

Computational Approach to Spatial Filtering

Andrzej Czyzewski, Jacek Lasecki, Bozena Kostek
Technical University of Gdansk
Sound Engineering Department,
Narutowicza 11/12, 80-952 Gdańsk, PL
& Institute of Physiology and Pathology of Hearing, Warsaw, PL
Phone: +48-58-347-23-01, Fax: +48-58-347-11-14
email:{andrzej, jacek, bozenka}@sound.eti.pg.gda.pl

ABSTRACT: Hearing impaired persons have difficulty in understanding speech in cocktail-party conditions. Spatial filtering may be very helpful for such people. This feature should be applied to the hearing aid, thus the computational complexity of spatial filtering-based algorithm must allow real time implementation. In order to meet this assumption some investigations were made and neural network-based algorithm was proposed. This algorithm is presented in this paper.

KEYWORDS: beamforming, spatial filtering, neural networks

INTRODUCTION

Spatial filtering of sounds is very important in many domains related to audio communications, especially in applications for hearing impaired people. Speech signals arriving from various directions not only interfere with the target signal but often obscure it. For the purpose of spatial filtration various algorithms were conceived. In the literature one may find many papers related to this subject (Griffiths & Jim, 1982) (Widrow & Stearns, 1985). Since the results obtained so far are not satisfactory yet, thus there is a need to improve already existed algorithms or to find a different approach to this problem. In this paper a system for spatial filter was proposed. It consists of a neural network used as the decision module. The main assumptions of this system are described, and additionally, details concerning computational complexity problem are discussed in the paper, since this system may be easily implemented on a signal processor.

SYSTEM ASSUMPTIONS

Bearing in mind that the proposed algorithm is aimed at being used in hearing prostheses, some assumptions should be made. First of all, the number of microphone channels should be limited to two. Since the microphones are placed close to eardrums so the distance between them equals approximately 17cm. Secondly, the main task to be performed by the proposed beamformer is to estimate signals coming from forward direction by attenuating signals coming from lateral directions. This feature can be represented by the following expressions (Lasecki, Kostek, Czyzewski, 1998):

$$l(t) = s(t) + n_l(t) \quad (1)$$

$$r(t) = s(t) + n_r(t) \quad (2)$$

where:

- $l(t)$ – signal from left channel,
- $r(t)$ – signal from right channel,
- $s(t)$ – desired signal from forward direction,
- $n_l(t), n_r(t)$ – noise coming from lateral directions

It is neither desirable nor possible to completely attenuate signals from lateral and backward directions. Therefore it is assumed that spectral components may represent signal coming from forward direction or other directions. It is also assumed that if a spectral component represents signals coming from not wanted directions, it should be attenuated by approximately 40 dB. If a spectral component represents the desired signal, it should be preserved in the output of the

algorithm. Authors decided to use the artificial neural network (ANN) as a learning signal processing algorithm making a decision whether spectral components should be attenuated. The signals after the attenuation of noisy components are described by the following expressions:

$$\tilde{L}(e^{j\omega}) = \sum_{i=1}^N g(i)L^i(e^{j\omega}) \quad (3)$$

$$\tilde{R}(e^{j\omega}) = \sum_{i=1}^N g(i)R^i(e^{j\omega}) \quad (4)$$

$$g(i) = \begin{cases} 1; i \in \text{Signal components} \\ 0.01; i \in \text{Noise components} \end{cases} \quad (5)$$

where:

- $g(i)$ – attenuation coefficient of noisy components (40dB),
- i – spectral component index,
- $\tilde{L}(e^{j\omega})$ – estimate of signal component in the left channel,
- $\tilde{R}(e^{j\omega})$ – estimate of signal component in the right channel.

There are a few possibilities to find out which spectral components are noisy and which represent signal. Classically, for that purpose adaptive filtering methods are used (Griffiths & Jim, 1982) (Widrow & Stearns, 1985). After some preliminary experiments authors decided to use artificial neural network (ANN) as a decision module. To reduce the amount of information contained in the signal, some parameters of sound are calculated. Therefore the input vector fed to the ANN consists of the following parameters.

First parameter, which expresses the interaural spectral magnitude ratio, is described by the following expression:

$$M^i = \frac{\min(|L^i|, |R^i|)}{\max(|L^i|, |R^i|)} \quad (6)$$

where:

- i – frequency component index,
- L^i, R^i – left and right signals for i -th frequency component,
- M^i – magnitude ratio for i -th frequency component.

The second parameter describing the interaural difference in phase, is expressed by the following term:

$$A^i = \left| \angle L^i - \angle R^i \right| \quad (7)$$

where:

- \angle – denotes the signal phase
- A^i – phase difference of i -th frequency component for left and right channels,
- The third parameter D^i , namely a relative ratio of i -th frequency component for left and right channels used in the learning phase is computed according to the following formula:

$$D^i = \frac{|L^i - R^i|}{|L^i| + |R^i|} \quad (8)$$

The main steps of the proposed algorithm are:

1. Calculate FFT of signals of both channels
2. For each spectral component use NN to decide whether this spectral component represents signal from lateral/backward directions or is coming from forward direction
3. Attenuate spectral components which represent signals coming from lateral or backward directions
4. Perform IFFT for signals of both channels

NEURAL NETWORK-BASED DECISION MODULE

In the algorithm presented above spectral components are classified by neural networks. Assuming that the system is deterministic, it is possible to find some hard-defined rules in order to classify spectral components according to the

directions of signal arrival. Instead of this authors decided to use NNs, especially because of their property to learn from real signal examples. This learning approach allows to represent some specific phenomena resulting from the propagation of acoustic waves in the natural acoustic environment, in which the sound wave can be reflected, its direction of propagation may be changed by some complex diffraction schemes and the spectra of signals arriving to eardrums may be modified by acoustic interference. These phenomena are quite difficult to model mathematically, thus the neural network was used as a tool for the modeling of multidimensional relations between signals, direction of their arrival and weighting factors needed to form the desired spatial characteristics of the microphone array.

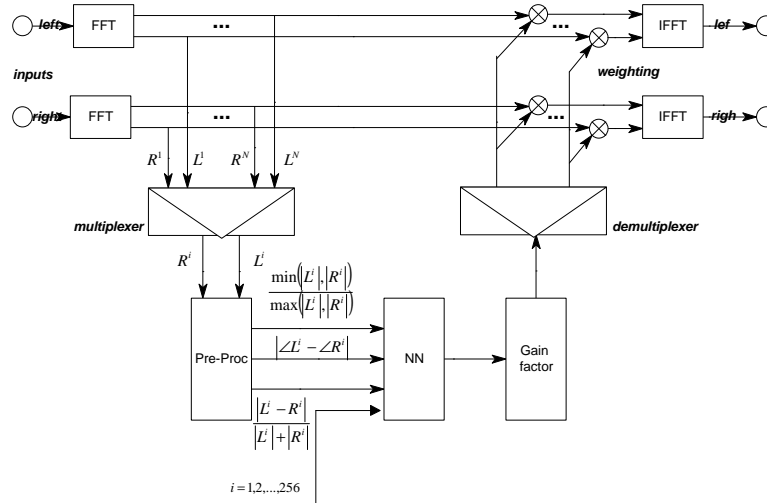


Figure 1. Structure of NN-based spatial filtering algorithm. The signal is buffered in FFT matrices.

There are two ways of realizing step 2 of the algorithm. The first, which employs as many NNs as is the number of spectral components. In this case each NN is responsible for the correct classification of a single spectral component. In the second approach only one NN is employed. In this case the NN is switched between spectral components and is used to classify all spectral components successively. The second solution proved to be equally efficient as the first one, and additionally it demanded much less computational power.

TRAINING PHASE

Training phase was divided into two phases. During first phase many networks were trained and the one with the lowest error was chosen. During second phase only one network was trained. The main reason for this approach was to find out which structure of the ANN will best suit decision tasks. Parameters of the learning process were as follows:

- $\mathbf{a} = 0.75$ - the learning rate,
- $\mathbf{b} = 0.02$ - the momentum coefficient.

In Table I structures of all networks used in first phase are presented. Additionally, the convergence of the learning process is shown in Figure 1.

Table I Network structures used in the learning phase

Net Name/Neurons	Net 1	Net 2	Net 3	Net 4	Net 5	Net 6	Net 7	Net 8
Hidden layers	1	1	2	2	3	3	4	4
Inputs	3	3	3	3	3	3	3	3
Hidden layer 1	5	9	5	9	5	9	9	11
Hidden layer 2			3	5	4	5	7	9
Hidden layer 3					3	3	5	7
Hidden layer 3							3	5
Output	1	1	1	1	1	1	1	1
Iterations	5000	5000	5000	5000	5000	5000	5000	5000
Error	0.993	0.15	0.283	0.234	0.563	0.857	1.257	0.932

The structure of the neural network No. 2 was converging most effectively, so it was chosen to be used in the second phase. This network was trained until the error value $e = 0.01$ was reached. The structure of this network along with the number of iterations needed to achieve the error value lower than 0.01 is presented in Table II.

Table II The structure of the chosen neural network

Layer	Number of neurons
Input	3
1 hidden	9
Output	1
Iterations	32456
Error	0.0098

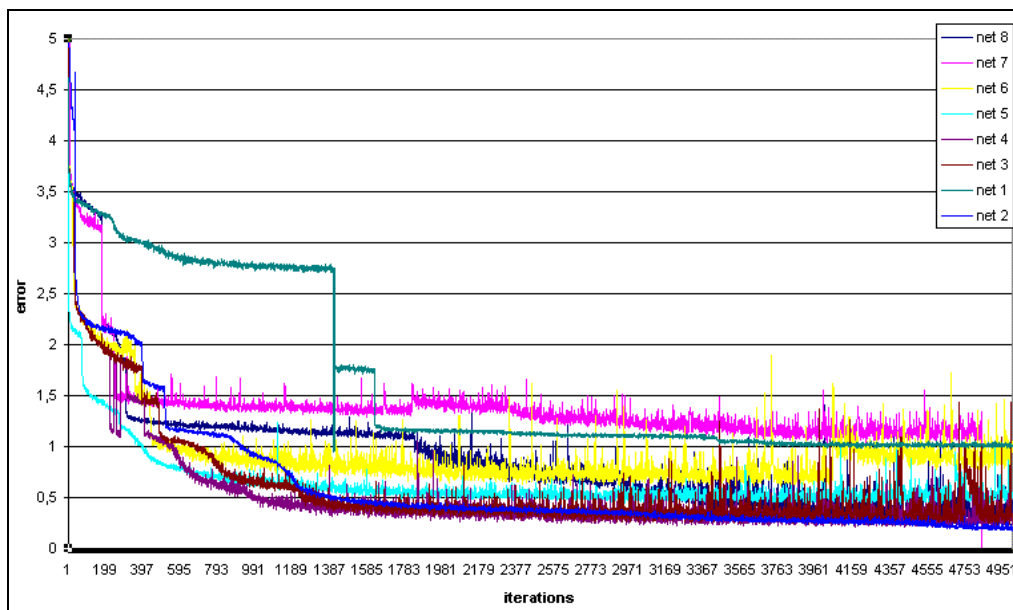


Figure 1: Convergence of the learning process

Additionally, in Fig. 2 both desired directional characteristic (the same for all frequencies) and spatial characteristics of NN-based filtration algorithm obtained with a multi-tone signal are presented.

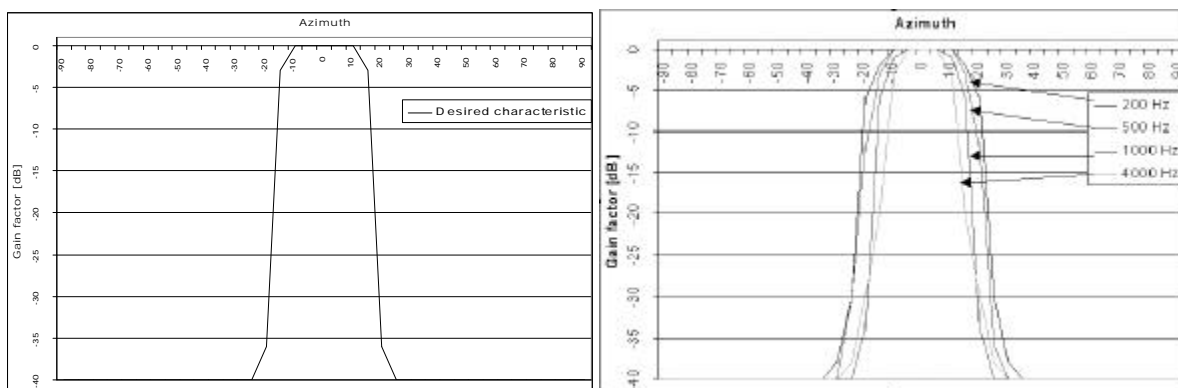


Figure 2: Desired directional characteristic (the same for all frequencies) and spatial characteristics of NN-based filtration algorithm obtained with a multi-tone signal

TEST PHASE

In experiments signal recordings made in MIT Media Laboratory were used (Gardner & Martin, 1994). The NN structure for which the smallest error was achieved was further tested during this phase. As is seen from Fig. 3 and in Fig. 4 the proposed algorithm causes small distortions, which are noticeable only when the signal is coming from directions other than forward. Fortunately, in this case the signal is strongly attenuated, so the distortions are not influencing substantially the overall quality of audio signal.

It can be noticed that definitions of parameters (6), (7) and (8) indicate that signals of the same spectrum composition coming from concurrent directions may not be effectively filtered by this beamformer algorithm. This most important drawback of the proposed method of spatial filtering occurring for low frequency components of similar signals coming from different directions is observed because phase shift is also small for this frequency range.

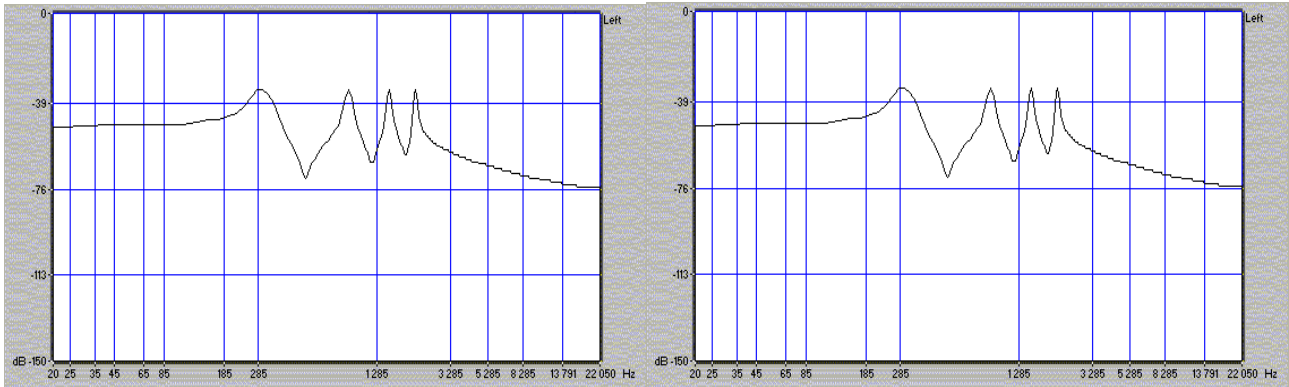


Figure 3: Spectral representation of multi-tone signal coming from azimuth 0^0 before and after processing

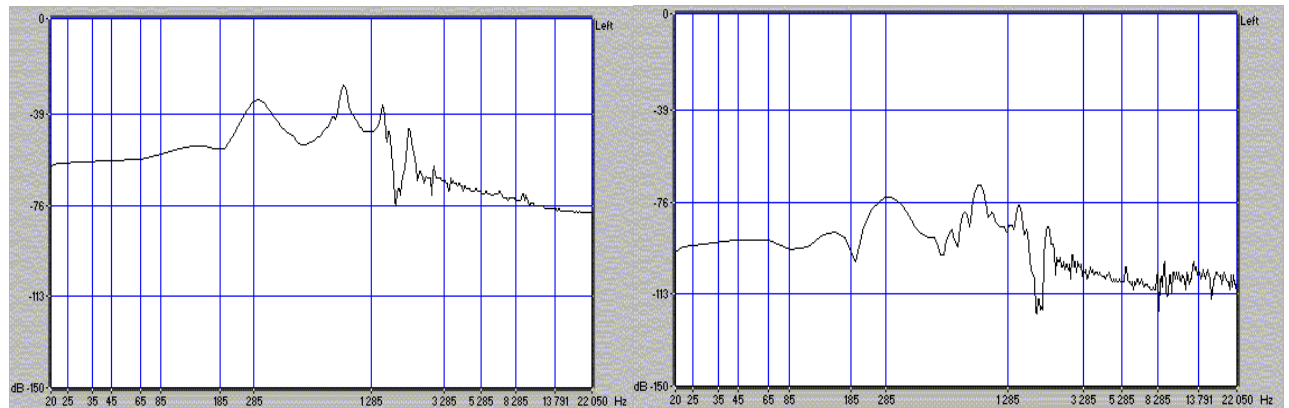


Figure 4: Spectral representation of multi-tone signal coming from azimuth 60^0 before and after processing

COMPUTATIONAL COMPLEXITY

In order to check whether proposed algorithm may be used in real time applications, the computational power needed to perform the algorithm was calculated (Hertz, Krogh, Palmer, 1991). It is obvious that ANN demand large computational power in the learning phase. During test phase an ANN with a small number of neurons can perform very fast. The computational complexity of the elaborated algorithm can be computed according to the following formula:

$$o(alg) = fps * (4 * o(FFT) + N/2 * (n * o(fu) + 2 * w)) \quad (9)$$

where:

- $o(alg)$ – computational complexity of the algorithm,
- fps – frames per second,

- $o(FFT)$ – computational complexity of the FFT algorithm,
- N – size of the FFT frame,
- n – number of neurons,
- $o(fu)$ – computational complexity of the unipolar transfer function,
- w – number of weights,

The following assumptions can be made:

- sampling rate – **44100** Hz,
- FFT size – **512** samples,
- overlap size – **256** samples,

Therefore:

$$N = 512$$

$$o(FFT) = 2 * N * \log_2 N = 9216,$$

$$fps = 44100 / N * 2 @ 173,$$

$$n = 10,$$

$$o(fu) = 20$$

$$w = 3 * 9 + 9 * 1 = 36,$$

$$o(alg) = 173 * (4 * 9216 + 256 * (10 * 20 + 2 * 36)) = \mathbf{18\ 423\ 808}$$

The overall number of multiplications and additions per second equals approximately 20MIPS. Therefore, the proposed algorithm can be implemented for example on the DSP processor Motorola DSP56002 and may run in real time.

SUMMARY

In this paper a system for spatial filtration of sounds based on neural network was described. The main assumptions of the system were also given. The neural network-based algorithm gains sharp directional characteristics in a wide range of frequency during the learning phase, preserving a good quality of speech signals coming from the forward direction. This demonstrates that neural networks can be implemented as a decision module in the spatial filtering algorithm. This kind of algorithm can be used in hearing aids, cochlear implants and in many other applications. The negative effects limiting the sharpness of directivity of the spatial filter occur only in the case of processing similar low frequency signals coming from concurrent directions. This is caused by the lack of differences in spectral magnitudes of similar sounds and by a too small phase shift of low frequency signals arriving to the microphone array.

ACKNOWLEDGEMENTS

The research was sponsored by the State Committee for Scientific Research, Warsaw, Poland. Grant No. 8 T11E 034 15

REFERENCES

1. Gardner B., Martin K., 1994, *HRTF Measurements of a KEMAR Dummy-Head Microphone*, MIT Media Lab Perceptual Computing – Technical Report #280, pp. 1-7.
2. Griffiths L. J., Jim C. W., 1982, *An alternative approach to linearly constrained adaptive beamforming*, IEEE, vol. AP 30.
3. Hertz J., Krogh A., Palmer R. G., 1991, *Introduction to the theory of neural computation*, Addison-Wesley Publishing Company.
4. Lasecki J., Kostek B., Czyżewski A., 1998, *Microphone Array for Improving Speech Intelligibility*, 20 Tonmeistertagung, Karlsruhe.
5. Widrow B., Stearns S.D., 1985, *Adaptive Signal Processing*, Prentice-Hall.