MICROPHONE ARRAY DESIGN WITH IMPROVED ACOUSTIC ECHO REJECTION

*Bernard Debail**

Harman Consumer International
2, route de Tours
72500 Chateau du Loir, France

André Gilloire

France Telecom R&D DIH/IPS,
Technopole Anticipa 2, Av. P. Marzin
22307 Lannion Cedex, France
andre.gilloire@rd.francetelecom.com

ABSTRACT

We consider the problem of acoustic echo control in a teleconference terminal, in which the sound pick-up device is placed on the top of the TV monitor at some distance of the conferees (2-3 m). In that case, the short distance between the sound pick-up device and the loudspeakers integrated in the monitor creates high acoustic coupling. We have designed a microphone array which provides an efficient solution to this problem, thanks to two improvements. Firstly, the use of specially designed low-noise sensors for the low frequencies allows superdirective design without the drawback of high noise in the low frequency range. The second improvement, which is discussed in details in the paper, comes from the use of an additional constraint in the optimization of the design of the filters behind the microphones of the array. This constraint aims at zeroing the short-range acoustic feedback from the loudspeakers to the array output, hence resulting in lower acoustic echo.

1. INTRODUCTION

Sound pick-up in teleconference systems is usually implemented with microphones placed at short distances of the conferees. Although this placement is favourable with respect to acoustic background disturbances (noise and room reverberation) and acoustic echo rejection, specific needs make the use of sound pick-up at larger distances (*e.g.* 2-3 m) more appropriate in some cases, for example if some freedom of motion is desirable, or if the wire connecting the microphone or the orientation of the microphone itself may cause trouble for the users. In that case the sound pick-up device may be integrated in a compact "set-top box" placed on the top of the TV monitor for easy installation and use. The resulting degradation of speech quality due to higher acoustic disturbances and much more severe acoustic coupling (especially if the loudspeakers are integrated in the monitor enclosure, which is often the case) needs appropriate solutions.

*formerly with France Telecom R&D DIH/IPS and AETA

It is well known that microphone arrays may provide much better sound pick-up performance than a single unidirectional microphone since the directivity of such arrays can be much higher [1]. Moreover, superdirective design has proven able to cope with poor low frequency directional characteristics of "standard" (delay ands sum) arrays, at the expense of some noise increase in the low frequency range [2]. We have designed a microphone array specifically for the considered application (see figure 1). The array incorporates low-noise sensors for the low frequency range (below 1 kHz). The filters behind each microphone have been designed to maximize the directivity factor of the array under several constraints, one of these being zero output for the acoustic waves coming from the loudspeakers either directly or after reflection on the obstacles in the vicinity of the terminal. The constraint takes into account real echo path measurements.

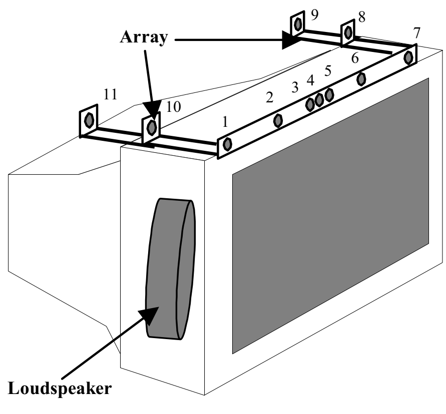

Figure 1: Arrangement of the array on the TV monitor.

We first recall briefly the design of low-noise sensors for the low frequency range; then we discuss the optimal design of the array with the appropriate constraints for the teleconference application. We give experimental results which show the performance of the array in real acoustic environment.

2. OPTIMIZED LOW-NOISE MICROPHONES

Independently of the acoustic background noise, the uncorrelated noise in microphone arrays comes from the electric noise generated by the microphones. Since superdirective techniques are more sensitive to incoherent noise than classical delay and sum arrays and since microphones used in a array are generally active in a limited frequency band, we concentrated on a proper design of the sensors for the low frequency band (below 1 kHz) where the superdirectivity principle is applied in the purpose of increasing the directivity of the array. We recall briefly the principle of the design of low-noise microphones that we have proposed in a previous paper [3].

We used electret microphones which yield excellent performance at a very low cost. Since the noise generated by electret microphones is essentially due to the self noise of the integrated FET (Field Effect Transistor) and the corresponding resistance at the FET input, higher intrinsic SNR can be obtained by increasing the microphone sensitivity by electroacoustic means which do not affect the noise. The study described in [3] shows how , starting with bi-directional (pure pressure gradient) electret microphones, the use of an additional back volume and a porous material, combined with the increase of the front to back distance by attaching a small tube at the back of the microphone, provides higher sensitivity though keeping unidirectional (cardioid) directivity pattern.

The figure 2 shows the significant SNR increase in the range 50 Hz - 1 kHz measured at the output of a typical superdirective array, due to the high sensitivity microphone design proposed. The spikes on the noise curves are measurement artifacts.

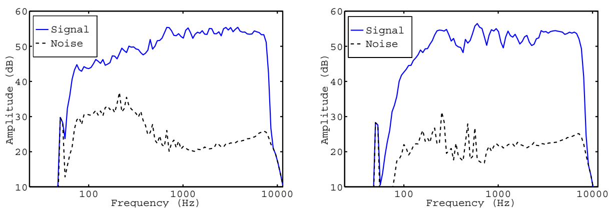

Figure 2: Signal and noise spectra for a typical superdirective array. Left: conventional sensors; right: high sensitivity sensors acting below 1 kHz.

3. ARRAY DESIGN FOR OPTIMAL ECHO REJECTION

3.1. Principles

Two principles have been used simultaneously for the optimization of the array in the low frequency range. The first one is the maximization of the near-field directivity factor under some usual constraints; the second one is the use of a specific constraint for maximum echo rejection. In the higher frequency range (above 1 kHz) a standard delay and sum design is implemented since the intrinsic directivity of the array is considered sufficient for the application.

3.1.1. Maximization of the near-field directivity

Experiments have shown that in the low frequency range, a superdirective array optimized with respect to the far-field directivity factor is very sensitive to noise sources placed in the near field; consequently a high disturbing effect due to the acoustic coupling between the loudspeakers an the microphones can be observed [4]. That is why we have chosen to maximize the near-field directivity factor, defined as follows:

$$F_{nf}(f) = \frac{1}{\frac{1}{4\pi}\underline{w}^H(f)\underline{D}_{nf}(f)\underline{w}(f)} \quad (1)$$

with the "directivity matrix" $\underline{D}_{nf}$ defined as:

$$\underline{D}_{nf}(f) = \int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi} W(f,\varphi,\theta). \\ .\underline{H}(f,r,\varphi,\theta)\underline{H}^H(f,r,\varphi,\theta)\sin\theta d\theta d\varphi \quad (2)$$

$\underline{w}$ is the vector of the weights applied at the microphones outputs for each frequency f; $\underline{H}$ is the free-field propagation vector from a source placed at a distance r from the center of the array in the direction (φ, θ). W is a spatial weighting function which attenuates the effect of the sound waves coming from the directions of the loudspeakers, and also may help obtaining adequate directivity patterns.

The optimization is made under several constraints on the weights $\underline{w}$. The purpose of the first one is to provide a desired response for the waves coming from the useful source, namely the speaking conferee; usually, this response is constant over the useful frequency range. This constraint is expressed as:

$$\underline{C}^H(f)\underline{w}(f) = \underline{s}(f) \quad (3)$$

where $\underline{C}$ contains the propagation vectors from the useful source to each microphone of the array (there may be more than one useful source as well) and $\underline{s}$ is the desired frequency response (typically phase shifts corresponding to pure delays).

The second constraint aims at limiting the incoherent noise amplification due to the superdirective design:

$$\underline{w}^H(f)\underline{w}(f) = \frac{1}{R_{Imin}(f)} \quad (4)$$

where R_{Imin} is the required minimum reduction factor of the incoherent noise.

A third constraint can be profitably used to control the width of the main lobe of the array as well as to limit the amplitude of grating lobes which may appear if some spatial aliasing is tolerated in the upper frequencies (this occurs if a small number of microphones is used in the superdirective part of the array, thus leading to large spacings between the microphones). This constraint is expressed as:

$$\underline{w}^H(f)\underline{H}_i^H(f)\underline{H}_i(f)\underline{w}(f) = g_i(f), \quad i = 0, ..., K \quad (5)$$

$\underline{H}_i$ is the vector of the transfer fuctions from a far-field source placed in the direction (φ_i, θ_i) to the microphones and g_i is the corresponding required gain.

These three constraints are incorporated in the directivity matrix according to:

$$\underline{D}_{nf}^{'}(f) = \underline{D}_{nf}(f) + \mu(f)I + \sum_{i=0}^{K} \alpha_i(f)\underline{H}_i(f)\underline{H}_i^H(f) \quad (6)$$

where μ and $\alpha_i, i = 0, ..., K$ are Lagrange multipliers which are ajusted to satisfy the constraints (4) and (5), respectively.

The optimal solution is then given by:

$$\underline{w}(f) = \underline{D}_{nf}^{'-1}(f)\underline{C}(f)(\underline{C}^H(f)\underline{D}_{nf}^{'-1}(f)\underline{C}(f))^{-1}\underline{s}(f) \quad (7)$$

The figure 3 shows the simulated directivity patterns obtained at 125 Hz and 344 Hz with this optimization. The simulation assumes perfectly matched sensors and free field propagation. It can be seen that the near-field directivity patterns (dotted curves, source at 0.5 m from the array) have similar main lobe widths as far-field patterns (continuous curves) and low sidelobes.

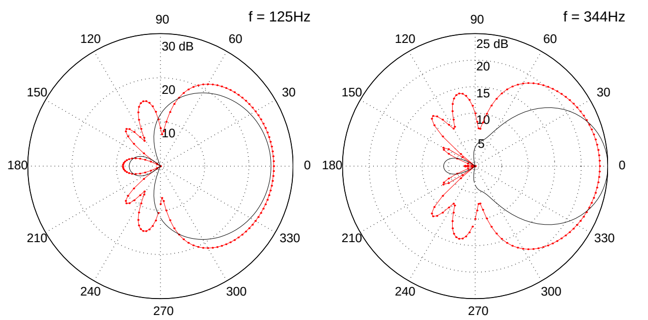

Figure 3: Simulated directivity patterns at 125 Hz (left) and 344 Hz (right) of the array optimized according to (7). Continuous curves: far-field source; dotted curves: near-field source.

3.1.2. Constraint for maximum echo rejection

The near-field optimization described above is fairly sensitive to sensors mismatch. With careful calibration of sensors in an anechoic room, we measured about 4 dB of echo reduction with a laboratory model of the array, compared with a single cardioid microphone placed at the center of the array. Since this figure of echo reduction can be found insufficient *w.r.t.* the expected performance of the array, we considered the use of measured transfer functions between the loudspeakers and the microphones of the array as an additional possibility to improve the echo reduction.

The principle that we tested is to add a constraint in the optimization process, which tries to null out the contributions of the waves coming from the loudspeakers and propagating to the microphones of the array either directly or after reflections on the short-range obstacles (*i.e.* localized within about 1 m around the array). It is assumed that the direct propagation as well as the reflections on these obstacles are essentially non-varying in time, which is a reasonable hypothesis. Practically, we use a time window which keeps the assumed stable part of the impulse responses (*i.e.* their first few milliseconds); the Fourier transforms of the windowed impulse responses corresponding to the different loudspeaker-to-microphone paths are then included in the additional constraint as follows:

$$\underline{M}^H(f)\underline{w}(f) = \underline{0} \quad (8)$$

The matrix $\underline{M}$ contains the transfer functions of the time-windowed measured loudspeaker-to-microphone paths. We can use only one time window (*e.g.* centered around the beginning of the impulse responses) as well as several ones to control the effect of high reflections coming from the surrounding environment.

The linear constraint (8) is similar to the constraint (3); therefore these two constraints are combined together for the computation of the optimal solution (7).

3.2. Implementation

The overall optimization algorithm is sketched figure 4. For each frequency f the first set of inputs is composed of the desired steering direction(s) corresponding to the conferees' locations, the constraint vector $\underline{s}(f)$, the minimum incoherent noise reduction factor R_{Imin} and the gains $g_i, i = 0, ..., K$ which control the width of the main lobe. In addition, the loudspeaker-to-microphones impulse responses $\underline{h}_{esti}, i = 1, ..., N$ (N being the number of microphones operating in the low frequency range, here $N = 5$) are measured (using a standard technique, for example loudspeaker excitation with a maximum length sequence when the conference equipment is turned on); then the impulse responses are time-windowed (the central position of the window is taken as the highest peak found in the beginning of the impulse response corresponding to the microphone closest to the center of the array). A FFT is performed on each windowed impulse response after padding with zeros to obtain the appropriate frequency selectivity. The obtained values of the loudspeaker-to-microphones transfer functions are stacked to form the vector $\underline{M}(f)$. All these inputs are used in the maximization algorithm to compute the optimal weights $\underline{w}(f)$.

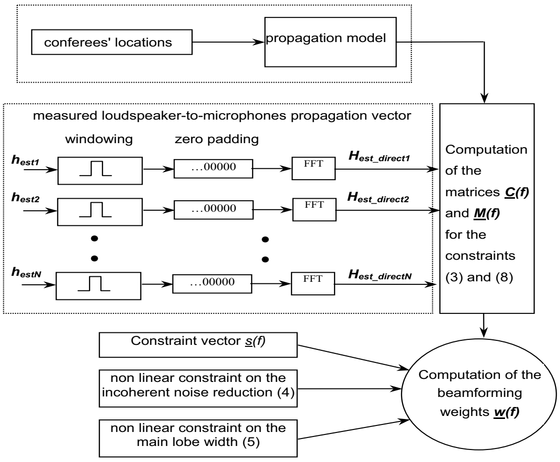

Figure 4: Block-diagram of the optimization algorithm.

4. EXPERIMENTAL RESULTS

We have performed the loudspeaker-to-microphones impulse responses measurements in a real conference room with the laboratory model shown figure 1. These impulse responses have been used to compute the weighting filters behind the low-frequency microphones according to the optimization procedure. A classical delay-and-sum broadside array design has been implemented above 1 kHz. The computed overall impulse response from the loudspeaker input to the array output is shown figure 5 for (b) 1 constraint (1 window) and (c) 2 constraints (2 windows). The windows locations are pointed by vertical arrows. The response (a) of a classical superdirective design is shown as a reference.

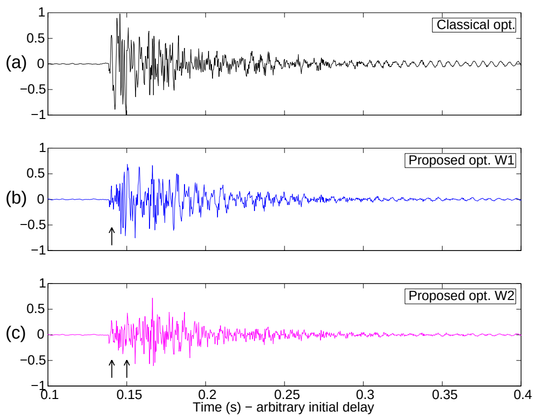

Figure 5: Overall coupling impulse response of the array. (b) 1 window, (c) 2 windows; the locations of the windows are pointed by the vertical arrows. (a) is a standard superdirective design reference.

The impulse responses (b) and (c) exhibit significant attenuation of the peaks at the windows locations. To get futher insight in the amount of echo rejection, the figure 6 shows the short term envelopes (echograms) of the first part (about 20 ms) of the impulse responses (a) and (c) shown figure 5. The echogram of the superdirective array optimized for acoustic echo rejection (Proposed opt.) exhibits much lower levels (by about 15 dB) than the echogram of the ”classical” superdirective design (Classical opt.). These results demonstrate the efficiency of the proposed tecnnique.

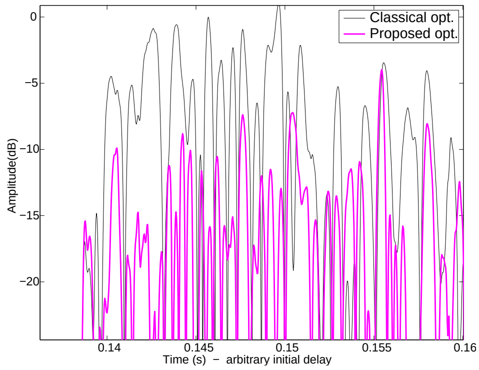

Figure 6: Echogram of the array optimized for high echo rejection (lower curve) compared with the echogram of the classically optimized array (upper, continuous curve). The overall pass-band of the arrays is 150 Hz - 7 kHz.

5. CONCLUSIONS

The combination of low-noise sensors and mixed near-field and far-field superdirective design with a specific constraint has proven able to yield high echo rejection in a microphone array for teleconference. Note that the assumed stationarity of short-range echo paths, although reasonable, can be somewhat relaxed by continuous learning of the corresponding impulse responses using subjectively hidden excitation.

6. REFERENCES

[1] Elko, G.W.: *Microphone Array Systems for Hands-Free Telecommunication.* International Workshop on Acoustic Echo and Noise Control, Roros, 1995.

[2] Tager, W.: *Near Field Superdirectivity (NFSD).* International Conference on Acoustics, Speech and Signal Processing, Seattle, 1998, pp. 2045-2048.

[3] Debail, B., Gilloire, A.: *Sound Pick-up Microphone Array Using Improved Acoustic Design Low Noise Sensors.* 5th French Congress on Acoustics, Lausanne, 2000, pp. 645-648.

[4] Debail, B.: *Etude de dispositifs de prise de son à faible bruit et à sélectivité spatiale controlée dans le cadre de la visioconférence de groupe.* Ph. D. Thesis, Université du Maine, Novembre 2000.