Real-Time Implementation of an Adaptive Beamformer-Postfilter System for Seat Belt Microphones on SHARC ADSP-21489

Matthias Herbert[1], Mohammed Krini[1], Zafar-Baig Mirza[2]
[1] *Hochschule Aschaffenburg, Signalverarbeitung, Würzburgerstr. 45, Aschaffenburg, Germany*
[2] *paragon AG, Division Acoustics, Schwalbenweg 29, Delbrück, Germany*
E-Mail: mohammed.krini@h-ab.de

Abstract

Belt Microphones are an elegant and interesting alternative to conventional microphones used for speech communications in automobile environments. Three microphones are integrated into safety belts of the car, which are placed around the shoulder and chest of the passenger after fastening the belt. Belt microphones offer an improved SNR (signal-to-noise ratio) due to small distance to the passenger's mouth compared to common microphone positions. State-of-the-art systems select only the microphone with the highest SNR on the belt as an input to speech enhancement schemes. The performance can be further improved by combining all belt microphones to generate an enhanced output. This contribution proposes a low-complexity and robust adaptive beamformer-postfilter system with special focus on its real-time implementation on a SHARC ADSP-21489 hardware. The system's overall performance designed for belt microphones is then evaluated and compared with the single best belt microphone in terms of SNR and SIR (signal-to-interference ratio).

Seat Belt Microphones

The seat belt microphones are invented and manufactured by paragon AG [10]. It consists of three omnidirectional microphones of diameter 10 mm and spaced 160 mm apart from each other, such that at least one microphone is optimally placed close to the passenger's mouth. In Fig. 1 an example of a belt microphone system installed in a car for driver seat is shown. All signal lines needed for signal transmission and voltage supply are invisibly integrated into the seat belt. In comparison with common microphone positions in automobiles like roof, steering wheel etc., belt microphones already exhibit the best signal quality for being closest to the passenger's mouth. The time-frequency analysis of the driver's belt microphone signals recorded at speed of 120 k.p.h. is plotted in Fig. 2. Since, all three microphones exhibit varying SNRs, they can be combined suitably to generate a single beamformed output. Hereby, all the inherent challenges associated with belt microphones have to be taken into account in order to achieve an optimal performance. These include changes in the position of the microphones when moved or being accidentally touched e.g. by fingers, ties and zips.

Figure 1: Driver seat equipped with belt microphones (marked in yellow boxes).

Hardware ADSP-21489 EZ-Kit Lite

Fig. 3 shows the ADSP-21489 EZ-Kit Lite. The digital signal processor (DSP) core on it is the ADSP-21489 from Analog Devices based on SHARC architecture. It's maximal operating frequency is 400 MHz, whereby it has an internal RAM memory of 5 Mbits. Two processing units are part of the ADSP-21489 that work in single-instruction, multiple-data (SIMD) mode to execute the same instruction on each processing unit but on different data. The DSP has two interfaces, the digital application interface (DAI) and the digital peripheral interface (DPI), that are connected to a corresponding signal routing units (SRU) of the processor for flexible routing. The EZ-Board uses the AD1939 codec. It has eight differential digital-to-analog converters and four analog-to-digital converters with a maximum sample rate of 192 kHz and is configured through serial peripheral interface (SPI). Correspondingly, eight RCA (Radio Corporation of America) connectors to measure the analog output and four RCA connectors for the analog input signals are available. Eight general-purpose LEDs and four general-purpose push buttons are also present on the EZ-Board.

Processing Framework

The signals from belt microphones are fed in analog form to the SHARC ADSP-21489 via the RCA inputs. The on-board codec AD1939 samples them at 48 kHz into digital domain. The Serial Ports namely SPORT0 and SPORT1 connected to the DAI port of ADSP-21489 are config-

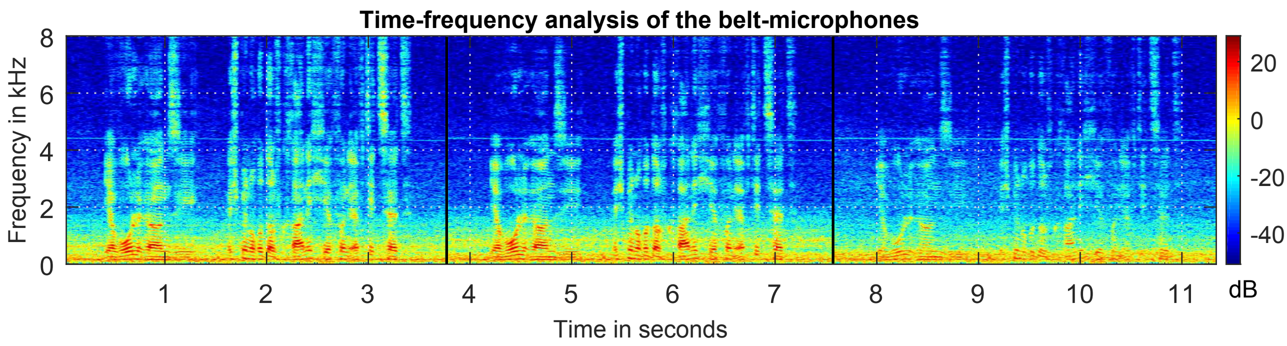

Figure 2: Time-frequency analysis of the bottom (left), middle (center) and top (right) microphones on the belt.

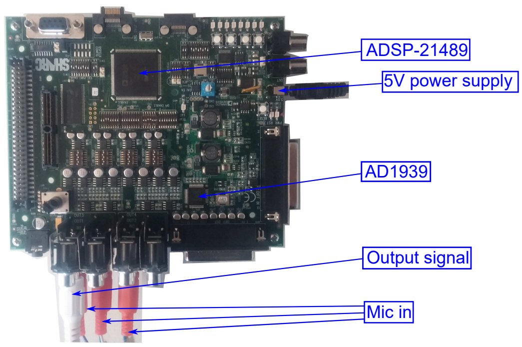

Figure 3: Evaluation Board ADSP-21489 EZ-Kit Lite with seat belt.

ured in multi-channel mode to transmit three channels and receive one channel respectively as a serial Time Division Multiplexed (TDM) stream. For efficient transmission of audio samples between the SPORTs and the processor, dedicated Direct Memory Access (DMA) is used. This operation offloads the processor for other critical signal processing tasks. An external signal block of length $r_{\mathrm{ext}} = 480$ samples at $f_{\mathrm{s,ext}} = 48$ kHz is used. The signals are downsampled by a factor 3 resulting in a sampling rate of $f_{\mathrm{s,int}} = 16$ kHz and internal block length $r_{\mathrm{int}} = 160$ samples. These are transformed into subbands by a Short-Term Fourier Transform (STFT) [2] also referred as the analysis filterbank. At this end, an FFT order of $N_{\mathrm{FFT}} = 512$ is chosen and the samples are weighted with a Hann window. The output of the analysis filterbank contains the spectra of the $M = 3$ microphone signals[1] which is fed as input to the beamformer-postfilter system. The output spectra from the system is transformed into time domain by the synthesis filterbank. It is then upsampled by factor 3 and sent to the codec via SPORT which converts it back into analog domain and routs it to the RCA outputs.

Beamformer-Postfilter System

The realized beamformer-postfilter system for belt microphones consists of an adaptive GSC-type beamformer [5] and a subsequent postfilter [7]. In the following we will give a short overview about all components that are necessary to perform a robust beamforming approach with seat belt microphones.

Source localization and signal selection: A priori knowledge about the direction of the beamformer is nearly impossible due to the flexible array geometry of the belt microphones. Before combining all microphones with adaptive beamforming, the moving signal source (passenger's mouth) has to be localized effectively. Source localization is realized based on the time-delay estimate between signals received at adjacent belt microphones. The best microphone channel used as the reference for the delay alignment is chosen depending on the smoothed time-delay estimate and on the highest long-term broadband SNR [7, 3]. *Delay alignment and equalization:* The distance between belt microphones is large and therefore a highly varying signal power exists at different microphones. In order to combine all microphones in a proper way the noise power in every microphone channel has to be corrected appropriately to the same level. For equalization a slowly varying correction factor per microphone using a multiplicative increment and decrement mechanism has been implemented to correct the belt microphones based on the average PSD of the noise.

Adaptive beamforming: The *first beamformer* in Fig. 4 is a modified filter-and-sum beamformer. It filters the equalized and time-aligned belt microphone signals in the subband domain based on a smoothed version of the normalized SNRs [9]. Due to the strong microphone SNR variations an NLMS-adaptive subband realization of an adaptive blocking matrix (ABM) as proposed in was implemented. By subtracting the filtered versions of the first beamformer output by the ABM from the equalized and time-aligned microphone signals a noise reference for the interference canceller (IC) is generated. Thus, by using ABM the SNR differences between belt microphones and the mismatch of the steering direction can be corrected. The IC adaptively removes signal components that are correlated to the interference input signals from the first beamformer output using the NLMS algorithm for filter adaptation. Whereas the control of step-sizes for the ABM and IC can be found in [9]. The adaptive beamformer output is determined by subtracting the magnitudes at the IC output from the first beamformer output.

Postfilter: In order to attenuate undesired signal components at the beamformer ouput while keeping the desired

[1]Since the input signals are assumed to be real, it is sufficient to store only the first frequency supporting points.

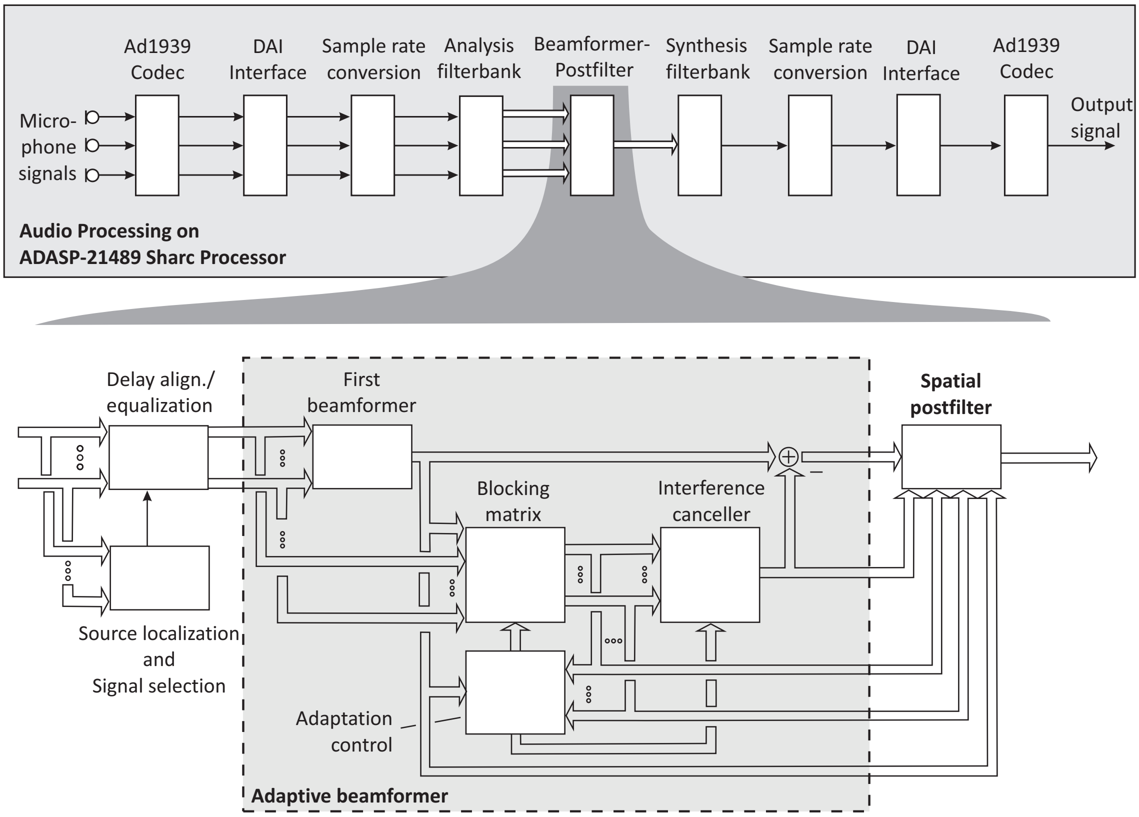

Figure 4: Audio processing on ADASP-21489 and overview of the proposed beamformer-postfilter system.

speech signal as natural as possible a spatial postfilter is utilized. Different methods based on spatial correlation or signal absence probability can be used. In this contribution a modified Wiener filter in the subband domain as presented in [7] has been implemented. The PSD of the non-stationary interference noise has been estimated based on the smoothed short-term powers at the adaptive beamformer output and at the IC output [7]. Whereas the PSD of the stationary noise at the output of the adaptive beamformer is determined analog to [4].

Implementation Aspects

The integrated development environment (IDE) used is Visual DSP++ from Analog Devices. In order to gain a higher performance of the beamformer, the algorithm C-functions of the run time library are utilized. Therefore the FFT and IFFT in the analysis and synthesis filterbank are substituded by an optimized function that uses the radix-2 algorithm. Additionally optimized C-functions to compute complex products as well as addition and subtraction between complex variables and vectors were utilized [11]. Also complex copy functions are incorporated. In order to efficiently compute the absolute value of a complex number, the absolute value of real and imaginary part are added instead of building the square root of real part plus imaginary part in brackets squared [8]. Instead of short-term powers often short-term magnitudes are implemented (to control e.g. the step-size for filter adaptation) to save computational cost.

Since the default segmentation of RAM on the ADSP-21489 was not sufficient to meet the memory requirements of the complete algorithm, the default address ranges were modified to allocate an additional heap segment. The default segments of the internal SRAM are *rth* (1.5 kB) located in Block 0, *intcode* (190.5 kB) located at Block 0, *dmda* (160 kB) located at Block 1, *stak* (32 kB) located at Block 1, *pdma* (128 kB) located at Block 2 and *heap* (128 kB) located at Block 3. Part *rth* includes the Interrupt table and the system initialization code. The *intcode* segment includes library code which adapts interrupt latch registers and the instruction code. The *dmda* segment contains global and static variables as well as string literals. The *stak* contains the run-time stack memory. The *pdma* segment holds program code, global and static data that are stored in the program memory and *heap* contains the run-time heap memory [11]. A new heap segment *heas* with the size of 128 kB was introduced at Block 1 whereas segment *dmda* was reduced with equal amount from 160 kB to 32 kB. After this modification, the heap memory has a overall size of 256 kB and which is equally allocated on Block 1 and Block 3.

In Visual DSP++ also SIMD mode instead of SISD (single input single data) mode was selected to minimize processing overhead.

Performance Measures

In order to determine the efficiency of the various algorithmic sub-modules, the required MIPS (Million Instructions Per Second) were measured. Since SIMD mode on the ADSP-21489 was utilized, the maximal count of

MIPS amounts up to 800 [1]. The MIPS of each sub-module was determined by equation 1. The variable m describes the computed MIPS of the algorithmic sub-module. v_{DSP} describes the maximal MIPS on offer by the processor core. d_{dur} describes the processing time in seconds and l_{in} describes the length of the input signal in seconds. In this case, the input signal length is 60 s.

$$m = \frac{v_{\mathrm{DSP}} \cdot d_{\mathrm{dur}}}{l_{\mathrm{in}}} \tag{1}$$

Table 1 shows the the result after optimization has been carried out. Furthermore, the required heap memory for

Part of the algorithm	Required MIPS
Source localization and selection	11
Delay align./ equalization	23
First beamformer	11
Adaptive blocking matrix	91
Interference canceller	57
Adaption control	31
Spatial postfilter	12

Table 1: Required MIPS of the several algorithm parts.

each sub-module is shown in Fig. 5. The overall count of the required heap memory is 119 kB whereby the required amount of stack memory counts up to 18 kB. A real world

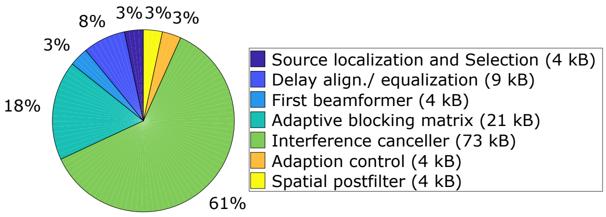

Figure 5: Pie chart of the required heap memory.

recordings were generated in the car at speed of about 130 k.p.h. on highways with the driver's seat belt to compare the performance of the implemented beamformer-postfilter system with the single best belt microphone on the belt. For the comparison the frequency based SNR and SIR averaged over all frequencies were used. SIR denotes the power ratio of signals coming from steering direction to undesired signals coming from other directions. As shown in Fig. 6 the non-adaptive first beamformer already exhibits a slightly better average SNR and SIR of about $2 - 4\,\mathrm{dB}$ to that of the best microphone.

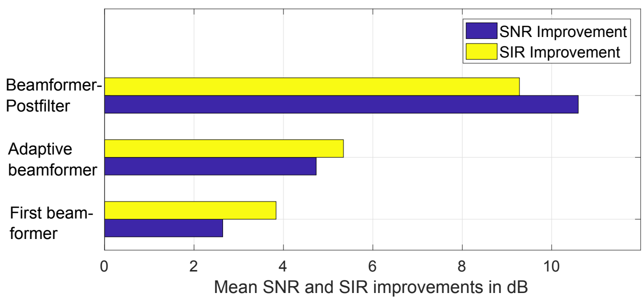

Figure 6: Average SNR and SIR improvements w.r.t. best belt microphone.

Using adaptive beamformer without a spatial postfilter an average SNR and SIR increase of more than 5 dB is achievable. Adding a postfilter to the adaptive beamformer an additional SNR improvement of more than 5 dB and an SIR improvement of about 4 dB on average can be obtained compared to the adaptive beamformer.

Conclusion and Future Scope

Real time implementation of beamformer-postfilter system on an embedded platform has been presented in this article. For this purpose, a floating point audio DSP (ADSP-21489) from Analog Devices has been chosen. The various modules of the algorithm are revisited and optimized in various aspects to suit the processor's architecture. Performance measures of the algorithm in terms of speed and memory have been evaluated and highlighted. The overall mean improvement achieved by the algorithm in terms of SNR and SIR to the best belt microphone is also presented at the end. In future, this algorithm can be further optimized to concurrently process signals belt microphones from other seats in the car. Furthermore, the proposed system can be integrated in a test car and its output can be fed to subsequent speech enhancement algorithms used for in-car communication or handsfree telephony. The overall performance can then be further tested and evaluated for different driving scenarios.

References

[1] ADSP-214xx SHARC Processor Hardware Reference, Revision 1.1, p.365, April 2013.

[2] J. B. Allen, "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform", Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 25, no. 3, pp. 235, 238, Jun 1977.

[3] G. C. Carter, A. H. Nuttall, P. Cable, "The Smoothed Coherence Transform", Proceedings of the IEEE , vol.61, no.10, pp.1497,1498, Oct. 1973.

[4] Y. Ephraim and D. Malah, "Speech Enhancement using a MMSE Short-Time Spectral Amplitude Estimator", IEEE Trans. on ASSP, vol. 32, no. 6, pp. 1109-1121, 1984.

[5] C. W. Jim, "A Comparison of Two LMS Constrained Optimal Array Structures", Proc. of the IEEE, vol. 65, no. 12, pp.1730, 1731, 1977.

[6] C. Knapp, G.C. Carter, "The Generalized Correlation Method for Estimation of Time Delay", Acoustics, Speech and Signal Processing, IEEE Transactions on , vol.24, no.4, pp.320,327, Aug 1976.

[7] M. Krini, Z. Mirza, K. Rodemer, "A Practical Beamformer-Postfilter System for Microphone Arrays on Seat Belts", 12. ITG Symposium Speech Communication (ITG'16), Paderborn, Germany, 2016.

[8] R. G. Lyons, "Unterstanding Digital Signal Processing", Addison Wesley, 1997.

[9] V. K. Rajan, M. Krini, K. Rodemer, G. Schmidt: "Signal Processing Techniques for Seat Belt Microphone Arrays", EURASIP Journal on Advances in Signal Processing, vol. 92, March 2016.

[10] paragon AG, URL: `http://www.paragon.ag/`.

[11] Visual DSP++ 5.0 Run-Time Library Manual for SHARC Processors, pp. 515,525, Revision 1.5, January 2011.