OPTIMAL STEERING OF A NEGATIVE BEAMFORMER FOR SPEECH ENHANCEMENT

Pedro Gómez, Agustín Alvarez, Rafael Martínez, Víctor Nieto, Victoria Rodellar
Departamento de Arquitectura y Tecnología de Sistemas Informáticos
Universidad Politécnica de Madrid, Campus de Montegancedo, s/n
28660 Boadilla del Monte, Madrid, SPAIN
Tel: +34.1.336.73.84; Fax: +34.1.336.66.01
e-mail: pedro@pino.datsi.fi.upm.es

ABSTRACT

This paper is devoted to show the possibilities for optimally steering a *Binaural Negative* Filter using two techniques based in the optimisation of a Cost Function. These filters may be used in source separation, speaker tracking, or speech enhancement with application in Robust Speech Recognition, Domotic Control, or Video-Conferencing, among other fields.

1. BINAURAL NEGATIVE BEAMFORMING

Speech Processing and Recognition is a field experiencing a rapid expansion. In certain situations it is of most relevance to detect the origin of different sound sources to enhance the reliability of speech recognition systems or for speaker tracking, signal selection or noise rejection. Classically *Microphone arrays* [4][8] have been proposed as a pre-processing technique to implement directional signal separation [8] of speech from noise or multiple sources (*cocktail-party effect* [3]) These systems, although efficient, require large sets of microphones equally calibrated and balanced, and suppose a high computational cost. Through this research [5][6][7], a simpler array have been proposed, based on *Negative Binaural Filtering* [2]. Its main advantage is a higher angular selectivity and a smaller complexity required. The proposed *Binaural Negative Filter* is shown in Figure 1.

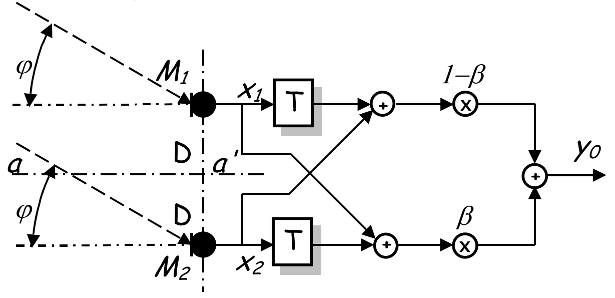

Figure 1. Structure of the *Binaural Negative Filter (BNF)*. The *angle of arrival* is φ. The separation between the two microphones is $2D$. The angular *steering factor* is β. The delay interval is $T=k\tau$, τ being the sampling period.

It may be shown that this structure presents a transfer function in the frequency domain given by:

$$|H(\alpha,\delta)| = 2|(1-2\beta)\cos\alpha \sin\delta/2 - \sin\alpha \cos\delta/2| \quad (1)$$

in terms of the angular shifts:

$$\alpha = \frac{2\pi fD}{c}\sin\varphi;\ \delta = 2\pi k f/f_s \quad (2)$$

where f is the *frequency* of a hypothetical sinusoidal plane wave reaching the array with an *arriving angle* φ relative to the main *array axis (a-a')*, f_s is the *sampling frequency*, c is the *speed of sound*, k is the *delay order* and β is the *filter steering factor*. This function shows a sharp notch at an angle given by:

$$\varphi_n = \arcsin\{\frac{c}{2\pi fD}\mathrm{arctg}[(1-2\beta)\,\mathrm{tg}(\pi k f/f_s)]\};\ f<f_s/2 \quad (3)$$

which may be observed in the behaviour of the *Negative Beamformer Transfer Function* as shown in Figure 2.

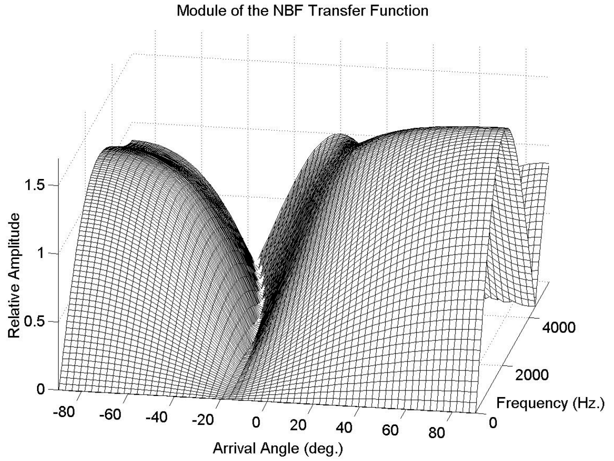

Figure 2. *Module of the Negative Beamformer Transfer Function* for $d=5$ cm, $f_s=11{,}025$ Hz, $\beta=0.25$ and $k=1$, showing two main lobes and a notch at a varying angle φ_n depending on frequency. Low frequencies are plotted in the front side, high frequencies in the rear side.

The behaviour of the notch is far from ideal, as its position will vary with frequency (see the valley in Figure 2). This poses an important problem when using the *NBF* with broadband signals, as is the case of speech. Therefore a correction of the steering factor to keep constant exploration angles is required.

2. SPATIAL FILTERING

The distortion produced by the non-linear relation between the notch angle and the steering factor requires a splitting of the two microphone inputs $x_1(n)$ and $x_2(n)$ into a set of narrow bandpass filters (*BPF$_{1\text{-}K}$*) as shown in Figure 3. Each pair of signals associated to the same frequency band, $x_{1k}(n)$ and $x_{2k}(n)$ will be treated by the same *Binaural Negative Filter* (*BNF$_{1\text{-}K}$*). The key to the successful processing is the assignment of the adequate value for the steering factor to each filter, β_k. This technique is based in signal sub-space tracking [1] and may be denominated *Frequency Domain Steering*. In preliminary papers [5][7] a technique to establish the best value for this parameter based on estimating the power of the output functions $y_k(n)$ from the array of *BNF's*, has been shown.

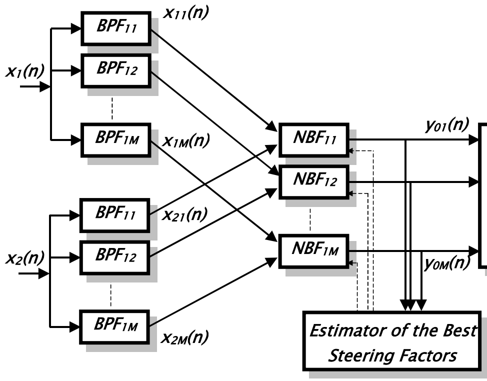

Figure 3. General structure of the *Frequency Domain Binaural Negative Beamformer.*

This technique consisted in evaluating the power of the *Negative Beamformer* output for at least three different bands, solving a set of nonlinear equations. It rendered good results to resolve the presence of two sources within the same band. Through the present paper a step-ahead will be given using a more precise technique based on the optimisation of a *Cost Function* derived from the operation of the array of *Negative Beamformers.*

3. FREQUENCY-DOMAIN OPTIMAL STEERING

Assuming that the output signals from the *Negative Beamformers* $\{y_m(n)\}$ contain information from all possible angular arrival directions except from the notch angle, comparing the output signals with the input signal, an error signal may be defined as:

$$\varepsilon_m(n) = \hat{x}_m(n) - y_{0m}(n) \qquad (4)$$

This signal, which may be called *Enhanced Error* will contain information from arrival angles given by the notch. Therefore we will introduce the following definition for a *Cost Function*, based on the ratio of the respective expectations of the *Enhanced Error* and the *BFN* output:

$$\rho_m = \frac{E\{\varepsilon_m^2(n)\}}{E\{y_{0m}^2(n)\}} \qquad (5)$$

It may be shown that this function shows local maxima, coinciding with the directions where important contributions of signal are present. An example of the Cost Function is given in Figure 4, corresponding to a pair of sinusoidal sources of *2,000 Hz* and *1,000 Hz* arriving to the array with angles of *22.5°* and *-22.5°* respectively. To implement the detection of the best arrival angles, the maxima of the function given in (5) will be traced as a function of the arrival angle. A plot of these maxima may be seen in Figure 5. This figure shows two main *maxima maximorum*, which clearly point out to the arrival angles of the mentioned sinusoidal functions.

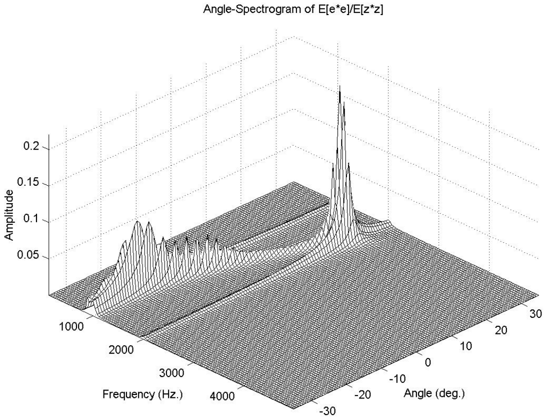

Figure 4. Estimation of the *Cost Function* in terms of angle and frequency. The left axis shows the frequency span from *0-5025 Hz*, the right axis shows the arrival angles from *–35°* to *35°*.

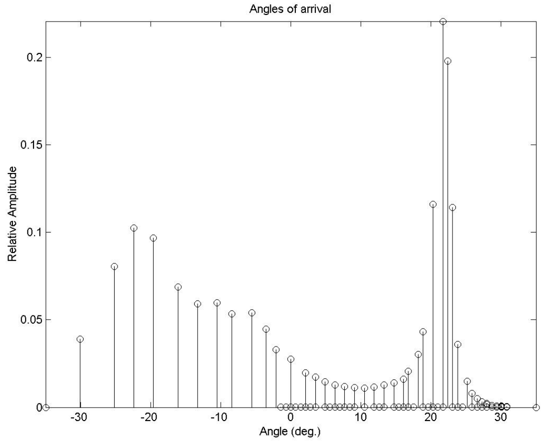

Figure 5. Angle of arrival as a function of frequency derived from the maxima of the *Cost Function* in Figure 4.

It may be seen that the low frequency maximum is more widespread than the corresponding one for high frequencies as a consequence of the lower frequency selectivity shown by the *NBF's*.

4. TIME-DOMAIN OPTIMAL STEERING

The *Frequency Domain Optimal Steering* shows that the hypothesized *Cost Function* defined in (5) reveals the interesting property of pointing out the angle and frequency position of individual sources. Nevertheless, its implementation requires the sweeping of the output signals $y_{0k}(n)$ for all the possible arrival angles within the observation span. Our interest now is to obtain a possible estimation of the best arrival angles directly from the input signals without sweeping the whole angular span. For such an expression for $\rho_k(\beta)$ will be obtained in terms of the intermediate signals:

$$u_m(n) = x_{1m}(n) + x_{1m}(n-k) \qquad (6)$$

$$v_m(n) = x_{2m}(n) - x_{1m}(n-k) \qquad (7)$$

$$w_m(n) = x_{1m}(n) + x_{2m}(n) - x_{1m}(n-k) - x_{2m}(n-k) \qquad (8)$$

It may be shown that the corresponding estimations of the *Enhanced Error* and the *BFN* output may be expressed as:

$$E\{\varepsilon_m^2(n)\} = R_{vv}^m + \beta^2 R_{ww}^m - 2\beta R_{vw}^m \qquad (9)$$

$$E\{y_{0m}^2(n)\} = R_{uu}^m + \beta^2 R_{ww}^m + 2\beta R_{uw}^m \qquad (10)$$

being in general:

$$R_{fg}^m = E\{f_m(n) g_m(n)\} \qquad (11)$$

Therefore, forcing:

$$\frac{\partial \rho_m}{\partial \beta} = 0 \qquad (12)$$

the best steering factor for each channel β_k may be found as a solution of the quadratic equation:

$$R_{ww}^m (R_{uw}^m + R_{vw}^m)\beta^2 + R_{ww}^m (R_{uu}^m - R_{vv}^m)\beta - \\ - R_{uw}^m R_{vv}^m - R_{vw}^m R_{uu}^m = 0 \qquad (13)$$

It is important to realize that to evaluate this steering factor it will only be necessary to use the correlations of the input signals, channel by channel, not requiring the *a priori* operation of the *NBF's*. This system is seen in Figure 6.

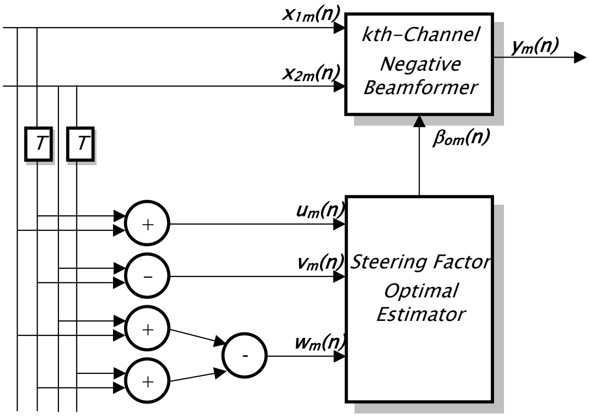

Figure 6. Optimal Steering Stragegy for the m^th frequency channel as derived from (6)-(13).

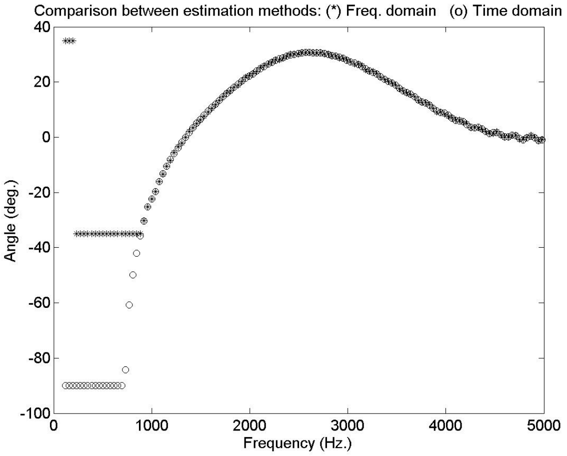

Figure 7. Angle of arrival as a function of frequency estimated from the maxima of the *Cost Function* in Figure 4 (marked with *) or from the method stated in (6)-(13) (marked with o).

The application of this methodology would require the estimation of several correlations on short-time windows. In the case commented before the estimation of the possible values of the steering factor given in Figure 7 are obtained with time windows of *N=256* samples. It may be seen in this figure that the values of the steering factor for each channel match closely the ones obtained from the function in Figure 4, this fact being of most importance to validate the *Time-Domain Method* as defined. The discrepancies between both methods for low frequencies are due to their respective angular span limits. The *Frequency-Domain Method* is subject to an angular span from *–35º* to *35º* (for *d=5 cm*, *f*$_s$*=11,025 Hz*, and *k=1*), whilst the limits for the *Time-Domain Method* are settled from *–90º* to *90º*.

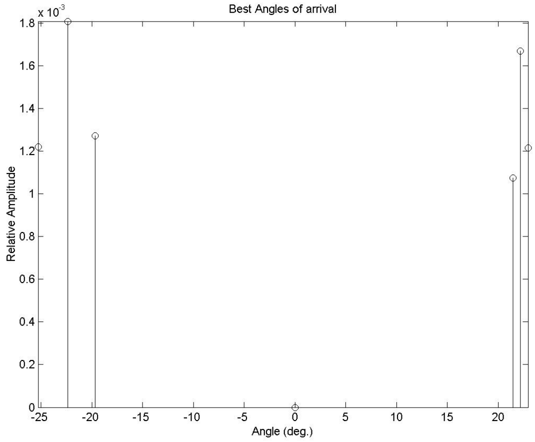

Figure 8. Best angles of arrival taking into account each channel input power. It may be seen that the main peaks point out to *+22.5º* and *–22.5º*.

As not all the arrival angles estimated by both methods are equally relevant, a weighting criterion is used to establish their relative importance. This criterion is the value of the $R^m{}_{XX}$, this function being the power of the equivalent input signal to the *m-th* channel. Using this criterion the selection of arrival angles given in Figure 8 may be obtained.

5. RESULTS AND DISCUSSION

The evaluation of the overall output implied in (4) will be implemented through *Spectral Subtraction* in the domain of the respective power spectra of the *BNF* input and output signals. This operation is carried out by the block diagram shown in Figure 9.

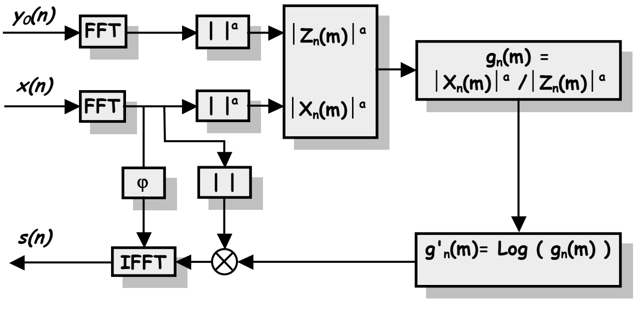

Figure 9. Evaluation of the overall output by *Spectral Subtraction*.

To implement the spectral subtraction the input *x(n)* and output *y*$_0$*(n)* to the *NBF* are Fourier-transformed, and the modules of their respective power spectra are divided. The logarithm of the resulting function is multiplied by the module of the input signal, and the resulting function is inversely Fourier-transformed assuming the phase of the input *x(n)*.

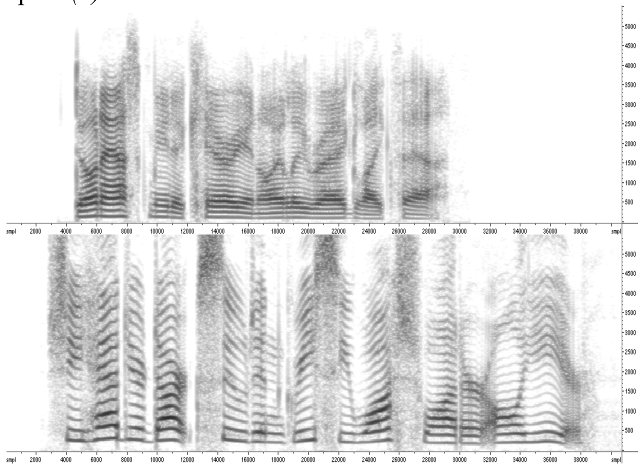

Figure 10. Upper part: Utterance of the sentence */Don't ask me to carry an oily rag like that/* (male speaker). Lower part: Utterance of the sentence */She had your dark suit in greasy wash water all year/* (female speaker).

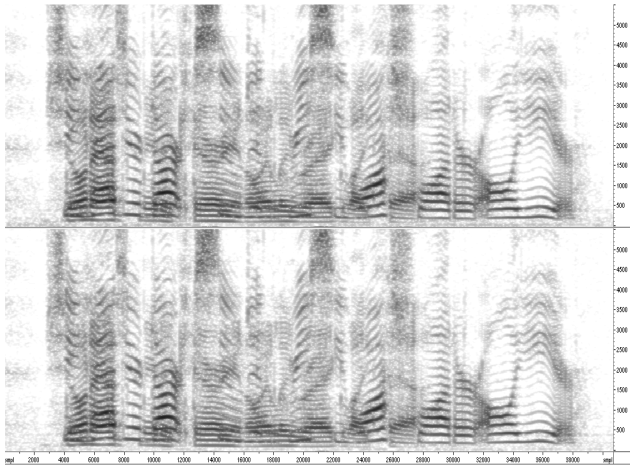

Figure 11. Power spectrum of the microphone inputs M_1 and M_2, respectively, showing a complete intermixing of both signals.

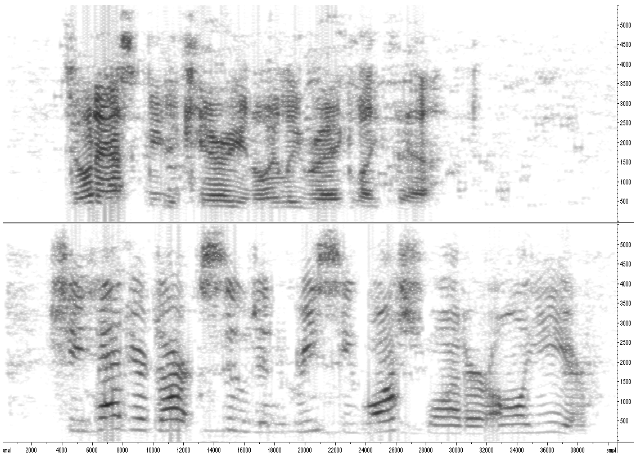

Figure 12. Power spectrum of the output. Both signals are reconstructed using the input channels and the *BNF* outputs.

The resulting output *s(n)* shows the characteristics of the signal arriving from the direction pointed by the *BNF* notches. The complete system has been checked using data from real recordings taken using two microphones (M_1, M_2) picking up the sound field produced by two loudspeakers (S_1, S_2), each one reproducing a different speech utterance from the database TIMIT. The original recordings and the resulting traces may be seen in Figure 10 to Figure 12. The operation of the full system is capable of separating both sources with a high degree of fidelity, and what is most important, the residual leftovers of the alien trace are low enough to not affect the intelligibility of the enhanced traces. Reductions in the power spectra of the contaminated signal as high as 30 dB are easily attainable.

6. ACKNOWLEDGMENTS

This research is being carried out under grant TIC99-0960 from the *Programa Nacional de las Tecnologías de la Información y las Comunicaciones (Spain)*, grant 07T-0001-2000 from the *Plan Regional de Investigación de la Comunidad de Madrid*, and a collaboration contract between *Universidad Politécnica de Madrid* and the *Centre Suisse d'Electronique et de Microtechnique*.

7. REFERENCES

[1] Affes, S., and Grenier, Y., "A signal sub-space tracking algorithm for microphone array processing of speech", *IEEE Trans. on Speech and Audio Proc.*, Vol. 5, No. 5, September 1997, pp. 425-437.

[2] Blauert, J., *Spatial Hearing*, MIT Press, Cambridge, MA, 1997.

[3] Bodden, M., and Blauert, J., "Separation of Concurrent Speech Signals: A Cocktail-Party-Processor for Speech Enhancement", *Proc. of the ESCA-Workshop on Speech Processing in Adverse Conditions*, Cannes, France, 10-13 November, 1992, pp. 1-4.

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

[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 IEEE International Symposium on Circuits and Systems ISCAS'2001*, Sydney, Australia, May 6-9, 2001 , pp. II.289-292.

[6] Gómez, P., Alvarez, A., Martínez, R., Nieto, V. and Rodellar, V. "Speech Enhancement through Binaural Negative Filtering", *Proc. of EUSIPCO 2000*, Tampere, Finland, 4-8 September 2000, Vol. I, pp. 187-190.

[7] Gómez, P., Álvarez, P., Nieto, V., Rodellar, V., and Martínez, R., "Multiple source separation in the frequency domain using Negative Beamforming", *Proceedings of the EUROSPEECH 2001*, Aalborg, Denmark, 3-7 September 2001, pp. 2619-2622.

[8] Jonhson, D. H., and Dudgeon, D., E., *Array Signal Processing: Concepts and Techniques*, Prentice-Hall, Englewood Cliffs, N.J., 1993.

[9] Yamada, T., Nakamura, S., and Shikano, K., "Robust Speech Recognition with Speaker Localization by a Microphone Array", *Proc. of ICASSP'97*, April 22-25, 1997.