NOISE REDUCTION FOR AUDIO IN REAL TIME AND WITH LOW POWER CONSUMPTION

Iko Pieper, Amelie J. Hintermaier, and Tamas Harczos

audifon GmbH & Co. KG, 99625 Kölleda, Germany

ABSTRACT

Two algorithms were developed for the reduction of static background noise in highly amplified audio signals. The targeted platforms are embedded systems with ultra-low power consumption. Algorithm A combines a fixed beamformer with an optimized version of the simple spectral subtraction algorithm by Leimeister et al. [1]. Adding a 14-channel compressor and a limiter results in algorithm B. We have submitted the algorithms to the SPEAR challenge to benchmark them against more complex and computationally expensive solutions including the baseline. They were tested on the PrimeHA development platform for hearing aids that features an ARM Cortex M4 microcontroller [2]. The processor load was 43% and the total system latency was less than 6 ms using a block length of 64 sound samples. However, the block length has been increased to 1024 samples in order to improve sound quality and make the algorithms applicable to the noise of background talkers in the data of the SPEAR challenge that is somewhat less static. The resulting theoretical latency still lies well below 50 ms.

***Index Terms*—** Noise reduction, power consumption, real time, embedded system, spectral subtraction

1. INTRODUCTION

It is not uncommon for hearing systems that are intended to compensate for hearing impairment, such as hearing aids and cochlear implants, to apply 50 dB or more gain to an audio signal that is captured with a microphone. For such high amplifications, the noise floor of the microphone is clearly audible, regardless whether an analogue or a more current pulse density modulation microphone pre-amp is used. Noise reduction algorithms can be applied to reduce the noise floor. But even if the processor is capable of performing these calculations in real time, high computational complexity results in high current draw and therefore in low battery life. Thus, less computationally complex algorithms are always favourable. Furthermore, it is tempting to solve multiple problems with a single algorithm or, at least, share intermediate results between algorithms.

The spectral subtraction algorithm has a low computational complexity. It requires a (frequency dependent) estimate of the power of the current noise floor. This estimate is the minimum in the short time integrated power spectrum of the signal at each frequency bin for the recently passed time period. Consequently, this algorithm works best if the (short time integrated) power of the noise floor does change slower than the (short time integrated) power of the targeted signal. This is usually given for the static noise that is caused by the electrical components of the audio signal chain. Environmental noise shares these characteristics to a certain extent. Thus, it is interesting to access whether the performance and sound quality of spectral subtraction is sufficient to be applied to both static and environmental noise.

2. ALGORITHM DESCRIPTION

2.1. Beamformer

Two microphones form a fixed beamformer that is aimed at sound waves with frontal incidence for each ear of the listener. The signal of the front microphone is delayed by D samples and added to the rear signal. Independent of the wavelength, constructive interference for frontal incidence is given if D is set to

$$D = \frac{f_\mathrm{s} l}{c}, \tag{1}$$

where f_s is the sampling frequency, l the distance between the microphones, and c the speed of sound.

2.2. Simple spectral subtraction

The suggested noise reduction is based on the simple spectral subtraction algorithm [1]. The audio signal blocks are processed in the frequency domain. Transformations between time and frequency domains are performed with the overlap-add method using a sine window for both analysis and synthesis.

Simple spectral subtraction basically determines the minimum short-time power at a frequency bin n (minimum short-time sub-band power $P_{n,\mathrm{min}}$) over a certain amount of previous time frames, where m denotes the current time frame. $P_{n,\mathrm{min}}$ can be regarded as an estimate for the signal power of the noise floor. If the current short-time sub-band power $P_{n,m}$

equals osub $\cdot P_{n,\min}$, the current signal $X_{n,m}$ is subtracted by itself yielding the output signal $Y_{n,m} = 0$. The higher $P_{n,m}$ is, the lesser is subtracted from $X_{n,m}$ to obtain $Y_{n,m}$:

$$Y_{n,m} = X_{n,m} - \sqrt{\text{osub} \cdot \frac{P_{n,\min}}{P_{n,m}}} \cdot X_{n,m}. \qquad (2)$$

The amount of spectral subtraction can be increased by increasing the value of osub.

The short-time sub-band power $P_{n,m}$ is calculated by short-time integration of the squared signal magnitude $|X_{n,m}|^2$ via first order low-pass filtering.

In the original algorithm, spectral flooring is used to reduce artefacts known as musical noise. This means the absolute value of the output signal $|Y_{n,m}|$ is ensured to be higher than the lower limit $\sqrt{\text{subf} \cdot P_{n,\min}}$, where subf is the spectral flooring parameter.

The phase of the signal is not altered by the algorithm. Thus, artefacts introduced by the algorithm can be masked with the input signal $X_{n,m}$ at the cost of less noise reduction by scaling it down and adding it to the output signal $Y_{n,m}$.

2.3. Optimizations (Algorithm A and B)

In the original algorithm, the integration time of the short-time sub-band power that is used to find $P_{n,\min}$ is longer than the integration time of $P_{n,m}$. Listening tests have revealed that sound quality and performance are maintained using the same value for both integration times. Thus, one instead of two low-pass filters per frequency bin is sufficient to calculate $P_{n,m}$ and $P_{n,\min}$.

Furthermore, the listening tests have revealed that the short window lengths used in the overlap-add framework call for a high amount of oversubtraction (high value for osub) but in turn make the spectral flooring unnecessary. Removing spectral flooring reduces the number of computationally costly square root operations.

Algorithm A is the optimized version of the simple subtraction algorithm. Algorithm B is an extension of A as described below.

2.4. Additional dynamic compression (Algorithm B)

In addition to spectral subtraction, Algorithm B introduces dynamic compression. The frequency bins are grouped into 14 channels on a logarithmic frequency scale. A compressor as well as a limiter is applied to each of the channels. An additional limiter is applied to the output signal in the time domain to ensure that remaining transients do not overshoot the full scale limit. Because of the dynamic compression, higher amplifications of the overall signal without uncomfortable loudness and therefore better speech intelligibility as for algorithm A might be achieved. Nevertheless, to ensure a fair comparison of the sound quality between the two algorithms in listening tests, the output of the signal is scaled down to obtain similar noise floor levels as for algorithm A.

3. MEASUREMENTS

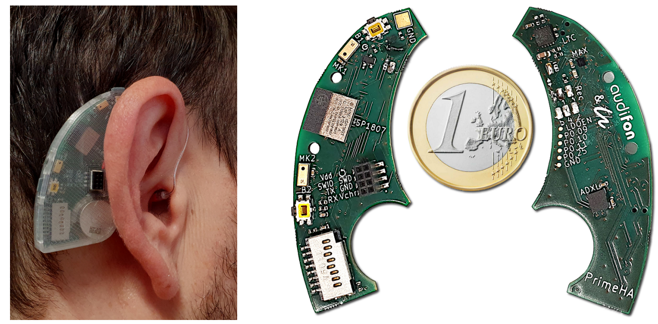

Fig. 1. The PrimeHA development platform for hearing aid algorithms, taken from [2].

In order to assess the practicability of the algorithms with respect to power consumption and latency, the computationally heavier algorithm B was implemented on the PrimeHA development platform for hearing aid algorithms (Fig. 1) [2]. The block size was set to 64 samples and the sampling frequency was set to 25 kHz. Acoustic signals were captured with the microphones, real time processed on the ARM Cortex M4 microcontroller, and output over the receiver (in ear speaker). Electrical parts that were not required (such as the SD card reader) were powered off.

3.1. Power consumption

Power consumption was assessed by measuring the battery life span for realistic conditions. The development platform is equipped with a rechargeable battery with a capacity of 77 mAh and a nominal voltage of 3.7 V. In order to protect the battery, the firmware shuts down and switches off the platform if the voltage drops below 2.7 V which is the lowest nominal voltage at which the development platform still operates as normal.

The battery was fully charged. The development platform was placed in front of a speaker that output an infinite loop of a long speech signal. The total processor load was about 43%. Repeated measurements gave an average duration of about 11 h and 30 min until the platform was switched off by the firmware.

3.2. Latency

To measure the total latency of the development platform a test signal was constructed from impulses, sine bursts with frequencies ranging from 50 Hz to 10 kHz, and a short speech signal. The test signal was output by the aforementioned speaker. The front microphone of the development platform was turned off. The development platform and a measurement microphone were placed next to each other in front of the speaker, so that the distance between speaker and measurement microphone capsule and the distance between speaker and rear microphone capsule of the development platform were identical. An additional measurement microphone was placed right in front of the receiver using a 2cc coupler. The signals of the measurement microphones, i.e., processed and unprocessed test signal, were recorded in stereo mode with an audio interface.

Cross correlation of the two stereo audio channels yielded a latency of slightly below 6 ms.

4. ENVIRONMENTAL NOISE

Algorithms A and B successfully reduce static noise caused by the electric components of the development platform (not shown). However, the algorithms have not yet been applied to environmental noise, and multiple background talkers in particular. In order to assess the performance of the suggested algorithms, they were submitted to the SPEAR challenge. There, listening tests and objective metrics are applied to the output signals and the results compared to likely more advanced but computationally more expensive algorithms with respect to the gained speech intelligibility and the sound quality.

Microphones 1 and 5 were used for the beamforming for the left ear, and 4 and 6 were used for the beamforming for the right ear. Since the microphone distances were not available, the microphone selection and the sample delay D was derived from the input signals of the SPEAR development data.

For the data of the challenge, computations were performed on a laptop. This additional computational power allowed us to increase the block size from 64 to 1024 samples. With the sampling frequency of 48 kHz used for the spear data and considering one block of recording delay and one block delay added by the overlap-add method, this yields a theoretical latency of $2 \cdot 1024/48\,\text{kHz} = 43\,\text{ms}$. An increased block size results in a better sound quality of the target signals, in particular at low frequencies, but increases the musical noise artefacts. Artefacts were masked by scaling the input signal $X_{n,m}$ of spectral subtraction with a factor of 0.25 and adding it to the output signal $Y_{n,m}$.

5. CONCLUSION

By successfully running the computationally heavier algorithm B of the proposed algorithms on a single core microcontroller with low latency and almost 12 h of battery life, it has been shown that the algorithms real time capability and power consumption are sufficient for practical use.

Further evaluations, including the SPEAR challenge, might clarify if the algorithms reduce not only static but environmental noise as well and thereby improve speech intelligibility.

6. REFERENCES

[1] M. Leimeister, C. Egervári, F. Kuhnke, A. Chilian, C. Voigt, and T. Harczos, "Simple spectral subtraction method enhances speech intelligibility in noise for cochlear implant listeners," in *Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies*, Nyborg, Denmark, 2015, 5th symposium on Auditory and Audiological Research, pp. 325–332.

[2] A.J. Hintermaier, *Entwurf einer kompakten, energieeffizienten, frei programmierbaren Hörgeräteplattform zur praxisnahen Erprobung audiologischer Algorithmen*, Dipl. thesis, Ilmenau University of Technology, Ilmenau, Germany, 2022.