www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

PAN ARCHIVES OF ACOUSTICS
Vol. **37**, No. 1, pp. 5–18 (2012)


*DOI: 10.2478/v10168-012-0002-y*

# Providing Surround Sound with Loudspeakers: A Synopsis of Current Methods

Jens BLAUERT[(1)], Rudolf RABENSTEIN[(2)]

[(1)] *Institute of Communication Acoustics, Ruhr-Universität Bochum*
Bochum, Germany; e-mail: jens.blauert@rub.de

[(2)] *Chair of Multimedia Communication & Signal Processing, Universität Erlangen-Nürnberg*
Erlangen-Nürnberg, Germany; e-mail: rabenstein@nt.e-technik.uni-erlangen.de

(*received November 4, 2011; accepted February 20, 2012*)

Available methods for room-related sound presentation are introduced and evaluated. A focus is put on the synthesis side rather than on complete transmission systems. Different methods are compared using common, though quite general criteria. The methods selected for comparison are: *Intensity Stereophony* after *Blumlein*, vector-base amplitude panning (*VBAP*), *5.1-Surround* and its discrete-channel derivatives, synthesis with spherical harmonics (*Ambisonics*, *HOA*), synthesis based on the boundary method, namely, wave-field synthesis (*WFS*), and binaural-cue selection methods (e.g., *DirAC*). While *VBAP*, *5.1-Surround* and other discrete-channel-based methods show a number of practical advantages, they do, in the end, not aim at authentic sound-field reproduction. The so-called *holophonic* methods that do so, particularly, *HOA* and *WFS*, have specific advantages and disadvantages which will be discussed. Yet, both methods are under continuous development, and a decision in favor of one of them should be taken from a strictly application-oriented point of view by considering relevant application-specific advantages and disadvantages in detail.

**Keywords:** surround sound, holography, wavefield synthesis, ambisonics, amplitude panning, summing localization.

## 1. Introduction

It is one of the goals of audio technology to present sound fields to listeners in such a way that they experience an auditory perspective, that is, perceive auditory events in various directions and distances, which may then form complex auditory scenes. If some mobility of the listeners in the synthesized sound fields is desired, loudspeakers at fixed positions in space are usually employed. This kind of sound-field presentation is called *room-related*, in contrast to the *head-related* one as used in *Binaural Technology*.

## 2. Intensity Stereophony

In this long established two-channel method (more precisely *amplitude-difference stereophony*) the horizontal angles of sound incidence are coded into amplitude differences of two loudspeaker signals. The auditory system, then, forms the direction of the auditory event from attributes of the two superposed sound fields of the two loudspeakers – a process which is called *summing localization*. This popular and surprisingly robust method is primarily restricted to 2-D presentation.

Figure 1 depicts a common coding scheme in this context (Blumlein, 1931; Blauert, Braasch, 2008). Two spatially coincident figure-of-eight micro-

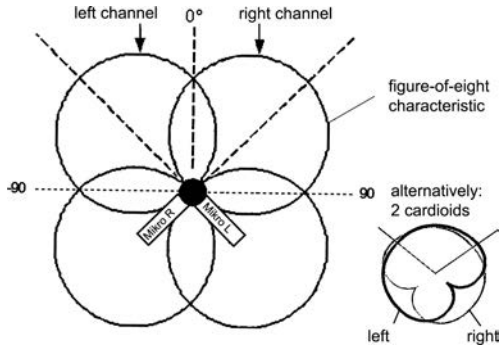


Fig. 1. Coding for *Intensity Stereophony* by use of two spatially-coincident directional microphones (Blumlein, 1931).

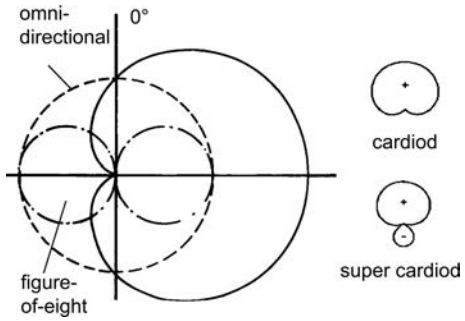


Fig. 2. Typical directional characteristics of microphones as can be formed by weighted superposition of an omni-directional with a figure-of-eight characteristic.

phones, arranged under a mutual angle of 90°, may be excited by one sound-source. They then produce two coherent microphone signals with a pure amplitude difference, the latter being unequivocally related to the angle of sound incidence. Instead of figure-of-eight microphones, cardioid or super-cardioid microphones are also in use. They are, for example, formed by means of a weighted superposition of an omni-directional characteristic upon the figure-of-eight one (Fig. 2). Figure-of-eight characteristics can be realized by pressure-gradient microphones, omni-directional characteristics by pressure microphones (see, e.g., Blauert, Xiang, 2009).

*Intensity Stereophony* has been a wide-spread spatial-reproduction method for more than 50 years now. The fact that it works so well is due to the following acoustic effect: In the frequency range of up to 1.5 kHz the two loudspeaker signals superpose in such a way that amplitude differences at the loudspeakers transform into arrival-time differences at the listener's ears (Fig. 3, Wendt, 1963). Interaural arrival-time differences, however, are the most robust attributes used

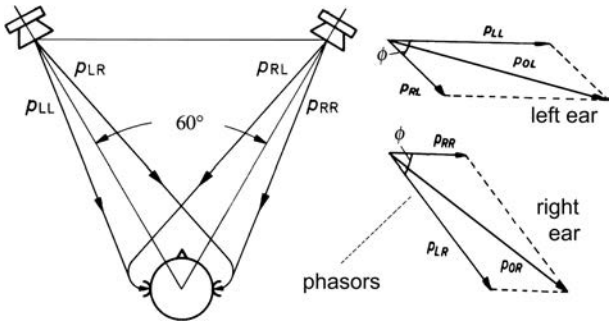


Fig. 3. Formation of interaural arrival-time differences of the signals at the two ears due to amplitude differences of the loudspeaker signals – effective at frequencies below about 1.5 kHz. A phase difference of the ear signals, $\Delta\phi$, is created due to the different delay times of each loudspeaker signal to the left or right ear, respectively. A larger amplitude of the left loudspeaker leads to a difference of the phase angles of the resulting phasors at the two ears, $p_{OL}$ and $p_{OR}$.

by the auditory system to form the directions of auditory events (Blauert, Braasch, 2008). Yet, this argument holds only up to about 1.5 kHz, as the human ear cannot detect the fine structure of ear-signal components of higher frequencies.

Besides the advantages of *Intensity Stereophony*, there are also significant disadvantages. One advantage is without doubt that loudspeaker signals stemming from the same source do not show any phase differences between them. Consequently, they can be electrically mixed without causing any coloration due to interferences - for instance, into a mono version. Further, proven microphone equipment is readily available and mixing rules (*panning rules*) are simple. In fact, they follow the relationships given by cardioid and/or figure-of-eight characteristics (Rumsey, 2001).

Disadvantages are as follows: Good reproduction of the directions requires a listener position as exactly as possible on the midline between the two loudspeakers (the so-called *sweet spot*), whereby the loudspeakers should be arranged under a horizontal angle of about 60° (Fig. 3). Such a standardized play-back arrangement is, to be sure, somewhat restrictive in terms of listener mobility. Yet, it facilitates the production of stereophonic program material.

*Intensity Stereophony*, unless special psychoacoustical "tricks" are employed, renders auditory events only in the horizontal sector between the loudspeakers. This is often described by saying "The orchestra comes into the living room". This saying also reflects the fact that this method can create spatial impression and ambience only in a very limited way. All auditory events appear predominantly at loudspeaker distance. Consequently, a convincing depth perspective is hard to achieve. Auditory events close to the listener are impossible. Auditory events at a larger distance than the loudspeaker can to a certain amount be simulated via the ratio of direct and reverberant sound, but the auditory perspective achieved in this way is not really perceptually convincing. Displacing the listener's head off the sweet spot gives rise to image shifts and coloration, due to interferences in the superposed sound field. Table 1 lists the advantages and disadvantages of *Intensity Stereophony*.

It should be noted at this point that there are many different methods for providing loudspeaker signals for stereophony, some of them employing inter-loudspeaker arrival-time differences solely or in addition to amplitude differences, for instance, by using spaced microphones at the recording end. For overviews and further discussion see, for example, Rumsey (2001), Theile (2001), Kamekawa *et al.* (2007), Blauert, Braasch (2008).

In any case, stereophony, that is, reproduction with only two frontal loudspeakers, does not allow for the provision of surround sound. Nevertheless, stereophony has been mentioned here, taking

Table 1. Pros (+) & Cons (−) of *Intensity Stereophony*.

| | |
|---|---|
| + | Simple panning rules (amplitude differences only) |
| + | Proven and readily available microphone equipment |
| + | Mono compatible |
| +/− | Standardized listening arrangement (restrictive) |
| − | Auditory events only in the frontal sector |
| − | Limited listening area (*sweet spot*) – image shifts outside |
| − | Auditory events appear predominantly at loudspeaker distance |
| − | Limited possibility to create room impression and ambience |
| − | Coloration possible due to interference by sound-field superposition |

*Intensity Stereophony* as an example, to introduce the psychoacoustic effect of summing localization, which is a basic effect for some surround-sound methods as well.

## 3. Amplitude-difference panning

*Intensity Stereophony* is based on the effect that the directions of auditory events are formed due to summing localization in a sound field which is superposed by more than one loudspeaker radiating coherent acoustic signals. More exactly, the loudspeaker signals are simultaneous in time but differ in terms of their amplitudes. In generalizing this basic idea, "domes" composed of triangles of loudspeakers have been built (Fig. 4, Pulkki, 2001).

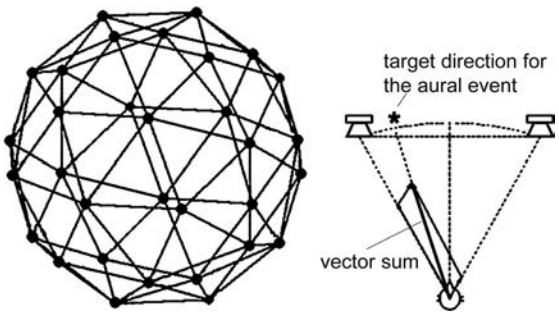


Fig. 4. Triangulation for amplitude-difference panning. In vector-base amplitude panning (*VBAP*), the vector sum is calculated with respect to a maximum of three loudspeakers that are adjacent in three dimensions. The simplified plot (right panel) only shows a two-dimensional case.

In order to achieve high accuracy when predicting the perceived directions of the auditory events, it is of advantage to employ as few loudspeakers as possible for each direction to be synthesized. For this reason, a maximum of three loudspeakers is used, namely, those three that are positioned closest to the target direction.

There is no immediate support for direct recording, therefore this method is primarily used for sound-field synthesis from parametric auditory-scene representations, for instance, *DirAC* (see Sec. 7). 3-D presentation is possible, yet not in a precise way for all directions. Adaptability to specific loudspeaker arrangements is fairly easy. Yet, there are massive problems regarding the quite narrow sweet spot and the inadequate production of perceived distances. Further, summing localization is rather unstable for sideward auditory directions, thus stable auditory events cannot be provided laterally (Plenge, Theile, 1977). In Table 2, advantages and disadvantages are given in more detail.

Table 2. Pros & Cons of Amplitude-difference Panning.

| | |
|---|---|
| + | 3-D presentation possible |
| + | Simple panning rules (amplitude differences only) |
| + | Easily adaptable to given loudspeaker arrangements |
| + | Low number of active loudspeakers per presented direction ($n < 3$) |
| − | No direct-recording technique available, thus for synthesis only |
| − | Limited listening area (*sweet spot*) – image shifts outside |
| − | Auditory events appear predominantly at loudspeaker distance |
| − | No precise localization in lateral directions |
| − | Serious problems with positioning auditory events at elevated directions (comb-filter effects) |
| − | Panning settings to be adapted to the specific loudspeaker arrangement used |
| − | Individual prediction of perceived direction hardly possible |
| − | Coloration possible due to interference by sound-field superposition |

In the nineteen sixties, the method of amplitude-difference panning was known as *synthetic sound field* and intensively used for scientific purposes, for example, at the Technical University of Dresden and the University of Göttingen (Meyer, Thiele, 1956), both Germany. The directions of sound incidence were visualized by so-called *hedgehog* plots (Fig. 5, left). Synthesis was performed with a dome of loudspeakers (Fig. 5, right). With a similar dome of loudspeakers, *Karl-Heinz Stockhausen* realized his famous performance of electronic music at the 1970 world fair in Osaka, Japan.

Lately, the panning rules as applied in 2-D as well as 3-D mixing of directions are preferably notated in vector form (Pulkki, 2001). Consequently, the method

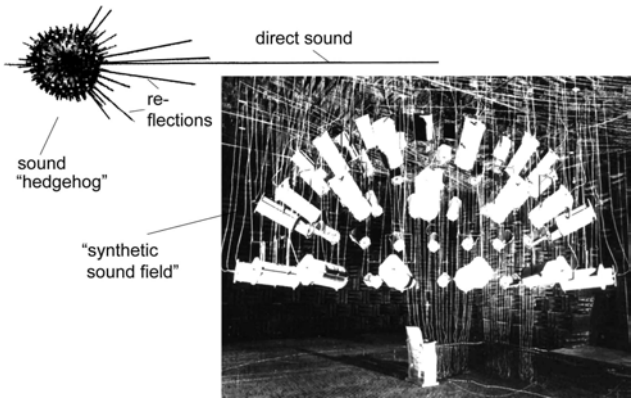


Fig. 5. A historic *synthetic-sound-field* arrangement, used in the early 1970s at the 3rd Physical Institute in Göttingen.

is currently often called *vector-base amplitude panning* (*VBAP*) – see Fig. 4.

The 2-D-panning rule as used by Pulkki (2001), is based on the so-called *tangent law* (compare, e.g., Rumsey, 2001). This law gives the panning functions, that is, normalized gain functions $g_l(\theta)$ and $g_r(\theta)$, for the left and the right loudspeaker positioned at azimuth angles of $\theta_1$ and $-\theta_1$ with respect to the listener and for an auditory event at the angle $\theta$ as follows,

$$g_l(\theta) = \frac{\sin(\theta - \theta_1)}{\sin(2\theta_1)}, \quad g_r(\theta) = \frac{\sin(\theta + \theta_1)}{\sin(2\theta_1)},$$

$$-\theta_1 < \theta < \theta_1.$$

Figure 6 (left) shows these gain functions for a pair of loudspeakers with $\theta_1 = 30°$ and , $-\theta_1 = -30°$ that is, for *Intensity Stereophony* and/or *VBAP* with a desired auditory event midway between $-30°$ and $30°$. The contribution of one loudspeaker positioned at 30 azimuth in a circular arrangement with seven loudspeakers is shown in Fig. 6 (right). The gain functions for the further loudspeakers are indicated by dashed lines.

## 4. Surround

A specific, very popular variant of the amplitude-difference-panning algorithm, is the standardized so-called *5.1-Surround* method. In this method, five loudspeakers are arranged according to Fig. 7 (Rumsey, 2001).

Formation of auditory-event directions is, again, based on summing localization. However, as has been mentioned before (Plenge, Theile, 1977; Pulkki 2001), the latter is rather unstable for lateral directions. In other words, precise synthesis of auditory events in lateral positions is hardly achievable. For this reason, in *5.1-Surround*, the two rear loudspeakers are positioned at 4 and 8 o'clock, respectively. This offers the possibility to create auditory events in a predictable way at least in these singular directions. Further, there is a frontal center channel (*dialogue channel*) which is used to stably position dialogues in front – even when the listeners move their heads out of the sweet spot. This is of particular importance with the method being used in connection with a TV-image (*home-theater* set-up). An optional sixth channel may, for example, be used for a sub-woofer or for other effects.

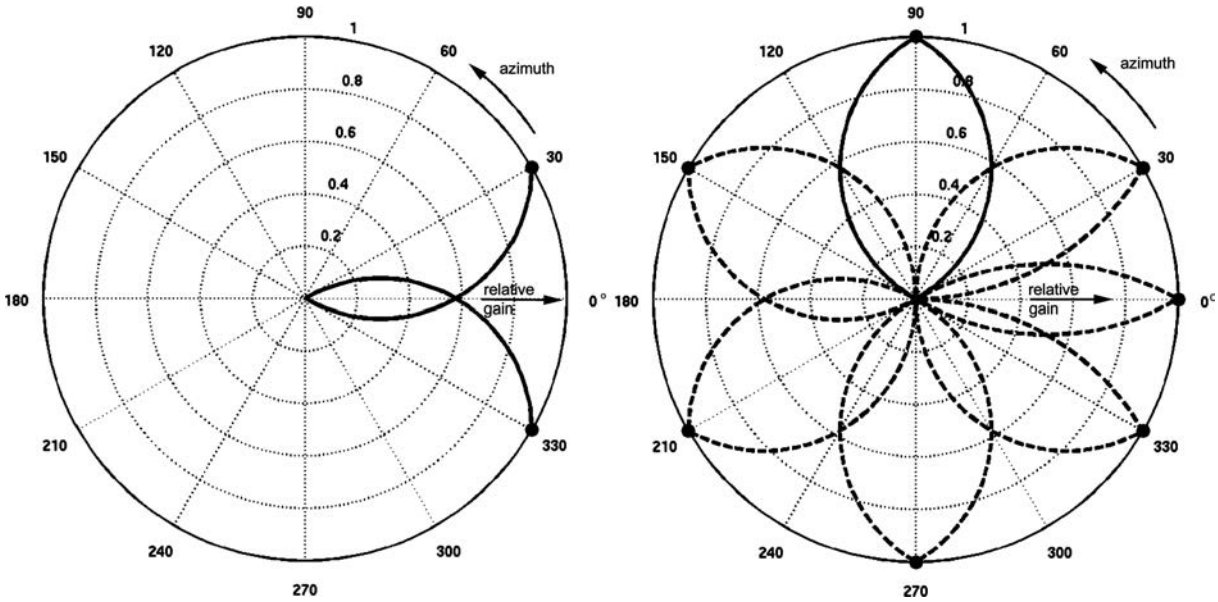


Fig. 6. Left panel: gain functions for panning an auditory event between two loudspeakers with the tangent law. Right panel: gain functions of one out of seven loudspeakers (solid line) and of the other six loudspeakers (dashed lines) for vector-base amplitude panning (*VBAP*).

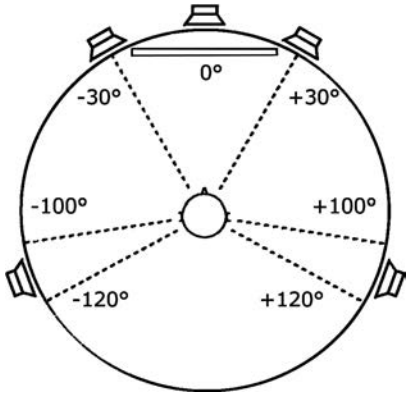


Fig. 7. Loudspeaker arrangement for *5.1-Surround*. There are one center, two front (left, right) and two rear (left, right) loudspeakers. A sixth channel is provided for, for instance, very-low-frequency or effect signals.

The most important advantage of *5.1-Surround* over conventional *Intensity Stereophony* is without doubt that this method can provide a sense of *immersion* – that is, *room impression* and a sense of *ambience*. Auditory events can be presented in all horizontal directions (surround!). Otherwise, all advantages and disadvantages as known from amplitude-difference panning remain valid. Program material that has been produced specifically for *5.1-Surround* often exhibits a typical "cinema sound" – which not everybody likes. The more important advantages and disadvantages of *5.1-Surround* are presented in Table 3.

Table 3. Pros & Cons of *5.1-Surround*.

| | |
|---|---|
| + | Less image shift for frontal direction due to center channel (*dialog channel*) |
| + | Simple panning rules (amplitude differences only) |
| + | Listening area broader than in intensity stereophony |
| + | Spatial impression and ambience can be experienced |
| + | Possibility to create special spatial effects |
| +/− | Standardized listening arrangement (restrictive) |
| − | Sweet spot limits possible listener positions |
| − | Auditory event predominantly in the horizontal plane |
| − | Auditory events appear predominantly at loudspeaker distance |
| − | No precise localization in lateral directions |
| − | Coloration possible due to interference by sound-field superposition |
| − | Often a characteristic "cinema sound" |

Following the idea of providing more loudspeakers than in stereophony to allow for a surround-sound impression, a number of further formats have been proposed, such as 7.1, 9.1, 10.1 and 11.1 (e.g., van Baehlen *et al.*, 2012) – up to 22.2 (Hamasaki *et al.*, 2005). Loudspeakers may not only be placed in the horizontal plane but also above and below it. However, since all of these methods are based on pure amplitude-difference panning, their pros and cons are implicitly included in the statements contained in Tables 2 and 3. Yet, the more channels are employed, the more auditory-event directions can be presented without having to make use of summing localization – and thus coloration can be avoided as may appear due to interference in superposed sound fields.

## 5. Spherical-harmonics synthesis

### 5.1. Classical Ambisonics

Looking back at Fig. 1 opens a possibility of constructing a set of four super-cardioid microphones by applying a suitable combination of omni-directional and figure-of-eight-characteristics – each super cardioid accounting for one of the main horizontal directions. By adding one more figure-of-eight microphone with perpendicular orientation, two further super-cardioids can be generated, one of them directed upward and the other one downward. The idea for this arrangement originates from Gerzon (1973) and was later dubbed *Ambisonics*. The set of four signals composed from the three figure-of-eight microphones plus an additional omni-directional signal is called *B-format*. The B-format is considered to be a "portable" signal format, since it can be adapted to a given loudspeaker arrangement by purely real factors, which cause appropriate shifts of the spatial characteristics. This kind of decoding can actually be performed by conventional mixing consoles.

Figure 8 depicts such a common decoding scheme, restricted here to the horizontal plane for simplicity. Each loudspeaker signal is weighted by a real factor that corresponds to the sensitivity of a super-cardioid receiver aiming at just that particular loudspeaker. It is apparent that all loudspeakers will be active

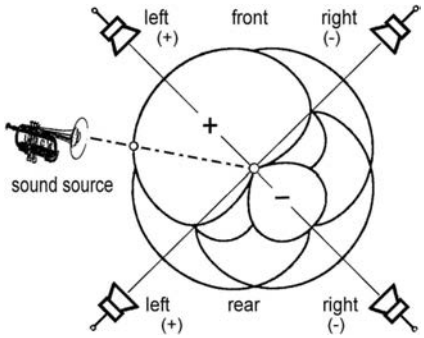


Fig. 8. A decoding scheme for *Ambisonics*. All four loudspeakers are active to reproduce a single sound source, where two of them radiate 180° out of phase in the example as plotted here.

in principle – though eventually with a 180° phase shift. By superposition of the sound fields of all active loudspeakers, a replica of the original sound field is achieved. Yet, this is only exactly true in the centre of the synthesis area. Unfortunately, this is precisely the position of the listener's head – which thus disturbs the sound field. Please recall at this point that *Ambisonics* is basically just another amplitude-difference panning method.

*Ambisonics*, in its classical form, never made it to wide application. Main reasons for this are listed in Table 4.

Table 4. Pros & Cons of Classical *Ambisonics*.

| | |
|---|---|
| + | 3-D presentation possible |
| + | Proven microphone equipment available |
| + | Simple panning rules (amplitude differences only) |
| + | Easily adaptable to given loudspeaker arrangements |
| – | Very narrow sweet spot |
| – | Auditory events appear predominantly at loudspeaker distance |
| – | No precise localization in lateral directions |
| – | Panning settings to be adapted to specific loudspeaker arrangement used |
| – | Listener's head disturbs sound field (directional errors and coloration) |
| – | High localization blur in general, particularly in lateral and elevated directions |

A closer analysis of *Ambisonics* reveals some interesting insights, for instance, that the method makes use of microphones with omni-directional and figure-of-eight sensitivity characteristics which, as is well known, can be described by spherical harmonics of the 0th and 1st orders. Acousticians are usually acquainted with spherical harmonics since these are also used to mathematically describe the radiation by spherical sound sources (see, e.g., Blauert, Xiang, 2009). Thus, a breathing sphere emits a spherical sound field of the 0th order, an oscillating, rigid sphere a 1st-order one. Higher orders are emitted when more complex radiation pattern on the surface of a sphere are given. Figure 9 shows examples up to the 2nd order. By linear superposition of spherical harmonics of the same order, these can be rotated – which is sometimes a useful feature. This is also how spatial characteristics are shifted into the actual loudspeaker directions for presentation – as is applied in Fig. 8.

The spherical harmonics are *eigenfunctions* of the sound-field equation, namely, they are its solutions in spherical coordinates. Actually, spherical harmonics represent an orthogonal system of functions in which all practically relevant sound fields can be developed – similar to the harmonics in the conventional

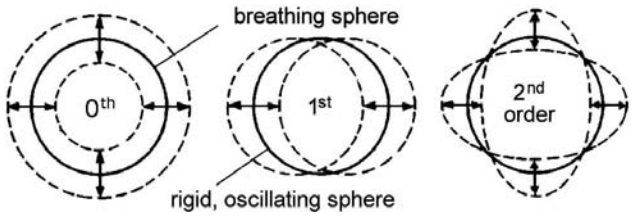


Fig. 9. Spatial-radiation patterns of a *breathing* sphere, an *oscillating* (rigid) sphere, and a sphere with a *clover-leaf* pattern of motion – examples of spherical sound-fields of the orders 0 to 2.

*Fourier* analysis of time functions (Rabenstein, Blauert, 2010; Rabenstein, Spors, 2008). Classical *Ambisonics* makes use of this possibility by involving spherical harmonics up to the 1st order. As applied in this paper, the term *Ambisonics* usually denotes this classical form. Figure 10 schematically depicts all spherical harmonics up to order 2.

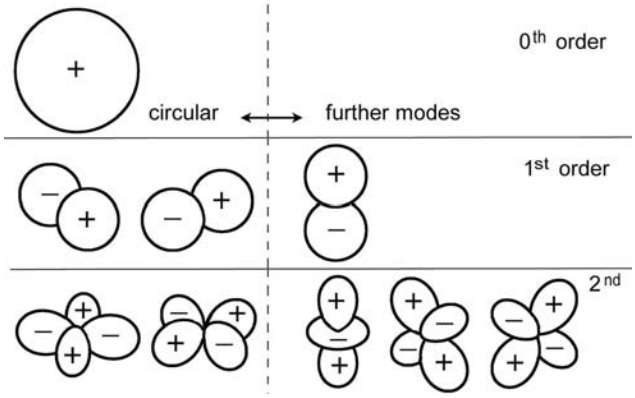


Fig. 10. All spherical harmonics of the orders 0 to 2. In many current installations the loudspeakers are arranged in a horizontal circle and, thus, only the circular modes are employed (left panel).

### 5.2. Higher-Order Ambisonics (HOA)

Higher-order ambisonics (*HOA*) represents a further development of the classical approach by involving spherical harmonics of higher order (Daniel *et al.*, 2003; Hollerweger, 2005; Nicol, 2010) as had already been suggested by (Gerzon, 1973). The spatial selectivity of the method increases with increasing order of the participating spherical harmonics. Accordingly, more loudspeakers are needed with increasing order. With $M$ being the highest order involved, the minimum number, $N$, of required loudspeakers corresponds to the sum of all linear independent spherical harmonics involved, namely,

- for spherical arrangements $N = (M+1)^2$,
- for circular arrangements $N = (2M+1)$.

This article is restricted to the discussion of planar arrangements, for example a set of loudspeakers sitting

on a planar circle (circular arrangement). Spherical, that is, 3-D loudspeaker arrangements for *HOA* have hardly been realized as of today, but are considered items of research.

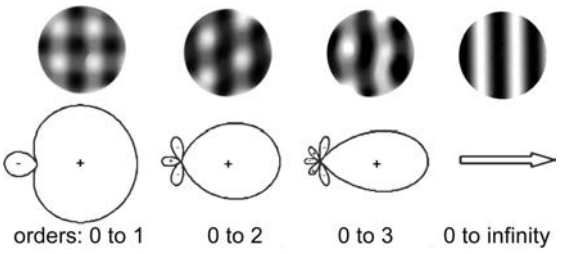


Fig. 11. Circular sound-field synthesis of a plane wave with increasing order of participating spherical harmonics. The higher the order involved, the larger the area in which the sound field is correctly reproduced. The lower panel shows the directivity patterns with which the sound field is spatially sampled.

All loudspeaker methods that have been dealt with so far in this article, namely, *Intensity Stereophony*, *5.1-Surround*, *Ambisonics* and, consequently, *HOA* are based on amplitude-difference panning and summing localization. It is thus to be expected that, also here, the formation of the direction of an auditory event will be the more precise, the smaller the number of active loudspeakers is at a time. The number of active loudspeakers for a specific direction decreases with increasing spatial selectivity of the method. This relationship is illustrated in Fig. 11 – compiled from simulation data by Faller (2004). It is evident that classical *Ambisonics* (orders 0-1) provides proper directional cues to the auditory system just in the very center of the synthesis area. Yet, with increasing participating order, the sweet spot becomes larger, until, for very high order, it finally extends over the complete synthesis area (Daniel *et al.*, 2003).

It is instructive to compare the gain functions of *HOA* to those of *VBAP* from Fig. 6. The same loudspeaker arrangement can be used also for *HOA*. The gain functions for second- and third-order *HOA* are shown in Fig. 12, left and right respectively. Similar to Fig. 6, the gain functions for one of the loudspeakers is indicated by solid lines and for the other ones by dashed lines. It is obvious that, for *HOA*, all loudspeakers contribute to the reproduction and not only two as for *VBAP*.

Microphone equipment for *HOA* is in the process of being developed in a number of laboratories (e.g., Moreau *et al.*, 2006; Meyer *et al.*, 2010). Commercially available models consist of a number of microphones in the surface of an as-small-as-possible rigid sphere. Other arrangements, such as directional microphones on an acoustically transparent spherical wire grid, are being tested. As of today, *HOA* recordings with up the 4th-order can be readily achieved. Regarding sound-field synthesis, there is no limitation in terms of the order. *HOA* is therefore a preferred format for many applications, for instance, for performances of electronic music. The portability of *HOA*-coded signals is considered beneficial in this context.

The original idea of spherical-harmonics synthesis was based on the assumption that plane waves ("sound rays") from all possible directions aim concentrically at the centre of the synthesis area. However, a sound source within the synthesis area cannot be rendered in this way. To enable this too, curved wave fronts must be synthesized, the amplitude of which naturally decreases with the assumed distance from the source. Ac-

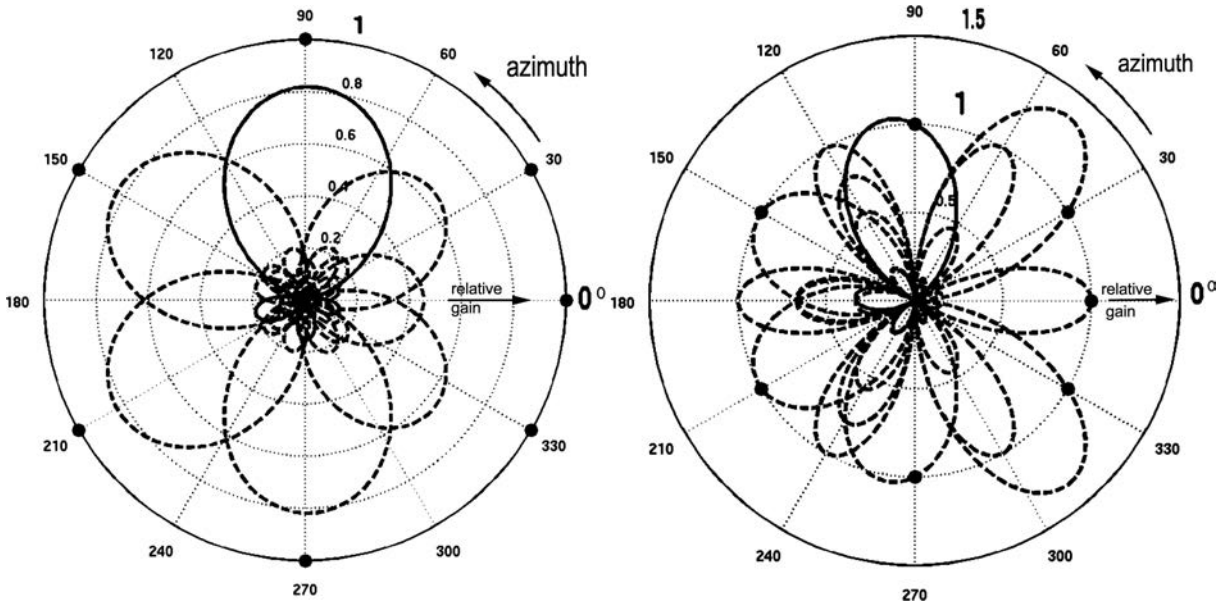


Fig. 12. Gain functions for *HOA*. Left: 2nd order, right: 3rd order.

tual loudspeakers emit such curved wave fronts anyhow – in fact, this leads to a well known boost of the low-frequency contents of the signal spectrum (see, e.g., Blauert, Xiang, 2009). By appropriate simulation of the low-end boost, the perceived distance of auditory events can be controlled. Suitable control enables virtual sound sources even within the synthesis area – so-called *focused* sources (Daniel, 2003; Nicol, 2010). However, for proper control of the perceived distances, the loudspeaker distances at the synthesis end must be known, such restricting the portability of *HOA*-encoded signals. Table 5 sums up most relevant advantages and disadvantages of *HOA*.

Table 5. Pros & Cons of Higher-order Ambisonics (*HOA*).

| | |
|---|---|
| + | 3-D presentation possible |
| + | Mathematically well defined by spherical-harmonics synthesis |
| + | Very broad sweet spot possible |
| + | Simple panning rules (amplitude differences only) |
| + | Easily adaptable to given loudspeaker arrangements |
| + | Localization blur decreases with increasing order of spherical harmonics |
| + | Graceful degradation at high frequencies (sweet spot becomes narrower but stays centered) |
| – | Higher-order microphones still under development |
| – | Auditory events appear predominantly at loudspeaker distance, unless compensated for |
| – | Panning settings to be adapted to specific loudspeaker arrangement used |
| – | Coloration possible due to interference by sound-field superposition |
| – | Heads shadow may cause localization errors and coloration – less with increasing participating order |

## 6. Wave-Field Synthesis (*WFS*)

Wave-field synthesis (*WFS*), is a method that – very much like higher-order ambisonics (*HOA*) – aims at synthesizing a sound field in a defined area such that it is actually a replica of a physically realistic sound field – be it real or conceptual. The theoretical approach to this problem, however, is significantly different. Namely, in *WFS*, arrival-time differences (that is, unwrapped phase differences) are applied in addition to pure amplitude differences of the loudspeaker signals. The theory of this method has been known for quite some time but can only be practically applied since the advent of micro-electronic signal processors (Berkhout, 1988). An early figure by Steinberg, Snow (1934) already illustrates the basic idea (Fig. 13). People sometimes talk of a transparent *acoustic curtain* in this context (e.g., Theile, 2005).

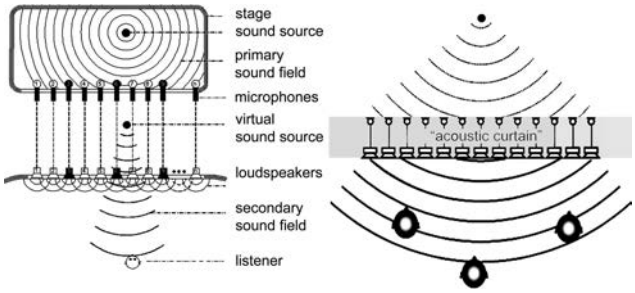


Fig. 13. A basic idea in wave-field synthesis: the *acoustic curtain* — that is, synthesis with a line-array of loudspeakers (left panel after Steinberg, Snow, 1934; right panel after Theile, 2001).

The mathematical calculation of such virtual sound fields is performed with superposition methods as are also applied for the calculation of the directions of line arrays of monopoles, for instance, *Rayleigh*'s integral equation, eventually complemented by *Fraunhofer*'s approximation (e.g., Blauert, Xiang, 2009). Figure 14 visualizes that various forms of superposed sound fields can be synthesized in this way. However, if the loudspeaker arrangement is not a line array but, for example, a rectangular (Fig. 15) or circular disposition, *Rayleigh*'s integral equation, which is based on elementary monopoles only, does no longer suffice. Instead, the *Kirchhoff-Helmholtz* integral equation is employed. This equation states that the sound field within a closed boundary is completely determined by both the sound pressure, $P(\mathbf{x},\omega)$, a scalar, plus the sound velocity or pressure gradient (both written in complex notation), respectively, a vector, everywhere on the boundary, namely,

$$P(\mathbf{x},\omega) = \int_{\partial V} \left( \frac{\partial}{\partial \mathbf{n}} G\left(\mathbf{x}|\boldsymbol{\xi},\omega\right) P(\mathbf{x},\omega) - G\left(\mathbf{x}|\boldsymbol{\xi},\omega\right) \frac{\partial}{\partial \mathbf{n}} P(\mathbf{x},\omega) \right) \mathrm{d}\,\boldsymbol{\xi}.$$

Here $\mathbf{x}$ denotes the position vector to any point within the closed boundary $\partial V$, $\boldsymbol{\xi}$ is an arbitrary point on this boundary with normal vector $\mathbf{n}$, and $\omega$ is the angular-frequency variable. The *Green*'s function from a surface point, $\boldsymbol{\xi}$, to an interior point, $\mathbf{x}$, is denoted by $G\left(\mathbf{x}|\boldsymbol{\xi},\omega\right)$. For a derivation of this equation see, for instance, Rabenstein, Blauert, (2010).

*Green*'s function describes the sound propagation from a source to a receiver and thus depends on the acoustic environment. In enclosures with low reverberation it can be approximated by *Green*'s function for the free field, with $c$ being the speed of sound

$$G\left(\mathbf{x}|\boldsymbol{\xi},\omega\right) = \frac{\exp\left(-j\frac{\omega}{c}|\mathbf{x}-\boldsymbol{\xi}|\right)}{|\mathbf{x}-\boldsymbol{\xi}|}.$$

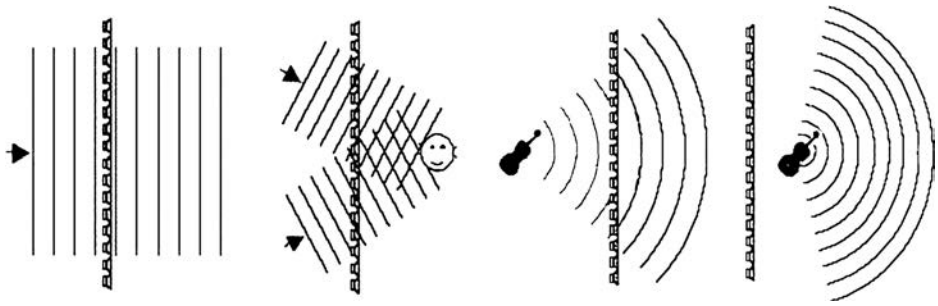

Fig. 14. Examples of *WFS* with line arrays (plots courtesy of *Sonic Emotion A.G.*). Sound sources can also be reproduced to be placed in front of the array – that is, as *focused* sources.

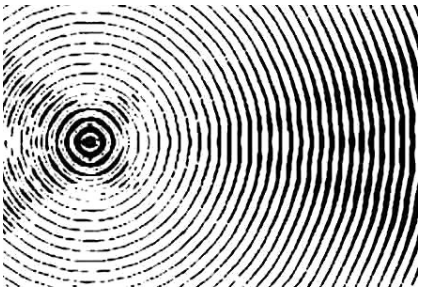

Fig. 15. *WFS* employing a rectangular loudspeaker disposition creating a *focused* source (simulation data courtesy of *ADA-AMC GmbH*).

Real synthesis equipment always embodies a limited number of loudspeaker channels only. Thus, the sound-field is sampled at a limited number of support positions. As known from quantizing time signals, at least two support positions are needed per period interval/length. If this condition is not fulfilled, distortions occur. In spatial sampling, the distortions are mirror images (*spatial aliasing*). In the mirror-image regions, a meaningful relationship of the directions of sound incidence and those of the auditory events is no longer given. Aliasing starts abruptly, right above a limiting frequency, $f_{\text{alias}}$, which is specific for the loudspeaker setting used. For line arrays of equidistant loudspeakers, it is, with $s$ being the inter-loudspeaker distance,

$$f_{\text{alias}} = \frac{c}{2\Delta s_{\text{loudspeaker}}}.$$

Actually most realized loudspeaker arrangement for *WFS* are planar, that is, linear, rectangular or circular – they may not even have a closed boundary. This leads to deviations from theory. Further, what is realized is hardly a (3-D) boundary – not to mention a closed one.

Restriction to planar disposition implies that cylindrical instead of spherical harmonics are engaged for the calculations (Rabenstein, Blauert, 2010). The fact that it is hard to build dipole sources in reality is less critical, because dipole and monopole signals are highly correlated in most cases, such that employing common (monopole) loudspeakers is sufficient for most practical implementations.

Microphone arrays for *WFS* recording are being studied but are not yet readily available. To compose auditory scenes for *WFS* reproduction, special algorithms are necessary and available in the form of special *WFS* mixing consoles. An example is a freely-available rendering software called *SoundScape Renderer*, which handles *WFS* among other spatial-reproduction methods (Geier, 2008; Geier *et al.*, 2012). Table 6 summarizes the most relevant advantages and disadvantages of *WFS*.

Table 6. Pros & Cons of Wave-field Synthesis (*WFS*).

| | |
|---|---|
| + | 3-D presentation possible |
| + | Mathematically well defined (*Kirchhoff-Helmholtz* integral equation) |
| + | Listening area not restricted within synthesis area (no sweet spot) |
| + | Panning possible, but more complicated (amplitude- plus arrival-time differences) |
| + | Localization blur decreases with increasing number of channels |
| – | Coloration above aliasing frequency |
| – | Proper directional information no longer available right above aliasing frequency |
| – | Suitable microphone equipment still rarely available |
| – | Panning settings to be adapted to specific loudspeaker arrangement used |

## 7. Binaural-cue selection

Application of all sound-field-synthesis methods dealt with above faces a severe problem, namely, data regarding actual auditory perception in synthetic fields are rare. In fact, even today, it is not understood in detail how summing localization really comes about. To avoid problems resulting from this lack of knowledge, it is often tried to synthesize sound fields in a physically as-authentic-as-possible way – the so-called *holophonic* approach. This requires a lot of effort. Alternatively, one can try to first identify those attributes

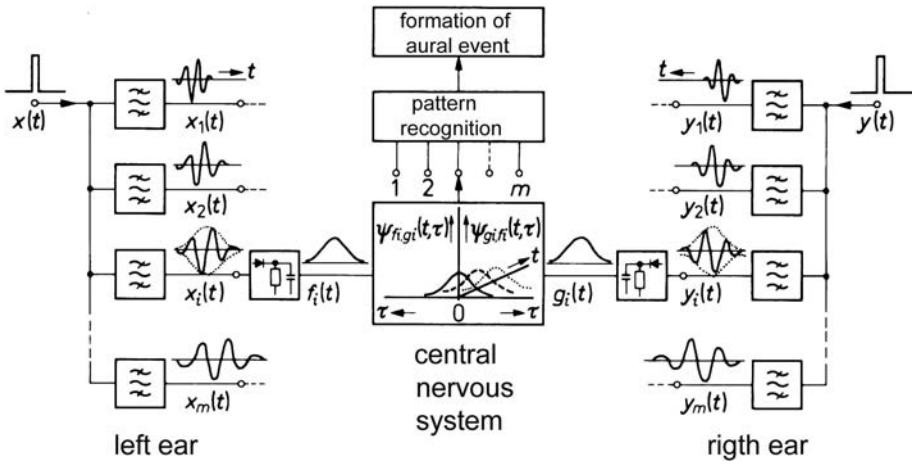


Fig. 16. Architecture of a model of binaural processing (schematic). The sound signals as received at the ears are decomposed in spectral bands. For spectral components above about 1.5 kHz, the envelopes of the band-pass signals are extracted. Then, a set of interaural cross-correlation-functions is calculated (*binaural-activity map*), which forms the basis of further evaluation.

of the sound field that are perceptually relevant. Once these "cues" have been identified, they are then treated with preference – irrelevant ones being neglected in the further course of synthesis.

For the identification of perceptually relevant cues, it makes sense to start from the binaural ear-input signals, that is, the sound signals at the entrances to the ear canals. A common assumption is that the binaural auditory system forms a kind of interaural cross-correlation on these signals – the process being carried out in parallel frequency bands. Figure 16 schematically depicts the architecture of a common model of binaural processing (Blauert, Braasch, 2008). A standardized output of such processing is called interaural coherence, $k$. In Fig. 17, an example of a time function in a specific auditory frequency band, $k(t, f_n)$, is given (Faller, 2004). By plotting the interaural cross-correlation function as a function of both the running time and the horizontal angle of sound incidence, one arrives at so-called *binaural-activity maps*.

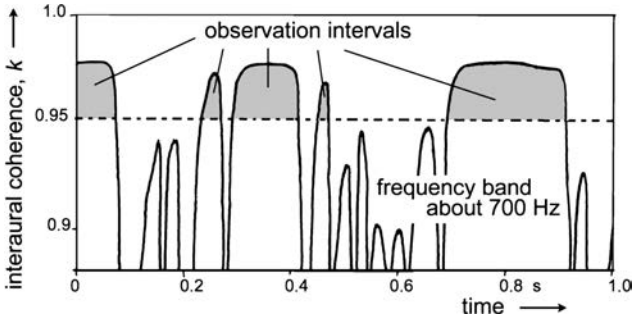


Fig. 17. Interaural coherence, $k$ – that is, a normalized interaural cross-correlation – as a function of time (after Faller, 2004). The amount of interaural coherence serves as a measure of confidence for detected interaural arrival-time differences, *ITD*s, and interaural level differences, *ILD*s, and, thus, for reliable spatial decomposition of auditory scenes.

Figure 18 displays such a map for a case where a distinct binaural impulse response as recorded in a concert hall is used as input to model of binaural processing (Lindemann, 1985; Gaik, 1990). Experienced room acousticians can judge upon the acoustics quality of halls by interpreting such maps.

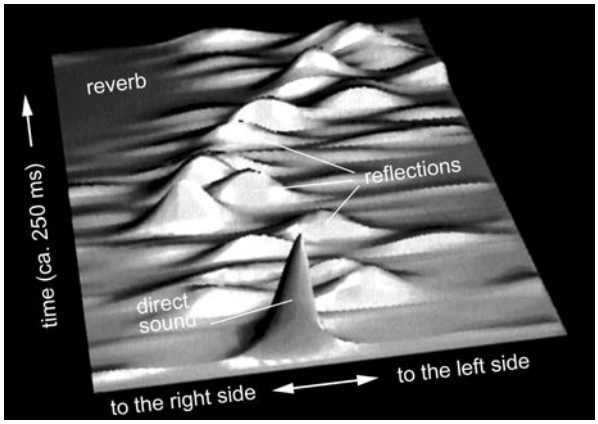


Fig. 18. Example of a *binaural-activity map*. The map shows the binaural impulse response of a concert hall as seen through a model of binaural processing.

It is assumed that the auditory system collects interaural attributes from the ear-input signals at just those moments where they are highly correlated, namely, when $k$ is relatively high (note the observation intervals in Fig. 17), because at those instances the interaural attributes can be related with confidence to particular directions of sound incidence. In Fig. 18, the instances of high coherence, and thus confidence, can be identified as peaks in the map. Assumingly, the auditory system determines the individual directions of sound incidence from the positions of these peaks.

Once the relevant perceptual attributes have been identified, a special focus can be put on them in the further process of auditory synthesis. This holds in particular, when the primary goal is not a physically authentic, but rather a perceptually plausible synthesis. Attributes of lower perceptual relevance may be added later with computationally less costly methods, such as artificial reverberation and back-ground noise (*ambience*).

Methods for sound-field synthesis with prior binaural-cue selection have gained in importance recently (Faller, 2004; Merimaa, 2006; Merimaa, Pulkki, 2005). A recent example in this context is the so-called *DiRAC* technique (Pulkki, 2006).

The scientific foundations in this area are in the process of being intensively investigated into (Blauert *et al.*, 2009). They are, in fact, important for a number of further applications as well, for example, spatial coding of binaural signals such as in *mp3*, measurement of the *quality of experience* in speech-dialogue and multimedia systems, planning processes in architectural acoustics, enhancement of speech intelligibility, and *ease of communication* in hearing aids and public-address systems.

## 8. Discussion and conclusions

There are different room-related methods available to generate spatial sound fields. In those cases where they are to be used for synthesis of auditory scenes only, problems regarding suitable recording techniques are completely avoided. When directional separation is the paramount issue, for instance, of speech and noise sources, the conventional *Intensity Stereophony* is fully sufficient – at least for directions in the frontal sector of the horizontal plane. If directions in further spatial directions are to be included, generalized amplitude-difference panning, such as *VBAP*, is adequate to create *synthetic sound fields*. However, if it is aimed at presenting a sound-field authentically in a spatially distinct synthesis area, methods like higher-order ambisonics (*HOA*) or wave-field synthesis (*WSF*) must be employed. Both these methods rest on exact solutions of the acoustical wave equation – this is why they are called *Holophony*. *HOA* and *WFS* can actually be transformed into one-another in a mathematically conclusive way (Nicol, 2010).

In practice, arrangements with a limited number of loudspeakers are used. This leads to approximation errors, which are specific for each of the two methods. Thus, which of the two methods is the best choice, depends on the specific use case that it is dedicated for. Figure 19 illustrates some characteristic differences of the two methods – plotted for a circular 32-loudspeaker disposition with data from Daniel *et al.* (2003). For further analysis of the approximation errors see Spors *et al.* (2008) and Ahrens *et al.* (2010).


Fig. 19. Examples for a comparison of *HOA* and *WFS*. The plots depict the reproduced sound fields (plotted from data by Daniel *et al.*, 2003).

For low frequencies with regard to $f_{\text{alias}}$, the sound-field is rendered in a largely correct way, whereby, in the case of *WFS*, there are no spatial distortions even in direct vicinity of the loudspeakers. Both methods are capable of generating focused sources, that is, sources within the synthesis area. Yet, *HOA* needs relatively high amplitudes for this purpose (not indicated in the plot). With increasing frequency the correct partition (sweet spot) in the synthesis area shrinks. In *HOA*, the sweet spot remains in the centre even above $f_{\text{alias}}$. In *WFS*, the sweet spot shifts rearward and dissolves abruptly with increasing frequency above $f_{\text{alias}}$, that is, when aliasing becomes effective.

There is the argument that the auditory system makes predominant use of directional cues in a frequency region below 1.5 kHz, which would imply that spatial distortions due to aliasing would not really matter perceptually. Yet, unfortunately, this assumption is not in accordance with common theories of auditory sound localization (see, e.g., Blauert, Braasch, 2008) and valid perceptual data regarding this not-yet-well-understood issue are hardly available at this time.

Regarding practical sound-field synthesis, the question is relevant of how to assign appropriate directions to each of the sound sources that finally comprise auditory scenes. Those methods that apply amplitude-difference panning – and all spherical-harmonics-synthesis variations belong to these – accomplish this by applying known amplitude-panning rules. In this context, it is a particularity of *Ambisonis* and *HOA* that synthesized sound sources can easily be rotated about the center. For *WFS* this issue turns out to be more complicated since both amplitude- and arrival-time differences must be implemented. As a tool for determining suitable panning settings, special mixing consoles with intuitive graphical user interfaces are provided (e.g., Geier *et al.*, 2008).

Spatial sound fields as produced by adequate loudspeaker arrangement can be combined with natural

sound fields, such creating a particular kind of *augmented* auditory reality. An application of this idea, among other ones, is its use for enhancing the quality of auditory experiences in performance spaces (e.g., Woszczyk, 2011).

Some shortcomings of the holophonic methods can be compensated by exploiting psychoacoustics effects, such as *spatial masking* and the *precedence effect* (e.g., Blauert, Braasch, 2008). Relevant research in this field is in progress (Blauert *et al.*, 2009). In principle, sound-field synthesis can be restricted to really audible binaural cues by making use of binaural-cue selection methods as the *DiRAC* method. Further suggestions include combinations with *Intensity Stereophony* and/or use of head-related transfer functions (*HRTF*s) (e.g., Lopez *et al.*, 2010) to provide further support.

Those who consider the expenditure of multi-loudspeaker systems to be too high, should consider *Binaural Technology* as an alternative. This technology is also capable of rendering auditory spatial scenes in a qualitatively excellent way. Instead of employing headphones to deliver the binaural signals directly to the listeners' ears, loudspeakers can also be used for this purpose. Figure 20 shows, as an example, a so called *transaural* system, where the cross talk between the adjacent loudspeaker and the averted ear is compensated by a dedicated electronic filter. In recent systems of this kind, the head position may even be tracked in order to continuously adjust the filter parameters. When the loudspeakers are positioned very close to the listeners' ears (neck-rest arrangement), cross-talk cancellation may even be skipped in less critical applications since interaural attenuation due to the head-shadow effect is substantial. Dispositions

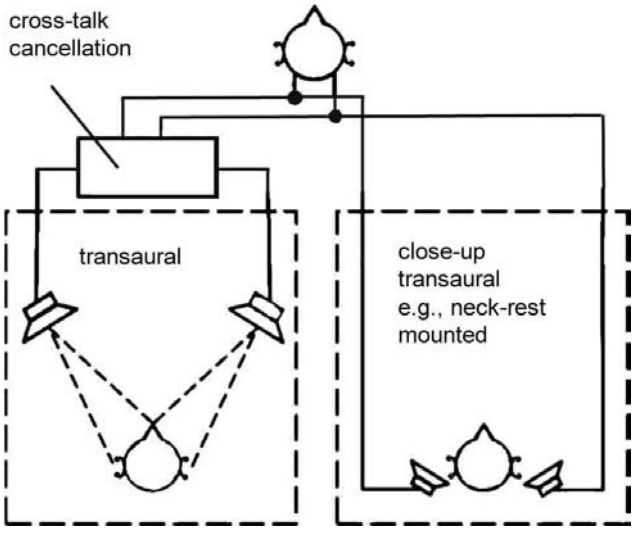


Fig. 20. Loudspeaker methods for binaural reproduction. The cross-talk cancellation may be omitted when the two loudspeakers are positioned close to the head, such as in a neck-rest kind of arrangement.

along this line of design are suitable for various applications, such as teleconferencing and video gaming (e.g., Kang, Kim, 1996; Menzel *et al.*, 2005).

## 9. Final remark

When deciding on purchasing hardware for *Holophony*, it is certainly a good advice to select the components in such a way that they can be utilized for different relevant reproduction methods – particularly for *HOA* and *WFS*. At this point in time, circular and spherical arrays offer the highest compatibility. Make sure to acquire loudspeakers and amplifiers of sufficiently high quality. The decision for a specific reproduction system will finally end up as being primarily an issue of software. In any case, the decision for a specific computational rendering algorithm should be taken from a strictly application-oriented point of view by considering the relevant application-specific advantages and disadvantages of the different methods in detail. The authors hope that this article may provide some initial guidance to this end.

## Acknowledgment

Earlier versions of this material have been presented at the *OSA'2010* Seminar in PL-Gliwice (in English) and at the 2010 *ITG-Fachtagung Sprachkommunikation* in D-Bochum (in German). The authors are indebted to various anonymous reviewers for constructive remarks.

## References

1. Ahrens J., Wierstorf H., Spors S. (2010), *Comparison of Higher-order Ambisonics and Wave-field Synthesis with respect to spatial-discretization artifacts in the time domain*, 40th AES Int. Conf., Audio-Engr. Soc, New York NY.

2. Berkhout A.J. (1988), *A holographic approach to acoustic control*, J. Audio-Engr. Soc. **36**, 977–995.

3. Blauert J., Braasch J. (2008), Räumliches Hören (spatial hearing), [in:] Handbuch der Audiotechnik, S. Weinzierl [Ed.], Springer, Berlin-Heidelberg-New York.

4. Blauert J., Xiang N. (2009), *Acoustics for Engineers*, 2nd ed., Springer, Berlin-Heidelberg-New York.

5. Blauert J., Braasch J., Buchholz J., Colburn H.S., Jekosch U., Kohlrausch A., Mourjopoulos J., Pulkki V., Raake A. (2009), *Auditory assessment by means of binaural algorithms – the* AABBA *project*, Int. Symp. Auditory Audiolog. Res., ISAAR'09, Danavox Jubilee Foundation, DK-Ballerup.

6. Blumlein A.D. (1931), *Improvements in and relating to sound transmission, sound recording and sound-reproducing systems*, British Patents #325 and #394.

7. Daniel J. (2003), *Spatial encoding including near-field effect: introducing distance-coding filters and a viable, new Ambisonics format*, 23rd AES Int. Conf., Audio-Engr. Soc, New York NY.

8. Daniel J., Nicol R., Moreau S. (2003), *Further investigation of high-order ambisonics and wavefield synthesis for holophonic sound imaging*, 114th AES Int. Conv., Audio-Engr. Soc, New York NY.

9. Faller F. (2004), *Parametric coding of spatial audio*, Doct. diss., EPFL, CH-Lausanne.

10. Gaik W. (1990), *Investigation regarding the binaural processing of head-related signals* [in German: *Untersuchungen zur binauralen Verarbeitung kopfbezogener Signale*], Doct. diss., Ruhr-Univ. Bochum, D-Bochum.

11. Geier M., Ahrens J., Spors S. (2008), *The SoundScape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods*, 124th AES Conv., Audio-Engr. Soc, New York NY.

12. Geier M., Ahrens J., Spors S. (2008), *The SoundScape Renderer*, www.tu-berlin.de/?ssr (last access Jan.2012).

13. Gerzon M. (1973), *Periphony: with-height sound reproduction*, J. Audio-Engr. Soc., **21**, 2–10.

14. Hamasaki K., Hiyama, R. Okumura (2005), *The 22.2 multi-channel sound system and its application*, 118th AES Conv., Audio-Engr. Soc., New York NY.

15. Hollerweger F. (2005), *An introduction to Higher-order Ambisonics*, www.create.ucsb.edu/wp/-FH_HOA.pdf (last access Febr. 2012).

16. Kamekawa T., Marui A., Irimajiri H. (2007), *Correspondence relationship between physical factors and psychological impressions of microphone arrays for orchestra recording*, 123rd AES Conv., Audio-Engr. Soc, New York NY.

17. Kang S.-K., Kim S.-H. (1996), *Realistic audio teleconferencing using binaural and auralization techniques*, ETRI J., **18**, 41–51.

18. Lindemann W. (1985), *Extension of the cross-correlation model of binaural signal processing by mechanisms of contra-lateral inhibition* [in German: *Die Erweiterung des Kreuzkorrelationsmodells der binauralen Signalverarbeitung durch kontralaterale Inhibitionsmechanismen*], Doct. diss., Ruhr-Univ. Bochum, D-Bochum.

19. Lopez J.J., Cobos M., Pueo B. (2010), *Elevation in wave-field synthesis using HRTF cues*, Acta Acustica united with Acustica, **96**, 340–350.

20. Menzel D., Wittek H., Theile G., Fastl H. (2005), *The Binaural Sky: A Virtual Headphone for binaural room synthesis*, Tonmeistersymposium. D-Hohenkammer.

21. Merimaa J. (2006), *Analysis, synthesis and perception of spatial sound – binaural localization modeling and multi-channel loudspeaker reproduction*, Doct. diss., Aalto Univ., FI–Helsinki.

22. Merimaa J., Pulkki V. (2005), Spatial impulse response rendering I: Analysis and synthesis. J. Audio-Engr. Soc. **53**, 1115–1127.

23. Meyer E., Thiele R. (1956), *Room-acoustical investigations in numerous concert halls and radio studios by means of novel measuring techniques* [in German: *Raumakustische Untersuchungen in zahlreichen Konzertsälen und Rundfunkstudios unter Anwendung neuerer Messverfahren*], Acustica, **6**, 425–444.

24. Meyer J., Elko G. (2010), *Analysis of the high-frequency extension for spherical eigenbeamforming microphone arrays*, J. Acoust. Soc. Am., **127**, 1979.

25. Moreau S., Daniel J., Bertet S. (2006), *3-D sound field recording with Higher-Order Ambisonics – objective measurements and validation of spherical microphones*, 120th AES Conv., Audio-Engr. Soc, New York NY.

26. Nicol R. (2010), *Représentation et perception des espaces auditifs virtuels (representation and perception of virtual auditory spaces)*, Habilitation thesis, Univ. du Maine, F-Le Mans.

27. Plenge G., Theile G. (1977), *Localization of lateral auditory events*, J. Audio-Engr. Soc., **25**, 196–200.

28. Pulkki V. (2006), *Directional audio coding in spatial sound reproduction and stereo upmixing*, 28th AES Int. Conf., Audio-Engr. Soc., New York NY.

29. Pulkki V. (2001), *Spatial sound generation and perception by amplitude-panning techniques*, Doct. diss., Aalto Univ., FI-Helsinki.

30. Rabenstein R., Blauert J. (2010), *Sound-field synthesis with loudspeakers, part II – signal processing*, [in German: *Schallfeldsynthese mit Lautsprechern II – Signalverarbeitung*], ITG-Fachtg. Sprachkommunikation, D-Bochum.

31. Rabenstein R., Spors S. (2008), *Sound-field reproduction*, [in:] Benesty J., Sondhi M.M., Huang Y. [Eds.], Springer Handbook of Speech Processing, 1095–1114, Springer, Berlin-Heidelberg-New York.

32. Rumsey F. (2001), *Spatial Audio*, Focal Press, GB-Oxford.

33. Spors S., Rabenstein R., Ahrens J. (2008), *The theory of Wave-field Synthesis revisited*, 124th AES Conv., Audio-Engr. Soc., New York NY.

34. Steinberg J.C., Snow W.B. (1934), *Auditory perspective – physical factors*, Electr. Engr., 12–17.

35. Theile G. (2001), *Multi-channel natural music recording based on psychoacoustic principles*, 19th AES Int. Conf., Audio-Engr. Soc., New York NY.

36. Theile G. (2005), *Spatial-audio presentation by means of wave-field synthesis* [in German: *Räumliche Tondarstellung mit Wellenfeldsynthese*], VDT-Magazin **2**.

37. Van Daele B., Van Baelen W. (2011), *Auro-3D: the advantage of channel-based sound in 3D*, Proc. Int. Conf. Spatial Audio, ICSA, D-Detmold.

38. Wendt K. (1963), *Directional hearing in two superposed sound fields as in intensity- and arrival-time stereophony* [in German: *Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie*], Doct. diss., RWTH Aachen, D-Aachen.

39. Woszczyk W. (2011), *Active acoustics in concert halls – a new approach*, Archives of Acoustics, **36**, 2, 379–393.