A SPEECH ENHANCEMENT SYSTEM BASED ON NEGATIVE BEAMFORMING AND SPECTRAL SUBTRACTION

A. Álvarez, R. Martínez, P. Gómez, V. Nieto
Universidad Politécnica de Madrid – Facultad de Informática
Campus de Montegancedo, s/n, 28660 Boadilla del Monte, Madrid, SPAIN
Tel.:+34.91.336.73.84, Fax: +34.91.336.66.01, Email: pedro@pino.datsi.fi.upm.es

ABSTRACT

Beamforming Filtering techniques are well known for their angular selectivity in speech enhancement applications with noisy backgrounds. Besides, *Negative Beamformers* achieve an important advantage compared with classical array structures: the lower number of signal sensors (microphones) required. The scheme that is proposed in this paper combines *Negative Beamforming* and *Spectral Subtraction* in order to obtain large gains in the *SNR* at a reasonable low computational cost. This method may be used to eliminate or enhance a specific signal using a *binaural array*. Applications of this technique may be found in *Surveillance Systems, Domotic Control* and also, to improve *Speech Recognition*.

1. INTRODUCTION

In speech-driven applications (e.g. *Speech Recognition*), a low signal-to-noise ratio produces a clear degradation in their performance. Unfortunately, a wide range of scenarios (e.g. conference rooms, hands-free telephones, speech recognizers in cars, etc.) may have several speakers or speech/sounds sources active at the same time. Besides, the utilization of proximity microphones it is not always desirable and also, moving speakers disable the use of highly directional microphones.

Array Beamforming [3] is a key technology allowing a flexible use of speech in cocktail party scenarios. In this sense, multimicrophone systems could be devices of choice for speech recording and enhancement in demanding environments, as they are well suited for the suppression of non-stationary interferences and have the great advantage that they perform dereverberation and noise suppression at the same time [9].

Through this paper, the application of *Negative Beamforming Filtering (NBF)* to speech signals is proposed. These structures achieve two main advantages: narrow negative beams and a lower number of processing elements required, as compared with classical beamformers [2],[8].

2. NEGATIVE BEAMFORMING FILTERING

The speech enhancement system is supported by the basic beamforming cell presented in Figure 1. That element may be seen as a hybrid unit falling between *Speech Beamformers* and *Source Separation Systems*. Speech beamformers [9] are similar to classical beamforming systems, used for phased arrays in radar applications, where the main goal is the reconstruction of the strongest source on average. In a different fashion, source separation applications [10] try to reconstruct all sources. The structure of a signal separation module includes some kind of channel mixing and adaptation in order to estimate isolated versions of incoming signals. In our case, the term *Negative* is related with the fact that sources of interest will be removed by the beamforming procedure.

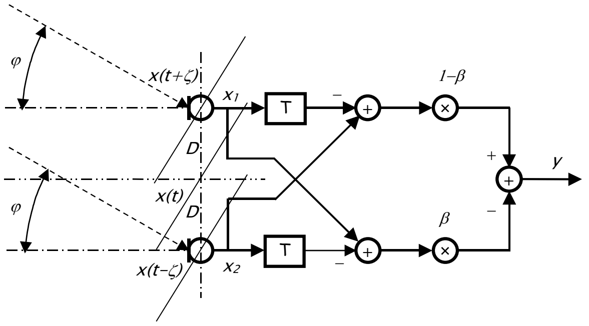

Figure 1. Elementary cell of the two-microphone negative beamformer. Microphones are separated a distance *d=2D*, being φ the angle of arrival for an incoming sound source. The angular tracking factor is modeled through parameter β with values included in the range *[0.0, 1.0]*. This processing element introduces a fixed delay interval $T=k\tau$, being τ the time delay unit.

The module, despite of its simplicity, recalls the behavior of a notch filter controlled by the β parameter. This mechanism shows a transfer function in the frequency domain, which may be formulated as [4]:

$$Y(\alpha,\delta)=2e^{-j(\pi-\delta)/2}\left((1-2\beta)\cos\alpha\, sin\,\delta/2-sin\alpha\cos\delta/2\right) \quad (1)$$

where:

$$\alpha = \omega\zeta = 2\pi f = 2\pi f\zeta = \frac{2\pi fD}{c} sin\,\varphi \quad (2)$$

$$\delta = \omega T = 2\pi fT = 2\pi fk\tau = 2\pi k\frac{f}{f_s} \quad (3)$$

φ being the *angle of arrival*, ζ *half the array travel* time, f the frequency of the signal, k the *delay order*, *d=2D* the *microphone distance*, and f_s the *sampling frequency*.

The notch appears at an angle given by:

$$\varphi_n= arcsin\{\frac{c}{2\pi fD} arctan[(1-2\beta)tan(\pi\, k\, f/f_s)]\}; \quad (4)$$

The evaluation of formula *(1)* for a fixed value of the steering parameter β is shown in Figure 2. As it may be noticed, an individual beta value not always implies the cancellation of source frequencies originated from a specific direction of arrival. In fact, that is only the case for only three beta values: *{0.0, 0.5, 1.0}*. Therefore, the value of β, which produces the highest degree of cancellation, depends not only on the value of the incoming angle φ, but on the signal frequency *f*, as well (see Figure 3). This property of the *Negative Beamformer* implies that for broadband signals like speech, the spectra of interest should be divided into different frequency bands [6], throughout the use of bandpass filters and, then replicating the beamformer cell previously presented (see Figure 4).

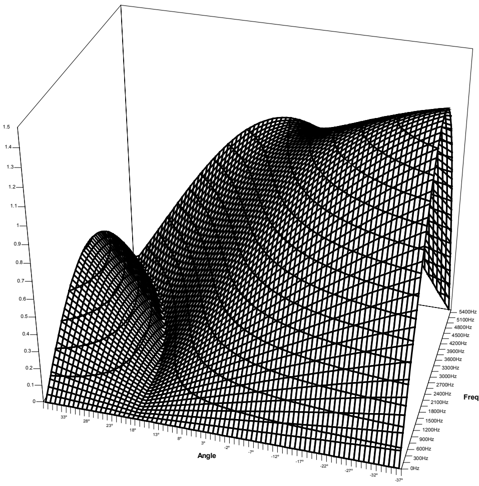

Figure 2. Module of the *Negative Beamformer* transfer function for *d=5 cm*, f_s*=11,025 Hz*, *β=0.25* and *k=1*.

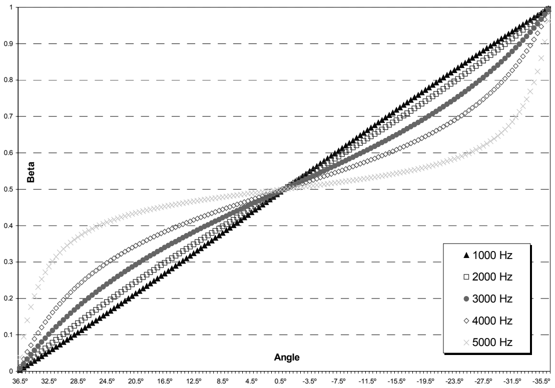

Figure 3. Mapping of several frequency values for combinations of angle of arrival *φ* vs. steering parameter *β*.

Finally, the relation among source angle of arrival *φ*, signal frequency *f* and values of *β*, is given by the following expression:

$$\beta = 1/2\left(1-\frac{tan\left[\frac{2\pi fD}{c}sin\varphi_n\right]}{tan(\delta/2)}\right) \qquad (5)$$

2.1 Source estimation

The other important aspect related to the use of the *Negative Beamformer* structure is the individual source tracking. This procedure involves finding the correct angle of arrival for a source present in the microphone inputs and its operation should be independent of the presence of other sources. The solution implemented in this system consist on tracking the presence of single sources in the different frequency subbands and, introducing a measure called *Groove Aspect Ratio (GAR)* [5] to determine whether there are one or more sources active at the corresponding band. The *Groove Aspect Ratio* may be seen as the result of a cost function and is based on the following formula:

$$c_j = \frac{2\left|Y_j(j_{min})\right|^2}{\left|Y_j(0.0)\right|^2+\left|Y_j(1.0)\right|^2} \qquad (6)$$

where $|Y_j(\beta)|$ is the module of the *Negative Beamformer* transfer function as given by *(1)*, *j* being the index of the corresponding filter bank, and:

$$\beta^j_{min} = \arg min\left\{\left|Y_j(\beta)\right|^2\right\} \qquad (7)$$

Acceptable *GARs* should accomplish a certain threshold that includes the ratios corresponding to surrounding bands. The idea behind this limitation in the number of valid hypothesis is to improve source detection accuracy, even though this may involve a high hypothesis rejection rate.

Finally, valid detections calculated for a subband (determination of the beta associated to an spatial origin) may be easily extended to the rest of the frequencies through the use of a *band cross-mapping* function, which will derive their particular *steering factors*.

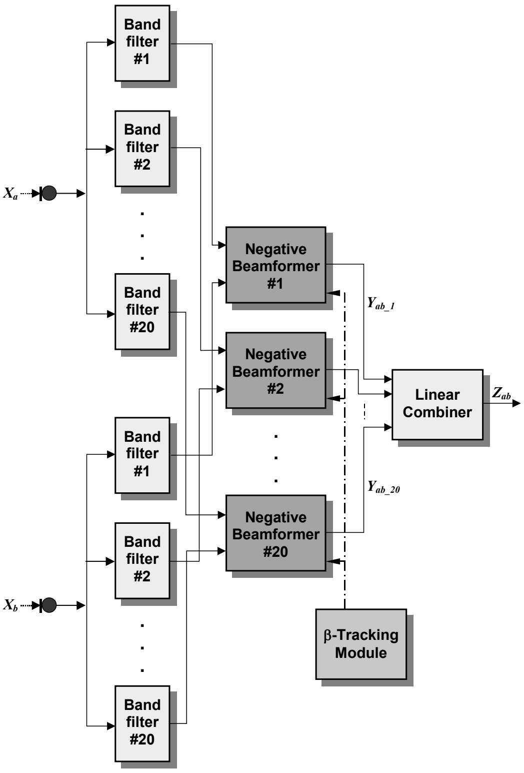

Figure 4. Complete structure of the *Negative Beamformer Filter* structure to manage a specific sound source.

3. SPECTRAL SUBTRACTION

The *Negative Beamforming* technique described may cancel an individual source but it is not able to enhance it by itself. However, it is feasible to combine this technique with spectral-domain filtering [1], as the beamformer output constitutes a valid noise-estimation for the *Spectral Subtraction* module. It is important to notice that beamforming-filter outputs z_{ab} contain frequency-decreased versions of original sources (x_a and x_b) pointed out and captured by the two microphones. In this sense an original input signal (x_a or x_b) is taken as the primary noisy channel and a beamformer output, once applied a suitable value for the steering factor, the reference one.

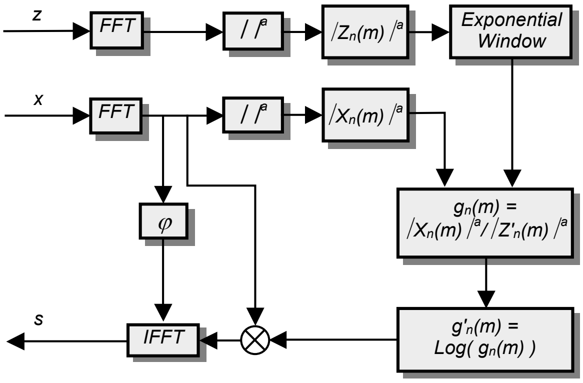

Figure 5. Spectral Subtraction method being used for signal recovery purposes.

The spectral subtraction procedure comprises several steps summarized in Figure 5. Initially, both signals are segmented in overlapped windows and transformed into the frequency domain using the short-time *Discrete Fourier Transform F{.}*:

$$Z = Z(m) = F\{z(n)w(n)\} \qquad (8)$$

$$X = X(m) = F\{x(n)w(n)\} \qquad (9)$$

where *M* is the size of the window used, *w(n)* is the window function, and *n* and *m* are the time and frequency indices.

In a following step the reference channel is applied a filter with an exponential decay:

$$Z'_n(m) = \alpha Z_n(m) + (1-\alpha) Z_{n-1}(m); \quad 0 \le m \le M/2 - 1 \qquad (10)$$

Then, the relationship between the power spectra of the primary and reference channel is calculated for every frequency channel:

$$g_n(m) = \frac{\|X_n(m)\|^a}{\|Z_n(m)\|^a}; \quad 0 \le m \le M/2 - 1 \qquad (11)$$

Now, the ratio is weighted using a logarithmic law before the subtraction is performed:

$$g'_n(m) = \log(g_n(m)); \quad 0 \le m \le M/2 - 1 \qquad (12)$$

$$\|S_n(m)\|^a = \|X_n(m)\|^a g'_n(m); \quad 0 \le m \le M/2 - 1 \qquad (13)$$

Therefore, a large cancellation is produced when the power of the beamformer-output is small compared with the original signal.

Besides, the phase of the enhanced signal is recovered from the primary trace *x*:

$$\varphi_{S_n}(m) = \varphi_{X_n}(m); \quad 0 \le m \le M/2 - 1 \qquad (14)$$

4. RESULTS

The framework of the experiments carried out, is shown in Figure 6. Left source $s_1(t)$ corresponds to the signal represented in Figure 7. Figure 8 contains the right source $s_2(t)$. The input recorded by the couple of microphones corresponds to a mixture of both signals (see Figure 9).

The beamforrmer is steered towards the two sources separately, producing two different outputs, as may be seen in Figure 10. As only two sources are present, the application of the *NBF* to the first one produces a partial cancellation of that signal and indirectly, that result may be consider the enhancement of the other one.

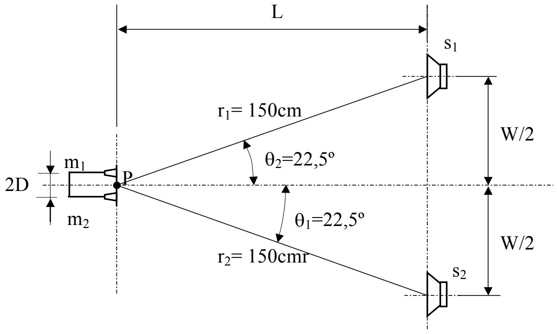

Figure 6. Framework for the recording of a two-source signal. Two sources (loudspeakers) $s_1(t)$ and $s_2(t)$ are placed on the same plane relative to the array microphone (m_1 and m_2).

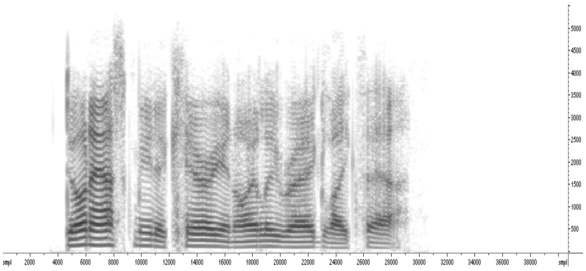

Figure 7. Utterance of the sentence */Don't ask me to carry an oily rag like that/* produced by a male speaker.

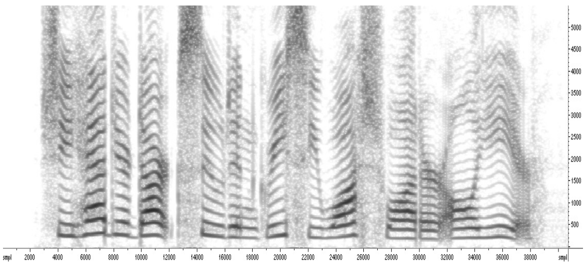

Figure 8. Utterance of the sentence */She had your dark suit in greasy wash water all year/* produced by a female speaker.

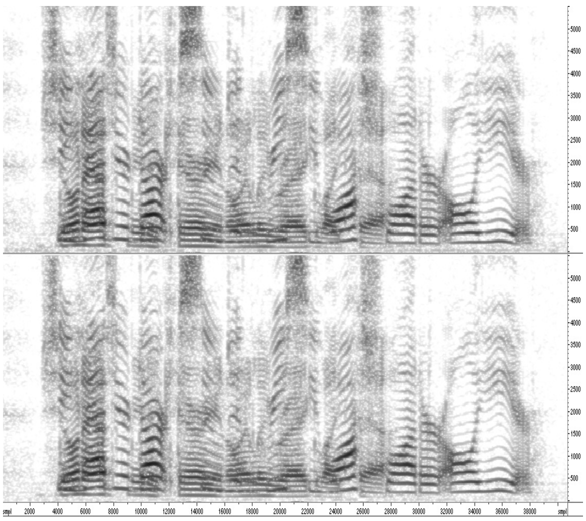

Figure 9. Power spectrum of the microphone inputs m_1 and m_2, respectively.

Finally, Figure 11 includes the enhancement of signals once the *Spectral Subtraction* module has been applied. These last results correspond to the recovery of the original sources.

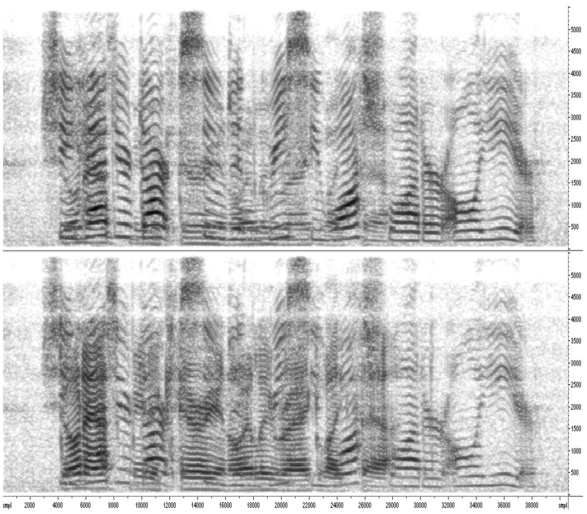

Figure 10. Power spectrum of the *Negative Beamformer* output when the *Band Steering Factors (β_j)* are tuned to an incoming angle of +22.5º and -22.5º respectively.

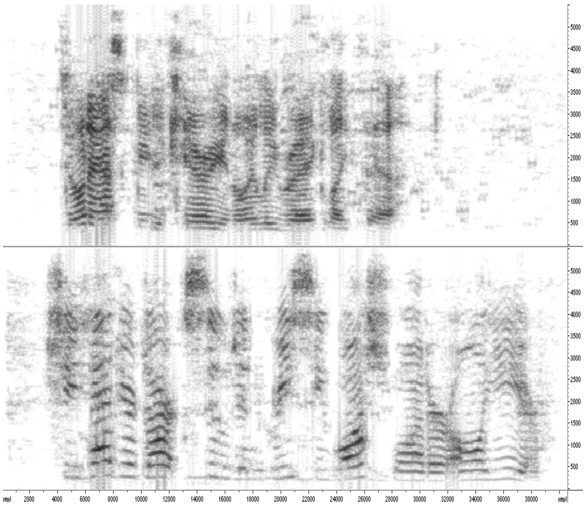

Figure 11. Power spectrum of the *Spectral Subtraction* module output. Both signals are reconstructed using one of the microphone input channels and its associate *Negative Beamformer* output.

5. CONCLUSIONS

The combination of *Negative Beamformer Filtering* (*NBF*) and *Spectral Subtraction* results in a high degree of source cancellation and enhancement at a reasonable computational cost, which allows building real-time cost-sensitive applications, as only two microphones are required.

These structures are rather selective in the angular domain attaining signal enhancement factors up to *20 dB*. Besides, several sources may be managed in parallel as the response of the system is controlled by the steering parameter β.

Important application fields are clean speech monitoring and noise removal in *Robust Speech Recognition Systems* and, source localization/tracking in *Combined Audio-Video Applications* [7].

6. ACKNOWLEDGMENTS

This research is being supported by Project TIC99-0960 (Programa Nacional de las Tecnologías de la Información y las Comunicaciones), Project 07T/0001/2000 (Plan Regional de Investigación de la Comunidad Autónoma de Madrid), and a Contract between the Centre Suisse d'Electronique et de Microtechnique and Universidad Politécnica de Madrid.

7. REFERENCES

[1] Álvarez, A, Gómez, P., Martínez, R.. and, Nieto V., "Combination of Negative Beamforming and Nonlinear Spectral Subtraction for Speech Enhancement and Source Tracking", *Proc. of 2001 IEEE- EURASIP Workshop on Nonlinear Signal and Image Processing*, Baltimore, Maryland, USA, June 3-6, 2001, pp. MonAmPO2.1.

[2] Fisher, S., and K. U. Simmer, "Beamforming Microphone Arrays for Speech Acquisition in Noisy Environments", *Speech Communication*, Vol. 20, 1996, pp. 215-227.

[3] Furui, S., "Recent Advances in Robust Speech Recognition", *Proc. of the ESCA-NATO Tutorial and Research Workshop on Robust Speech Rec. for Unknown Comm. Channels*, Pont-à-Mousson, France, 17-18 April 1997, pp. 11-20.

[4] Gómez, P., Álvarez, A, Martínez, R, Nieto, V., and Rodellar, V., "Speech Enhancement through Binaural Negative Filtering", *Proc. of X European Signal Processing Conference*, Tampere, Finland, September 4-8, 2000, pp. 187-190.

[5] Gómez, P., Álvarez, A., Martínez, R., Nieto V. and, Rodellar, V., "Frequency-Domain Steering for Negative Beamformers in Speech Enhancement and Directional Source Separation", *Proc. of the IEEE ISCAS'2001*, Sydney, Australia, May 6-9, 2001, vol. II, pp. 289-292.

[6] Schmidt, G., "Acoustic Echo Control in Subbands – an Application of Multirate Systems", *Proc. of the EUSIPCO'98*, Rhodos, Greece, September 8-11, 1998, pp. 1961-1964.

[7] Strobel, N., Spors, S., and Rabenstein, R. "Joint Audio-Video Object Localization and Tracking. A Presentation of General Methodology", *IEEE Signal Processing Magazine*, January 2001, pp. 22-31.

[8] Sullivan, T. M., *Multi-Microphone Correlation-Based Processing for Robust Automatic Speech Recognition*, Ph.D. Thesis, Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, August 1996.

[9] Van Compernolle, D. and Van Gerven, S. "Beamforming with Microphone Arrays", *Applications of Digital Signal Processing to Telecommunications*, pp. 107-131, E.U. 1995. COST 229.

[10] Van Gerven, S. *Adaptive Noise Cancellation and Signal Separation with Applications to Speech Enhancement*, *PhD thesis*, K.U.Leuven, ESAT, March, 1996.