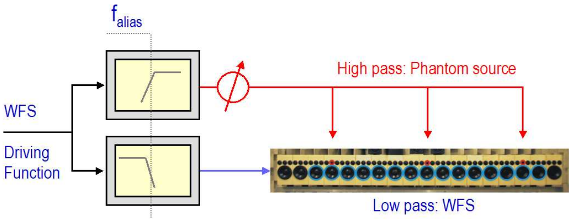


Figure 1-2: Creation of OPSI signals: The WFS array is fed with the low-passed WFS signals only. The stereo loudspeakers are fed with the high-passed WFS signals after a level adjustment. The signals are split at the crossover frequency f_{cross}.

Simulations of the OPSI localisation error show that the error depends on the position of the virtual source. Also the size and position of the stereo loudspeaker setup influence the performance. In order to optimise for a minimum OPSI localisation error, the choice of the suitable stereo loudspeakers depends on the synthesised virtual source distance. In the case of plane waves or sources with larger distances, more than just a few stereo loudspeakers have to be used.

2. EXPERIMENT: TIMBRAL FIDELITY

2.1 EXPERIMENTAL SETUP

The experiment aimed to compare the sound colour properties of different WFS systems as well as OPSI and stereo. It was performed with a modified MUSHRA method after ITU-R BS.1534 which utilised fixed anchors exhibiting spectral alterations of different degrees. In pilot experiments a suitable anchor was found, being the reference signal, processed with sine-ripple spectra of different ripple depths (see [10]). The ripple depth (amplitude of the sine) is defined as the difference between the maximum of the first half wave and zero. Five anchors were utilised; the ripple depth was 0, 1, 2, 3 and 4 dB. The utilised anchors are shown in Figure 2-1.

It was decided to assess the system colouration which was defined as an intra-system parameter. This means that in each trial only one system has to be assessed which makes the experiment design easier. In each trial 9 different signals were reproduced. These are:

- the reference (direction -5°)
- three stimuli in other directions than the reference (-10°, 3°, 15°). The directions were chosen such that the differences between reference and stimulus direction were unequal for all stimuli.

- the hidden reference
- four anchors (see Figure 2-1)

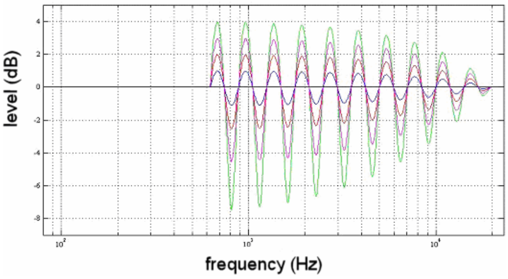
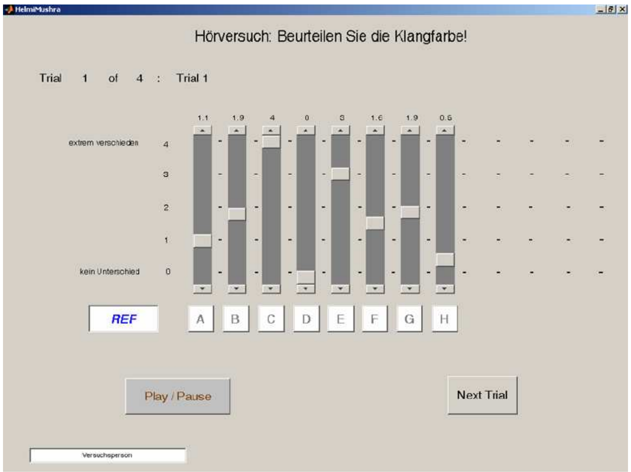


Figure 2-1: Anchors of the experiment: Sine-ripple spectra from 625 to 20.000 Hz. The ripple density is 2 (ripples per octave). The ripple depth is 0, 1, 2, 3 and 4 dB. The anchors were designed in order to sound not dissimilar to spatial aliasing. The graph shows the level in dB vs. the frequency.

Figure 2-2 shows the MUSHRA interface of the experiment. The subjects could switch between the 9 different stimuli by clicking on the buttons “REF” or “A”-“H”.



Hörversuch: Beurteilen Sie die Klangfarbe!

Trial 1 of 4 : Trial 1

Figure 2-2: Screenshot of the MUSHRA-like interface of the experiment. The software was programmed in MATLAB.

For practical reasons a virtual reproduction based on a binaural system utilising headphones and head-tracking was used. This system, called BRS (Binaural Room Synthesis, [11][12]) was developed at the IRT and recently was realised as a VST plug-in which can be run within the host software Steinberg Nuendo. The BRS filters were changed after each trial in order to assess the next system.

BRS is a convolution-based system, i.e. it needs binaural measurements of BRIRs (Binaural Room Impulse Responses) to correctly reproduce the virtual sources. The filters used in the BRS system were produced by utilising a database of BRIRs measured in the listening room of the IRT. By using this database a natural BRIR of any system can be produced. The resulting BRIRs are produced by superimposing the BRIRs of the single loudspeakers of the system according to the driving function of the system.

The BRS system produces out-of-head-localisation, which is the prerequisite of any serious experiment about spatial and timbral attributes. Only successful out-of-head-localisation avoids a system-inherent colouration of the binaural reproduction [12].

The stimuli of the experiment were dry pink noise bursts of the length 800 ms with a fade-in and fade-out time of each 50 ms. The dry stimuli were convolved in real-time with the corresponding BRIRs of the current system and azimuth to result in the binaural stimuli assessed by the subject.

System	No. of speakers	f_{alias} [Hz]	f_{cross} [Hz]	$f_{alias} > f_{cross}$
Real sources	1	-	(24000)	
Stereo	2	-	(0)	
OPSI 3	200	9600	750	↑
OPSI 3	200	9600	1500	↑
OPSI 3	200	9600	3000	↑
OPSI 3	200	9600	6000	→
WFS 3	200	9600	(24000)	↓
OPSI 12	50	2400	750	↑
OPSI 12	50	2400	1500	→
OPSI 12	50	2400	3000	↓
OPSI 12	50	2400	6000	↓
WFS 12	50	2400	(24000)	↓
OPSI 24	24	1200	750	→
OPSI 24	24	1200	1500	↓
OPSI 24	24	1200	3000	↓
OPSI 24	24	1200	6000	↓
WFS 24	25	1200	(24000)	↓
WFS 48	12	600	(24000)	↓

Table 2-1: Systems under test. The WFS and OPSI systems are labelled by the loudspeaker spacing which is 3, 12, 24 or 48 cm. The last column denotes the systems by their aliasing as described above.

The subjects were sitting such that the visual scenery matched as closely as possible the acoustic scenery produced by the virtual acoustics system. This means that they were positioned in front of the WFS test array which was built of 32 small broadband speakers. This array was not active in the experiment, but it served as a visible anchor to support the acoustic illusion.

The subjects were asked to grade the perceived colouration between the stimuli in the interface on a 5-grade scale.

Only the extremes of the scale were labelled with verbal descriptions:

Is there a timbral difference between the reference and the stimulus?

0="no difference" - 4="extremely different"

The attribute *colouration/timbral difference* was defined as the sound colour difference between the reference and the chosen stimulus. A small training phase introducing possible colourations was performed before the experiment.

2.2 SYSTEMS UNDER TEST

Table 2-1 lists the system assessed in the experiment together with the relevant system parameters. By this table, the WFS and the OPSI systems can be classified into three categories, marked with arrows in the last column:

- no aliasing, but the crossover frequency is unnecessarily low (↑)
- no aliasing, the crossover frequency is optimal (→)
- the crossover frequency is too high, there is aliasing in the signal (↓)

2.3 RESULTS AND DISCUSSION

Figure 2-3 and Figure 2-4 show the results of the experiment. In Figure 2-3 the perceived colouration is shown for each system and OPSI crossover frequency f_{cross}. Figure 2-4 sorts the WFS results by the OPSI crossover frequency f_{cross}.

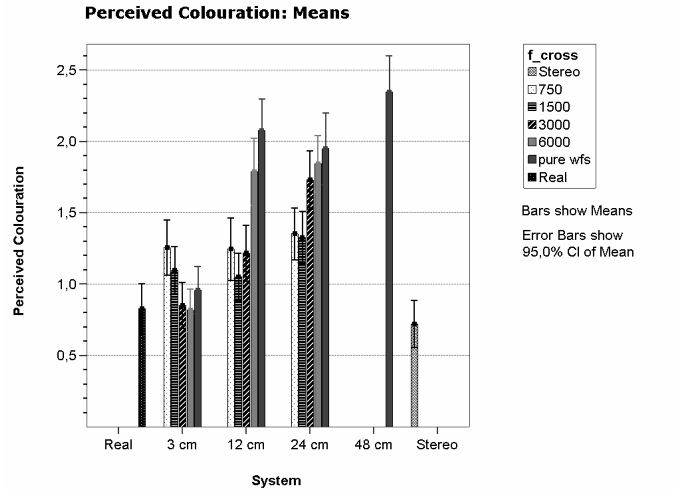


Figure 2-3: Results from the experiment: the perceived colouration is shown for all systems of the test. The category is the OPSI crossover frequency in Hz.

The results can be depicted by these main observations:

1. The real source and the phantom source as well as the optimal WFS3/OPSI3 systems have the best colouration grades.
2. The perceived colourations of the WFS and OPSI systems generally increase with increasing loudspeaker spacings.
3. The more the aliasing frequency exceeds the crossover frequency, the larger the colouration is. In other words: the colourations increase with the amount of aliasing in the signal.
4. When the crossover frequency is reduced below 3000 Hz, the colourations also increase.

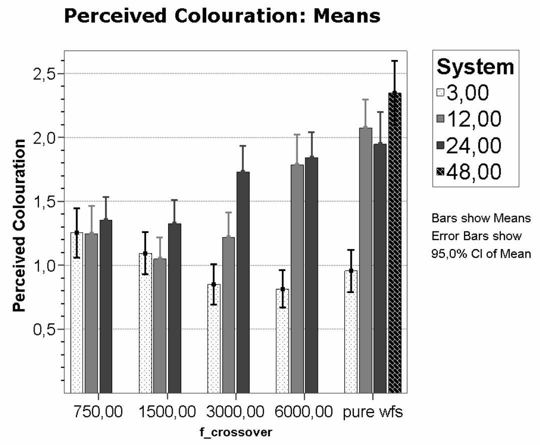


Figure 2-4: Results from the experiment: the perceived colouration is shown against the OPSI-Crossover frequency in Hz. The category is the WFS loudspeaker spacing in cm.

The discussion is based on the above-mentioned observations:

Observation 1: It was expected that the real sources as well as the WFS3 sources would have the best colouration grades. The OPSI3 systems partly have the same good results. The spatial aliasing in the WFS3 system may even lead to a small degradation in its sound colour performance, whereas the OPSI systems applying WFS3 may have better grades. It may be surprising that also the phantom sources achieve the same optimal grades.

Observation 2: As expected, the sound colour performance deteriorates with increasing WFS loudspeaker distance and thus with an increased aliasing in the signal. The experimental results show that only the WFS24 system achieves slightly better grades than the WFS12 systems which could indeed be explained by the objective predictors.

Observation 3 and 4: An OPSI system applying the optimal crossover frequency always achieves the best possible result for any WFS loudspeaker distance. When the crossover frequency is too high (aliasing in the signal) or too low (too little WFS information) the sound colour performance is degraded. Note that the OPSI systems OPSI3/750Hz, OPSI12/750Hz and OPSI24/750Hz produce a similar signal as there is a perfect WFS signal below the same crossover frequency and rather similar stereo loudspeaker setups for the stereo part. It plays no role whether this unaliased WFS signal is produced by a 24cm-spaced, a 12cm-spaced or a 3cm-spaced array. The same is true for the systems OPSI3/1500Hz and OPSI12/1500Hz.

A reduction of the crossover frequency below *3000 Hz* leads to a deteriorated sound colour performance. This can be read from the results of the OPSI3 system. The same rule is valid for the other systems, only that the existence of aliasing in the signal gives even worse results. This also means that the crossover frequency cannot be reduced further than the aliasing frequency to give the best results.

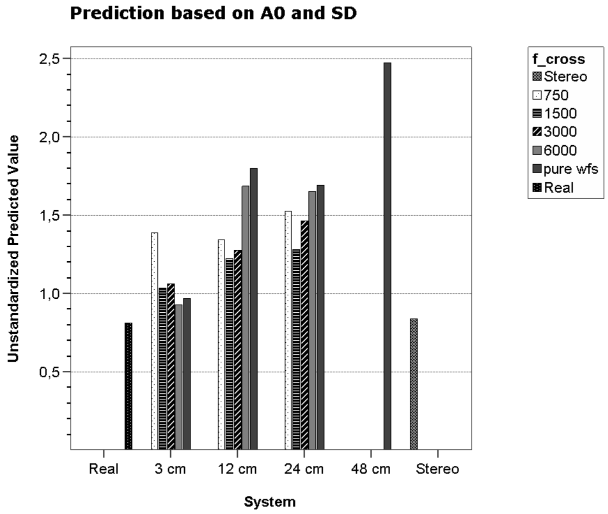


Figure 2-5: Results of the experiment as predicted by combined predictors SD and A_0-measure. Compare with Figure 2-3.

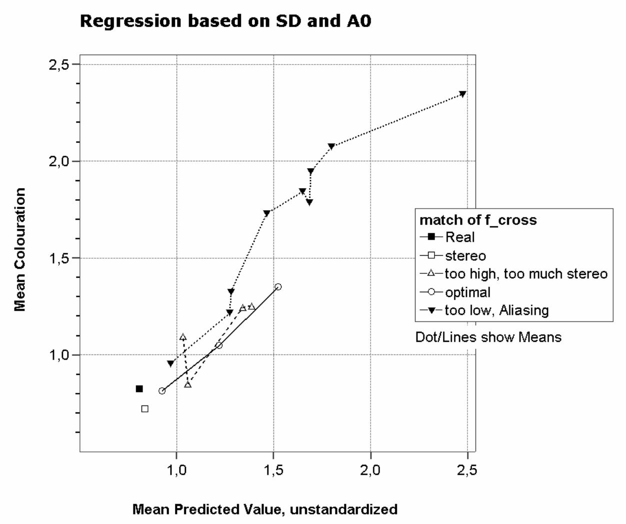


Figure 2-6: Regression analysis: The mean colouration grades of the experiment are drawn against the mean predicted values. Systems indicated with white circles and triangles (no aliasing) are predicted and graded better than systems indicated with solid triangles (with aliasing), see Table 2-1. Systems containing stereo are rather overestimated, aliased systems (solid triangles) are rather underestimated in their colouration by the prediction.

2.4 PREDICTION OF COLOURATION PERCEPTION

A prediction of the perceived colouration based solely on the spectral alterations may lead to different results for the different system types because it does not take into account the hypothesised listener's ability to segregate or decolour. Hence, the prediction can identify systematic differences in the perception mechanism.

This investigation adopts the *"central spectrum"* approach to predicting the perceived colouration [4][6][7][8][9]. The central spectrum is generated by averaging the power density spectra of the left and right ear signals. Furthermore, a critical-band filtering (Patterson) is performed before the averaging to simulate the frequency analysis properties of the auditory system. The intra-system spectral differences in the binaural transfer functions between the reproduced sources, processed after the central spectrum theory, will be referred to as "*spectral alterations*".

The prediction is attempted by performing a regression analysis. It is based on the measured spectral alterations and the perceived colouration gathered in the experiment. These predictors were chosen after comparing different alternatives described in the literature:

A_0-criterion: measure defined by Atal et al. [13]. Renewed definition by Salomons [9]: "coloration is perceptible if the maximum modulation depth (i.e. the level difference between maxima and minima) of the spectrum convolved with auditory filters exceeds a certain threshold A_0".

Spectral deviation (SD): Standard deviation of the spectral alterations on a log scale. The standard deviation measures the mean deviation from the spectrum to the mean value, calculated from the graph in logarithmical representation in order to correspond to auditory perception.

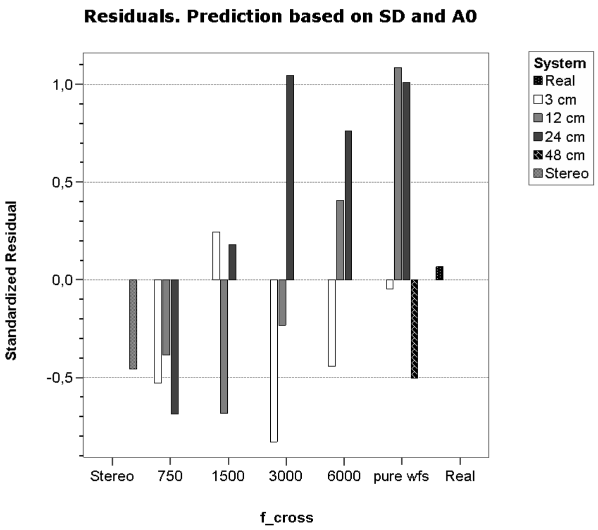


Figure 2-7: Standardized residuals of the regression based on SD and A_0-measure.

These predictors are rather crude measures that are defined without a very precise psycho-acoustic justification. It cannot be expected that they fully agree with actual perception. However, it can be seen from Figure 2-6 that the regression based on predictors A_0-measure and SD produce rather good results in terms of the qualitative distribution of the results. The squared correlation coefficient is $R^2 = 0.76$ for the multiple regression based on both predictors.

Figure 2-6 shows the prediction against the mean colouration grades of the experiment. The standardized residuals of the regression based on predictor SD are shown in Figure 2-7, where positive residuals correspond to an underestimation of the colouration and negative residuals to an overestimation of the colouration. The results show that the prediction quality apparently is dependent on the type of system. The systems containing aliasing (solid triangles in Figure 2-6) are mostly underestimated in their colouration, whereas the systems without aliasing but with a stereophonic contribution are mostly overestimated. This means that the perceived colouration of the stereophonic systems is lower than was predicted by the spectral alterations. This leads in to confirming the hypothesis that stereophonic perception is different from conventional auditory perception.

3. CONCLUSIONS

WFS properties compared to stereo

The experiment revealed the sound colour properties of WFS. Spatial aliasing introduces colouration to the virtual sources. The colouration can be rather well predicted by an analysis of the spectral alterations of the ear signals. Colouration generally is dependant on the aliasing frequency and the shape of the aliasing, i.e. the peak/notch distance and the spectral deviation. Stereo showed the least colouration of all systems.

OPSI technique

The sound colour properties of a WFS virtual source can be optimised by avoiding aliasing. Both theoretical and practical investigations have shown how an introduction of stereophonic techniques to WFS can help avoiding colouration artefacts. The idea of the OPSI concept is the coexistence of the low frequency WFS localisation properties and the high frequency sound colour properties known from stereo. Experiments that are not described in this paper [15] revealed that apart from a certain reduction of the listening area, the localisation properties of an OPSI system are similar to a corresponding WFS system.

Perception of stereo

Stereo showed the least colouration of all systems. Also the spectral alterations were smaller for the stereo sources. This means that the spectrum of the source is often more flat in the case of stereo compared with typical WFS.

However, both the stereo and the OPSI sources were graded better than it was predicted from the spectral alterations of the ear signals. Hence, the experimental results show that the perception of stereo is to be regarded differently from WFS perception. This suggests the existence of some kind of decolouration which leads to an improvement of sound colour perception in stereo reproduction similar to the decolouration of a sound source in a reflective environment. The segregation cannot be considered as leading to an ideal separation with regard to localisation and sound colour perception. This leads to the hypothesis of a partial decolouration in stereophonic perception. It can be considered an interpretation of Theile's association model.

REFERENCES

[1] Theile, G., 1980: *"On the localisation in the superimposed sound field"*, Dissertation Technische Universität Berlin, available: www.hauptmikrofon.de/theile.htm .

[2] Theile, G., 1991: "On the Naturalness of Two-Channel Stereo Sound", *Journal of the Audio Engineering Society*, Vol.39, No.10, pp.761-767

[3] Wittek, H., 2001: "JAVA-Applet *Image Assistant 2.0* and documentation" on www.hauptmikrofon.de/ima2.html

[4] Zurek, P.M., 1979: "Measurements of binaural echo suppression", *Journal of the Acoustical Society of America* Vol.66, pp.1750–1757.

[5] Brüggen, M., 2001: "Coloration and binaural decoloration in natural environments", *Acustica* 87, 400–406.

[6] Brüggen, M., 2001: *"Sound coloration due to reflections and its auditory and instrumental compensation"*, Dissertation Ruhr-Universität Bochum

[7] Bilsen, 1977: "Pitch of noise signals - Evidence for a 'central spectrum' ", *Journal of the Acoustical Society of America*, Vol.61, No.1, January 1977.

[8] Kates, J.M., 1985: "A central spectrum model for the perception of coloration in filtered Gaussian noise", *Journal of the Acoustical Society of America*, Vol.77, No.4, pp.1529-1534

[9] Salomons, A. M., 1995*: "Coloration and binaural decoloration of sound due to reflections"*, Dissertation, Delft University.

[10] Supin, A., et al., 1999: "Ripple depth and density resolution of rippled noise", *Journal of the Acoustical Society of America*, Vol.106 (5), November

[11] Horbach, U., Pellegrini, R., Felderhoff, U., Theile, G., 1998: "Ein virtueller Surround Sound Abhörraum im Ü-Wagen", *Proceedings 20th Tonmeistertagung*, Karlsruhe.

[12] Rathbone, B., Fruhmann, M., Spikofski, G., Theile, G., 2000: "Untersuchungen zur Optimierung des BRS-Verfahrens (Binaural Room Scanning)", *Proceedings 21st Tonmeistertagung*, Hannover, pp. 92-106.

[13] Atal, B. et al., 1962: "Perception of coloration in filtered gaussian noise-short-time spectral analysis by the ear", *Proceedings 4th ICA*, Copenhagen.

[14] Raatgever, J. and Bilsen, F.A., 1986: "A central spectrum theory of binaural processing. Evidence from dichotic pitch", *Journal of the Acoustical Society of America*, Vol. 80(2), p.429-441

[15] Wittek, H., Rumsey, F., Theile, G., 2007: "Perceptual enhancement of wavefield synthesis by stereophonic means". *Journal of the Audio Engineering Society*, Vol.55, No.9, September 2007, pp 723-751.