Audio Engineering Society

Convention Paper 5721

Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

Microphone array beam forming for multichannel recording

Juha Backman[1],[2]

[1]Nokia Oyj, 02150 Espoo, Finland

[2]Helsinki University of Technology, 02150 Espoo, Finland

ABSTRACT

A new multi-channel microphone technique, based on using a microphone array with some inherent directivity, combined with DSP beamforming, is introduced. The advantage of the new method over purely acoustical or simple analog polar pattern control is added freedom in defining the polar pattern shape, resulting in more precise control of panning laws and side-lobe behaviour. Alternative microphone arrangements, e.g. directional microphone arrays or sphere-mounted arrays, are discussed.

1. INTRODUCTION

Multi-channel sound reproduction undeniably has great potential for rendering realistic acoustic spaces. However, microphone techniques still pose a limit for capturing the original event. The channel separation achievable by simple acoustic means is often insufficient, and providing adequate front-back separation is especially problematic. Another problem in using simple microphone techniques is that typical (e.g. ITU-R) loudspeaker placements call for unequal angular separation between channels, but maintaining similar timbre between microphones with widely differing polar patterns is difficult using only acoustical or simple analog matrixing methods.

The technique presented in this work provides one possibility for constructing a near-coincident microphone system that can yield the desired properties with both polar pattern control and frequency response.

2. CURRENT MULTI-CHANNEL MICROPHONE TECHNIQUES

2.1. Short overview of microphone techniques

Here is only a very short summary of most common multi-channel microphone techniques, with emphasis on their correspondence to microphone design and room acoustics, is presented here. For a more complete overview of microphone techniques, especially from the application point of view, the author would like to refer the reader towards Rumsey[1].

Coincident microphone arrangements rely on creating the desired sound image using the acoustical directional properties of microphones to yield the amplitude differences for creating phantom images in reproduction.

Ambisonics yields a theoretical re-creation of the original field at the point of capture, but minor errors (e.g. the presence of the listener's head) remove the realistic chance of accurately re-creating the field, so the system, from the practical point of view, reduces to a form of amplitude panning.

A common practice in recordings made in concert halls is to use widely spaced microphones at least for rear channels, and also possibly for the front channels (e.g. derivatives of the Decca tree). This ensures that diffuse-field signals picked up by the rear-channel microphones are essentially uncorrelated, due to the statistical properties of the reverberant field, with each other and the front channels. This, in turn, alleviates potential problems in downmixing the signals to two-channel stereo or mono (reduced comb filter effects as compared to closely spaced microphone arrangements), but unfortunately any chance of stable, well-defined phantom images is almost unavoidably lost.

2.2. Limits of simple polar pattern control

All the simple acoustic or analog matrix techniques share about similar fundamental constraints in determining the polar patterns. This observation is justified by noting that any first-order gradient microphone can be formed from spaced omnidirectional (i.e. zeroth-order) microphones with frequency-independent delay and simple summation and first-order filtering operations, which are also the key elements of the analog beam pattern control methods. This provides insufficient means of precise polar pattern control, since in acoustical polar pattern control the parameters are not independently adjustable over very large ranges, and in any analog control scheme the system parameters at different frequencies cannot be adjusted independently. All this implies a reduced chance of achieving both the desired panning law and good channel separation simultaneously.

Higher-order gradient microphones can be used to produce narrower beams with good front-to-back separation and constant low-frequency directivity. In multi-channel recording these have been proposed for second-order Ambisonics. However, even in these applications beam shape control is limited, and the large amount of acoustical cancellation results in reduced signal-to-noise ration, especially at low frequencies, where signal level is greatly reduced.

3. PRINCIPLES OF THE NEW METHOD

The method proposed in this paper is based on using digital signal processing to sum the outputs of a closely spaced microphone group to create the desired channel separation characteristics. The microphone array consists of directional microphones arranged approximately corresponding to target loudspeaker directions, and more precise adjustment of the polar pattern is achieved by summing the weighted microphone outputs. As the weighting coefficients have to be both frequency-dependent and complex, this essentially translates into defining a *n*m* filter matrix for summation, where *n* is the number of microphone channels and *m* is the number of target reproduction channels.

Unfortunately, most of the traditional beamforming theory [2] concentrates on main lobe characteristics with linear or rectangular arrays, so the detailed results have only limited use, but the general principles of beamforming algorithms are readily transferable.

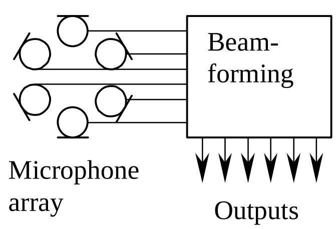

Figure 1 Principle of combining a directional micrphone array with postprocessing for beamforming.

If an identical number of microphones and target channels are used, it would be possible to solve the summation coefficients to yield polar patterns that would have a null towards all other channels but the target channel. However, this cannot be regarded as an optimal solution, since adhering to one strict constraint would imply that all control over panning law and side lobe behaviour for phantom images between the loudspeakers would be lost. A preferable approach is to use numerical techniques for obtaining a reasonable compromise between main lobe, side lobe, and channel separation performance. Optimisation is performed in frequency domain, a single frequency at a time, using an appropriate combination of criteria discussed below.

Using non-identical microphones (e.g. narrower polar pattern, such as hypercardioid or higher-order gradient microphones) or uneven angular separation between the microphones would be possible. The benefit of non-identical microphones is limited, since then the microphone frequency responses may have large differences (especially at low frequencies), resulting in summation coefficients with large magnitude and thus possibly some lost signal-to noise ratio. Also, care must be taken that individual microphone polar patterns should not be significantly narrower than the target polar pattern, since otherwise getting a smooth main lobe behaviour might be challenging. If identical microphones are used, then having equal angles between all the microphones maximises the minimum value of channel separation.

The low frequencies (below about 80 Hz) could in principle be also be allocated to the individual stereo channels; however, there are some issues related to both beamforming and practical domestic audio systems that justify allocating the low-frequency content to only the LFE channel using a cross-over network. If directional control is attempted at the microphone system, low-frequency response falls off in a manner similar to normal gradient microphones. As musically useful directional information is scarce at those frequencies, and difficult to reproduce in small rooms (albeit there is some evidence that these effects could contribute to spatial impression in concert halls), the benefit would hardly justify the penalty from the lost low-frequency signal-to-noise ratio. Also, many domestic multi-channel systems are reasonably well equipped to handle single extended low-frequency channel, but low-frequency response of the main channel loudspeakers is limited, and bass management (i.e. allocating the low-frequency contents of the main channels to the LFE channel) is often poorly defined.

As the optimisation procedure for summation coefficients concentrates only on ensuring proper polar pattern, post-equalisation for magnitude and phase is needed. This can be easily defined as an inverse filter of the on-axis (nominal direction) response of each output channel. The optimisation does not guarantee by any means that the original responses would be causal, but this is easily ensured by adding enough phase shift to guarantee at least minimum-phase behaviour. It is also important to maintain identical phase behaviour for the possible LFE channel and the stereo channels, as otherwise post-processing (e.g. mixing the LFE back to main channels, for instance for two-channel use) will suffer from out-of-phase cancellation.

It is however important to remember that a small number of microphone channels will have inevitable limitations in the beam shaping, arbitrary beamwidth or side-lobe attenuation is not feasible. Only improvement when complex digital filters are used instead of purely acoustical or simple analog methods is that some practical design constraints causing compromises in achieving good results over a wide frequency range are removed.

4. DEFININING OPTIMISATION CRITERIA

4.1. Panning law

First question regarding the choice of panning law is whether the transition between adjacent channels should occur so that the in-phase summation amplitude remanis constant (implying e.g. that the cross-over point should be 6 dB down as compared to single-channel level), or whetheer the summed output power should remain constant. There are several justifications for constant-power approach: the reverberant field, which should with typical speaker placement consist of uncorrelated signals, would then have constant amplitude regardless of phantom source position, which would keep the room colouration constant. At the lowest frequencies still fed to the stereo speakers, however, the wavelength cannot be any more regarded as small compared to the speaker separation. When the speaker separation goes below about half-wavelength, then the summation starts to aquire the features of coherent summation, and at lowest frequencies thus a constant-amplitude panning law should be advisable. The power radiated by a pair of omnidirectional loudspeakers behaves in a same way as the output power of a source and its perferct reflection, as described by Morse and Ingard [3]. Also, dividing the signal between more than two channels or using opposite-phase outputs [4], although having potential for creating a good, consistent imaging for the direct sound, will result in non-constant overall acoustic power output.

Proper front-back panning is, for all practical purposes, an impossible task. The phantom images are perceived as unstable and experiments [5], [6] indicate that the perceived direction remains with the louder source (front or back) until the levels are almost equal, and then the movement of the image is rapid and not well defined. The conclusion from the experimental data is that to avoid unnecessary coloration or image instability it is best to define the transition between front and back channels to be rapid, with minimum overlap.

4.2. Non-adjacent channel separation

The requirements for non-adjacent channel separation (or side lobe attenuation) stem from reasonably low colouration and image shift. If colouration from the leaked signals should be below 1 dB, then the sum of non-adjacent signals should in the worst case (exactly in-phase or out of phase) be 20 dB below the desired signal. Also, in actual listening conditions the signal level always varies due to room reflections and reverberation, and precise match between levels and frequency responses is certainly not guaranteed, so some additional headroom is recommended; about 5 dB additional channel separation should be required. However, achieving such levels of worst-case channel separation is difficult with a small number of microphones, and thus a reasonable compromise has to be accepted. Minimising the maximum sidelobe amplitude will tend towards having all the sidelobes at equal level.

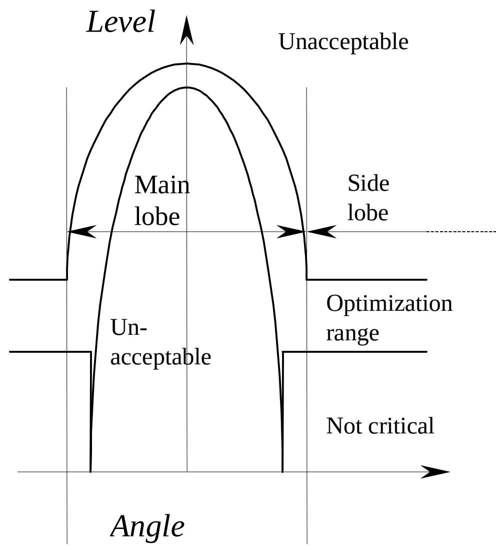

Figure 2 Tolerance window for defining beam shape.

The criteria can be defined as tolerance windows around nominal angle: unacceptable region, region around nominal for optimisation, and non-critical region (e.g. sidelobe behaviour below a defined level). These regions can be realised as weighting coefficients applied to the distance measure between actual and target polar patterns. The use of a non-critical region appears to go against a mathematically acceptable criterion of a measure that is appropriate for optimisation, i.e. the measure should be zero only when two data vectors are identical. However, in this case, when the optimisation criterion is a combination of several subcriteria, this kind of non-critical region allows the optimiser to concentrate on the still relevant criteria, if some aspects of the performance are unlikely to yield any further subjective improvement.

4.3. Additional criteria

Noise gain

To ensure good noise behaviour cancellation of out-of-phase cancellation of about equal magnitude should be kept to minimum. Noise gain G_n is defined as the ratio of actual sum of complex amplitudes a_i to the sum of the magnitudes of the amplitudes

$$G_n = \frac{\sum |a_i|}{\sum a_i} \tag{1}$$

Some noise gain is unavoidable, as some cancellation out-of-phase cancellation of signals is needed for beamforming, but minimising the additional noise helps to prevent sub-optimal solutions.

Elevation effects

Most direct sound (in typical acoustical performances) approximately in horizontal plane; first reflections in a concert hall may have some elevation, but podium/orchestra shell design aims often at constraining first reflection. If the microphone arrangement is two-dimensional, rigorous optimisation with respect to off-plane arrival will result in a compromise for in-plane performance (a problem with almost every multi-channel microphone arrangment, except Ambisonics, which works explicitly with three-dimensional space). Channel separation for sources far from horizontal plane will be worse than for in-plane sources. Important optimisation criterion (although rather time-consuming due to large number of computation angles needed) is that array gain outside horizontal plane may not exceed that in horizontal plane. A significant advantage of inherently directional microphones (gradient microphones, omni pairs) is that off-plane signal is already acoustically attenuated.

Coefficient constraints

If the coefficients are left completely unconstrained, then due to the non-uniqueness of the solutions the optimiser can converge towards a solution where the microphone pointing closest to the source contributes rather little to the output signal, yielding probably non-optimal noise performance. More reasonable behaviour can be ensured by a few constraints: the gain of the microphone closest to the source direction can be, without any loss of generality, be set to unity, and the magnitude of the maximum gain of the adjacent microphones can be set to a reasonable value, e.g. 3 dB, and the gain of the microphones pointing further away to a smaller value, e.g. 6 dB.

Frequency response smoothness

Completely independent optimisation can result in the system finding different local minima for the optimisation measure, so one possible additional criterion would be the deviation of the coefficients at one frequency from the coefficients at the frequency below (magnitude only, phase is bound to vary due to transducer separation). An alternative would be to perform simultaneous optimisation with respect to both angle and frequency, with normalised off-axis response smoothness as an additional criterion, but this leads easily into excessive memory requirements, since for n microphones there are n 1 complex coefficients to be optimised at each frequency; otherwise computational complexity is not much different from optimising a single frequency at a time.

5. ALTERNATIVES FOR THE MICROPHONE ARRANGEMENT

5.1. Closely spaced omnidirectional microphones

A group of closely spaced omnidirectional transducers is the usual starting point in conventional beamforming. For multi-channel audio there are problems, however. At low frequencies (when the array size is small as compared to wavelength) the array could perform well, but it is difficult to design a high-performance microphone array that would fulfil the requirement of size small as compared to wavelength at the highest audio frequencies. This causes spatial aliasing, making controlling the beam shape (especially in three dimensions) very difficult.

5.2. Cardioid microphones

A simple way of achieving the spatial filtering to ensure that beamforming is possible at all frequencies is to use a group of closely spaced cardioid microphones. Especially when constraints on the summation weights are employed to ensure sensible use of transducers, the polar pattern of individual microphone capsules prevents the formation of strong sidelobes.

An additional omnidirectional microphone could be used for low frequencies, primarily because in typical condenser microphones the low-frequency response of an omnidirecitonal microphone can be designed to be essentially flat.

5.3. Omnidirectional microphones on a sphere

Another way of constraining the polar pattern of individual microphones at high frequencies is to provide a baffle for the microphones by mounting them on surface of a sphere. There are some advantages to this: spatial aliasing control can be even more efficient than with cardioid microphones, extending low-frequency response is easier, and reflections between microphone capsules, causing both coloration and polar pattern irregularity, can be completely ignored. The size of the sphere can be chosen rather freely, there is no need to e.g. approximate the behaviour of the human head; for practical designs a sphere diameter of about 20 cm appears a good starting point.

Systems with similar physical appearance are manufactured for two-channel use by Schoeps, with two capsules placed on opposite sides; also Gerzon [7] suggested a three-microphone arrangement for three-channel recordings, with capsules placed at 0˚ and ±120˚. These systems rely only on the shadowing effect of the sphere, without any attempt on analog or digital beam control, limiting their efficiency to high frequencies only.

The sphere-mounted microphone system behaves at low frequencies essentially as a single-point microphone, and the movement of the apparent acoustic centre with angle is very small as compared to the wavelength. At higher frequencies, however, the acoustic centre will, due to the shadowing effect, will be closer to the sphere surface and more dependent on the incident angle; also the acoustic centres for different output channels will be separated. This can imply some coloration if signals need to be downmixed to e.g. two-channel stereo or matrixed analog surround, and if these applications are considered to be critical, then separate beam-forming filter matrices should be defined for them.

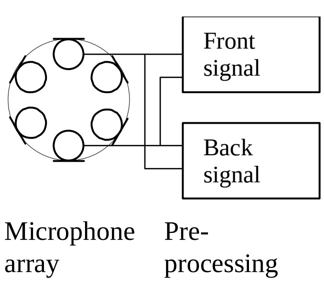

Figure 3 Pre-processing of individual microphone pairs.

To help the optimisation process, the microphone signals can combined pairwise from opposing microphones to yield an approximately cardioid response; this operation can be relatively coarse, just using unity gain for the front microphone and appropriate gain for the back microphone to yield a null towards back. At low frequencies this yields precisely the ideal cardioid polar pattern, but at higher frequencies the polar pattern will deviate from the ideal. Of course, in the actual signal processing there is no need to apply a two-layer method with the pre-processing and beamforming as separate steps, but rather the two processes should be convolved into a single filter.

Low-frequency signal for mono channel easy to separate as a simple sum of the six microphone outputs. An additional benefit is an improvement of low-frequency signal-to-noise ratio (with six microphones $10 \cdot \log_{10} 6 \approx 7.7$ dB).

Any practical design should be based on measured impulse response (or complex frequency response) data from the actual microphone assmbly, but design pre-studies can be performed by using a numerical model for the polar pattern of a piston cap on a sphere. The high-frequency polar pattern of an actual microphone deviates from this simplified model due to non-rigid membrane and possible front grid effects, but this approximation illustrates the shadowing effects of the sphere well. Morse and Ingard [8] give the polar pattern as a series expansion using Legendre functions P_m. (Their analysis is for radiation from a piston in a sphere, but the polar pattern behaves in a similar manner for a receiver.)

If the angle υ_0 taken up by the vibrating piston is small, then the velocity distribution can be written (using the notation of Morse and Ingard) simply as

$$U(\vartheta) = \begin{cases} u_0, & 0 \le \vartheta < \vartheta_0 \\ 0, & \vartheta_0 < \vartheta \le \pi \end{cases} \tag{2}$$

This yields the *m*th component of the expansion of the surface velocity as

$$U_m = \tfrac{1}{2} u_0 \left[P_{m-1}(\cos \vartheta_0) - P_{m+1}(\cos \vartheta_0) \right] \tag{3}$$

This can be substituted in the general expression for the angular dependence of the sound pressure,

$$\psi(\vartheta) = \frac{1}{ka}\sum_{m=0}^{\infty}\frac{U_m}{U_0 B_m}P_m(\cos\vartheta)e^{-i\delta_m-\frac{1}{2}i\pi(m+1)} \tag{4}$$

where U_0 is the average velocity and the coefficients B_m and δ_m can be solved from the pair of equations

$$\begin{aligned} &mn_{m-1}(ka)-(m+1)n_{m+1}(ka) \\ &=(2m+1)B_m\cos\delta_m \\ &(m+1)j_{m+1}(ka)-mj_{m-1}(ka) \\ &=(2m+1)B_m\sin\delta_m \end{aligned} \tag{5}$$

where n_m and j_m are the spherical Neumann and Bessel functions of mth order. To simplify the appearance of the solution we introduce definitions

$$\begin{aligned} \alpha &\stackrel{def}{=} \frac{mn_{m-1}(ka)-(m+1)n_{m+1}(ka)}{2m+1} \\ \beta &\stackrel{def}{=} \frac{(m+1)j_{m+1}(ka)-mj_{m-1}(ka)}{2m+1} \end{aligned} \tag{6}$$

The equations (5) have four possible solutions, but B_m to has to be constrained to positive values to yield a positive amplitude, and phase angle δ_m can be chosen to correspond to the positive solution; choice of either sign is not critical, as in computing the polar pattern the amplitudes are anyhow to be normalized to on-axis value $\psi(0)$. With these choices the solutions are

$$\begin{aligned} B_m &= \sqrt{\alpha^2+\beta^2} \\ \delta_m &= \arccos\frac{\alpha}{\sqrt{\alpha^2+\beta^2}} \end{aligned} \tag{7}$$

Morse and Ingard discuss asymptotic approximations of the term B_m, but it appears that the exact expressions cause no problems in computation in ka range of interest. B_m grows relatively rapidly, so about 10 terms are sufficient when evaluating the series expansions.

6. REFERENCES

[1] Rumsey, Francis, “Spatial Audio,” Focal Press, Oxford, 2001, pp. 187 - 209.

[2] Ziomek, Lawrence J., “Fundamentals of Acoustic Field Theory and Space-Time Signal Processing,” CRC Press, Boca Raton, 1995, pp. 497 - 579.

[3] Morse, Philip M., K. Uno Ingard, “Theoretical Acoustics,” McGraw-Hill, New York, 1968, pp. 372 - 374.

[4] Gerzon, Michael A., “Panpot Laws for Multispeaker Stereo,” presented at the AES 92nd Convention, Vienna March 24 - 27, 1992, preprint 3309 (4PS2.04).

[5] Rumsey, ref. [1], p. 210.

[6] Blauert, Jens, “Spatial Hearing,” revised edition, MIT Press, Cambridge, Massachussets, 1997, p. 326.

[7] Gerzon, Michael A., “Microphone techniques for 3-channel stereo,” presented at the AES 93rd Convention, San Francisco, October 1-4, 1992, preprint 3450 (S-3).

[8] Morse, Ingard, ref. [3], 332 - 346.