Filter-Based Model of Multimicrophone Array in an Adverse Acoustic Environment

Rafał Samborski and Mariusz Ziółko

Abstract—**In this paper filer-based model of sound propagation combined with a model of multimicrophone array is presented. Simple operations employed in the model such as sum and linear convolution allows one to implement it easily using numerical methods. Additionally, described solution was discussed in connection with filter-based model of multimicrophone array. The usefulness of the models was proved in the context of filter-and-sum beamformer synthesis. Practical methods of calculating filters for 1-D multi-microphone arrays were stated. Then the results were employed to synthesis filters for wideband four-microphone array.**

Index Terms—**multimicrophone array, multipath propagation, speech enhancement, beamforming.**

I. Introduction

WITH the rapid development of Automatic Speech Recognition (ASR) systems it was necessary to create sophisticated speech enhancement systems [1], [2]. Nowadays, it is possible for computers or mobile devices to recognize speech effectively in extremely difficult acoustic conditions including traffic jam, conference hall, car interior. Speech enhancement can be based on single-channel (e.g. spectral subtraction) or multi-channel techniques such as adaptive noise cancellation [3], blind source separation or beamforming. Despite quite impressive performance presented by all these methods we are still far away from the efficiency with of human in terms of speech recognition. The reason of that is noise. It can be assumed that the effectiveness of the recognition ratio of the single phonemes decreases by 5%-10% with every 3 dB of SNR [1]. That means that the difference of SNR which is just noticeable for human (3 dB) is still significant for a computer system. This is the reason for the development of advanced systems of the speech enhancement. All of the techniques mentioned above base on some model of sound propagation. The choice of the proper mathematical model is extremely important especially in the adverse acoustic environment. The basic models come from general wave equation. The model presented in this paper bases on experimentally calculated Room Impulse Response (RIR) and it is a combination with filter-based model of multimicrophone array. The paper is divided as follows. Section 2 covers some necessary theory on sound propagation and methods of finding the relation between sound emitted by the source and wave reaching the receiver. Section 3 presents model of multimicrophone array in acoustic environment. The practical method of designing the beamformer using described method is provided in Section 4. The paper is summed up with the conclusions.

R. Samborski and M. Ziółko are with AGH University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, Department of Electronics, al. Mickiewicza 30, 30-059 Krakow, Poland, e-mail: sambo@agh.edu.pl.

II. Sound propagation

Sound is a mechanical wave traveling from one location to another through a medium such as air or water. As a wave, sound have certain properties such as refraction, attenuation, interference, reflection and diffraction. The speed c at which sound propagates through the material depends on physical properties of the medium such as density and elasticity. In this article the air will be default medium. However, described model can be used for any type of material.

$$\nabla^2 u(x,t) = \frac{1}{c^2}\frac{\partial u(x,t)}{\partial t^2} \tag{1}$$

where $u(x;t)$ is a potential state in the point x in the moment t.

To find the relation between sound wave leaving the source and one reaching the receiver is to find the solution of the equation (1). From the computational point of view, wave equation can be solved using finite element methods (FEM) by expressing as a linear equations in a discrete domain [4].

Employing equations which can describe the effect of diffraction brings about reasonable accuracy. In most cases only the generation of the second harmonic wave is considered, with the higher order harmonics being neglected. The acoustic analysis available in FEM software, is able to modeling the fluid medium and the surrounding structure, studying the pressure distribution in the fluid at different frequencies, pressure gradient, particle velocity, the sound pressure level, as well as, scattering, diffraction, transmission, radiation, attenuation, and dispersion of acoustic waves.

Despite the continuous growth of computing resources available to engineers, FEM performs well only at low frequencies and in not sophisticated environments. The compute time increases dramatically with frequency and with every object appearing in the sound scene. The difficulty of this challenge increases when it comes to multipath propagation, scattering and interference of sound waves etc. In most of the indoor situations the analytical calculation of the sound waveform in the point of interest, even approximately, becomes impossible.

A different way to determine the relation between the source and the sound in the point of interest (e.g. next to the microphone) in the indoor acoustic scenes is to determine RIR. It is an analogy of the impulse response well known from linear signal systems. The information about propagation of impulse can be used to derive information about propagation of any arbitrary signal [5].

Let $s'(t)$ be the detection of the signal $s(t)$ emitted by the source, $\delta(t)$ be the impulse signal and $r(t)$ be the RIR. The process of propagation can be described as a convolution of emitted signal and some function $d(t)$.

(Advance online publication: 21 November 2012)

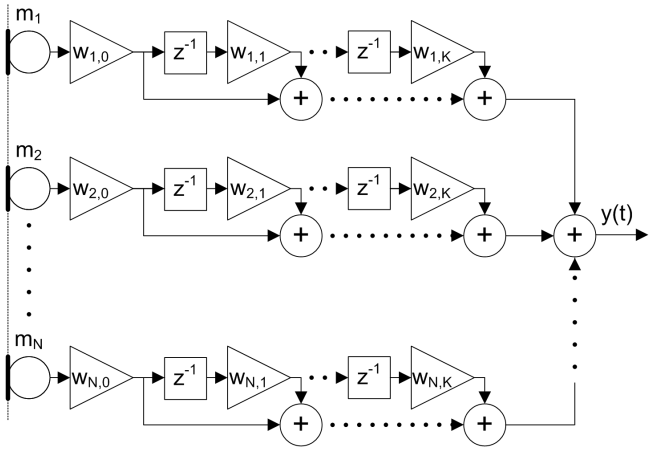

Fig. 1. A general structure of filter-and-sum beamformer.

$$s'(t) = d(t) * s(t) \tag{2}$$

$$s'(t) = \delta(t) * d(t) * s(t) \tag{3}$$

$$\begin{aligned} s'(t) &= s(t) * (d(t) * \delta(t)) \\ s'(t) &= s(t) * r(t) \end{aligned} \tag{4}$$

The experimental methods of RIR measurements include employing impulsive sources (such as balloons or blank pistol) or generators of Maximum Length Sequence (MLS). MLS is a binary signal obtained with suitable shift-register, which contain equal amounts of energy for all frequencies. Most important attribute of MLS is the fact that its autocorrelation is the Dirac delta function. Recording containing MLS can be emitted by dodecahedron loudspeaker and traveling to the microphone. Afterwards, RIR can be calculated by computing the cross-correaltion between MLS and its recording. This method was used in our experiments.

III. Filter-based model of multimicrophone array

Let us consider the general structure of filter-and-sum device [6] as it is presented in Fig. 1. This structure is a generalization of the delay-and-sum beamformer which operates properly only for narrowband signals. Since the wideband signal (such as speech) is composed of infinite number of narrowband components, the weights w should differ for every such component. This can be achieved by employing subband beamforming or by employing filters instead of scalar weights [6]. The output is described by (5).

$$y(t) = \sum_{n=1}^{N} \sum_{k=0}^{K} w_{n,k} x_n(t - kT_s) \tag{5}$$

where T_s is sampling period, K is order of the FIR filters in beamformer and $w_{n,k}$ are filter coefficients for the nth microphone.

To combine the model from (5) with RIR let us analyze the situation presented in Fig. 2. When the source is in far field we can assume that RIR for every of sensors is the same. Then sound coming to the microphone can be considered as a plane wave reaching every microphone with some delay. For uniformly spaced array with spacing d the output can be expressed as

$$y(t) = \sum_{n=1}^{N} \left((s(t) * r(t))\, e^{-j\omega\tau_n} \right) * h_n(t) \tag{6}$$

where $\tau_n = nd \sin\theta c^{-1}$ is the propagation time between the first and nth sensor with the sound velocity c.

Finally the response of RIR-beamformer system can be expressed as

$$P(f,\theta) = \sum_{n=1}^{N} R(f) H_n(f) e^{-2\pi f j n d \sin\theta c^{-1}} \tag{7}$$

where $R(f)$ and $H_n(f)$ are Fourier transforms of $r(t)$ and $h_n(t)$ respectively.

IV. Experiment and results

To design beamformer is to choose the appropriate coefficients $w_{n,k}$ in such a way that the output signal $y(t)$ match the input signal $s(t)$ as closely as possible. There are several different methods of adjusting such a system. One of them is to employ adaptive filtration techniques. This usually leads to linearly constrained minimum variance beamformer or to reference signal based adaptive beamformer. These classes of devices can adapt themselves when data statistics keeps changing. In this paper, however, we will assume that the data statistics are time invariant. Then the weight coefficients can be fixed during operating of the beamformer. This may be the cause for SNR of the output signal being lower than in the adaptive system. On the other hand, such a beamformer has lower computational complexity and can be implemented in less sophisticated hardware.

We begin with the minimization of an objective function $\Phi(w)$ given as a sum of the squares of the difference (error) between $P_d(f,\theta)$ and the designed response $P(f,\theta)$ over the ranges of interests (f_d and θ_d):

$$\Phi(w) = \int_{f_d} \int_{\theta_d} \left[P_d(f,\theta) - P(f,\theta) \right]^2 \mathrm{d}f \mathrm{d}\theta \tag{8}$$

Since $\Phi(w)$ is nonlinear objective function in $N(K+1)$ dimensional space, one of the commonly used nonlinear optimization techniques was employed. We employed the Nelder-Mead method (unconstrained nonlinear optimization) [7] to optimize the coefficients of the filters.

In the presented experiment we assumed the $\theta = 0$ (the source was placed directly in front of beamformer). Having RIR calculated, the iterative optimization was conducted to build the beamformer with wide-band characteristic and maximally possible damping in all directions but θ. The calculated filter coefficients were placed in real hardware and the measurements were conducted. The results are shown in the Fig. 3.

V. Conclusions

The model described above combines filter-based model of sound propagation determined by the RIR with general structure of filter-and-sum beamformer. Presented model was successfully employed to develop filters for four microphone 1-D beamformer (Fig. 4 in given acoustic scene. The results were satisfactory, especially when it comes to side source

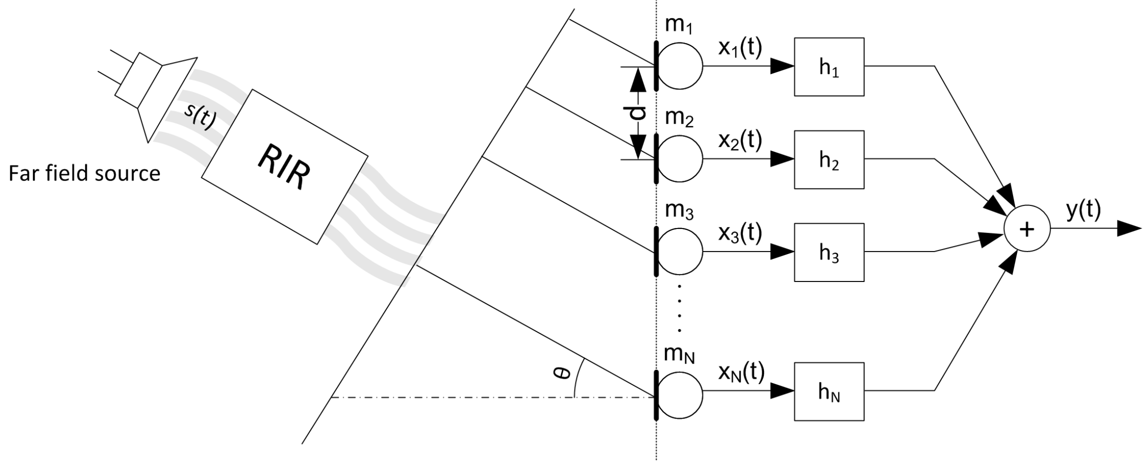

Fig. 2. 1-D beamformer placed in acoustic environment with far field source modified by RIR.

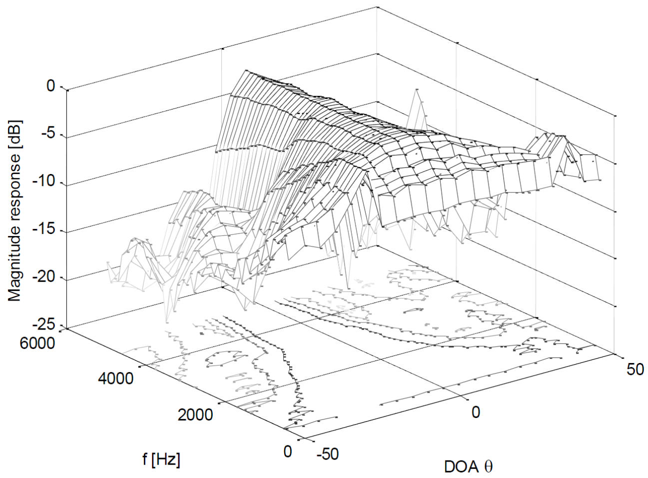

Fig. 3. Results of the iterative optimization of beamformer in adverse acoustic environment having the desired system response.

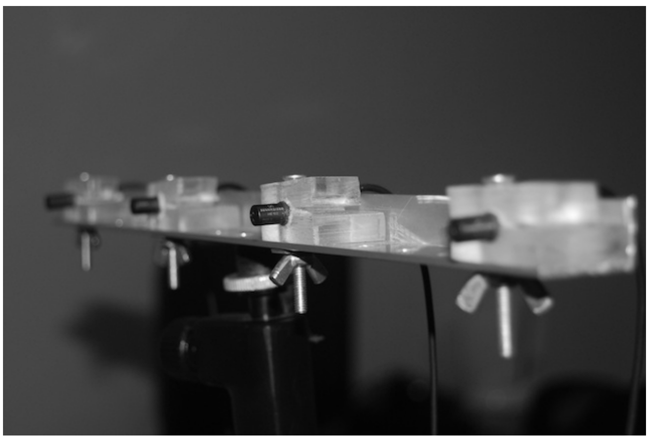

Fig. 4. The four-microphone, 1-D beamformer during the experiments. The devices is a part of Automatic Speech Recognition system.

cancellation (small side-lobes), especially when compares with naive delay-and-sum system. This result is extremely important as the developed device is a part of the Automatic Speech Recognition system. The further experiments are conducted to increase the SNR ratio and efficiency of speech recognition using the 2-D, 16-microphone device.

REFERENCES

[1] B. Kollmeier and M. Wesselkamp, "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," *The Journal of the Acoustical Society of America*, vol. 102, pp. 2412–2421, 1997.

[2] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons, 2006.

[3] R. Samborski, M. Ziółko, B. Ziółko, and J. Gałka, "Wiener filtration for speech extraction from the intentionally corrupted signals," *2010 IEEE International Symposium on Industrial Electronics*, p. 1698 1701, 2010.

[4] P. Filippi, D. Habault, J. Lefevre, and A. Bergassoli, *Acoustics, Basic Physics, Theory and Methods*. Academic Press, 1999.

[5] C. Bourget and T. Aboulnasr, "Inverse filtering of room impulse response for binaural recording playback through loudspeakers," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1994.

[6] W. Liu and S. Weiss, *Wideband Beamforming. Concept and Techniques.* John Wiley & Sons, 2010.

[7] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.