Audio Engineering Society

Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

Modifications on NIST MarkIII array to improve coherence properties among input signals

Luca Brayda[1], Claudio Bertotti[2], Luca Cristoforetti[2], Maurizio Omologo[2], Piergiorgio Svaizer[2].

[1]*Institut Eurecom, Sophia Antipolis, 2229 route des Cretes, 06904, France*

[2]*Istituto Trentino di Cultura (ITC)-irst, Povo, Via Sommarive 18, 38050, Trento, Italy*

Correspondence should be addressed to Luca Brayda or Maurizio Omologo (brayda@eurecom.fr, omologo@itc.it)

ABSTRACT

This work describes an activity that led to the realization of a modified NIST Microphone Array MarkIII. This system is able to acquire 64 synchronous audio signals at 44.1 kHz and is primarily conceived for far-field automatic speech recognition, speaker localization and in general for hands-free voice message acquisition and enhancement. Preliminary experiments conducted on the original array had showed that coherence among a generic pair of signals was affected by a bias due to common mode electrical noise, which turned out to be detrimental for time delay estimation techniques applied to co-phase signals or to localize speakers. A hardware intervention was realized to remove each internal noise source from analog modules of the device. The modified array provides a quality of input signals that fits results expected by theory.

1. INTRODUCTION

Nowadays, the microphone array technology has an important impact for a variety of applications, including human-computer interaction and hands-free telephony. In particular, research studies are being conducted inside the European Project CHIL (see [1] and [2]), with the purpose of applying this technology to acoustic scene analysis in meeting and lecture scenarios and, more in general, in any smart room equipped with multi-channel acoustic sensoring. In order to pick-up speech from a distance of 5-6 meters as well as to apply effectively enhancement techniques based on filter-and-add beamforming, the NIST MarkIII array [3] was considered as the best device available for research purposes. This array consists of eight microboards, each having eight microphone inputs and related amplification and A/D conversion stages. The whole digital stream is eventually made available to the user through a very effective interface to Ethernet.

In general, using a 64-microphone array and an accurate time delay estimation technique, as that based on Generalized Cross-Correlation (GCC) PHAse Transform (PHAT) described in [4] and [5], one can solve the speaker localization problem and provide enhanced speech in a very effective way. However, system performance can highly depend on the quality of the input signals.

One of the key points to derive excellent results from the above mentioned techniques is that input channels be independent each other. For instance, if a synchronous common-mode noise occurs in two microphones, a time delay estimation technique will reveal an artificial coherence at zero sample delay. The latter fact is equivalent to have an active noise source in front of the array, which actually does not exist.

The present work[1] was conducted starting from a preliminary observation that a 50 Hz interference was evident in all the input channels of the MarkIII array. Once eliminated that source of noise in the easiest way, i.e. by replacing in-house alimentation with rechargeable batteries, a consistent synchronous interference was still present in the input signals. Although this interference had a rather small dynamics, the coherence between two signals was still biased at zero samples. To remove completely or to deviate the given electrical interference, the hardware of the device was changed, based on some substitutions of electrical components (e.g. polarized capacitors, tension regulators, etc.) as well as on modifications of the power supply ground stage in order to feed each microboard and each microphone circuit with an independent power supply. The modification process was conducted in several steps, each revealing an objectively quantifiable improvement with respect to the previous one.

In the remainder of this work, the basic theory of microphone array processing will be introduced together with a detailed description of the array device. In particular, the computation of the coherence between microphone signals will be described with a technical discussion and related figures. Secondly, the basic hardware changes will be discussed with suitable details for a possible intervention on the circuitry in order to fix a similar platform. More details about this activity can be found in [6].

2. MICROPHONE ARRAYS

2.1. Microphone arrays

The use of a microphone array for distant-talking interaction is based on the potentiality of obtaining a signal of improved quality, compared to the one recorded by a single far microphone [7, 8, 9]. A microphone array system allows the talker message to be enhanced as well as noise and reverberation components to be mitigated, so that it can be used to achieve a hands-free human-machine voice interaction.

A microphone array consists of a set of acoustic sensors placed at different locations to spatially sample a sound pressure field. Using a microphone array it is possible to selectively pick-up a speech message, while avoiding the undesirable effects due to distance, background noise, room reverberation and competitive sound sources. This objective can be accomplished by means of a spatiotemporal filtering approach.

The directivity of a microphone array can be electronically controlled, without changing the sensor positions or requiring the talker to speak close to the microphones. Moreover, detection, location, tracking, and selective acquisition of an active talker can be performed automatically to improve the intelligibility and quality of a selected speech message in applications such as teleconferencing and hands-free communication (e.g. car telephony).

2.1.1. Beamforming

A beamformer exploits the spatial distribution of the elements of a microphone array to perform spatial filtering [10]. The microphone signals are appropriately delayed, filtered and added to constructively combine the components arriving from a selected direction while attenuating those arriving from other directions. As a consequence, signals originating from distinct spatial locations can be separated even if they have overlapping bandwidths.

[1] This work was partially supported by the EU under the Integrated Project CHIL and by the French MESR (Ministère de l'Enseignement Supérieur et de la Recherche).

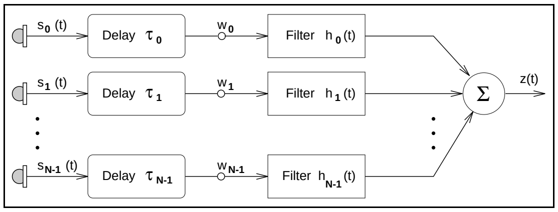


Fig. 1: Filter and sum beamformer.

Delay-and-sum beamforming is the simplest and most straightforward array signal processing technique. A proper delay τ_n is applied to each microphone signal $s_n(t)$, $n = 1..N$, in order to cophase the desired component by compensating for the difference in path lengths from the source to each microphone. Each signal amplitude can be weighted by a coefficient w_n in order to shape the overall polar pattern, and finally the N signals associated to N microphones are summed together as follows:

$$z(t) = \sum_{n=0}^{N-1} w_n s_n(t - \tau_n). \tag{1}$$

By adjustment of the delays, the array can be electronically steered towards different locations. Desired main beam width and sidelobe level can be obtained with a proper choice of the coefficients w_n. A delay, in the time domain, is equivalent to a linear phase shift in the frequency domain. Therefore the beamformer can also be implemented by a proper phase alignment in the frequency domain according to the relationship:

$$Z(f) = \sum_{n=0}^{N-1} w_n S_n(f)\; e^{\,-j2\pi f \tau_n}. \tag{2}$$

A more general beamforming structure is the *filter-and-sum* beamformer described by the equation:

$$z(t) = \sum_{n=0}^{N-1} w_n\; h_n(t) * s_n(t - \tau_n). \tag{3}$$

In this case, an additional linear filtering with impulse response $h_n(t)$ is inserted in each channel prior to summation, according to the scheme of Figure 1, in order to perform more complex spatiotemporal filtering and directivity shaping, especially when dealing with broadband signals (e.g. speech). Moreover, the transfer functions of the filters can be chosen according to the statistical characteristics of the desired signal and interfering noise.

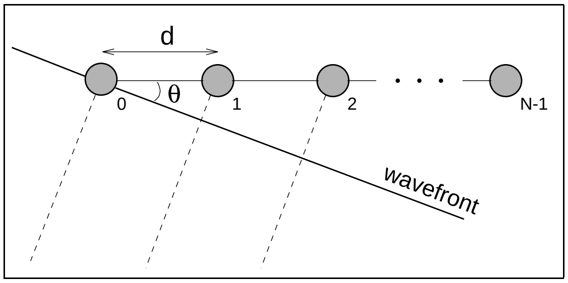


Fig. 2: Uniform linear array in far-field conditions.

The directional characteristics of the microphones, the locations of the array elements, and the overall array geometry provide additional degrees of freedom in designing the directivity pattern of a beamformer.

2.1.2. Uniform linear array

The uniform linear array is the most commonly used sensor configuration in multichannel signal processing. It consists of N transducers located on a straight line and uniformly spaced by a distance d.

If a source is far from the array (in the so called "far field" region), then the arrival direction of the sound wavefronts is approximately equal for all sensors, and the propagating field can be considered to consist of plane waves. If a wavefront reaches the sensors from a direction forming an angle θ with the normal to the array, as depicted in figure 2, then the relative delay τ_a between wavefront arrivals at two adjacent microphones is given by:

$$\tau_a \;=\; \frac{d}{c} sin\theta. \tag{4}$$

If the source is located close to the array (in the "near field" region), then the wavefronts of the propagating waves are perceivably curved with respect to the dimensions of the array and the arrival direction differs from an element to another. In this case, the relative delays of wavefront arrival at successive sensors lie on a hyperbolic curve.

More details on this topic can be found in [8]. In the following of this work we refer to the case of a medium size uniform linear array consisting of 64 sensors, with 2cm inter-microphone distance.

3. THE MARKIII MICROPHONE ARRAY

The NIST Microphone Array MarkIII [3] is an array of microphones composed by 64 elements, specifically developed for voice recognition and audio processing. It records synchronous data at a sample rate of 44.1 kHz or 22.05 kHz with a precision of 24 bits.

The particularities of this array are the modularity, the digitalization stage and the data transmission via an Ethernet channel using the TCP/IP protocol.

The array uses 64 electret microphones installed in a modular environment. Two main components constitute the system: a set of microboards for recording the signals and a single motherboard to transmit the digital data over the network. There are eight microboards in the array, and every microboard is connected to eight microphones. The first step done by the microboard is the polarization of the microphones and the amplification of the signals. Electret microphones need a phantom power to work properly and provide a low voltage signal. So the microboard adapts the signals to be converted in the digital format. The digitalization of the audio signals is done on each microboard, using four dedicated stereo analog to digital converters. The choice of putting the A/D converters as close as possible to the microphones reduces the possibility of having the analog signal disturbed by electrical interferences.

The task of the motherboard is to collect all the digital signals from the single microboards, multiplex them and pack all the data in a format suitable for being sent over the network. The motherboard uses an Ethernet channel to transmit the digital signals: it gets an IP address via a DHCP service and sends broadcast data on the network. If a PC needs audio signals from the array, it has just to contact the array using a certain protocol and read the data from the network card. Due to the huge amount of material (64 ch × 44100 samples/sec × 3 bytes = 8.07 MB/sec), it has been chosen to use the UDP protocol. This allows to transfer a big quantity of data, but lacks of integrity checks. If the receiving computer is momentarily not fast enough to read all the packets, some packets are simply lost and the recorded signal will contain discontinuities. A software protocol to resend the lost packets has been implemented but is not encouraged for the high chances to lose data again.

The weak part of the chain is the storing of the data on the computer. In theory it could be possible to connect the MarkIII array to a switch and then listen to the data from a generic computer on the network. But since the transmission volume is very high, a computer with a single network interface card is not able to get all the data and loses packets. This is a crucial aspect since missing samples in the signal lead to worse performance of any of the above mentioned technologies. The solution is to install a dedicated network card on a PC and connect the array directly to that machine. This leads to the loss of flexibility guaranteed by the Ethernet protocol, but at least allows to record seamless data. However, there is the necessity to tune the operating system for receiving a lot of UDP packets. This tuning could not be done for Microsoft Windows machines, forcing the array to be used only with UNIX/LINUX operating systems. The machine connected to the array has to be in any case pretty fast, able to store data without losing incoming packets. This leads to the necessity to have a dedicated machine only for data recording, while real-time processing seems not feasible at the moment.

4. THE CROSS-POWER SPECTRUM PHASE TECHNIQUE

A microphone array performs a spatial sampling of the acoustic wavefronts propagating inside an enclosure. It is often of interest the capability of comparing the signals captured by different microphones in order to calculate a degree of similarity between them as a function of their mutual delay. Given two microphones and their related signals s_i and s_k, it is possible to define a Coherence Measure (CM) function $C_{ik}(t, \tau)$ that expresses for each delay τ, the similarity between segments (centered at time instant t) extracted from the two signals. While the two microphones are receiving the wavefronts generated by an active acoustic source, this function is expected to have a prominent peak at the delay corresponding to the direction of wavefront arrival (e.g. positive if the source is on the left and negative if it is on the right). For each microphone pair a bidimensional representation of the CM function can be conceived. In this representation horizontal axis is referred to time, vertical axis is referred to delay and the coherence magnitude is represented by means of a "heat" palette (see for example Fig. 6 and Fig. 8). A particularly convenient CM function can be obtained starting from a Crosspower Spectrum Phase (CSP) analysis [5, 11], also known as PHAT transform, a particular case of Generalized Cross Correlation [4]. The procedure for estimating a CSP-based Coherence Measure (CSP-CM) starts from the computation of the spectra $S_i(t,f)$ and $S_k(t,f)$ through Fourier transforms applied to windowed segments of signals s_i and s_k centered around time instant t. Then these spectra are used to estimate the normalized Crosspower Spectrum:

$$\Phi_{ik}(t,f) = \frac{S_i(t,f) \cdot S_k^*(t,f)}{\|S_i(t,f)\| \cdot \|S_k(t,f)\|} \tag{5}$$

that preserves only information about phase difference between s_i and s_k. Finally the inverse Fourier transform of $\Phi_{ik}(t,f)$ is computed:

$$C_{ik}(t,\tau) = \int_{-\infty}^{\infty} \Phi_{ik}(t,f) e^{j2\pi f\tau} df \tag{6}$$

The resulting function (considered as dependent of the lag τ) is the transform of an all-pass function and has a constant energy, mainly concentrated on the mutual delays at which there is high correlation between the two channels. The CM representation is very useful when it is necessary to locate the acoustic source or to analyze the multipath propagation inside a room, as delays associated to direct wavefront and to principal reflection are easily detectable. The same representation turns out to be extremely advantageous also to analyze the mutual "independence" between the acquisition channels of an array. In the fact problems as cross-talk or common mode noise components generated within the acquisition device are clearly put into evidence by the appearing of graphical patterns (i.e. lines) in the CM that otherwise, in a quiet environment, should be rather uniform along the τ coordinate.

5. THE MARKIII/IRST BASED ON BATTERIES

In this section we describe the problems we encountered with the array, originally designed at NIST [3], and how we solved them. It is worth noticing that, for project constraints, the underlying purpose of this initial improvement was to obtain a performant device in a short time, no matter how complex, costly or reproducible the solution would be. This improvement led us to obtain a first new prototype of MarkIII, from now on called "MARKIII/IRST". Each of the following subsections describes how the disturbances were eliminated one by one. In some cases the final solution was obtained after many subsequent trials, fully described in [6]. Of all the problems solved in [6], only a subset related to the quality of the speech signals acquired is presented here.

5.1. Early saturation effect of microphones

It was observed that when a speaker was near the array the microphone signals immediately saturated. One could guess that the Panasonic microphones were too sensitive or the OPAMPs were pushed to the limit. In any case, the device did not allow to control input levels. Moreover, it is worth noting that some microphones were more sensitive than others. The biggest ratio from the most sensitive (ch 35 and ch 8, respectively, in the array available at ITC-irst) was of 2:1, i.e. 6 dB in amplitude. Since no trimmer or other regulations of the input level were available, we eventually decided to physically bypass the first amplification stage as described in the following and shown in Figure 3. For comparison purposes, one can find in [3] the original layout of the NIST device.

The two capacitors C1 and C6, placed at the very beginning and at the very end of the amplification stage, were 1 μF polarized capacitors in the original design. They were substituted with two polyester 0.47 μF capacitors, which generate much less noise. The first amplification stage was then bypassed via a 0.47 μF capacitor, keeping the second stage polarization to the phantom GND with a 100 kΩ resistor. As a result, the original gain of 68 was reduced to 6.8, which is suitable to avoid any signal clipping.

5.2. 50 Hz disturbance

In our preliminary recordings (done in an insulated room) we observed the presence of a perceivable 50 Hz interference. We realized the disturbance was

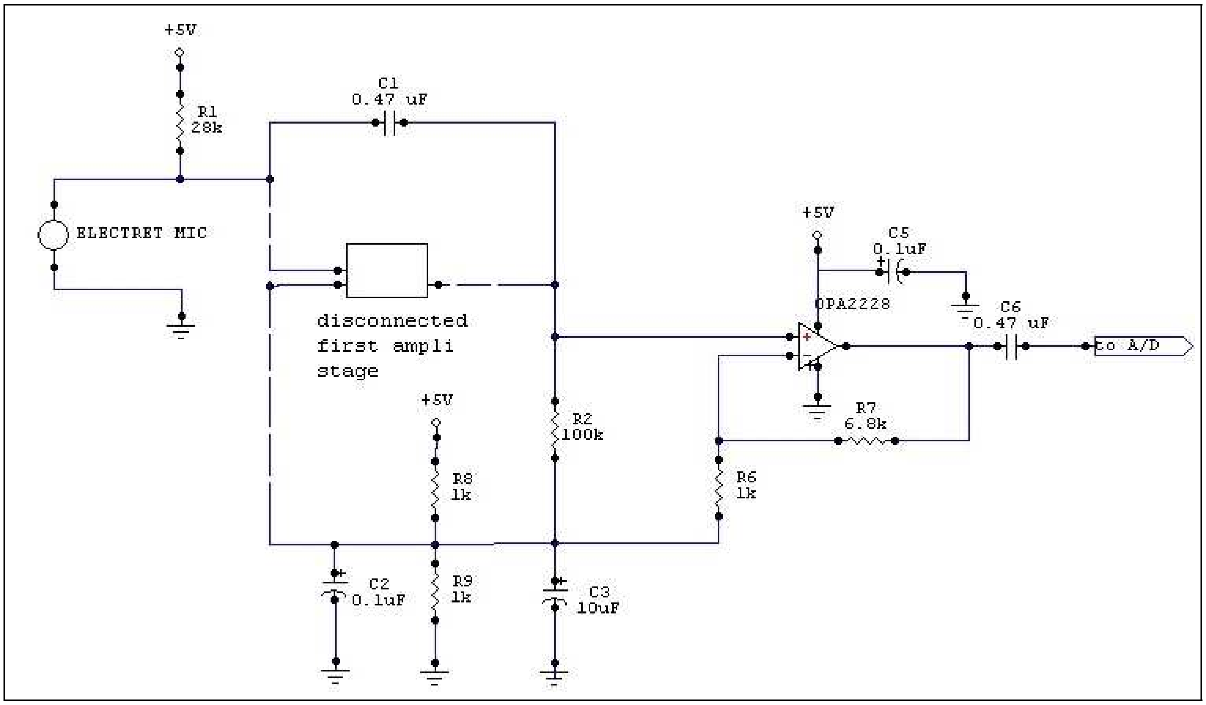


Fig. 3: Modifications of the amplification stage in the first prototype (MarkIII/IRST).

due to the power supply: this problem was solved by substituting the 220V-AC to 9V-DC power adaptor, provided with the array, with a Pb rechargeable battery. This was not the final solution, as in a second step we solved the device noise problem (see Section 5.3) by switching to a battery power supply for the whole analog part. It is worth noting that, even with the best battery-based power supply available, still a light 50Hz disturbance persisted: it was much lower than the one coming from the AC current and it was totally due to environmental electromagnetic fields. By consequence it was definitely eliminated by surrounding the MarkIII with a Faraday cage.

5.3. Device noise

The device noise represents the major obstacle to the use of the MarkIII for speaker localization and beamforming purposes. It is also subtle to detect, as this problem is neither perceivable in normally reverberant rooms nor evident through waveform or spectral analysis of a single channel.

The device noise problem was evident once eliminated the 50 Hz interference (see Section 5.2). In other words, the following experiments regard the use of the MarkIII array powered by a rechargeable battery and installed in a very quiet insulated room. The room is characterized by less than 30 dBA background noise level (that is very close to the acoustics of an anechoic chamber) and a reverberation time lower than 100 ms. Recordings were done at 44.1 kHz. As discussed below, the electrical problem can be revealed both at single channel level (perceptually evident through listening tests) and at inter-channel correlation level (through inter-channel coherence measurements) analysis.

5.3.1. Single channel analysis

The device noise can be perceptually detected only in recordings taken in a very silent room, because in this condition it can be distinguished from real background noise. Alternatively it can be detected, without the need of an anechoic chamber, by manually detaching the microphones from the boards: the signals acquired from the array is then only pure noise coming from the devices. The effect of the device noise can be observed in Figure 4, where two

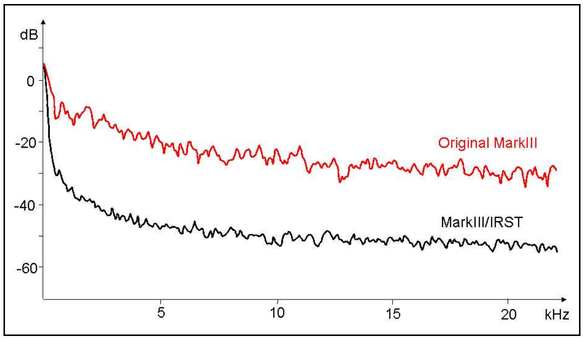
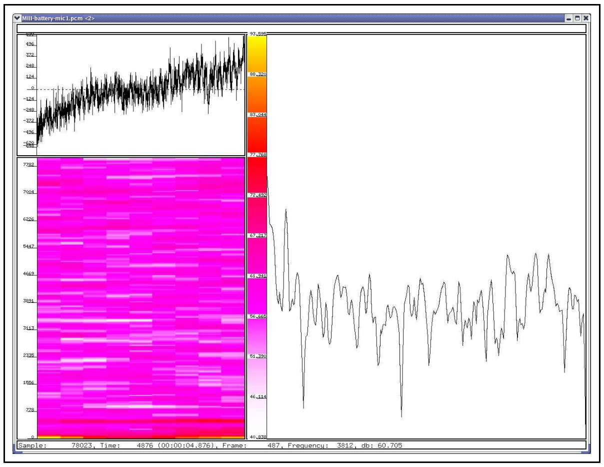


Fig. 4: Spectra corresponding to a 600 ms of background noise. The red, upper line hints at the signal quality of the original MarkIII, while the black, lower one hints at the signal quality of the MarkIII/IRST. A reduction of 20 dB is evident at most of the frequencies.

average spectra of 600 ms of silence sequence are provided. The red, upper line is relative to a single channel of the original MarkIII array. The black, lower line is relative to a silence sequence of the same length recorded with the MarkIII/IRST. The environmental conditions were approximately the same, but clearly the device noise affects the whole spectrum. According to the given figures, more than 20 dB noise reduction was obtained at almost all the frequencies. Another very detailed analysis was done by shortcutting each microphone input in order to measure only the board circuitry noise and, also in this case, a noise reduction of about 15-20 dB was observed. To better understand the entity of the noise, Figure 5 is related to some silence collected in the ITC-irst insulated room. From Figure 5, one can observe that the noise dynamics (between -300 and +300) involves about 9 bits. It was clear that losing 9 bits out of the first 16 most significant ones was a heavy limitation to the potential of this array.

5.3.2. Cross-channel analysis

An analysis of the CSP (described in Section 4) between pairs of channels put into evidence other problems related to the so called "device noise". This noise component, which can be observed in all the channels, is neither acoustic noise nor transduction noise of the microphones. It dominates over acoustic background noise of a relatively quiet environment.



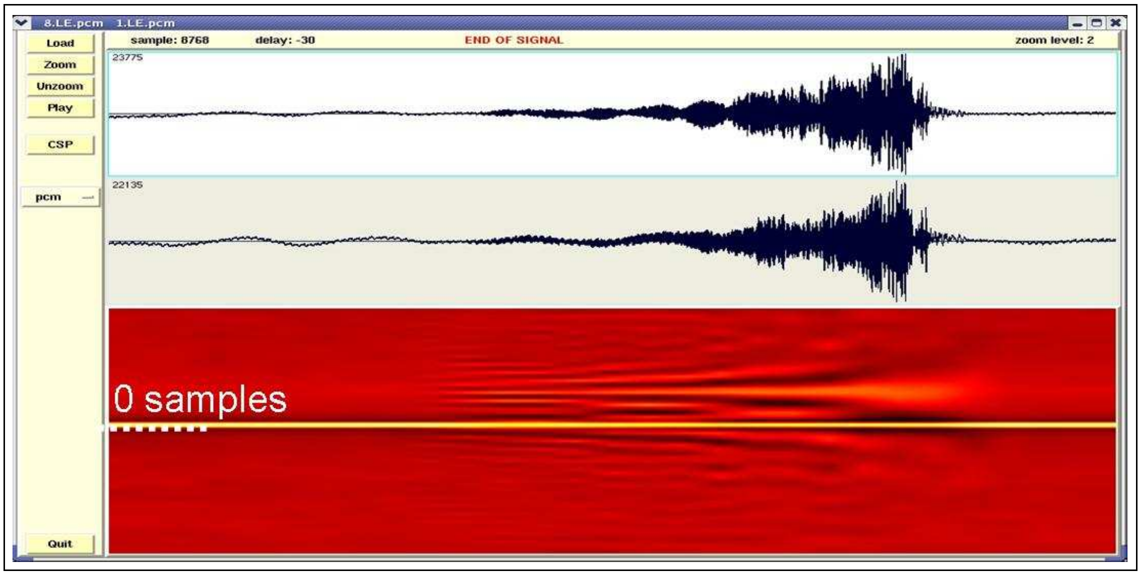
```
Sample:      78023, Time:    4876 (00:00:04.876), Frame:     487, Frequency:  3812, db: 60.705
```

Fig. 5: Analysis of a background noise sequence of 32ms length. The lower left part of the figure reports the spectrogram. The log power spectrum is given in the right part. The device noise is here more evident both in its dynamics and in its spectral characteristics. Note that the slope of the signal is due to a 2.5 Hz interference characterizing the given recordings.

It exhibits a "common mode" within the 8 channels of each array microboard. Different modules (e.g. from channel 1 to channel 8, and from channel 9 to channel 16) have different and uncorrelated noise components. This is evident on the basis of a CSP analysis.

Figures 6 and 7 show the noise coherence between channels 1 and 8, which was derived from the analysis of a chirp-like signal reproduced through a hi-fi loudspeaker placed at the left side of the array: in this case, a strong coherence is evident between the (mainly electrical) noise sequences. A strong coherence at 0 samples is equivalent, for any localization algorithm, to determine a direction of arrival from an acoustic source right frontal to the array. In practice, the device noise takes all the energy of the CSP and concentrates it where no sources actually exist. Figure 7 specifically shows how the artificial peak, at 0 samples, dominates the secondary, true, peak located at +5 sample delay.

On the other hand, the same analysis repeated on channels 1 and 9, which are on two different microboards and therefore have no common mode noise, demonstrates the absence of any coherence at


8.LE.pcm 1.LE.pcm

sample: 8768 delay: -30 END OF SIGNAL zoom level: 2

Fig. 6: Chirp signals acquired in an insulated room before the intervention on the device. As the two channels belonged to the same microboard, there is a high peak of CSP function at 0 samples inter-microphone delay, which masks the true peak: this means a strong coherence between the device noise sequences.

any particular delay.

5.3.3. Device noise removal

The single and cross-channel analysis clearly show the effect of the device noise. We describe in the following how we detected its origin and how we eliminated it. It is worth noting that a better solution was found with the next prototype, the MarkIII/IRST-Light. The device noise was caused from the tension regulator LM2940 (see technical documentation of Mark III in [3]). There is one such a regulator for each of the 8 microboards. This tension regulator provides the operation voltage to 8 Panasonic microphones, to 4 A/D converters and to 8 OPAMPs. As mentioned in Section 5.3.2, the device noise has a common mode within the 8 channels of each array microboard.

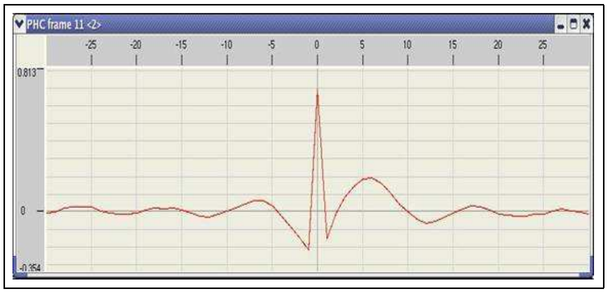


Fig. 7: A slice of the CSP-gram in a fixed instant shows the artificial peak of the CSP, which masks the true one, located at a 5 samples delay.

In order to keep the original device layout, the problem was solved by physically removing such regulators and feeding the analogue part of every board directly with a circuit of battery designed ad hoc, while the digital part remained fed with a new transformer stabilized and filtered ad hoc. It is worth noting that part of the device-noise is caused by the LM2940 and part by the surface-mounted polarized capacitors, which should theoretically remove the regulator noise. These capacitors have an inner leakage current which creates the necessary oxide between the armors, thus generating a disturbance. Hence, they were substituted with polyester capacitors, which are bigger but generate much less noise. An effective solution was to feed the analogue part of each microboard with 4 × 1.2V, 5Ah batteries (a total of 32 batteries), so to guarantee the galvanic

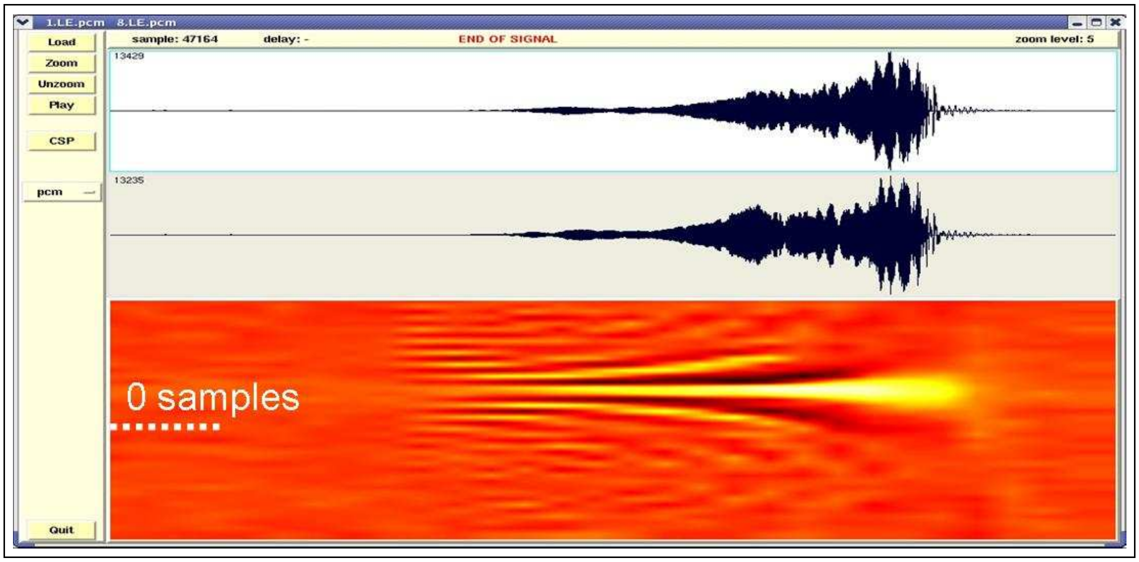


Fig. 8: Signals extracted from Channel 1 and Channel 8 after our intervention. The peak of the CSP function reported in the lower part of the figure shows a strong coherence only when the chirp is played.

de-coupling of each power supply source. Nevertheless the best solution (see Section 6) turned out to be rising the power supply ground of each microphone with respect to the real ground. The use of external batteries required an analysis of power consumptions prior to any decision about the components to buy. This analysis, together with a history of our several trials and the corrected layouts of the circuitry around the removed tension regulators, is detailed in [6].

After our intervention the device noise disappeared: Figures 8 and 9 have to be compared with Figures 6 and 7 respectively. The delay in samples at which the chirps arrive at the two microphones is clearly detected. In fact, by comparing the CSP coherence measures of Figures 6 and 8, it is evident that the constant yellow stripe at zero samples, caused by the device noise, has disappeared completely. With the new device, the coherence representation is now highlighting the true interchannel delay (i.e. +5 samples). For a single frame, this fact is evident in the main peak depicted in Figure 9.

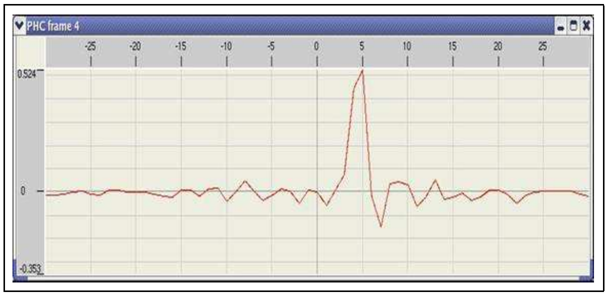
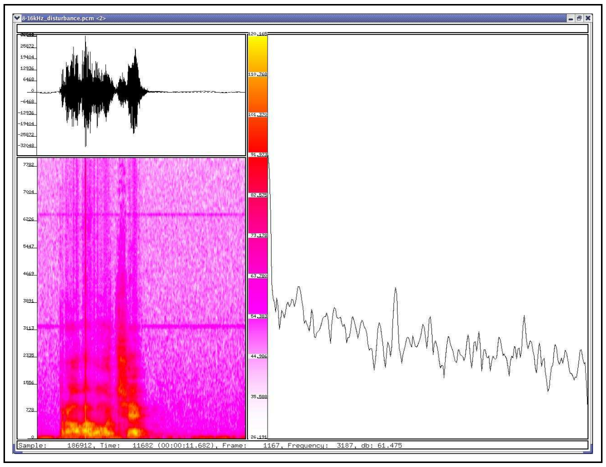


Fig. 9: A slice from the CSP-gram in a fixed instant reveals now the true peak at a 5 samples delay. The device noise is totally absent.

5.4. 8 kHz and 16 kHz common ground noise

A further problem was observed by analyzing the spectrogram of some utterances. This problem became evident once both the 50 Hz and the device noise problems were solved. Two disturbances at about 8 kHz and 16 kHz appeared in the spectrogram, as shown by Figure 10: two relatively strong stripes appear in red and violet in the spectrogram on the left part of the picture, which correspond to the two peaks evident in the right part.

Though the disturbance was present at frequencies not closely related to the speech signal, it was verified that it did not come from the environment and it was then worth to investigate, as it represented another common mode noise component across differ-



```
Sample:     186912, Time:   11682 (00:00:11.682), Frame:     1167, Frequency:  3187, db: 61.475
```

Fig. 10: The 8 kHz and 16 kHz disturbance peaks are evident in the right part of the picture, where the spectrum of a silence segment is taken after the utterance depicted in the left part. Notice the absence of the device noise, removed as described in Section 5.3

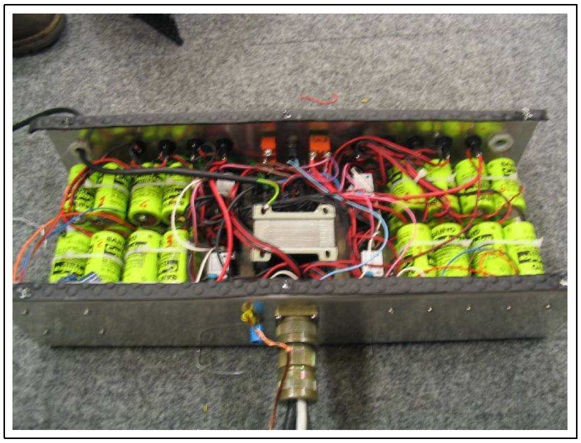


Fig. 11: Inside of the power supply box: from the 8 groups of 4 batteries the power supply passes through the red and violet cables, placed on purpose in those positions. The transformer, which provides power supply for the digital part (in acquisition state) and recharges the batteries (in recharging state), appears in the center of the box.

ent channels preventing a clean data collection. We discovered it was due to the coupling between the digital and the analog ground. This coupling was made around the A/D converter PCM1802: the device was originally provided with two separate pins for the two grounds. In the original project of the MarkIII the two pins were connected via a short circuit. This makes the analog ground, which the audio signal relies upon, coincident with the digital ground, which collects the noise coming from the various integrated devices, such as the A/D converter and the two tension regulators. The final solution consists in avoiding the common ground by feeding each microboard separately with an independent group of batteries, thus obtaining 8 groups of 4 x 1.2V, 5Ah batteries. Figure 11 shows the battery box entirely built at ITC-irst. More pictures and details are available in [6].

6. THE MARKIII/IRST-LIGHT

This section reports on further improvements of the MarkIII/IRST. For the purpose of making the modifications we did in the MarkIII/IRST easily reproducible by an expert in electronics in every laboratory of the CHIL project [1], we were motivated to find another solution. The new prototype, from now on called “MARKIII/IRST-Light”, solves the same problems reported in section 5 in a very efficient, cheap and replicable way. It even performs better than the MARKIII/IRST in terms of SNR and coherence measures: see details in [12]. The multichannel corpus being collected at University of Karlsruhe [1], is based on the use of this improved release of the device.

6.1. Manual gain correction

In order to better exploit the acquired signal dynamic range, in the new layout (Figure 12) we chose to keep both the amplifiers while reducing the total gain and making it tunable: the potentiometer R11 allows the total gain to be in the range 12÷16.7 (R11’s nominal value gives a total gain of 15), which is both a compromise between good amplification and clipping avoidance, and a way to cope with the different sensitivity of the electret Panasonic microphones. Notice that R11 must be of high quality (possibly of plastic-film type).

6.2. High impedance microphone power supply

The main purpose of the MarkIII/IRST-Light is to reduce complexity and cost while keeping, and possibly improving, performance. It was realized that

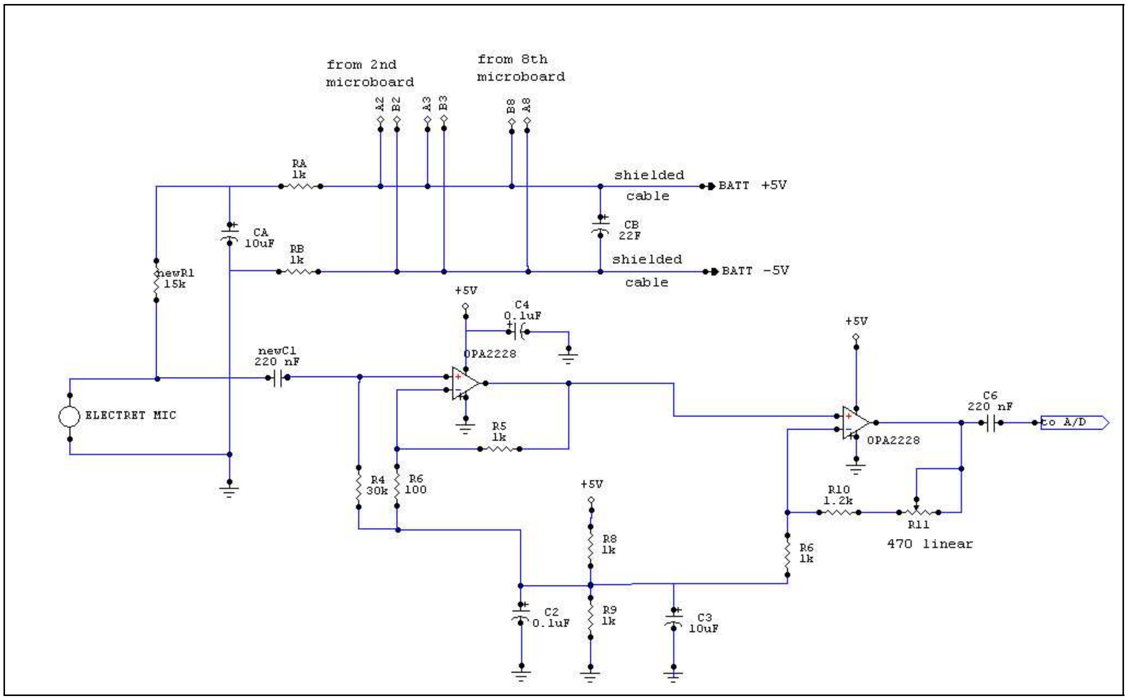


Fig. 12: Modifications of the amplification stage in the MarkIII/IRST-Light. Notice the high impedance power supply stage, which connects each group of 8 microphones on the same microboard to a dual positive-negative power supply.

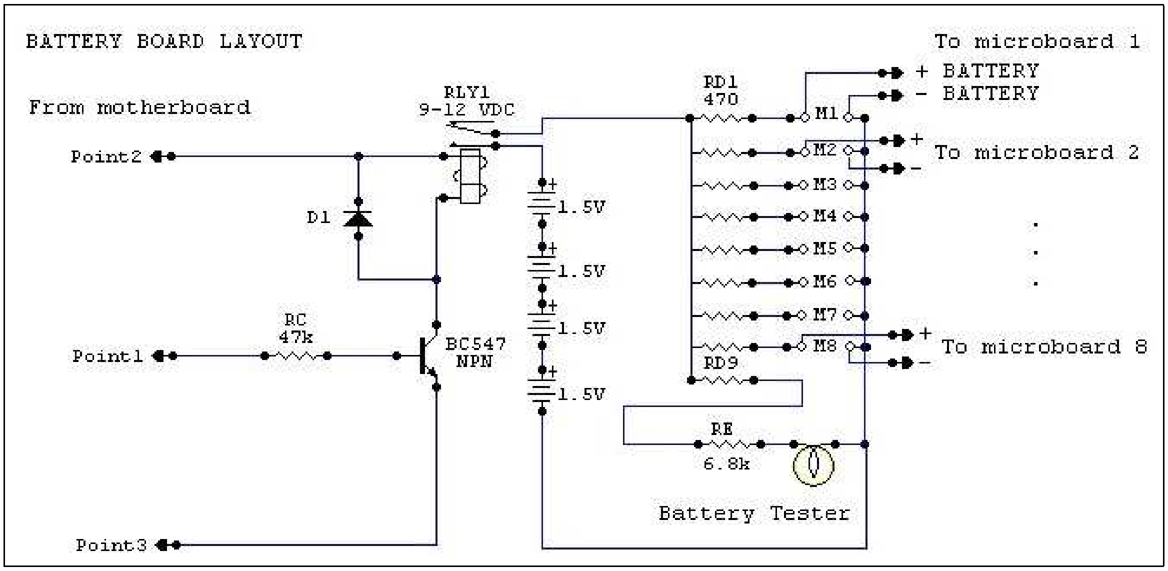


Fig. 13: Battery saver microboard layout.

performance could even be improved with a different approach, taking into account that noises, circulating both on the analog and on the digital ground, could be deviated instead of suppressed. In order to feed microphones with a very clean supply, a high impedance path was designed for the DC coming from the batteries and each microphone power supply ground level was rised with respect to the real ground. A typical π RC cell scheme was built via R1, R2, R3 and C1. This is feasible because the electret microphones power consumption is very low. Notice that:

- C1 and CB are preferably of Tantalium type;
- C2 and C7 are preferably of Polyester type;
- there is one CB every 8 microphones, i.e. one per microboard.

6.3. Battery saver microboard

We built and inserted into the Faraday cage a further microboard (Figure 14) to power the microphones only when the MarkIII is acquiring signals: this is simply done by letting this microboard be driven by the Capture Led ([3], page 40). The purpose is to let batteries last as long as possible: we placed a series of 4 Alcaline batteries. The new microboard amplification stage layout is depicted in Figure 13. We estimated the MarkIII/IRST-Light can continuously acquire for 150 hours with this configuration, but one could freely make the series voltage be in the range [4,5 - 9V] or different combinations series-parallel to increase the duration. The battery saver microboard needs three signals from the motherboard: "Point 1" is the signal coming from the Capture Led, "Point 2" is the motherboard power supply for the relais, "Point3" is the motherboard GND. A small battery tester was added to check the batteries state.

7. CONCLUSIONS

This work reported on a recent activity conducted at ITC-irst laboratories which allowed us to realize a new release of the NIST MarkIII microphone array.

The current prototype is able to provide clean signals that are suitable for speech enhancement as well as for automatic speaker localization purposes, thanks to improved characteristics in terms of coherence among different channel signals. The MarkIII/IRST Light prototype is presently used at the University of Karlsruhe to record a large corpus of seminars and meetings for benchmarking of various speech and acoustic related technologies under study inside the Integrated European CHIL project [1].

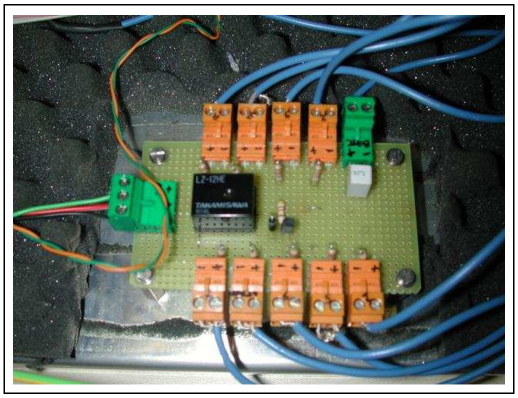


Fig. 14: Battery saver microboard, inserted in the Faraday cage of the array.

Moreover, the new hardware layout is being used at NIST to produce a new generation of MarkIII arrays, on the basis of the above described interventions. The resulting analogue circuitry, together with a very effective digital section formerly designed and realized by NIST, makes the new device a very useful tool for future research and prototyping in the field of microphone arrays and distant-talking interaction.

8. REFERENCES

[1] URL: http://chil.server.de

[2] D. Macho et al. "First experiments of automatic speech activity detection, source localization and speech recognition in the CHIL project", *Hands-Free Speech Communication and Microphone Arrays Workshop*, CAIP, Piscataway, 2005.

[3] Cedrick Rochet,
URL: http://www.nist.gov/smartspace/toolChest/cmaiii/userg/Microphone_Array_Mark_III.pdf

[4] C.H. Knapp and G.C. Carter, “The generalized correlation method for estimation of time delay”, IEEE Trans. on Acoustics, Speech and Signal Processing, vol. 24, n. 4, pp. 320–327, 1976.

[5] M. Omologo, P. Svaizer “Acoustic event localization using a Cross-power Spectrum Phase based technique”, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Adelaide, 1994.

[6] C. Bertotti et al.,
URL: http://www.eurecom.fr/∼brayda/MarkIII-IRST.pdf

[7] J. Flanagan, D. Berkley, G. Elko, J. West and M. Sondhi, “Autodirective Microphone Systems”, *Acustica*, vol. 75, pp. 58–71, 1991.

[8] M. Omologo, P. Svaizer, R. De Mori, ”Acoustic Transduction”, in R. De Mori (ed.) “Spoken Dialogues with Computers”, pp. 23-67, Academic Press, London, UK, 1998.

[9] M. Brandstein and D. Ward (eds.) “Microphone Arrays Signal Processing Techniques and Applications”, Springer-Verlag, 2001.

[10] B.D. Van Venn and K.M. Buckley, “Beamforming: a versatile approach to spatial filtering”, *IEEE ASSP Magazine*, April 1988.

[11] M. Omologo, P. Svaizer, “Use of the cross-power-spectrum phase in acoustic event location”. IEEE Trans. on Speech and Audio Processing, vol. 5, n. 3, pp. 288–292, 1997.

[12] C. Bertotti et al.,
URL: http://www.eurecom.fr/∼brayda/MarkIII-IRST-Light.pdf