Application Report

April 2001

Noise suppression and signal enhancement for speech

CASTELLI Eric *CLIPS/IMAG*

ISTRATE Dan *CLIPS/IMAG*

NGUYEN Quoc-Cuong *CLIPS/IMAG*

VAUFREYDAZ Dominique *CLIPS/IMAG*

ABSTRACT

Nowadays, one challenge for hands-free mobile phone and/or for man-machine dialog (especially speech recognition) is the environmental conditions. In one hand, rooms produce reverberations, which alter the signal to be recognised. In an other hand in noise conditions the rate of recognition strongly decreases.

Our application is a part of project RESIDE-HIS who is a telemonitoring system in a habitat equipped with physiological sensors, position encoders of the person, and microphones. Its principal goal is telemonitoring activity of a patient in order to decide, if necessarily, to announce the emergencies.

In the framework of this project, because the parts of the apartment have an important and variable influence on the signals collected by the microphones, degrading them, we decide to use two signal preprocessing modules : echo cancellation module and blind separation of sources module.

The influence of the room can be modelled using the transfer function *h(n)* as the response of a filter. The CMS (Central Mean Substraction) method allows evaluating this function by deconvolution through cepstres. After evaluation of *h(n),* deconvolution proceeded on the signal eliminates the echo.

In order to separate speech and noise, we use a blind separation algorithm of acoustic sources. Our hypothesis is that the output signal of a microphone is the result of an unknown convolute mixture of unknown primitive signals (sources). The signals mixture is modelled by causal FIR filters. The algorithm is based on crossed cumulants of order 4 and it tries to identify the coefficients of the source mixture in order to separate them by a neuromimetric filter.

These two pre-processing modules were designed with the floating point DSP **TMS320C6701.** This DSP give us a sufficient precision especially for our blind separation of sources. The fact of being in floating point gives facility of programming, and decreases the computation time. We use the **TMS320C6701EVM** board, who allow a real time data exchange with host PC. The two modules were designed in order to work independently or coupled.

This document was an entry in the DSP Challenge 2000, an annual contest organized by TI to encourage students from around the world to find innovative ways to use DSPs. For more information on the TI DSP Challenge 2000, see TI's World Wide Web site at www.ti.com/sc/dsp_challenge.

Key words: enhancement of speech signal, blind source separation, elimination of the echo, cepstre, neuromimetric filter.

Contents

Figures

Tables

Introduction

The application that we present is a part of project RESIDE-HIS of our laboratory with the goal to realize of a smart home for the patient or eardly people monitoring. Our application is a subsection of a telemonitoring system.

Presentation of entire project

Introduction

Information technologies made enormous progress these last years and they are used in all industrial or public fields. Thus, it is not astonishing that medicine is more and more interested in these information technologies and proposes a new form of "remote" medicine, or telemedecine. This one is announced as a significant reform of the medical care because it allows to improve the response time of the specialists face to the emergencies. They could be informed about a medical emergency and react as the first symptoms appear or they could intervene without waste of time as soon as a serious situation is announced. The telemedecine also allows a significant reduction of the costs of public health by avoiding the hospitalization of the patients for long period of time.

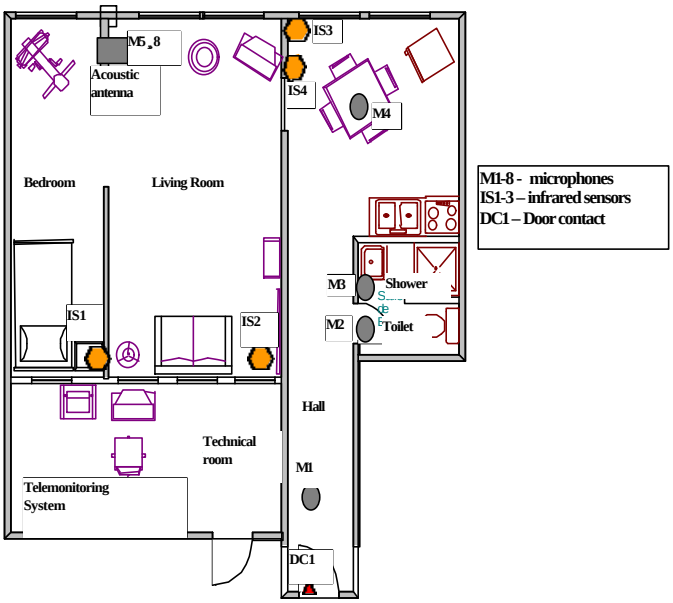
Telemedecine consists in associating electronic techniques of monitoring, with computers' "intelligence" and with the speed of telecommunications (established either through network or radio connections). A patient in convalescence, a pregnant woman carrying a difficult pregnancy, a person suffering from a chronic disease or the elderly, are examples which can be treated by telemedicine. In this way, they are constantly connected to a medical center charged to analyze the information collected by the telemonitoring system and to take the decision to act if needed [1]. The system we work on is conceived for monitoring the elderly. Its main goal is to detect serious accidents as falls or faintness (which can be characterized by a long idle period of the signals) at any place in the apartment [2].

We noted that the elderly had difficulties in accepting a monitoring by video camera, because they consider that their constant recording is a violation of their privacy. The originality of our approach consists in replacing the video camera by a system of multichannel sound acquisition charged to analyze in real time the sound environment of the apartment in order to detect abnormal noises (falls of objects or of the patient), calls for help or moans. These could characterize a situation of distress in the habitat and could indicate to the emergency services that they have to intervene.

Presentation of telemonitoring system

The habitat we used for experiments is situated in the TIMC laboratory buildings, at the Michalon hospital of Grenoble. The diagram of the apartment and the position of the sensors are given in Figure 1:

Figure 1. Position of the sensors inside the apartment



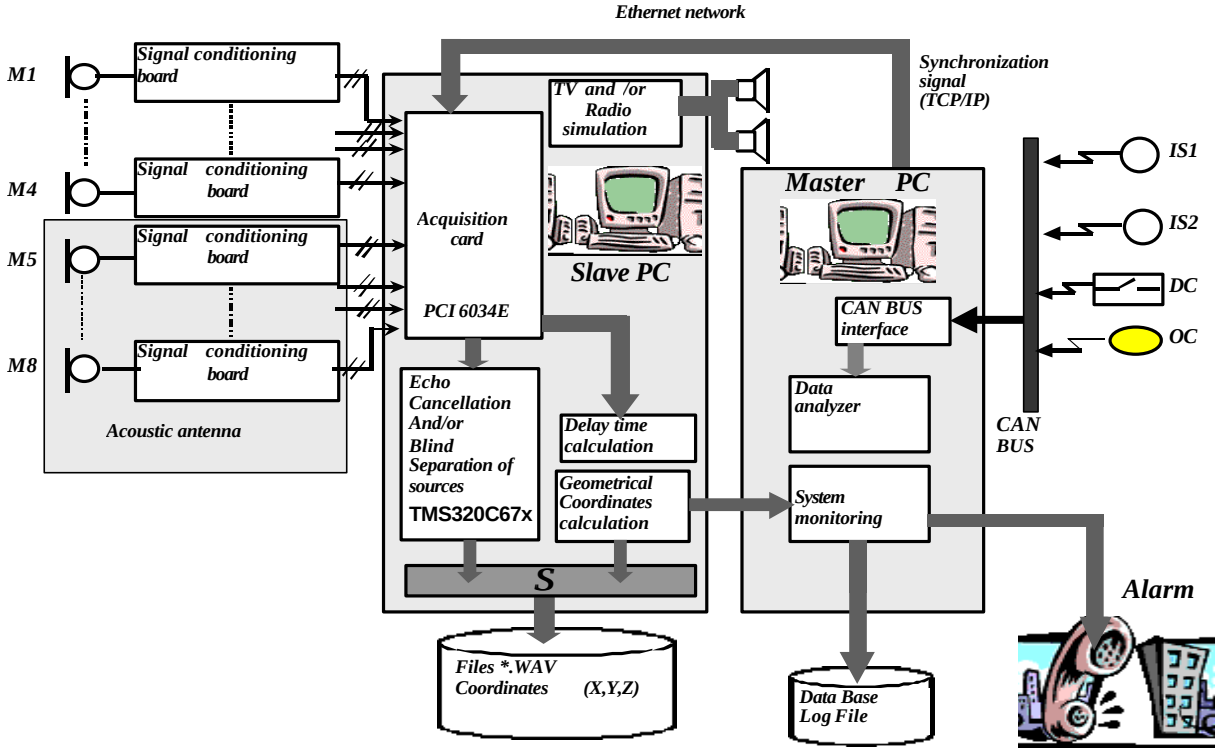
The multi-sensor system consists of several types of sensors. One or several sensors are carried by the person and provide information about his/her activities, i.e. vertical position (standing) or horizontal (lying) of the person and fast movements of the person (possible fall). It is possible to add to these sensors a heart monitoring (ECG)[3], [4]. The localization sensors with infra-red radiation are installed in each part of the apartment in order to establish where the person is at any moment. These sensors communicate with the acquisition system by radio waves. The control of the activities sensors is ensured by a PC using a monitoring software programmed in JAVA.

The sound sensors are represented by eight microphones. Taking into account the configuration of the apartment we decided to install an acoustic antenna composed of four microphones, as indicated in Figure 1, in order to observe in the same time the living room and the bed room. The acoustic antenna will be used as well to localize the person in these two rooms and to analyze the sound sources. As the other rooms are much smaller, only one microphone per room is sufficient: in the middle of the kitchen, at the beginning of the hall, in the toilet and the shower-room. The microphones used are omni-directional, condenser type, of small size and low cost. A signal conditioning card, consisting of an amplifier and an anti-aliasing filter is associated to each microphone. The connections transmitting the audio signals are carried out by soundproofing differential pairs of professional audio quality.

The principal part of the acquisition system consists of a multi-channels acquisition card PCI 6034E of National Instruments, installed inside a second computer. This card has an analog-to-digital converter on 16 bits, it has eight differential inputs and a maximum speed sampling rate of 25 KHz on each channel. However, we work at a sampling rate of 16 KHz, a frequency usually used in speech applications.

After digitalization the sound data is saved in real time on the hard disk of the host PC in a temporary file of binary type (the only limitation being the space on the hard disk). However, we use signal preprocessing procedures, as the cancellation of the room reverberation, when the audio signals are degraded by the environment of the apartment. The PC used for the acquisition of the medical activity is the " master " and the PC used for the sound acquisition is its " slave ". The two computers are connected between them by an Ethernet network. Synchronization between the two systems is obtained by TCP/IP protocol. The process of acquisition works within a thread, parallel to that of synchronization. However, it is also possible to launch the synchronization of acquisition by an external signal TTL compatible. The general outline of the acquisition system is presented in Figure 2.

Figure 2. Diagram of telemonitoring system



Modules of signal preprocessing

Introduction

As we already mentioned, because the parts of the apartment have an important and variable influence on the signals collected by the microphones, degrading them, we use two preprocessing signal module implemented on TMS320C6701 DSP of Texas Instruments : echo cancellation and blind separation of sources. These two modules can work independently or coupled (first the blind separation of sources module and after the echo cancellation module).

Our two algorithms for signal preprocessing require a longer processing time (because of their complexity) and a good numerical precision. Taking into account these facts, we chose the TMS320C6x family from Texas Instruments. This signal processor family is one of the most powerful on the market, with a processing capacity of 1.5GIPS. We had two options: TMS320C6201 and TMS320C6701, these DSP being the most powerful of the family. We finally chose the DSP in floating point because of its high accuracy.

Because our research does not aim serial production, we decided to use the development card of the processor and not a card designed by ourselves. Besides, the development card from Texas Instruments is provided together with a modern and efficient development software (CCS). This software allows, among others : to debug in real time, to observe on the screen the loading of the processor, to exchange in real time data with the host PC etc.

The first module : Blind separation of source

Introduction

Blind source separation is a subject which has been much studied by the scientific community [7],[8]. Many papers have been published, and numerous solutions have been proposed. Our module built upon this knowledge base.

Theory

We consider a system of N independent and unknown sources, with P recording microphones. If we assume an convoluted mixture of sources, this system can be modelled mathematically as such : $Y(t)=A(t)*X(t)+N(t)$ where A(t) is an unknown matrix of linear filters, and N(t) is the additive noise vector. The star "*" denotes convolution.

If the filters of A(t) can be considered as finite impulse response filters (FIR) of order L, the mixture can be described in the z-plane by the equation : $Y(z)=A(z)\cdot X(z)+N(z)$ where the elements of the matrix A are defined by: $A_{ij}(z)=\sum_{k=0}^{L} a_{ij}(k)\, z^{-k}$

For the simplified model we make the same simplifying assumptions as before ([8]) :

- The transmission channel, mixtures to microphones, is noiseless : N(t) = 0
- The number of sources and microphones is fixed to 2 : N = P = 2

With the convoluted mixture, we will continue to work in the z-plane. We want a vector S, whose components are the two separated sources. This vector will be obtained by linear filtering of the observations vector Y. The two order L linear filters which will be used can be described :

$$S_1(n)=Y_1(n)-\sum_{k=0}^{L} c_{12}(k)S_2(n-k) \text{ and } S_2(n)=Y_2(n)-\sum_{k=0}^{L} c_{21}(k)S_1(n-k)$$

Giving, in the z-plane:

$$S(z)=\left(I+C(z)\right)^{-1}Y(z)=(I+C(z))_{-1}A(z)X(z)=H(z)X(z)$$

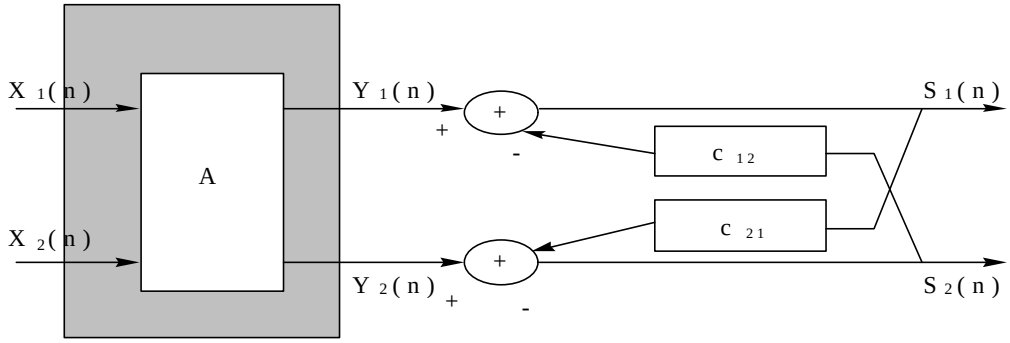
$$H(z)=\frac{1}{1-C_{12}(z)C_{21}(z)}\begin{pmatrix}1-C_{12}(z)A_{21}(z) & A_{12}(z)-C_{12}(z)\\ A_{21}(z)-C_{21}(z) & 1-C_{21}(z)A_{12}(z)\end{pmatrix}$$

Using a normalised A(z) matrix, we have two possible conditions for separation :

$$C_{ij}(z)=A_{ij}(z) \quad \text{or} \quad C_{ij}(z)=\frac{1}{A_{ij}(z)} \quad \text{with} \quad i\neq j$$

However, if we wish C_{ij} (z) to be an finite impulse filter, only the first solution is possible.

Figure 3. The diagram of blind separation algorithm



Once again, the matrix A is unknown, and so the filters C_{IJ} must be evaluated using an adaptive algorithm. In this more complicated model, 2L as opposed to 2 coefficients must be calculated as each new sample arrives. By analogy to the instantaneous case, the adaptation rule is :

$$C_{ij}(n+1,k)=C_{ij}(n,k)+\mu\, Cum_{31}\left(s_1(n),s_2(n-k)\right) k\in[0..L-1]$$

with $s_i=S_i(n)-E[S_i(n)]$

Using an approximation for the cumulant ([8]) , we obtain:

By varying μ, the adaptation gain, the performance of the algorithm can be improved through faster convergence and lower residual noise. For non stationary signals, μ must follow this rule :

$$0<\mu(n)<\frac{1}{MP_S(n)}$$

Here, P_s is the average power 4 of the separated output signal, and M is the number of samples over which this average has been calculated. The power 4 is used as a power 4 is being used to achieve convergence :

$$P_S(n)=\frac{1}{M}\sum_{m=n}^{m=n-(M-1)} S_i{}^4(m)$$

For our system then, we use the following rule :

$$\mu_{ij}(n)=\frac{\mu_0}{P_{Si(n)}} \quad \text{where} \quad 0<\mu_0<\frac{1}{M}$$

μ_0 is called the amplitude of the adaptation gain.

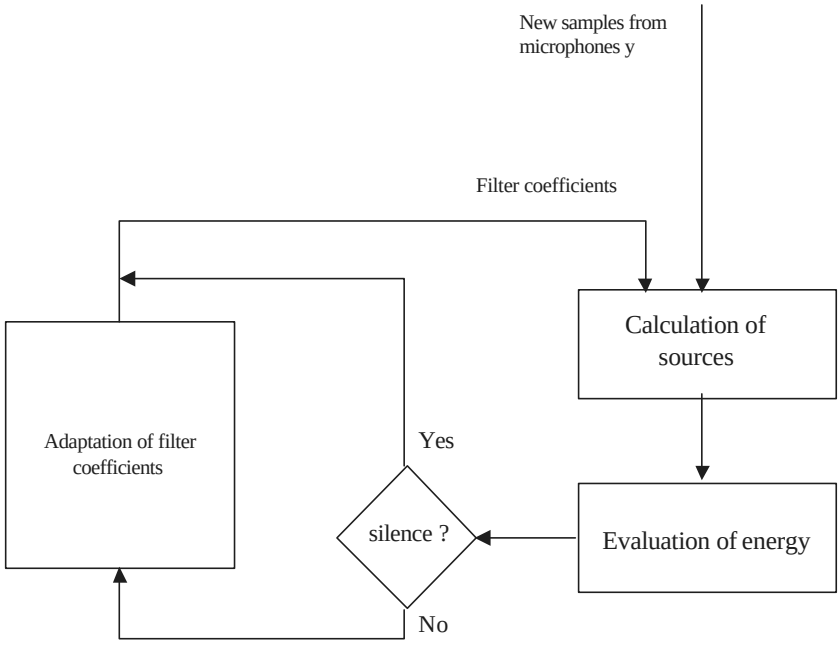
To achieve more precise convergence, a further modification can be made to the adaptation gain :

$$\mu_{ij}(n)=\frac{\mu_0}{P_{Si(n)}}\left(1+\frac{\gamma}{n}\right)$$

Addition of the parameter γ gives a rapid initial convergence followed by more precise convergence later on.

Speech signals are inherently non-stationary, as there are frequent periods of silence. If the system is to work correctly, it is crucial to stop the adaptation of parameters during such periods. As each sample is received therefore, the energy of each microphone output is evaluated, and compared to a set threshold. A decision to pause or continue adaptation is then made.

Figure 4. A diagram for speech/silence detector



Implantation of the source separation system on the TMS320C67x

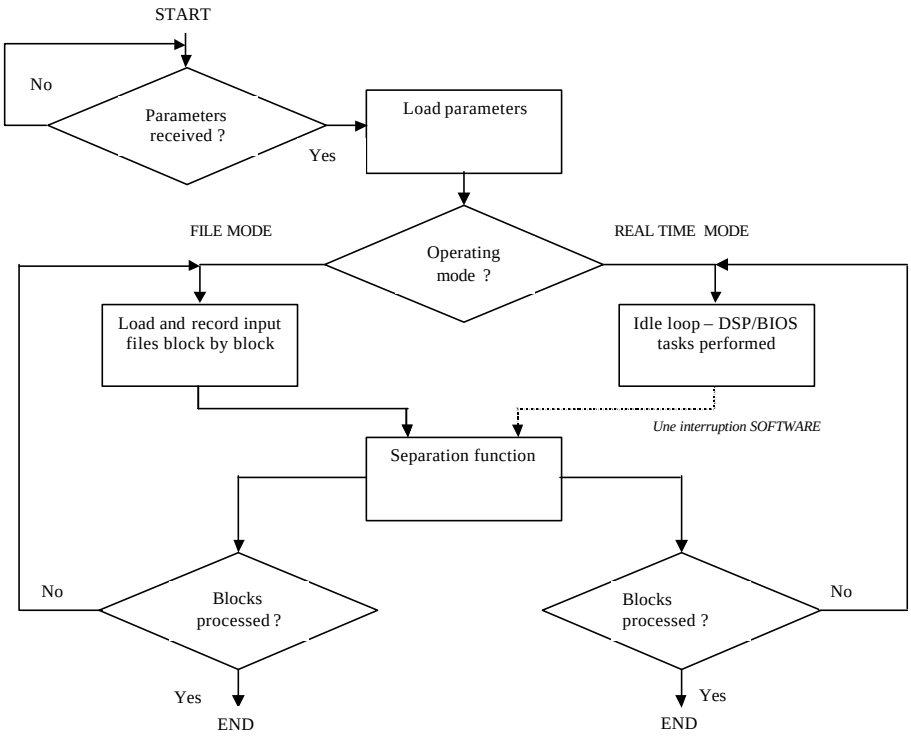
A blind source separation system has been created which runs on the TMS320C6701 EVM. The program works in two modes : file mode and real time mode.

The file input/output functionality was included to allow clean testing of the algorithm. Two Microsoft format '.wav' files (mono, 16 bit) are used as inputs to the DSP. These inputs, representing the two microphones, are loaded directly into the DSP memory. The DSP then processes the loaded data, and outputs directly to two output files on the host, of the same format. The motivation behind this mode is to avoid any problems of acquisition or restitution when testing the algorithm. It is clear that no noise or distortion is added to the inputs, apart from any precision errors due to the processing. At the end of each block of 80 samples, the filter coefficients are written to files on the host for later analysis. This mode gives best-case results for any particular mixture.

The other execution mode available is real-time mode. In this mode, the system uses the audio samples coming from acquisition card by the PCI port. Due to the processing power constraints of the C6701 DSP, the available sample rates are limited at: 8 kHz, 9.6 kHz, 16 kHz.

Note that the same separation processing is used in the two different modes, file input/output (for testing reasons) and real-time.

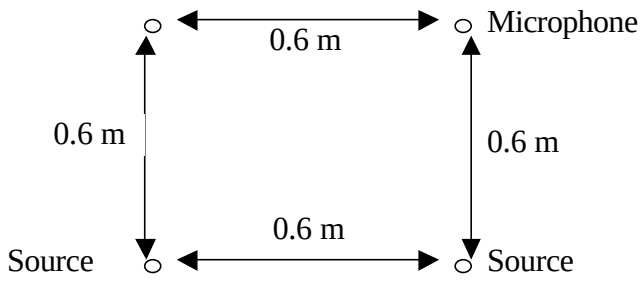
Figure 5. Block diagram of separation system's operation



The tests of blind separation of source module

The recordings have been realised in the sound studio of the Laboratory CLISP-IMAG. The geometrical arrangement of the microphones and the two sources is described in the diagram below. The mixtures were recorded via the microphones and a mixing desk as stereo .WAV files, of sample frequency 16 000 Hz.

Figure 6. The position of the sources and the microphones in our experiments



Four separate mixtures were recorded :

- A speech signal and white noise

  Eight recordings of twenty seconds were initially made. One source was a person speaking. The second was recorded white noise, created using the program Goldwave, being played through a loudspeaker. Four recordings consisted of a Vietnamese male counting from one to twenty, and the other four consisted of an English male counting similarly. Of these eight recordings, one was selected on the basis of quality of recordings, and then cut down to ten seconds for later testing.

- A speech signal with coloured noise

  Four recordings of twenty seconds were initially made. One source was a person speaking. The second was recorded coloured noise (created by band-pass filtering white noise using the program GoldWave) being played through a loudspeaker. Two recordings consisted of a Vietnamese male counting from one to twenty, and the other two consisted of an English male counting similarly. Of these four recordings, one was selected on the basis of quality of recordings, and then cut down to ten seconds for later testing.

- Two speech signals

  Six recordings of twenty seconds were initially made. Two were made with a Vietnamese male and a Romanian male counting from one to twenty. Two were made with a Vietnamese male and an English male, and the final two were created using an English male and a Romanian. Of these six recordings, one was selected on the basis of quality of recordings, and then cut down to ten seconds for later testing.

- A speech signal and music

  Eight recordings of twenty seconds were initially made. Half of these recordings were made using twenty seconds of largely wordless music, the other half with music with singing. For the speech, in each case, half were made with a Vietnamese male counting, and the other half with an English male counting. Of these eight recordings, one was selected on the basis of quality of recordings, and then cut down to ten seconds for later testing.

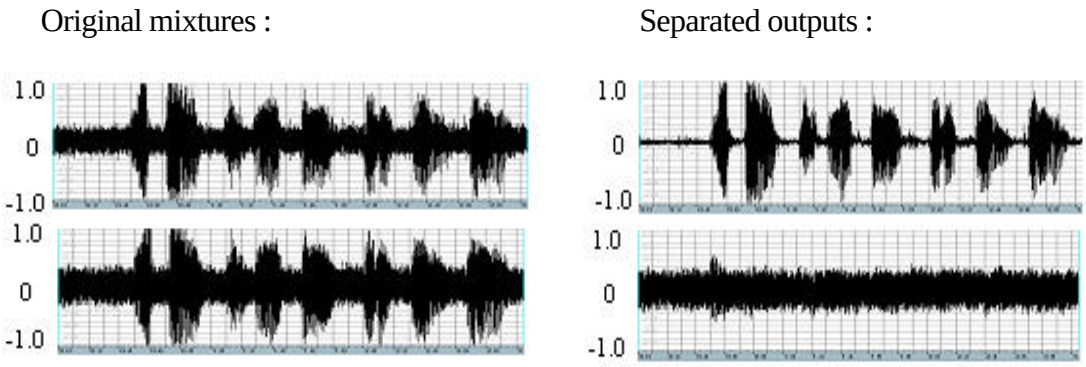
The testing of the separation system using the test files was divided further into two stages. First it was tested in file input/output mode, and then in real-time mode.

The separation test files that are used are :

- The four real mixtures recorded by ourselves(speech + white noise, speech + coloured noise, speech + music, speech + speech) like the above description
- Two mixtures recorded by Te-Won Lee, Research Associate at the Computational Neuroscience Laboratory, the Salk Institute, La Jolla (speech + music, speech + speech)
- One simulated mixture of white noise and speech. Source signals available
- Two convoluted mixtures recorded by Dominic Chan, a Cambridge University research student (music + music, speech + speech)

Some results in the file input/output mode of the separation system are showed below.

Figure 7. The results of separation on a French speaker with white noise



File mode results : French Speaker + White noise (simulated mixture)

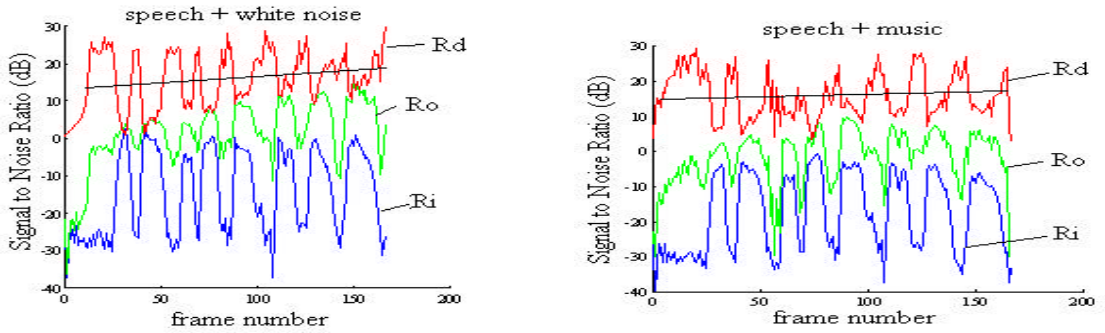
For the simulated mixture, where the original sources are available, the Signal to Noise Ratio (SNR) is defined as:

$$SNR=10\log\left(\frac{\sum_{n=0}^{N-1}(X_i(n))^2}{\sum_{n=0}^{N-1}(S_i(n)-X_i(n))^2}\right)$$

Where S represents the separation output, and X represents the corresponding *Source*.

Some results about SNR in the file input/output mode for the simulated mixture are showed below. Using the separation system, SNR have been gained.

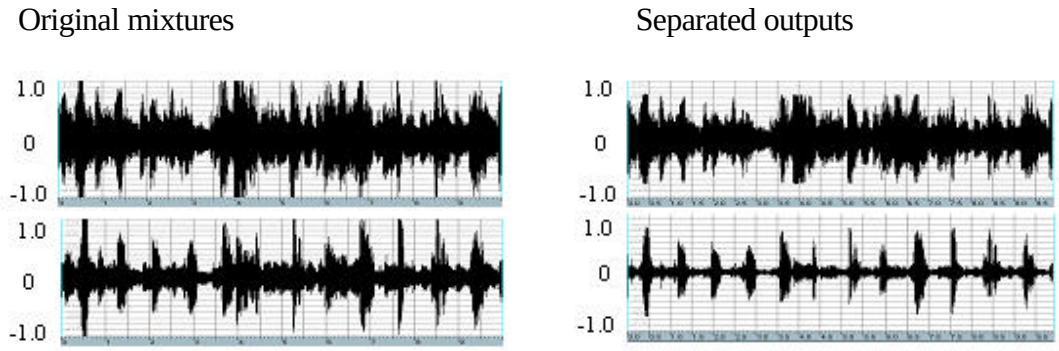
Figure 8. The gain in signal to noise ratio for our module



Ri : SNR between source and simulated mixture
Ro : SNR between source and output of separation system
Rd : delta of Ro-Ri

For real-time mode, the test signal have been created by using one sound card with the files WAV. The output of the sound card is connected with audio input of the DSP card. The output signal of the separation system on Audio output of the DPS card is connected with the input of sound card. Some the results are showed below .

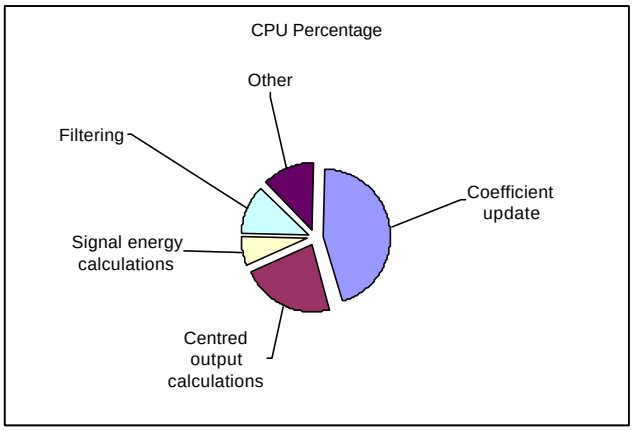
Figure 9. The results of separation of a speech and music mixture



Real time results : Speech + Music (real mixture recording by Tee-Won)

The system works well in real-time mode at a sample frequency of 9600 Hz. Unfortunately, the DSP was not sufficiently powerful for the algorithm to work equally effectively at the higher sample rate of 16000 Hz. With such a sample frequency, separation orders approaching 20 are possible (the value is not fixed as other separation parameters affect this limit). Such a sample frequency is adequate for speech recognition systems. To discover what was taking up all the calculation time, I did a small analysis on the make-up of the available calculation time. We can see that the separation function takes up by far the largest part of the time available, with acquisition/restitution taking up a mere 10% (fs = 16 000 Hz, Order = 9).

Figure 10. The loading time of the DSP



The second module : Echo cancellation

Theory

The reverberation (that is the acoustic influence of the room) can be modelled through an impulse response filter h(n): $x(n)=s(n)*h(n)$ where:

- s(n) : neat speech signal
- h(n): impulse response of the room and the microphone
- x(n): recorded signal

Let's remind that, if $x(n)=x_1(n)*x_2(n)$, then the cepster of the two convoluted signals ([5]) is:

$$\hat{x}(n)=IFFT\{\log(X_1(n)\cdot X_2(n))\}=IFFT\{\log(X_1(n))+\log(X_2(n))\}=\hat{x}_1(n)+\hat{x}_2(n)$$

(In the cepstral domain, the convolution operation becomes an addition operation)

The CMS algorithm ([5],[6]) that we decided to use allows to estimate the impulse response of the room through the average of the cepsters. If several signals are recorded in the same room, the cepsters of the recorded signals are: $\hat{x}_1(n)=\hat{s}_1(n)+\hat{h}_1(n)$ …….. $\hat{x}_J(n)=\hat{s}_J(n)+\hat{h}_J(n)$ and if we calculate the average of all the signals, we obtain the following average of the cepsters:

$$\hat{m}(n)=\frac{1}{J}\sum_{i=1}^{J}\hat{x}_i(n)=\frac{1}{J}\sum_{i=1}^{J}\hat{s}_i(n)+\frac{1}{J}\sum_{i=1}^{J}\hat{h}_i(n)=\hat{m}_s(n)+\hat{m}_h(n)$$

that is, the average of the voice signals added to the average of the impulse responses of the room.

As the impulse response of the room does not change in time (or changes very slowly), we approximate that: $\hat{m}_h(n)\approx\hat{h}(n)\rightarrow\hat{m}(n)=\hat{m}_s(n)+\hat{h}(n)$. We also consider that the speech signal is a noise of an average equal to 0. Then, we have: $\hat{m}(n)\approx\hat{h}(n)$.

Once we obtained the cepster of the impulse response, it is simple to eliminate the influence of h(n): $\hat{y}(n)=\hat{x}(n)-\hat{m}(n)=\hat{s}(n)+\hat{h}(n)-\hat{h}(n)=\hat{s}(n)$

Description of the algorithme for DSP TMS320C6701

The program has two steps:

- The initialisation step gives a first approximate estimation of the impulse response by the calculus of the cepsters average on two cepsters;
- The next step is an infinite loop, where, for every N sample, we de-convolute the signal, dividing the FFT of the entering signal by the FFT of h. In the mean time, the value of the impulse response is recalculated for every new cepster.

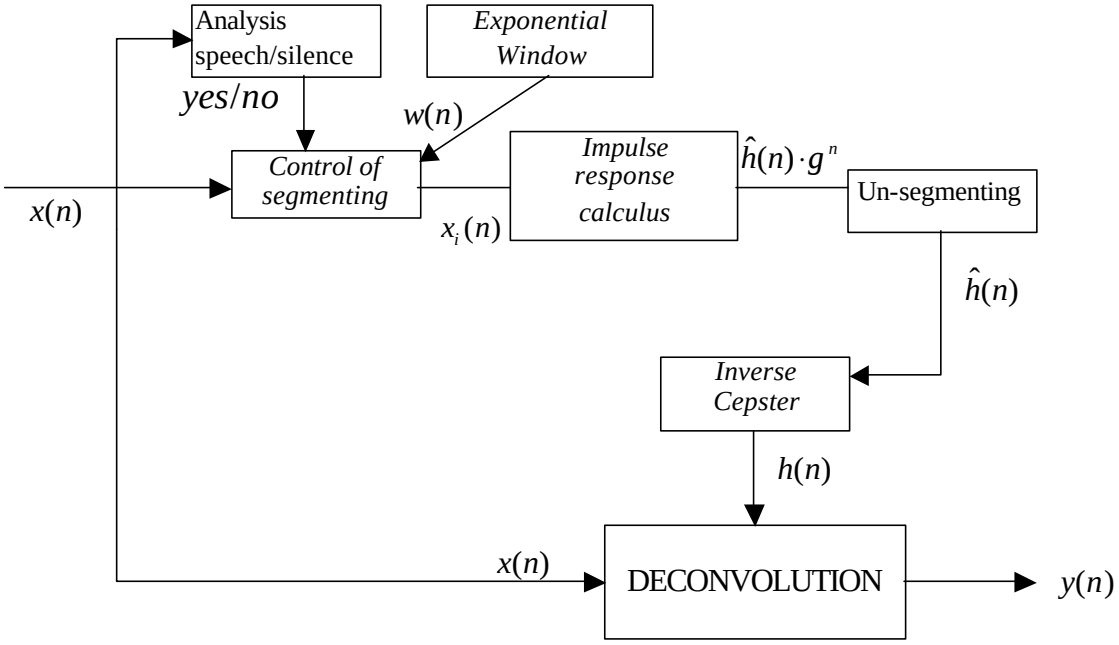
We used the following method: first, we calculate the impulse response making the average of the cepsters and next we pass to un-reverberation through de-convolution and not through subtraction. The treatment diagram is below.

The cepsters are calculated on segments obtained through a well chosen mathematical window. Indeed, segmenting means cutting the original signal s(n) into parts s_i (n):

$$x_i(n)=s_i(n)*h(n)\rightarrow\hat{x}_i(n)=\hat{s}_i(n)+\hat{h}(n)$$

The problem comes from the fact that there is always an error $x_i(n)=s_i(n)*h(n)+e_i(n)$, which is: $e_i(n)=v_i(n)-u_i(n)$. The first item is given by the previous segment $x_{i-1}(n)$ and the second is given by the samples of the next segment $x_{i+1}(n)$. Indeed, the result of a convolution operation is longer than the initial signal. That's why the result of the convolution of a $x_i(n)$ segment with the impulse response is longer than the segment $x_i(n)$ with a value indicated by $u_i(n)$ (wherefrom the "minus"). Besides, $x_i(n)$ is added to $v_i(n)$, given by the result of the convolution of the previous segment $x_{i-1}(n)$ with h(n). In order to minimize this error, the choice of the window is essential: it can be rectangular, Hamming type, or exponential.

Figure 11. The echo cancellation treatment diagram



The exponential window has the two required advantages:

- it reduces the amplitude of the samples at the end of the window (because $|\gamma| < 1$)
- its influence over the calculus of the cepster is already known.

It is important to notice that the use of this window needs un-segmenting because it modifies the net impulse response to find.

However, the exponential window produces a quite important error $v_i(n)$ (because of the important value it has at the beginning). Anyway, the error $u_i(n)$ can be as low as we choose (thanks to the low values of the last samples of the exponential window).

In order to reduce as much as possible the error $v_i(n)$, it is necessary to place the beginning of every window in a silence interval, where $s_{i-1}(n)\approx 0 \rightarrow x_{i-1}(n)\approx 0$, and $v_i(n)$ are very small.

So, we have to use a silence/speech detector. The detection of silence or speech segments is based on energy calculus. We calculate a threshold energy $E_t=K*E$ on the first acquired block,

with $E = \sum_{n=0}^{Ne-1} s2(n) = \sum_{k=0}^{Ne-1} S2(k)$, and K=1,5. This factor is a "precaution" we take and it ensures us that, if E>E_t, then the segment is a speech one, if not, it is a silence one.

We make the average of the cepsters on several segments in order to calculate the impulse response of the room. Besides, a real cepster average is sufficient. Indeed, thanks to the segmentation by exponential window, our system becomes a minimal phase system: the multiplication with γ, whom module is inferior to one, moves the poles and the zeros of the system inside the unit cercle. Or, if a signal x(n) is in minimal phase, then its cepster $\hat{x}(n)$ is causal. Besides, the Z transformation of a causal sequence is entirely determined by the real part of its FFT. Therefore, because x(n) and $\hat{x}(n)$ are causal, the formula $\hat{X}_R(e^{jw}) = \log\left|X(e^{jw})\right|$ entirely defines and allows to obtain $\hat{x}(n)$.

In order to de-convolute the signal, we have just to move it into the frequency domain through a FFT. Hence, we can obtain the estimated original signal $\tilde{s}(n)$, using the following formulas:

$$\tilde{S}(e^{jw}) = \frac{X(e^{jw})}{\tilde{H}(e^{jw})}$$

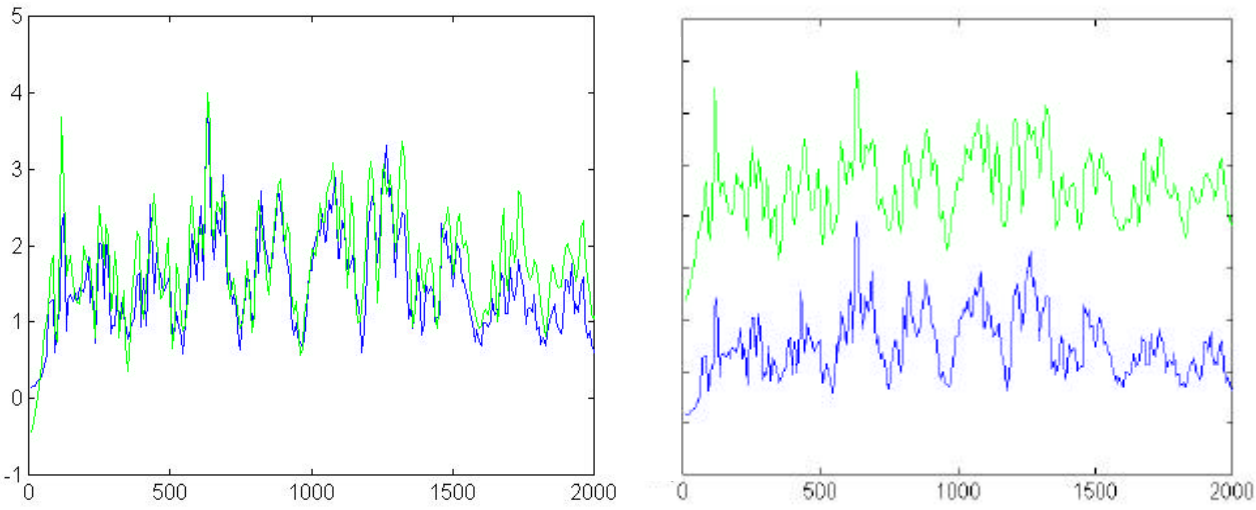
$$\tilde{s}(n) = IFFT\left[\tilde{S}(e^{jw})\right]$$

In order to avoid discontinuities caused by the signal segmentation, we make a Bézier interpolation between two windows of the output signal.

The tests of echo cancellation module

To validate the algorithm, we calculated the impulse response of the pseudo-anechoidal room of our laboratory using a classical method (recording with a white noise excitation and calculus of an average of FFT on a long time – 20s). The figure below shows the superposition of the reference transfer function (in black) with the estimated transfer function (in gray). In the b part of the figure, the two curves are voluntarily separated in order to compare them.

Figure 12. The reference and estimated transfer fonction of our pseudo-anechoidal room



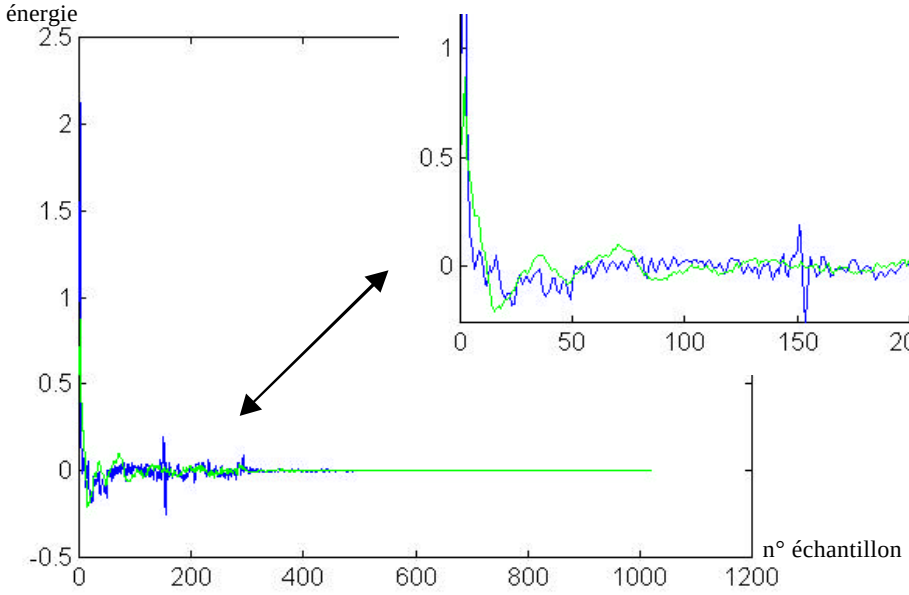
a b

It is obvious that the two curves are similar: the peaks are positioned on the same frequencies and their amplitudes are coherent. The differences between the two curves can be explained by some estimation errors of our algorithm.

We have just shown that the estimation of the acoustic characteristics of the room, calculated through our algorithm is true for an ideally sound source of white noise type. We shall show below that this estimation is true when the sound source is a sound more œmplex that speech. The figure shows the superposition of the impulse response of the room, estimated by our algorithm for a recorded speech signal (gray curve) with the same impulse response estimated by our algorithm for an entering signal of white type noise (black curve).

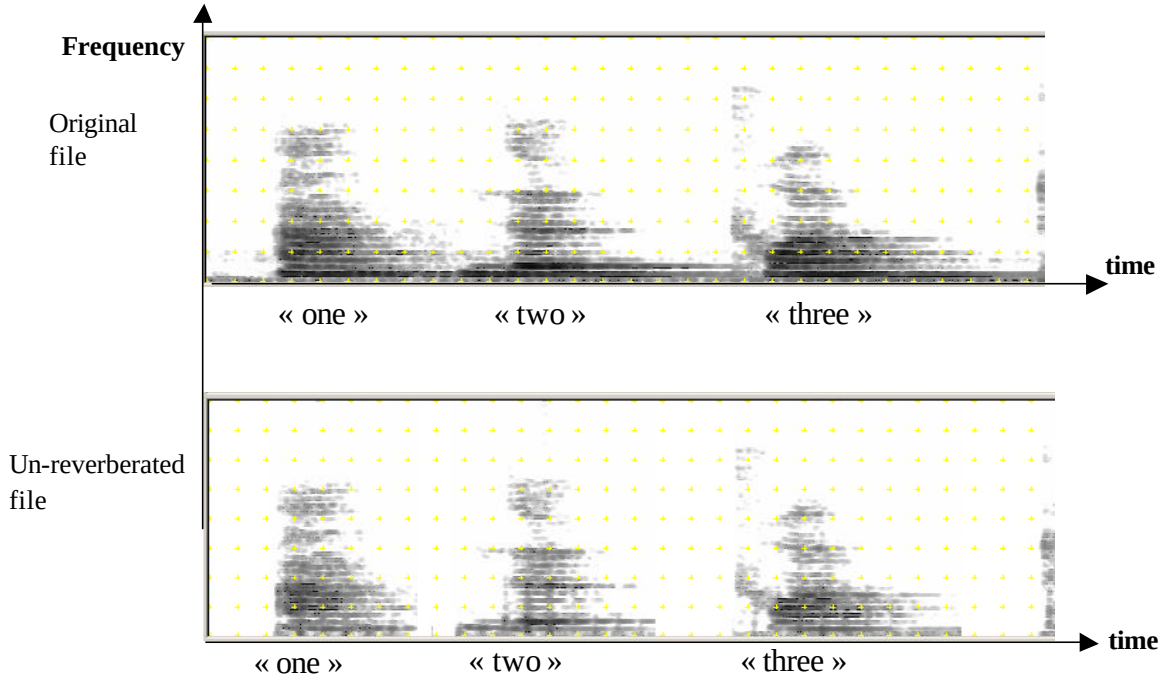
A brief comparison of the two curves shows that their general shape is similar, but a focus on the low frequencies shows notable differences. The average oscillations of the “white noise” response are less obvious than those of the speech response. However, the curve of the speech response is smoother than the white noise’s.

Figure 13. The estimated transfer fonction for a white noise and a speech signal of entry



In the case of a speech signal, the reverberation generally leads to a growth of the low frequencies and to a longer time of the words. The output file of the recording analysed in the figure below seems to have no reverberations when listened. The figure makes a comparison between the input and the filtered output: the low frequencies have lower intensities and time, comparing to the original file.

Figure 14. The sonograms of a reverberated and unreverberated file



The last test that we have done consisted in submitting reverberated and un-reverberated files to our ASR. The language module is reduced for this test and it considers only the numbers, as we adapted it to our test files. The recognition rates are presented in the table below. We compare the recognition rate of our ASR for the reverberated files (called "original files") with the rate for the files obtained after pre-treatment ("un-reverberated" files).

Table 1. The recognition rate for reverberated and unreverberated files

File	Original	Un-reverberated	Gain
1	61,54%	63,64%	2,10%
2	38,46%	60,00%	21,54%
3	42,86%	63,64%	20,78%
4	63,64%	66,67%	3,03%
5	70,00%	80,00%	10,00%
6	60,00%	60,00%	0,00%
7	70,00%	90,00%	20,00%
8	37,50%	37,50%	0,00%
9	50,00%	62,50%	12,50%
10	88,89%	80,00%	-8,89%
Average	58,29%	66,39%	8,11%
Minimum			-8,89%
Maximum			21,54%

Out of the ten files analysed, two have identical recognition rates. Seven other files have better recognition rates after un-reverberation, and only one file has a smaller rate. The average gain is 8%, and the maximum 21%. Hence, the algorithm has in most of the cases a positive effect upon the recognition rates.

Conclusion

We presented both signal preprocessing module : blind separation of source and echo cancellation used in our medical telemonitoring project. They were designed in C using the software development program CCS offered by Texas Instruments with his DSP TMS320C6701. We integrated these two signal pre-processing modules in the chains of our signal processing system of our project in order to have clean signals necessary to recognition system.

Now, we record a corpus of life situations in our study apartment. The corpus is made up of a set of 20 scripts reproducing a string of events, either voluntary actions of the patient inside the apartment, or unexpected events which could characterize an abnormal or distress situation. Below we present an example of script :

*"<**Script no. A0**>*

*<**description**> Walking in the house.Normal situation (no alarm) </**description**>*

*<**time**>0</**time**><**Position**>Entry</**Position**><**Action**>Person speaking</**Action** >*

*<**time**>4</**time**><**Position**>Hall</**Position**><**Action**>Person speaking</**Action** >*

*<**time**>10</**time**><**Position**>Toilet</**Position**><**Action**>Person speaking</**Action** >*

*<**time**>12</**time**><**Position**>Shower</**Position**><**Action**>Person speaking</**Action** >*

*<**time**>17</**time**><**Position**>Toilet</**Position**><**Action**>Person speaking</**Action**>*
*<**time**>20</**time**><**Position**>Hall</**Position**><**Action**>Person speaking</**Action**>*
*<**time**>21</**time**><**Position**>Kitchen</**Position**><**Action**>Person speaking</**Action**>*
*<**time**>25</**time**><**Position**>Livingroom</**Position**><**Action**>Person speaking*
*</**Action** >*

*<**time**>30</**time**><**Position**>Bedroom</**Position**><**Action**>Person speaking</**Action** >*
*</**Script no. A0**>"*

References

[1] N. Noury et al. , "A telematic system tool for home health care", Int. Conf. IEEE-EMBS, Paris, 1992, part 3/7, pp 1175-1177

[2] V. Rialle et al., "A smart room for hospitalised elderly people: essay of modeling and first steps of an experiment", Technology and Health care, vol7, 1999, pp. 343-357

[3] M.Ogawa and al., "Fully automated biosignal acquisition in daily routine through 1 month", Int.Conf. IEEE-EMBS, Hong-Kong, 1998, pp. 1947-50

[4] G.Williams and al., "A smart fall and activity monitor for telecare application", Int.Conf IEEE-EMBS, Hong-Kong, 1998, 1151-1154

[5] D.Bees, M.Blostein and P.Kabal, "Reverberant speech enhancement using cepstral processing", ICASSP 1991, pp.977-980

[6] S.Chuicchi, F.Piazza, "V-Stereo system with acoustic echo cancellation", EURASIP Conférence, 1999, Krakow

[7] C.Jutten,J.Herault, « Blind separation of sources.An adaptive algorithm based on neuromimetic architecture », Signal processing, Vol.24, 1996, pp.1-20

[8] H.L. Nguyen Thi, "Séparation aveugle de sources à bande large dans un mélange convolutif, application au rehaussement de la parole », Thèse INP Grenoble, 1993

[9]CHARKANI EL HASSANI (Ahmed Nabil),Séparation Auto-adaptative de sources pour des mélanges convolutifs.Application à la téléphonie mains-libres dans les voitures.Thèse, INP Grenoble, Novembre 1996 (276 pages)