Directional Microphones in Automotive Applications (Adaptive Beamforming)

Patrick Vicinus

Peiker acustic GmbH & Co. KG, D-61381 Friedrichsdorf/Ts., Germany, Email: patrick.vicinus@peiker.de

Introduction

With silicon microphones there exists one more possibility for placing a capsule in a microphone. This paper discusses the capability to place these capsules in small size microphone-arrays. It separates the two application scenarios "speech recognition" and "handsfree phoning", which are usually important for an inner car application. It analysis the noise coherence matrix in a car cabin and evaluates the performance of a beamformer algorithm under different noise conditions.

Silicon Microphones

Especially the better soldering capabilities of silicon microphones as SMD part makes them suitable for small size arrays, where the capsules are placed directly to the PCB (printed circuit board). This enables an integration of the microphone array in a small environment, where is not enough space for unidirectional electret microphones. The directivity is introduced by a beamformer algorithm, which places the zero(s) to an optimal direction, which maximizes the SNR of the processed signal. Refer to later section about beamformer for details.

Speech recognition

A huge application area for microphones in cars are speech recognition engines. They are used in navigation-systems, handsfree car kits and as control system for electrical inner car components. In all use cases the speech recognition shall make the usage of the components easier, which allows the driver to spent more attention to the traffic of the road. This increases the safety and let the driver feel more comfortable by using the car. This is the reason, that the performance of a system with integrated speech recognizer is primary measured by the error rate of the recognition engine. This makes the enhancement of the recognition rate to a big point of interest. The recognition rate is correlated to the SNR (signal to noise ratio) of the input signal. More than the human ear (which corrects missing or wrong samples, missing letters or even missing words, by interpretating the sense of letter, word or sentence) the speech recognizer is very sensitive about a distortion of signal. Therefore for SNR enhancement, algorithms should be used, which minimize the need of adaptivity during active speech. Those algorithm are primary beamformer and (adaptive) linear filter, but also model based noise cancellation methods.

Handsfree Communication

Another field of microphone usage in car is the handsfree phoning. Not like a speech recognition engine, a noise cancellation for handsfree communication may influence some non audible artifacts, which are corrected by the human brain. Simple algorithms, which are based on spectral subtraction, will fulfill these requirements, by being very cost effective (hardware requirements).

Beamformer

The behavior of a beamformer is very similar to an analog unidirectional microphone. A dual element array can directly be compared with a gradient microphone of first order. These microphones are available with basically a directivity of a cardioid, super cardioid and a hypercardioid. The cardioid is optimized for a high suppression for signal arriving from the backside of capsules; the hyper cardioid is optimized for a cancellation of diffuse noise field. Both directivity patterns can also be achieved by the beamformer. An adaptive one adapts to a directivity which achieves the optimal degree of noise cancellation.

The presented beamformer in this paper is based on a "generalized sidelobe structure". It gets robust by applying an additional decorrelation filter behind one microphone. A transformation into a complex frequency domain by usage of a polyphase filterbank gives us the possibility to process a beamforming-algorithm even for very short microphone distances (smaller a half of minimum wavelength λ_{min}) by small filter sizes.

Blocking the signal Principal a blocked signal can be achieved by just delaying and subtracting the microphone signals:

$$b = x_1 - d \cdot x_2 \tag{1}$$

There are x_1 and x_2 the both microphone signals and d a delay which optimally consists of a simple phase shift, which corresponds to the estimated acoustic delay between the two microphone signals, when the signal arrives from speaker speaking in direction of the DOA (Direction of arrival). In praxis it covers also short time reflections and should be implemented in an adaptive way.

Attenuation of White Noise By delaying and adding (delay 'n sum) the two microphone signals, a minimization of power can be reached, in case, that the two microphone signals are uncorrelated. It is:

$$\mathbf{E}\left\{\frac{(N_1+N_2)^2}{2}\right\} = \frac{\mathbf{E}\left\{N_1^2\right\}+\mathbf{E}\left\{N_2^2\right\}}{4} = \frac{E\{N^2\}}{2} \tag{2}$$

If both sources of noise have the same power, the noise is attenuated by $3dB$ by the "Delay 'n Sum"-Path.

Interference Cancellation Filter To reduce also correlated signal components effectivly, the blocked signal path will be subtracted by using a Wiener filter from the "delay 'n sum"-path. This filter will be norm-contrained to prevent a target signal cancellation. The correlation of the two input signals decreases with the number of uncorrelated sources of noise. Following equation enables a good estimation about the expected performance of the beamformer. It shows a strong dependence between the performance and the location of noise sources. We calculate the degree of noise cancellation in dependence of the location of noise.

To make the equations more understandable, it is reduced first to only two uncorrelated sources of noise. Afterwards it will be expanded to a variable number of sources. We are talking in the following about two signals: first of the "blocked signal" (b) and second of the "delayed and summed signal" (d). All equations are written in the frequency domain and are complex. The dependence of the signal to the frequency ω is not explicit written.

It is

$$b = n_1 + n_2 \tag{3}$$

and

$$d = a_1 n_1 + a_2 n_2 \tag{4}$$

Here a noise transfer function a_n was introduced. The Wiener filter estimates:

$$\tilde{d} = \frac{E\left\{S_{db}\right\}}{E\left\{S_{bb}\right\}} \cdot b \tag{5}$$

The estimated degree of noise cancellation can be calculated as:

$$\epsilon = \frac{d - \tilde{d}}{d} = 1 - \frac{n_1 + n_2}{a_1 n_1 + a_2 n_2} \cdot \frac{a_1 m(|n_1|^2) + a_2 m(|n_2|^2)}{m(|n_1|^2) + m(|n_2|^2)} \tag{6}$$

The power of the uncorrelated noise sources will in the following estimated to be equal. This means:

$$m(|n_1|^2) = m(|n_2|^2) \tag{7}$$

and as estimated suppression of power follows:

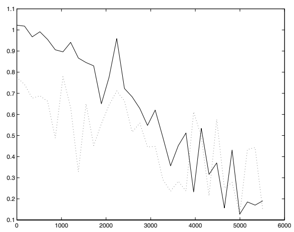
$$\begin{aligned} \epsilon^2 &= 1 - E\left\{\frac{|n_1|^2+|n_2|^2}{|a_1|^2|n_1|^2+|a_2|^2|n_2|^2} \cdot \left(\left|\frac{a_1+a_2}{2}\right|\right)^2\right\} \\ &= 1 - \frac{1}{2} \cdot \frac{|(a_1+a_2)|^2}{|a_1|^2+|a_2|^2} \end{aligned} \tag{8}$$

This equation can now be expanded to a variable number of noise sources. It follows:

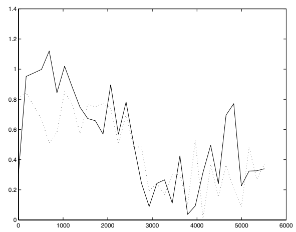
$$\epsilon^2 = 1 - \frac{1}{N} \cdot \frac{\left|\sum_{n=0}^{N-1} a_n\right|^2}{\sum_{n=0}^{N-1} a_n^2} = 1 - \frac{|E\{a\}|^2}{E\{|a|^2\}} \tag{9}$$

Interpretation Equation 9 shows the degree of noise cancellation in dependence of the noise transfer function a. If there is a low variance of a the beamformer reaches his maximum performance. This means, that a low distance of microphones, which minimizes the variance , will optimize the performance of beamformer in case of uncorrelated noise sources are existing. Further more the sum over all a_n should never converge to zero and no main noise should arrive close to the DOA (high values of a_n), which would causes a high variance.

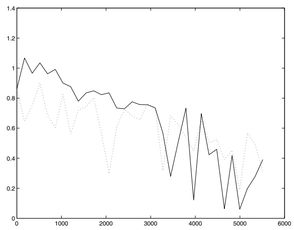
Measurements A common statement about the noise-field in cars can not be given. But normally there exists following main sources of noise: engine, fan, open window and diffuse noise of the wheels. Figure 1(a) - 1(d) shows the estimated coherence function of the inputs and the signals d and b. It can be well seen, that this sample array is not performance optimized. Optimally, in case of noise, both functions should be identical. For high frequencies the endfire array is performing optimal, a further improvement is only possible by changing the location of the array. The large distance between the lines in deeper frequencies identify range for more improvements, by changing only the steering direction.



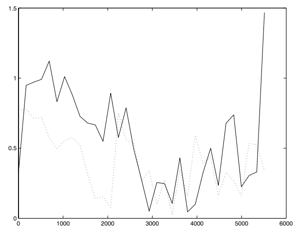
(a) Endfire, signal of interest



(b) Endfire, noise



(c) Broadside,signal of intersest



(d) Broadside, noise

Figure 1: Performance Check: Coherence index versus frequency. Dotted: Coherence of blocked signal and DSB; Solid: microphone 1 and microphone 2

Results This paper presented a possibility to estimate the performance of a beamformer by determine main sources of noise. In car environments the performance can be optimized by minimizing the variance of the noise transfer values of the main sources.