COMBINED NOISE SUPPRESSION SYSTEM FOR MONAURAL COCHLEAR IMPLANTS

Milan Svoboda, Pavel Sovka, and Petr Pollák

Czech Technical University in Prague,
ČVUT FEL K331, Technická 2, Praha 6, Czech Republic
E-mail: svobodam@feld.cvut.cz, sovka@feld.cvut.cz, pollak@feld.cvut.cz

10.21437/Eurospeech.1999-581

ABSTRACT

This contribution is devoted to the noise suppression system for monoaural SPEAK-type cochlear implants. The system is based on the combination of the superdirective two-microphone array and the noise suppression post-processing, both implemented in the frequency domain. The relative insensitivity of the two-microphone superdirective system to the spectral magnitude differences in both channels enables the two-channel noise suppression post-processing. The new criterion originating from the SPEAK coding strategy is used for the evaluation of the whole system.

1. INTRODUCTION

Noise cancellation systems play the important role in the communication between people. For the hearing impaired people using either the hearing aids or the cochlear implants the noise cancellation is even more needed to ensure the ability of proper hearing. This contribution is devoted to the application of the noise suppression system for the monaural SPEAK-type cochlear implants. The SPEAK strategy uses dividing the input signals into 20 frequency bands, and selecting 6 (in average) bands with the largest speech power. Although this strategy is relatively noise insensitive the high-level noisy environment can cause the erroneous choice of frequency bands due to the changes in the spectral shape of the input signal. At the same time the chosen bands containing speech can be masked by the noise.

The speech pre-processing system using two microphones is suggested for the noise reduction to ensure the proper function of the SPEAK strategy in noisy environments. As mentioned before the selection of the frequency bands is determined by the spectral shape of the input signal. Therefore to recover the original spectral shape it is desirable to suppress the noise without the speech cancellation, and with the minimum of strong residual noises.

2. DESCRIPTION OF THE NOISE SUPPRESSION SYSTEM

The presented system consists of the combination of the microphone array and the noise suppression algorithm using the Wiener-filter (see Fig. 1).

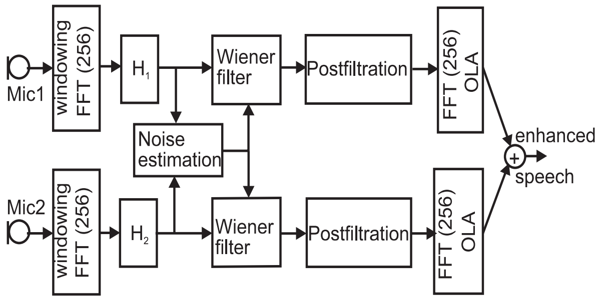

Figure 1: Structure of the proposed noise reduction system

The reason why the microphone array was chosen for the pre-processing is the fact that it achieves the directional selective receiving of the input signal. Therefore the person wearing the cochlear implant with this pre-processing method is able to select the direction from which he (or she) wants to receive sounds.

The proposed system includes the fixed two-microphone end-fire array designed as the superdirective array. This array offers a good directionality even for small number of microphones in contrast with the conventional delay-and-sum arrays. Taking into account the comfort of wearing and the low price the small number of microphones is important in hearing-aid speech pre-processing. The configuration of the microphones in our system can be seen from Fig. 2. They are configurated as the end-fire array with the distance between the microphones $d\_mic = 6.17\ cm$.

Although the directivity of the superdirective microphone array is relatively high the noise suppression is not sufficient for the cochlear implant pre-processing. That is why the combination of the microphone array with the further noise reduction post-processing is proposed.

There are two possible ways how to combine the microphone array with the noise reduction post-processing:

- to combine both channels of the microphone array and then to use the one-channel noise suppression system,
- to apply the noise suppression post-processing to

ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on Speech Communication and Technology (EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

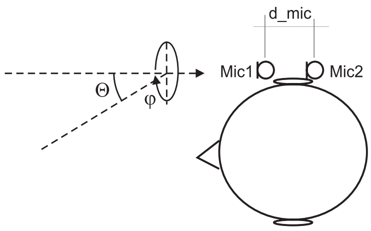

Figure 2: Configuration of the microphone array: end-fire array placed on one side of the head. $d\_mic = 6.17\ cm$

both channels of the microphone array and then to combine their outputs.

Both solutions have advantages and disadvantages. The performance of the microphone array is sensitive to the differences of the spectral magnitudes and the phases between both channels. If the two-channel noise suppression post-processing is used the signals in both channels are treated independently. This post-processing may cause the differences between the frequency responses of the channels. As the result the deterioration of the directivity properties of the microphone array can be expected.

Analysis and experiments given in [2] show that the 10 % differences in magnitude spectra and 0.03π differences in phases of the two-microphone array cause only low deterioration of the properties of the array.

We measured the impact of the differences between the magnitude spectra on the directivity index $D(f)$:

$$D(f) = \frac{|G(f,\Theta_0)|^2}{\frac{1}{4\pi}\int\limits_0^{2\pi}\int\limits_0^{\pi}|G(f,\Theta)|^2 \sin\Theta \, d\Theta \, d\varphi}, . \quad (1)$$

where $G(f,\Theta)$ is the transfer function of the microphone array for the sounds coming from the angle Θ (see Fig. 2), Θ_0 is the angle representing the straight direction ($\Theta = 0$).

Our simulations showed that 50% differences in the magnitude spectra cause the decreasing of $D(f)$ of about 1 dB (see Fig. 3):

Under this condition it is possible to use the two-channel noise suppression system. The system suggested in [4] was taken as the base for the proposed noise suppression system. Some modifications described below were suggested with the aim to improve this method. The modified system offers the noise suppression with relatively small speech cancellation and very low residual noises.

2.1. Superdirective two-microphone array

In the superdirective array the input signals are filtered by filters with time-invariant impulse responses that are designed with respect to the maximum sensitivity of the array for sounds coming from straight direction Θ_0. The fact that the design of the array is made in the frequency domain is advantageous in

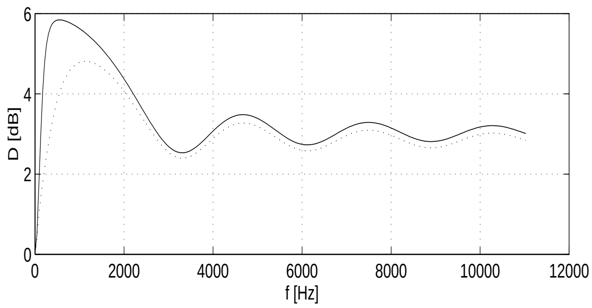

Figure 3: Directivity index of the microphone array with no differences between spectra in both channels (solid line) and with the 50% differences (dotted line)

our application since the rest of the proposed system works in this domain too. In the frequency domain the spectrum of the output signal of the array is computed as the sum of the input signal spectra multiplied by the complex transfer functions $H_n(f)$ ($n = 1, 2$):

$$Y(f) = \sum_{n=1}^{N} H_n(f) X_n(f). \quad (2)$$

The transfer functions $H_n(f)$ are computed by solving the set of equations (3) for discrete frequencies $f = kf_s/M$, $0 \leq k \leq M/2$, given by the Fourier transformation of the input signal:

$$\sum_{m=1}^{N} h_{nm}(f) H_m(f) + \mu H_n(f)$$
$$= \exp(-j\beta d_{mic}(\frac{N+1}{2} - n)\cos(\Theta_0)) \quad . \quad (3)$$

The functions $h_{mn}(f)$ are computed from equation (4), ($m = 1, 2,\ n = 1, 2$).

$$\begin{aligned} h_{mn}(f) &= \frac{1}{4\pi}\int\limits_0^{2\pi}\int\limits_0^{\pi} e^{(j\beta d_{mic}\cos\Theta)} \sin\Theta \, d\Theta \, d\varphi \\ &= \begin{cases} \frac{sin(\beta d_{mic})}{\beta d_{mic}} & \text{for } n \neq m \\ 1 & n = m \end{cases}, \end{aligned} \quad (4)$$

where μ is the Lagrangian multiplier that control the superdirectivity (it was set to $\mu = 0.02$)

The disadvantage of the array consisting of filters with the transfer functions $H_n(f)$ designed according to the equations above is the fact that it does not have flat magnitude response for signals coming from straight direction and so it causes distortion of the input signals. That is why this design was modified.

In the two-microphone array the transfer functions $H_1(f)$ and $H_2(f)$ have the same magnitudes and it can be shown that the directivity index $D(f)$ of the array does not change if both the magnitudes are

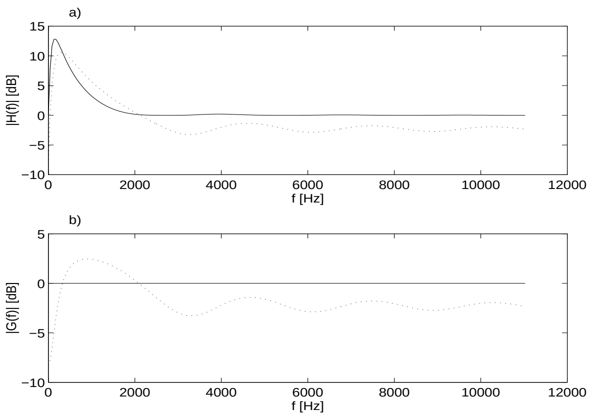

Figure 4: The comparison of the filter transfer function magnitudes $H_n(f)$ a), and the array transfer function magnitudes $G_n(f)$ b) of the original (dotted line) and the new (solid line) method

changed in the same way. So the new transfer functions may be computed by combining the phases of the original transfer functions $\angle H_n(f) = \arg(H_n(f))$ with the new magnitude $M(f)$:

$$H_{nNEW} = M(f)\exp\bigl(j\angle H_n(f)\bigr). \qquad (5)$$

The new magnitude $M(f)$ ensuring the flat transfer function of the array is obtained by solving the equation:

$$\left|\sum_{n=1}^{2} M(f) e^{(j\angle H_n(f))} e^{(-j\beta d_{mic}(\frac{N+1}{2}-n))}\right| = 1. \qquad (6)$$

Fig. 4 shows the comparison between the original and the new filter magnitudes and between the transfer functions of the array using the filters with these magnitudes.

2.2. Noise suppression post-processing

2.2.1. Noise estimation

Noise estimation suggested in [4] is replaced by the estimation presented in [9] which is more suitable for coherent noises. The supposition of this method – the spatial stationarity of the noise can be used in our application because the noise of moving sound sources (cars, etc.) are not suppressed by this approach. This approach requires the voice activity detector (VAD). Two VADs using the *magnitude squared coherence* were tested [12], [7]. The first batch-processing based VAD can be optimized to give the small variance of results (especially for lower signal to noise ratio-SNR) but the latter VAD using recursive processing is more suitable for the real-time implementation.

2.2.2. Noise suppression

Instead of the *Minimum Mean Square Error Short-Time Spectral Amplitude Estimator* (MMSE) suggested in [5], [6] the Wiener filter [1] was used. The reasons for this choice can be summarized as follows: the MMSE has the tendency to attenuate speech [4], MMSE assumptions are not always fulfilled in real conditions [11], the Wiener filter proposed by [1] is easier to implement.

2.2.3. Residual noise postprocessing

The impact of the residual noise postfiltration suggested in [4] on the performance of the SPEAK system was tested.

3. SYSTEM IMPLEMENTATION

The whole system was realized in the frequency domain using 256-points FFT with 50% overlap and the overlap add synthesis. Sampling frequency 22050 Hz meets requirements of the SPEAK strategy. Some parts (VAD, Wiener filter) of the whole system were implemented on DSP TMS320C30 board.

4. EXPERIMENTS AND RESULTS

4.1. Evaluation criteria

The noise reduction system influences the behaviour of the SPEAK strategy in two ways:

- it ensures that the powers in all frequency bands are close to the powers of the clean speech,
- it forces the SPEAK coding strategy to select the proper frequency bands as for the clean speech.

To describe this influence the suitable criterion must be used. The speech distortion and the noise suppression evaluated using the segmental signal-to-noise ratio (SSNRE) approach suggested in [8] were used first. This criterion seems not to be suitable in our case because it is not able to estimate the results of the SPEAK coding. Better criterion based on the preceding discussion is the evaluation of the signal power in:

- 20 frequency bands corresponding to the SPEAK processor filter bank,
- 6 frequency bands with the highest power selected according to the SPEAK strategy[1].

Both the criteria can be used for the numerical evaluation of spectral distances between clean, noisy and enhanced speech and for the graphical comparison of the modified spectrograms of the signals. The spectrograms of noisy, processed and clean speech can be seen from Fig. 5. The comparison of spectrograms a) and d) reveals that the the noise may cause the erroneous choice of frequency bands. The comparison of spectrograms b) and d) reveals that the noise reduction system (without residual noise processing)

[1]In fact, the number of bands selected by SPEAK coding strategy varies from 4 to 10 according to the input signal characteristic. For the lack of precise information about the details of the SPEAK strategy the rough approximation using 6 bands is used in this study.

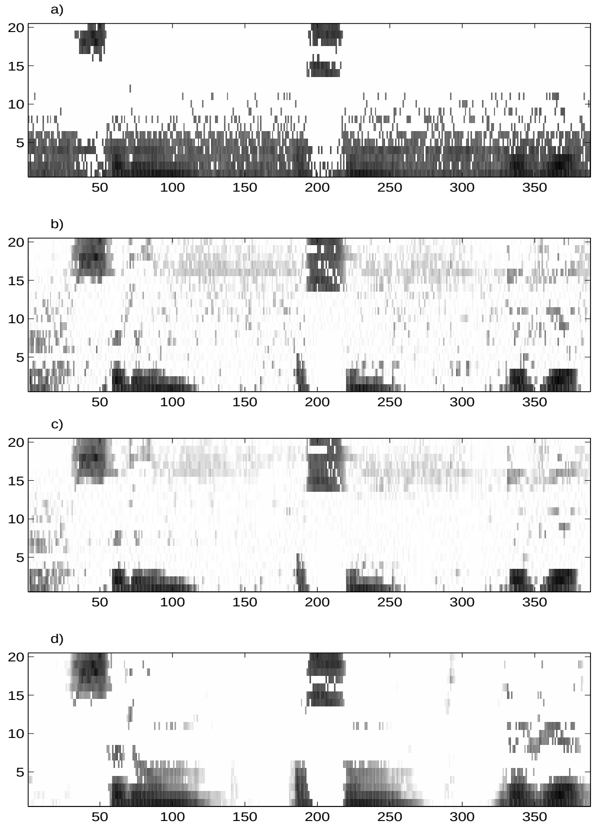

Figure 5: SPEAK coding strategy simulation. a) signal corrupted by street noise, b)signal processed by system without residual noise postprocessing, c)signal processed by system with residual noise postprocessing, d) clean signal

generates residual noises. Finally, the comparison of spectrograms c) and d) shows the decreasing of the residual noise in the enhanced speech (with the residual noise processing). The usefulness of both criteria were verified by the SPEAK strategy simulation. The final verification based on the subjective tests with hearing impaired people are being prepared.

4.2. System testing

The proposed system was tested using real-life noises and speech signals picked up on streets, underground, railway stations, office rooms, etc. The results of experiments confirmed the ability of the method to recover the original speech spectra in most cases [13].

ACKNOWLEDGEMENTS

This project was supported by the grant GACR 102/96/k087 "The theory and application of speech communication in Czech", and it was solved in the collaboration with the Laboratory of Electronic Sensory Substitutions in Prague.

The authors would also like express thanks to Mr. Pham Quang Hung for the implementation of some parts of this system on DSP TMS320C30 board.

REFERENCES

[1] A. Akbari Azirani, R. Le Bouquin Jeannes, and G. Faucon:. Speech enhacement using a Wiener filtering under signal presence uncertainty. In *EUSIPCO'96*, pages 971–974, Trieste, Italy, 1996.

[2] M. Dörbecker:. Small microphone arrays with directivity for speech enhancement. In *EUROSPEECH'97*, pages 327–330, Rhodos, Greece, 1997.

[3] M. Dörbecker:. Speech enhancement using small microphone arrays with optimized directivity. In *IWAENC'97*, pages 100–103, Imperial College London, UK, 1997.

[4] M. Dörbecker and S. Ernst:. Combination of two-channel spectral subtraction and adaptive Wiener post-filtering noise reduction and deverberation. In *EUSIPCO'96*, pages 995–998, Trieste, Italy, 1996.

[5] Y. Ephraim and D. Malah:. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. Vol. ASSP-32(6):1109–1121, December 1984.

[6] Y. Ephraim and D. Malah:. Speech enhancement using a minimum mean-square log-spectral amplitude estimator. Vol. ASSP-33(No. 6):443–445, 1985.

[7] R. L. Bouquin Jeannes and G. Faucon. Voice activity detector based on the averaged magnitude squared coherence. In *The Int. Conf. on Sig. Proc. Appl. and Technology ICSPAT'95.*, pages 1964–1968, Boston, USA, October 1995.

[8] R. Martin:. Design and optimization of a two microphone speech enhancement system. In *EUROSPEECH'95*, pages 2009–2012, Madrid, Spain, 1995.

[9] J. Meyer, K. U. Simmer, and K. D. Kammeyer:. Comparison of one- and two-channel noise-estimation techniques. In *IWAENC'97*, pages 17–20, Imperial College London, UK, 1997.

[10] Spectra 22. *Technical reference manual.* Cochlear Pty. Limited, May, 1994.

[11] P. Scalart, J. V. Filho, and J. G. Chiquito:. On speech enhancement algorithms based on MMSE estimation. In *EUSIPCO'96*, pages 471–474, Trieste, Italy, 1996.

[12] P. Sovka and P. Pollák:. The study of speech/pause detectors for speech enhancements methods. In *EUROSPEECH'95 - Proceedings of the 4th European Conference on Speech Communication and Technology.*, pages 1575–1578, Madrid, Spain, September 1995.

[13] M. Svoboda and P. Sovka. Two-channel method for speech enhancement combining microphone array with Wiener filtration. Technical Report #R99 - 1, FEE CTU in Prague, Prague, Czech Republic, January 1999.