DIFFERENTIAL MICROPHONE ARRAYS FOR SPECTRAL SUBTRACTION

Marc Ihle

Muehlstrasse 11
D-89077 Ulm, Germany
Email: marc.ihle@ieee.org

ABSTRACT

In [1] and [2] the author presented for the first time a differential microphone array, which is used to calculate an accurate estimate of the power spectral density (PSD) of the ambient noise – excluding the desired signal – and the PSD estimate of the noisy input signal – including the desired signal and the ambient noise. With both PSD estimates high-quality noise suppression can be achieved by using the spectral subtraction technique.

This article describes a modified array geometry, which can also be used for this purpose. The differences in signal processing and performance are shown; the advantages and drawbacks of both systems are discussed.

1. INTRODUCTION

G. W. Elko et. al. [3], [4], [6] combined the signals of closely spaced microphones to achieve a steerable cardioid polar pattern. With this method the directivity index of the array is improved against the index of each microphone capsule. Increasing the order of the differential microphone array, thus using more microphone capsules, further increases the directivity index.

In contrast to Elko's approach, a first order differential array comprising three microphone capsules is described in this paper. Instead of steering the angle of highest sensitivity towards the desired signal source the null is directed towards the desired source. Like this the desired signal is cancelled whereas the ambient noise persists in the signal. It will be described later how this signal is used for spectral subtraction to achieve a very strong directivity towards the desired signal source for the overall system.

This paper describes two similar systems comprising a differential microphone array and signal processing for noise suppression using spectral subtraction, each. Due to the different geometries of the two microphone arrays the required pre-processing differs. Also the properties of the overall systems are not the same.

To understand the array geometry and the array processing proposed in this paper a short explanation of the spectral subtraction method is required. An in depth explanation can be found in e.g. [5].

Spectral subtraction is a well-known method for noise suppression working in the frequency domain. The noisy input signal is therefore either transformed using the Short Time Fourier Transform or a filter bank. Then the amplitude of each sub-band is attenuated to approximately the level the input signal would have if not being corrupted by noise. At last the modified signal is transformed back into time domain. To calculate the appropriate attenuation an estimate of the power spectral density (PSD) of the noisy input signal and an estimate of the PSD of the noise (excluding the desired signal) is needed.

Many proposals exist to estimate the noise PSD with only one microphone. Most of them work well with stationary noise, some even with slowly varying non-stationary noise [7] whereas all single-microphone systems suffer for high estimation errors when the noise PSD changes simultaneously with the PSD of the desired signal. This problem can be solved with microphone arrays, only, exploiting spatial information of the desired signal source and the ambient noise. This article shows two systems with different microphone arrays capable of accurately estimating the noise PSD even when the noise is varying very fast.

The next section briefly describes the geometries of the two microphone arrays, the pre-processing and the processing for spectral subtraction. In section 3 differences between both geometries are discussed. In addition the limitations of both systems are described.

2. SYSTEM OVERVIEW

The two systems presented next have in common that they exploit spatial information to estimate the noise PSD. As this estimate shall not be influenced from the signal coming from the direction of the desired signal the polar pattern of the estimated noise PSD shall have a null in the direction of the desired sound source. For all other directions the sensitivity shall be approximately the same as the sensitivity of the signal capturing the noisy signal.

The two systems under investigation have in common that each microphone array consists of three closely spaced microphone capsules. For the first system the capsules are

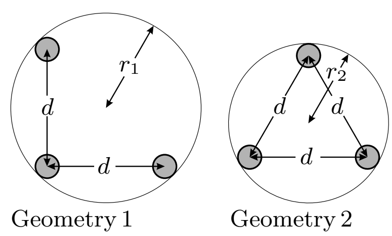

Figure 1: Geometries and sizes of both arrays

arranged in the corners of a right-angled triangle (geometry 1). The hypotenuse has the length $\sqrt{2}d$, the other sides have the length d. For the second system (geometry 2) the capsules are arranged in the corners of a triangle with a length of d for all sides (see figure 1).

For both systems pairs of microphone signals are subtracted from each other to calculate gradient signals. These gradient signals are then filtered to adjust the frequency response, transferred into the frequency domain and then squared (block PSD est. in figure 2 and 3) to get an estimate of the gradient PSDs. With two different gradient PSDs for geometry 1 and three different gradient PSDs for geometry 2 the PSD of the ambient noise signal is then estimated by simply adding the squared gradient PSDs.

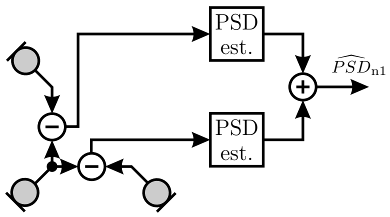

Figure 2: Pre-processing for geometry 1

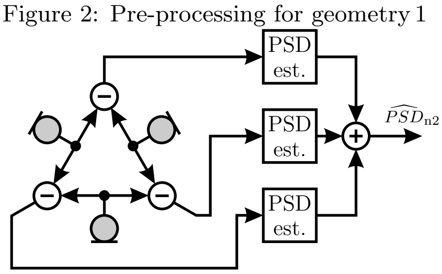

Figure 3: Pre-processing for geometry 2

For further definitions and calculations polar coordinates are used where φ is defined as the azimuth angle between the x-axis and the projection of the signal direction vector to the x/y-plane. Θ is the angle between the z-axis and the signal direction (see figure 4).

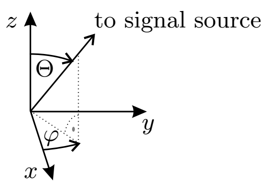

Figure 4: Definition of the angles φ and Θ

Although the two geometries differ in the position of one microphone capsule, only, the processing required is not the same for both systems to achieve an adequate signal to estimate the noise PSD. As the goal is to achieve a cardioid polar pattern with its null towards the desired signal source (φ_0, Θ_0) the figure eight polar patterns of the differential pairs have to sum up to a smooth torus for $\Theta_0 = 0$. We will see later why this leads to differences in signal processing for the array geometries under investigation.

2.1. Directivity pattern for low frequencies

A sound source from the direction (φ, Θ) generates the time varying power spectral density $\mathrm{PSD}_{\mathrm{in}}(f, k)$ on each of the microphone signals. f denotes the frequency index and k denotes the time index.

The PSD of the gradient signal calculated by subtracting microphone signal i from microphone signal j is then

$$\begin{aligned} \mathrm{PSD}_{ij}(f,k) &= 2 \cdot \mathrm{PSD}_{\mathrm{in}}(f,k) \cdot (1 - \cos\psi_{ij}) \\ &= 4 \cdot \mathrm{PSD}_{\mathrm{in}}(f,k) \cdot \sin^2\left(\frac{\psi_{ij}}{2}\right) \end{aligned} \tag{1}$$

with ψ_{ij} being the phase shift between the two microphone signals

$$\psi_{ij} = 2\pi f \left(\frac{\Delta d}{v_{\mathrm{Air}}} - \tau_{ij}\right). \tag{2}$$

v_{Air} indicates the propagation speed in air.

$$\Delta d = d \cos(\varphi - \varphi_{ij}) \sin\Theta \tag{3}$$

indicates the distance between the two microphones d as seen from the direction (φ, Θ) with the microphones placed in the x/y-plane with orientation φ_{ij}.

$$\tau_{ij} = \frac{d}{v_{\mathrm{Air}}} \cos(\varphi_0 - \varphi_{ij}) \sin\Theta_0 \tag{4}$$

denotes any parasitical time delay differences between the individual microphone capsules i and j or any intentionally implemented delay that can be used to tilt the beam of the array in the direction (φ_0, Θ_0). With (2), (3) and (4) the resulting phase shift is

$$\psi_{ij} = \frac{2\pi f d}{v_{\mathrm{Air}}} \left(\cos(\varphi - \varphi_{ij}) \sin\Theta - \cos(\varphi_0 - \varphi_{ij}) \sin\Theta_0\right). \tag{5}$$

With (1) the PSD of the noise reference is

$$\mathrm{PSD}_{\mathrm{n}}(f,k) = 4 \cdot \mathrm{PSD}_{\mathrm{in}}(f,k) \cdot \sum_{ij} \sin^2\left(\frac{\psi_{ij}}{2}\right) \tag{6}$$

for both geometries. For geometry 1 the summation term comprises the two elements with indices {21} and {31} whereas for geometry 2 the three elements {21}, {31} and {32} must be considered.

With the first order Taylor approximation of $\sin x \approx x$ for $x \ll \pi$ we get

$$\mathrm{PSD}_{\mathrm{n}}(f,k) \approx \mathrm{PSD}_{\mathrm{in}}(f,k) \cdot \sum_{ij} \psi_{ij}^2 \tag{7}$$

for low frequencies.

With $\varphi_{21} = 0$ and $\varphi_{31} = \pi/2$ for geometry 1 and

$$\cos^2(x) + \cos^2(x + \frac{\pi}{2}) = 1 \tag{8}$$

the noise reference PSD (1) gets independent from φ for $\Theta_0 = 0$.

$$\mathrm{PSD}_{\mathrm{n1}}(f,k) \approx \mathrm{PSD}_{\mathrm{in}}(f,k) \cdot \left(\frac{2\pi f \cdot d}{v_{\mathrm{Air}}}\right)^2 \cdot \sin^2\Theta \tag{9}$$

This corresponds to a polar pattern with the shape of a torus symmetrically to the z-axis. The dependency of the frequency f can easily be cancelled by low-pass filters (i.e. by using integrators in the PSD est. blocks).

With $\varphi_{21} = 0$, $\varphi_{31} = 2/3\pi$ and $\varphi_{32} = 4/3\pi$ for geometry 2 we get the same result except for a higher gain level by 3/2. This is because

$$\cos^2(x) + \cos^2(x + \frac{2}{3}\pi) + \cos^2(x + \frac{4}{3}\pi) = \frac{3}{2} \quad (10)$$

for all x. Thus, both systems are equivalent for low frequencies. It can be shown that this also holds for $\Theta_0 \neq 0$. With (8) and (10) we see why one additional summation term is required for (6) to achieve the desired torus shaped polar pattern.

3. RESULTS

3.1. Differences directly related to the geometry

In figure 1 the radiuses for both geometries

$$r_1 = \frac{d}{\sqrt{2}} + r_c \quad (11)$$

$$r_2 = \frac{d}{\sqrt{3}} + r_c \quad (12)$$

are indicated. With r_c indicating the radius of the capsules it can be easily seen that the space needed for geometry 2 is smaller by a factor of 2/3 if the radius of the capsules is neglected.

On the other hand geometry 1 with its 90° angle may have advantages from an optical design aspect when placed in a corner of a right angled box (e.g. in mobile phones). In addition, for geometry 1 slightly less computational effort is required compared to geometry 2 as only two PSD calculations have to be calculated. Compared to the effort needed for the transformation into the frequency domain and back this is only a little improvement.

To obtain best quality with differential microphone arrays the distance d must be selected carefully. If the distance is too small, noise problems and phase errors occur [2], [8]. On the other hand the directivity index decreases if the wavelength is near or below d.

Figures 5, 6 and 7 show the polar pattern of both array geometries in the direction of the x/y-plane. The direction of the null is also set in the x/y-plane with $\Theta_0 = \pi/2$ and φ_0 set to 0°, 30° and 45°, respectively. The microphone distance d was set to 2 cm for both geometries. Each figure shows 9 polar patterns for frequencies starting from 500 Hz up to 11.3kHz in steps of half an octave. The plot shows the intended cardioid slopes for frequencies up to 5.6 kHz. For higher frequencies up to 8 kHz the geometry 2 performs slightly better. If the frequency is even higher both geometries show undesired side-lobes and nulls. In addition the sensitivity decreases towards higher frequencies. This can easily be adjusted by filtering.

3.2. Capabilities and Limits of the proposed systems

For one specific recorded signal figure 8 shows how much the portion of the desired signal and the noise signal are sup-

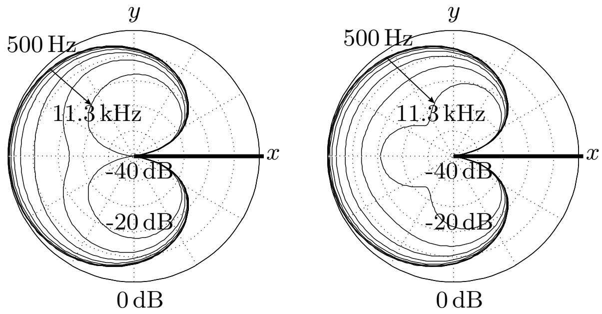

Figure 5: Polar Pattern of Array 1 and 2 for $\varphi_0 = 0°$

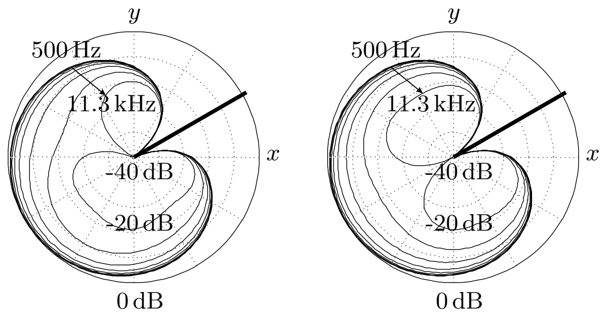

Figure 6: Polar Pattern of Array 1 and 2 for $\varphi_0 = 30°$

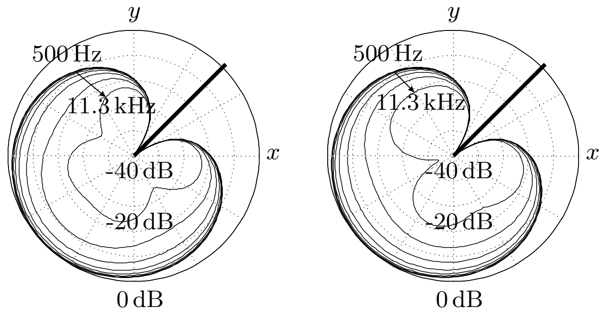

Figure 7: Polar Pattern of Array 1 and 2 for $\varphi_0 = 45°$

pressed in the output signal after spectral subtraction in respect to the input signal-to-noise ratio (SNR). The slopes vary very much dependent on the properties of the record to be analyzed and the parameters of the spectral subtraction algorithm. Thus, absolute values were omitted for the axes of figure 8. The figure clearly shows that for low noise levels the noise is attenuated quite well whereas the desired signal is attenuated only little. With increasing noise level the noise *and* the desired signal are attenuated more often and more severely. This results in an increased signal-to-noise ratio enhancement (SNRE) for moderately disturbed signals. If the SNR decreases significantly the attenuation of the desired signal becomes too strong. In this case the system attenuates all signal portions and does not discriminate between desired signal and noise any more. Therefore the SNRE drops to 0 dB for very low SNR.

This general behavior of the spectral subtraction algorithm has to be considered also if the polar patterns are analyzed. Even though the directivity of the array is very strong (see [1]), the system does not perform as well as a

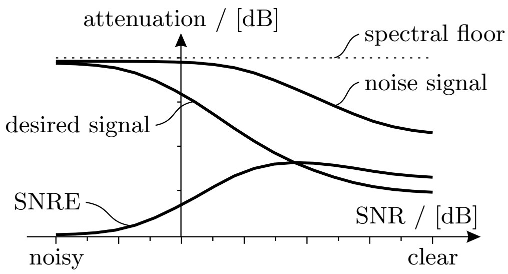

Figure 8: Suppression capability in respect to the input SNR

directional microphone when applied to very noisy environments.

3.2.1. Musical tones

Spectral subtraction often suffers from distortions known as 'musical tones' or 'musical noise'. These distortions mainly result from estimation errors regarding the PSD of the noisy signal and the noise PSD, respectively. Single-microphone systems are in principle unable to estimate the noise PSD of non-stationary noise in an accurate way. Most changes in frequency or amplitude therefore result in musical tones. With the microphone arrays described in this paper this source of estimation errors is removed. Nevertheless, not all musical tones can be avoided. To achieve the accurate noise PSD it would be required to average over all stochastic realizations of the spectra of the noise reference signal at each time. This is not possible as only one realization is available at a time. Therefore – assuming an ergodic process – averaging over time is used. Thus, a trade-off in the averaging time has to be chosen. A long averaging time will result in low estimation errors for stationary signals whereas shorter averaging must be chosen for non-stationary signals.

3.2.2. Influence of reverberation

If a noise reduction system consisting of one of the two array geometries, PSD estimation and spectral subtraction is used in a moderately noisy environment in which the desired signal and the noise sources are clearly spatially separated this system will perform very well. This also holds for non stationary noise including reverberation and echoes. If the noise also comes from the direction of the desired signal the attenuation of the noise will decrease. This is valid for the system under investigation and normal uni-directional microphones. With respect to reverberation both systems perform well if spaced closer than the reverberation radius to the signal source.

3.2.3. Influence of echoes

For the system under investigation no difference exists between noise, reverberation and echoes. Thus, echoes are also perceived to be significantly reduced if coming from other directions than the desired signal. But this only holds for human perception. Any echo canceller of the LMS-type applied to the output of the system under investigation will attenuate the echo by reducing the correlation between the loudspeaker signal and the output signal of the above system. Such canceller will encounter a nearly unchanged correlation. This is because the spectral subtraction will attenuate the amplitude of each spectral band whereas the phase remains unchanged. The echo canceller will then subtract the estimated echo from the noise-reduced signal. As the echo estimation is normally adapting very slowly this will lead to fluctuations in the residual echo. Therefore it is advantageous to apply any echo canceller in front of the noise reduction system.

4. CONCLUSIONS

In this paper two geometries for differential microphone arrays were compared, which both can be used as front-end for spectral subtraction. While array 1 needs slightly less computational power array 2 performs better for high frequencies. With both geometries it is possible to suppress noise, reverberation and echoes if the signal-to-noise ratio is not too high.

5. REFERENCES

[1] Ihle, M. et. al.: *A Novel Noise Suppression Algorithm Using a Very Small Microphone Array.* 109th Convention of the AES, Los Angeles, 2000 (Preprint).

[2] Ihle, M. et. al.: *Properties of a Surface Mountable Sub-Wavelength Array.* Seventh International Workshop on Acoustic Echo and Noise Control, IWAENC 2001, Darmstadt, 2001

[3] Elko, G.W.: *Adaptive noise cancellation with directional microphones.* IEEE Applications of Signal Processing to Audio and Acoustics, ASSP, 1997, pp. 4-7

[4] Elko, G.W.: *Superdirectional microphone arrays in Acoustic Signal Processing for Telecommunication.* Kluwer Academic, Dordrecht, 2000

[5] Vaseghi, S. V.: *Advanced Signal Processing and Digital Noise Reduction.* Wiley/Teubner, Chichester, New York, Brisbane, Toronto, Singapore, Stuttgart, Leipzig, 1996.

[6] Teutsch, H. and Elko, G. W.: *First- and Second-order Adaptive Differential Microphone Arrays.* Seventh International Workshop on Acoustic Echo and Noise Control, IWAENC 2001, Darmstadt, 2001

[7] Martin, R.: *An efficient algorithm to estimate the instantaneous SNR of spech signals.* Proceedings of the EUROSPEECH'93, Berlin, 1995, pp. 1093-1096

[8] Buck, M. and Roessler, M.: *First order differential microphone arrays for automotive applications.* Seventh International Workshop on Acoustic Echo and Noise Control, IWAENC 2001, Darmstadt, 2001