PROPERTIES OF A SURFACE MOUNTABLE SUB-WAVELENGTH ARRAY

Marc Ihle

Siemens AG., ICM MP PO2 ULM 23
Lise-Meitner-Str. 5-7
D-89081 Ulm, Germany
marc.ihle@ieee.org

Kristian Kroschel and Dirk Bechler

Institut für Nachrichtentechnik
Universität Karlsruhe, Kaiserstr. 12
D-76128 Karlsruhe, Germany
kroschel@int.uni-karlsruhe.de

ABSTRACT

This paper presents some new results that prove the usability of a Surface Mountable Sub-Wavelength (SMSW) Array for low-cost, hands-free applications and as a front-end for speech recognition systems. A noise suppression system using an SMSW-array shows a strong directivity index and improves the recognition rates of the speech recognizer used for our investigations.

1. INTRODUCTION

The Surface Mountable Sub-Wavelength (SMSW) Array was developed at the Institute for Communication Technology, University Karlsruhe, Germany. It is designed as a front-end for noise suppression systems based on spectral subtraction. It generates a precise noise estimation that instantly reacts to any changes of the ambient noise. Thus, the overall system shows fewer artifacts when tested with non-stationary noise.

In this paper, we show the properties of the SMSW-array with respect to realization problems when low-cost microphone capsules are used. In addition the performance of a speech recognizer is analyzed which optionally uses a noise canceller with SMSW-array as front-end.

In the next section 2, we will describe the architecture and the most important properties of the SMSW-array. In section 3 we present some simulation results that prove the high directivity of the system. Thereafter we show the properties of an SMSW-Array built up using low-cost microphone capsules (section 4). In the last section, we present the results of a speech recognition system with an SMSW-Array as optional front-end.

2. DESCRIPTION OF THE SYSTEM

In this section, we explain the SMSW-array only very briefly. A comprehensive description can be found in [1].

Noise suppression algorithms based on spectral subtraction use an estimate of the signal-to-noise ratio of the input signal to calculate how much the noisy input signal must be attenuated. This is normally done individually for each frequency band. Therefore, a good estimate of the noise level in each frequency band is essential. For relatively stationary background noise, such a noise reference can be achieved, for example, using the minimum statistics approach [2]. Especially for non-stationary background noise, the SMSW-array can be used to calculate the power spectral density (PSD) of the noise more precisely.

Three small microphone capsules are placed according to Fig. 1, e.g. in a plane surface. The distance d must be smaller than half the shortest wavelength transmitted by the overall system.

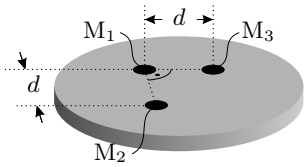

Figure 1: Geometry of the proposed microphone array

First, the microphone signals are amplified, digitized and equalized. Then, the short-time Fourier transforms (STFT) of the gradient signals

$$R'_{\mathrm{x}}(n,k) = \mathrm{STFT}(r'_2(k) - r'_1(k)), \quad (1)$$

$$R'_{\mathrm{y}}(n,k) = \mathrm{STFT}(r'_3(k) - r'_1(k)) \quad (2)$$

are calculated, where $r'_i(k)$ represents the i-th equalized microphone signal. n represents the discrete frequency index, whereas k indicates the time index of the signal frame under consideration. (Instead of a short-time Fourier transform, any other filter bank, for example wavelets can be used, too.) A sound source which is placed in the direction (φ, ϑ) and which generates the power spectral density $\mathrm{PSD}_{\mathrm{in}}(n,k)$ on each of the microphone signals will generate the PSDs

$$\mathrm{PSD}_{\mathrm{x}}(n,k) \approx \mathrm{PSD}_{\mathrm{in}}(n,k) \cdot \sqrt{2}\pi \frac{n}{N} f_{\mathrm{A}} \cdot \left| \frac{d}{v_{\mathrm{L}}} \cos\varphi \sin\vartheta + \tau_{21}(n) \right| \quad (3)$$

and

$$\mathrm{PSD}_{\mathrm{y}}(n,k) \approx \mathrm{PSD}_{\mathrm{in}}(n,k) \cdot \sqrt{2}\pi \frac{n}{N} f_{\mathrm{A}} \cdot \left| \frac{d}{v_{\mathrm{L}}} \sin\varphi \sin\vartheta + \tau_{31}(n) \right| \quad (4)$$

on the gradient signals, respectively. In these equations $\tau_{i1}(n)$ represents the delay difference between the microphone signal $r'_i(k)$ and $r'_1(k)$. It is introduced by mismatches of the capsules, the amplifiers and, intentionally, by a phase shifter or equalizer. The latter compensates for the non-ideal properties of the capsules and amplifiers and steers the directivity of the overall system. φ and ϑ are defined as the declination angle between the x-axis and the projection of the signal direction vector to the x/y-plane and as the azimuth angle between the signal direction and the z-axis, respectively (see Fig. 2). $\mathrm{PSD}_{\mathrm{x}}(n,k)$ and $\mathrm{PSD}_{\mathrm{y}}(n,k)$ mainly depend on the angles φ and ϑ and

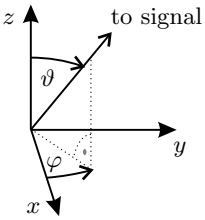

Figure 2: Definition of the angles φ and ϑ

the frequency $f = \frac{n}{N} f_\mathrm{A}$ with f_A representing the sampling frequency and N the STFT's frequency resolution.

The noise estimation required for the noise reduction system is then calculated by

$$\widehat{\mathrm{PSD}}_\mathrm{N}(n,k) = \left(|R'_\mathrm{x}(n,k)|^2 + |R'_\mathrm{y}(n,k)|^2\right) \cdot H_\mathrm{I}^2(n) \qquad (5)$$

with

$$H_\mathrm{I}(n) = \begin{cases} \dfrac{v_\mathrm{L} N}{\sqrt{2}\pi f_\mathrm{A} d\, n} & , \quad 0 < n \le N/2 \\ \dfrac{v_\mathrm{L} N}{\sqrt{2}\pi f_\mathrm{A} d\,(N-n)} & , \quad N/2 < n < N \end{cases} \qquad (6)$$

to compensate for the frequency dependency of equations (3) and (4).

This noise reference shows a directivity that is similar to the directivity of each of the microphone capsules alone with the exception of a null in the direction (φ_0, ϑ_0). If this null is pointed towards the direction of the desired signal a very good noise estimation is achieved.

(φ_0, ϑ_0) can be set to any direction when appropriate delays

$$\tau_{21} = \frac{d}{v_\mathrm{L}} \cos\varphi_0 \sin\vartheta_0 \qquad (7)$$

$$\tau_{31} = \frac{d}{v_\mathrm{L}} \sin\varphi_0 \sin\vartheta_0. \qquad (8)$$

are introduced by a phase shifter or an equalizer.

3. DIRECTIVITY INDEX

A noise suppression system using an SMSW-array shows a strong directivity. The directivity index depends on the parameterization of the noise reduction system applied. On the other hand it is independent from φ_0 and shows only weak dependency on ϑ_0. If d is much smaller than half the wavelength of the highest frequency to be considered, the directivity is even independent from frequency. All these properties hold, of course, only if omni-directional microphone capsules with excellent properties are used.

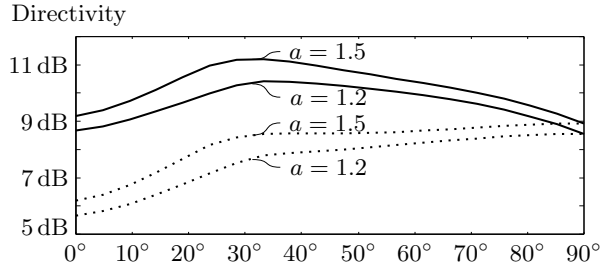

Figure 3: Directivity index with respect to ϑ_0; solid lines: surface mounted array, dashed lines: free-field environment

Fig. 3 shows the directivity index of a noise suppressor in respect of the angle ϑ_0 with the overestimation factor a as parameter. Figures for ideal microphone capsules placed in a free-field environment and for capsules mounted on a large plane surface are shown.

4. INFLUENCE OF NON-IDEAL MICROPHONES

The usability of the system strongly depends on whether the system is usable with low-cost microphones. Because three microphone capsules are needed to built up an SMSW-array, the price for these capsules has a big influence on the overall cost.

Microphone capsules of the same type differ individually in their sensitivity and frequency response. In addition, they suffer from phase errors and a relatively high noise level.

To verify the properties of the simulation results we set up two microphone arrays with three low-cost Hosiden KUB3323 capsules, each. These omni-directional capsules are designed for usage in cellular phones and other low-cost devices. They are specified to have a gain tolerance of up to +4/-6 dB in the narrow band frequency range for telephony (200 Hz – 3.4 kHz). In addition the overall sensitivity may vary by ±3 dB. The phase tolerance is not specified at all for these capsules. We built the capsules into a wooden disc of 14 cm diameter, according to Fig. 1. The distance d between the capsules was set to 2 cm. All recordings for the frequency response, directivity and noise measurements we performed in an anechoic chamber. For the investigations described below, we did not compensate for amplitude or phase differences between the capsules.

4.1. Amplitude Sensitivity

The noise reference signal is derived from the differences between the input signals. Therefore, amplitude mismatches caused by differences of the microphone capsules frequency responses and sensitivities increase the level of the noise reference signal. As a consequence, the achievable directivity index decreases. Especially for low frequencies, any mismatch causes severe degradations. Fig. 4 shows the influence of amplitude mismatches of a single microphone to the directivity index.

In Fig. 5, the polar plot of one of the SMSW-arrays used for our investigations is shown. The degradation for low frequencies is clearly visible. The slope for 400 Hz shows only very little directivity towards the region of interest ($\vartheta_0 = 0°$). For higher frequencies, the directivity fits quite well to the theoretical achievable slope, although the frequency responses of the capsules differ.

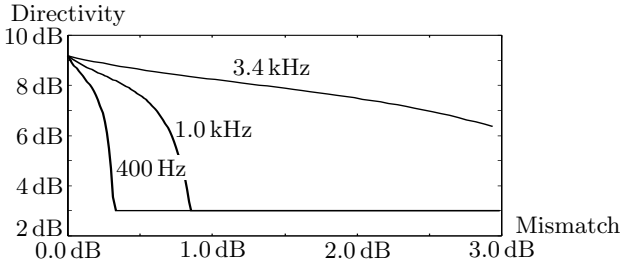

Figure 4: Degradation of the directivity index in respect to amplitude mismatches ($\vartheta_0 = 0$, $a = 1.5$)

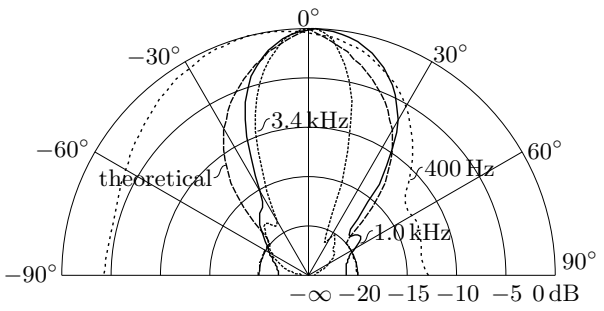

Figure 5: Directivity of an SMSW array built up with Hosiden KUB3323 capsules

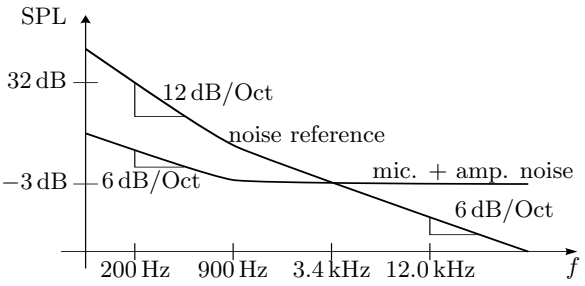

Figure 6: Power Spectral Density of the microphone capsules and the noise reference signal (schematic)

The system can be approved, of course, for low frequencies. As the microphone capsules of the array are located very near to each other, their mean input pressure spectral density level over time is identical. Thus, some adaptive gain control algorithm realized in the frequency domain can be applied. It can significantly reduce the gain and frequency response differences and therefore will improve the directivity index.

4.2. Phase Sensitivity

In contrast to amplitude mismatches phase errors have no direct influence on the directivity index. Because the signal delay between the equalized microphone signals is used to steer the null of the noise reference signal to the region of interest, a mismatch of the capsules in the group delay will tilt the beam of the overall system away from its intended direction. The influence of non-equalized group delay difference can be easily derived from equations (7) and (8).

The arrays we built up with the Hosiden capsules showed no phase errors that degraded the overall performance of the system. In Fig. 5 the slopes for all frequencies show their maximum very near to the intended angle $\vartheta_0 = 0°$. Within the anechoic chamber we measured a maximal directivity error of $\Delta\vartheta_0 = \pm 10°$ for both arrays tested.

When an SMSW-array is used in some reverberant environment with strong early reflections, (for example if the array is placed near a window), additional phase errors are introduced by the environment. Another problem might be that the steering direction must be adapted over time when the sound source is moving around. For these applications blind equalization techniques [3] can be used to automatically adjust the delays $\tau_{21}(n,k)$ and $\tau_{31}(n,k)$ over time and frequency.

4.3. Noise Sensitivity

Electret microphones normally show a uniform noise level over a wide frequency range. In contrast to that microphone amplifiers suffer for low-frequency noise which decreases by 6 dB per octave towards higher frequencies. The Hosiden capsules together with the microphone amplifiers used for our experiments (OP-amp. NE5534) show exactly this behavior. For frequencies below 900 Hz the noise of the amplifier dominates. Above this frequency a constant noise level of about -3 dB SPL was measured (see Fig. 6).

The low frequency noise of the amplifiers has some minor impact on the noise reference of the SMSW-array. The noise floor of the noise reference signal rises by 6 dB/Octave towards lower frequencies due to the impact of the filter $H_I(n)$ used to compensate for the frequency response of the gradient signals. Fig. 6 also shows the resulting noise level of the noise reference signal. In quiet situations, the noise reference signal contains more low frequency noise than the input signal. This leads to a somewhat too strong suppression of the low frequencies in the output signal.

The properties of the SMSW-array described in this section show that there is room for improvements. Algorithms for amplitude and phase adjustments are currently under investigation. We assume that the quality of the overall system can be enhanced significantly, especially for low frequencies.

5. SPEECH RECOGNITION EVALUATION

Speech recognition has reached a high level of quality if the systems are applied in an environment with a low noise level. In this case, recognition rates over 90% are a standard. In a noisy environment, a significant degradation of the recognition rate can be observed so that the customer is bothered when communicating with the system. Hence, means for noise reduction are required.

5.1. Description of the real-time demonstrator

To evaluate the performance of the proposed noise reduction method for speech recognition, the system has been implemented in real-time on Motorolas DSP MC 96002. The sampling frequency is 8 kHz. The recorded data is processed in block segments of 256 samples (32 ms) with a 50% overlap. For the SMSW-array low-cost electret microphone capsules are used. The main beam of the SMSW-array is steered orthogonally to its plane surface with $\vartheta_0 = 0°$. Thus, the delays of equation (7) and (8) are $\tau_{21} = \tau_{31} = 0$. For the overestimation a value of 1.2 and for the spectral floor a value of 0.1 were chosen as parameters of the spectral subtraction. The demonstrator provides two output signals: the SMSW-array noise suppressed signal can be compared to the non-processed signal from a single microphone.

5.2. Experimental setup

The speech data was recorded in an office room (5 m × 5 m). To ensure reproducible recording conditions, both the speech signal and the noise signal were pre-recorded and played back by loudspeakers L1 and L2. The distance between L1 (speech signal) and the SMWS-array is 40 cm. The main beam of the array is steered to the direction of L1. At an angle of 60° to the main beam direction, L2 (noise signal) is placed at a distance of 2 m from the array.

5.3. The speech recognizer and its training

The commercial speech recognizer *Dragon Naturally Speaking 5 Preferred* of *Dragon Systems (L&H)* [4] was used for the performance evaluation. The active (passive) vocabulary is about 270 000 (340 000) words.
For each of the two output signals, the system was trained 30 minutes for a single German native speaker. To obtain this output data, the pre-recorded noiseless training data was played back by L1. During the training, no noise signal at L2 was replayed. Only the typical low-frequency environmental noise in an office (fans, etc.) was present.

5.4. Test data and background noise

The type of speech is continuous but non-spontaneous, i.e. the speaker reads a pre-formulated text. This text consists of 40 test sentences (319 words) for speech recognition according to Beckmann and Schilling, which were uttered three times from the single speaker for whom *Dragon* was trained. Thus, statistics on word error rates in speech recognition were made with $3 \times 319 = 957$ words. It was ensured that all words of the test sentences were included in the active vocabulary of *Dragon*.

The influence of the system's performance for different noise types was evaluated. A vacuum cleaner was used for the stationary noise case, as non-stationary noise signal babble noise has been chosen. The noise sources are replayed from L2.

5.5. Recognition results

To evaluate the performance of the proposed system, the recognition rates with and without the SMSW-array noise suppression system are compared as a function of the signal-to-noise ratio (SNR). Fig. 7 shows the results for the stationary, Fig. 8 for the non-stationary noise case. In Tab. 1 some exemplary word error rates for different SNRs are listed. The performance for both noise types is comparable. For decreasing SNRs the recognition rate enhancement for the SMSW-array processed speech data increases by 6.42% for stationary noise and by 5.02% for non-stationary noise. Thus, the method can be considered as noise-type independent.

SNR	17 dB	13 dB	8 dB
		Stationary	
SMSW-array	10.03%	45.77%	79.62%
without	13.01%	49.53%	86.05%
		Non-Stationary	
SMSW-array	11.02%	45.56%	79.31%
without	12.54%	48.69%	84.23%

Table 1: Word Error Rates for stationary and non-stationary noise

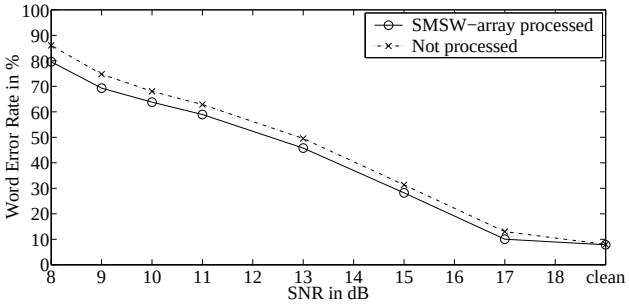

Figure 7: Word Error Rates for stationary noise

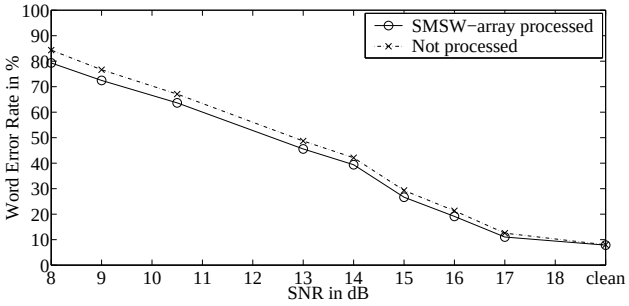

Figure 8: Word Error Rates for non-stationary noise

6. CONCLUSIONS

The results we have presented show that the Surface Mountable Sub-Wavelength (SMSW) Array can be built up using low-cost microphone capsules. An SMSW-array based noise suppression algorithm can be used as a front-end to a speech recognition system especially for the use in environments that are strongly interfered by non-stationary noise.

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

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

[4] *http://www.dragonsys.com*