# ENHANCEMENT OF A SPEECH SIGNAL EMBEDDED IN NOISY ENVIRONMENT USING TWO MICROPHONES

*Allan Kardec Barros*[1,5], *Hideki Kawahara*[2], *Andrzej Cichocki*[3]
*Shoji Kajita*[4], *Tomasz Rutkowski*[3], *Mitsuru Kawamoto*[6] *Noboru Ohnishi*[1,4]
1 : BMC, RIKEN, 2 : Wakayama University, Japan, 3 : BSI, RIKEN, Japan,
4 : Nagoya University, 5 : UFMA, Brazil, 6 : Shimane University, Japan.
E-mail: akbarros@ieee.org

## ABSTRACT

In this work we develop a system for enhancement of a speech signal with the highest energy from a linear convolutive mixture of independent sound sources (interferences) recorded by two microphone signals. In this system we use the concept of independent component analysis (ICA) together with auditory filter banks, pitch tracking, adaptive band pass filters and masking. Preliminary computer simulations experiments confirm the validity of the proposed algorithm.

## 1. INTRODUCTION

One of the most well known problem in auditory scene analysis is that of *cocktail party*. This is generally related to the problem of *selective attention*: how humans can select the voice of a particular speaker in a noisy environment where there are many other speakers and interferences.

Independent component analysis (ICA) appears as an important technique to help solving this problem. Standard ICA is based on the following principle. Assuming that the original (or source) signals have been linearly mixed, and that these mixed sensor signals are available, ICA finds a linear combination of the mixed signals, which recovers the original source signals, possibly re-scaled and randomly arranged in the outputs.

However, there are at least two difficulties related with the *cocktail party* problem: firstly, due to reverbation effect we observe convolutive mixture; secondly, in practice we have smaller nubmer of microphones than unknown acoustic source signals, so standard ICA can not be directly applied. It is interesting to notice that humans can deal with this problem by using only two ears.

Similarly to humans, our aim here is not to recover simultaneously all the original acoustic signals. The reason is that we believe that the human hearing system is not using the so-called *room inversibility*, which for the real world environment seems impossible, due to the various factors which should be taken into account.

In other words, our aim here is rather to turn a specific speech signal more intelligible than available microphone signals. As our auditory systems, we try to enhance the signal nearest to the microphones, i.e., the signal with highest energy. We realize this by emulating some properties of human auditory system. This is carried out by (a) mimicking the inner ear, through the use a bank of self-adaptive band-pass wavelet filters (b) the tracking of the speech fundamental frequency (f0) and; (c) by masking some parts of the speech with lower energy.

## 2. ON THE PROPAGATION OF SOUND

Consider $n$ source signals at time $t$, $\mathbf{s} = [s_1(t), s_2(t), \ldots, s_n(t)]^T$ arriving at $m$ receivers $\tilde{\mathbf{x}}(t) = [\tilde{x}_1(k), \tilde{x}_2(t), \ldots, \tilde{x}_m(t)]^T$. In the general model of cocktail party, each receiver gets a non-linear combination of the source signals, so that we have a Volterra series,

$$\tilde{\mathbf{x}}(t) = \int_{-\infty}^{\infty} \mathbf{H}_1(\tau_1)\mathbf{s}(t+\tau_1)d\tau_1 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{H}_2(\tau_1,\tau_2)\mathbf{s}(t+\tau_1)\mathbf{s}(t+\tau_2)d\tau_1 d\tau_2 + \ldots, \quad (1)$$

where $\mathbf{H}_i$ is a linear filter operator.

However, generally it is assumed that the propagation of sound is linear, thus the above equation can be simplified to,

$$\tilde{\mathbf{x}}(t) = \int_{-\infty}^{\infty} \mathbf{H}(\tau)\mathbf{s}(t+\tau)d\tau \quad (2)$$

The interesting point about the above equation is that the source signals, besides being mixed, are also arriving at the microphones after reverberating in the environment. It is also important to notice that in an actual environment, $\mathbf{H}$ is a non-minimum phase low-pass filter, which turns the task of recovering the original signals very difficult.

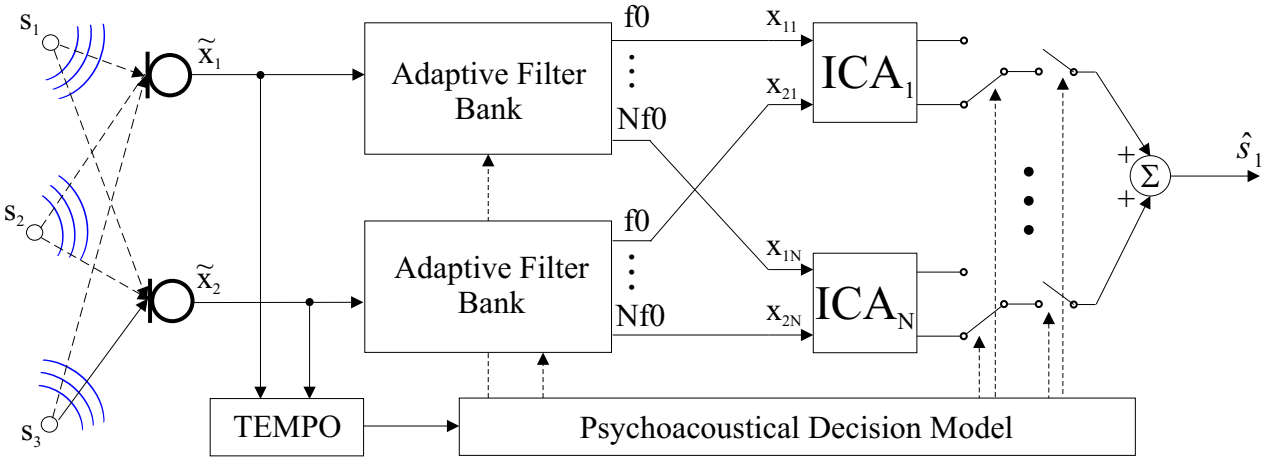


Figure 1: Block diagram of the algorithm which mimicks the auditory system. First it tracks the fundamental frequency (f0) using TEMPO. Then, it process the mixed signals using a bank of band-pass filters (such as the inner ear). After, it process each mixed/convolved signal by an ICA algorithm. Finally, it carries out masking by turning switches *on* or *off*.

## 3. AUDITORY FILTER BANKS

It is well known that the human auditory system can roughly be described as a nonuniform bandpass filter bank, consisting of strongly overlapping bandpass filters with bandwidth in the order of 50 to 100 Hz for signals below 500 Hz and up to 5000 Hz for signal at high frequency. The hearing system performs a spectrographic analysis of any auditory stimulus at the cochlea, which can be regarded as a bank of nonuniform self-adaptive filters whose outputs are ordered tonotopically. Recently, many sophisticated and biologically plausible models of such auditory filters banks have been proposed. In this paper we employ the fundamental frequency extractor developed by Kawahara et. al.[15], along with a bank of wavelet band pass filters[5].

Our final objective is to develop an algorithm whose output signal $y(t)$ is a modified version of a given source signal $s_i(t)$, i.e. signal of interest $y(t) = g(s_i(t))$, where $g(\cdot)$ can at the same time be a filter and a non-linear transformation operator.

An important property of speech signals is that they are non-stationary, but can be regarded as locally stationary. Roughly speaking, speech can be divided in the time domain into voiced and unvoiced sounds, the first ones having more structure in the frequency domain than the later. Indeed, voiced sounds are regarded in general as *quasi-periodic*. In this matter, some experiments pointed that humans may be using this voiced structure to separate sounds from the background. Moreover, humans can more easily understand a voiced than a unvoiced sound[23].

One open problem is how the auditory system segregates those sounds in a higher level (at the cortex). In this work, we suggest that this can be carried out by exploiting the local statistical independence of a pair of sub-band sounds for each frequency sub-band (bin).

Also in our algorithm we included the temporal masking characteristic of the auditory system. This is managed by a switch which is *on* in the voiced part, and turns *off* in the unvoiced part.

Fig. 1 shows the conceptual block diagram of the system. It is composed of four parts. The first one is the TEMPO algorithm[15], which extracts the fundamental frequency from the mixed signals as well as it shows which part of the speech is voiced. The second is a bank of adaptive band-pass wavelet filters, which process the signals around the fundamental frequency and around its harmonics. The third is a bank of ICA algorithms for enhancing the desired signal for each frequency sub-band. The final one is a bank of switches which performs temporal masking. Let us give a brief description for each of them.

### 3.1. Extraction of the Fundamental Frequency

In TEMPO[15], the fundamental frequency is extracted as the instantaneous frequency of the fundamental component of the signal. This is carried out by introducing a measure called *fundamentalness*, which gives *a-priori* knowledge about the fundamental frequency. The *fundamentalness* have its maximum value when FM and AM modulation magnitudes are minimum.

Using an analyzing wavelet $g_{AG}(t)$ constructed from

a complex Gabor function having a slightly finer resolution in frequency (*i.e.* $\eta > 1$), the input signal can be divided into a set of filtered complex signals $D(t, \tau_c)$.

$$D(t,\tau_c) = |\tau_0|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t)\overline{g_{AG}\left(\frac{t-u}{\tau_c}\right)}du \quad (3)$$

$$g_{AG}(t) = g(t-1/4) - g(t+1/4) \quad (4)$$

$$g(t) = e^{-\pi\left(\frac{t}{\eta}\right)^2} e^{-j2\pi t}. \quad (5)$$

The characteristic period $\tau_c$ of the analyzing wavelet is used to represent the corresponding filter channel.

The *fundamentalness* index $M(t, \tau_c)$ is calculated for each channel $c$ based on the output. The definition of the index is given as follows.

$$\begin{aligned} M(t,\tau_c) = & - \log\left[\int_\Omega \left(\frac{d|D|}{du}\right)^2 du\right] \\ & + \log\left[\int_\Omega |D|^2 du\right] \\ & - \log\left[\int_\Omega \left(\frac{d^2 \arg(D)}{du^2}\right)^2 du\right] \\ & + \log\Omega(\tau_c) + 2\log\tau_c, \end{aligned} \quad (6)$$

where the integration interval $\Omega(\tau_c)$ is set proportional to the size of the corresponding analyzing wavelet and is a function of $\tau_c$.

Extracting fundamental frequency $f0$ simply means finding the maximum index of $M_c$ in terms of $\tau_0$ and calculating the average (or more specifically, interpolated) instantaneous frequency using the outputs of the channels neighboring $\tau_0$.

The instantaneous frequency $f0$ of fundamental band-pass filter output signal $D(t, \tau_0)$ is calculated using the following equation.

$$f0(t) = \frac{f_s}{\pi} \arcsin\frac{|y_d(t)|}{2}$$

$$y_d(t) = \frac{D(t+\Delta t/2, \tau_0)}{|D(t+\Delta t/2, \tau_0)|} - \frac{D(t-\Delta t/2, \tau_0)}{|D(t-\Delta t/2, \tau_0)|} \quad (7)$$

It should be noted that TEMPO algorithm tracks the fundamental frequency of speaker who is closer to the microphones, which means that its signal energy is much higher. This assumption preserves TEMPO from confusing speaker's pitch with frequency components from interfering signals.

### 3.2. The Bank of Adaptive Band-pass Filters

We use here the concept of harmonicity of the voiced sounds that was exploited in some models of *computational auditory scene analysis* (CASA) which group together spectrotemporal regions that are modulated by the same period [7, 22].

The idea is to use a bank of band pass filters centered at the fundamental frequency f0 and its harmonics, as used in [5]. To this end, the following mother wavelet is used,

$$\Psi(t,k) = \frac{1}{2\pi}\frac{d}{dt}\left[e^{-\pi\left\{\frac{kf0^{\Omega}(t)}{6}\right\}^2} \cos\left(2\pi t \int_t kf0(\tau)d\tau\right)\right],$$

$$\overline{f0^{\Omega}} = \frac{1}{\Omega}\int_\Omega f0(\tau)d\tau. \quad (8)$$

where $\Omega$ is a short time interval.

From this, we obtain intermediary signals $r_{i,k}(t)$ which are $\tilde{x}_i(t)$ filtered around frequency $kf0(t), (k = 1, 2, \ldots, N)$, given by

$$r_{i,k}(t) = \int_{-\infty}^{\infty} \Psi(t,k)\tilde{x}_i(t)d\tau$$
$$\text{for } k = 1, \ldots, N,\ i = 1, 2, \ldots, m. \quad (9)$$

where $N$ is the number of harmonics (and therefore of sub-bands).

Then, we find the instantaneous amplitude of each $r_{i,k}(t)$ by the following operation,

$$\hat{a}_{i,k}(t) = |H[r_{i,k}(t)]| \quad (10)$$

where $H[r_{i,k}(t)]$ is the Hilbert transform of the signal $r_{i,k}(t)$.

At this point, however, we have no phase information about the signal we want to estimate. Thus, we generate from $\hat{a}_{i,k}(t)$ and $f0(t)$ a set of orthogonal signals,

$$z_{q,i,k} = \hat{a}_{i,k}(t)e^{qj2\pi tkf0(t)}, \quad (11)$$
$$q = -1, 1 \text{ and } k = 1, \ldots, N.$$

In order to obtain the phase information of the signal, we use the Wiener theory. In this case the output of the $k$-th sub-band will be,

$$x_{i,k} = \mathbf{c}_{i,k}^T \mathbf{z}_{i,k}(t),\ i = 1, 2, \ldots, m. \quad (12)$$

where $\mathbf{z}_{i,k}(t) = [z_{1,i,k},\ z_{-1,i,k}]^T$. In Wiener theory, given the signal $\hat{x}_{ik}(t)$, the weight vector $\mathbf{c}_{i,k}$ which gives the minimum mean squared error between the estimated signal $p_{i,k}$ and $\hat{x}_{ik}(t)$ is given by[12],

$$\mathbf{c}_{i,k} = \mathbf{R}^{-1}\mathbf{P}$$
$$= E[\mathbf{z}_{i,k}(t)\mathbf{z}_{i,k}(t)^T]^{-1}E[\hat{x}_{ik}(t)\mathbf{z}_{i,k}(t)]. \quad (13)$$

Since the elements of $\mathbf{z}_{i,k}(t)$ are mutually orthogonal, matrix $\mathbf{R}$ is diagonal. Thus, it is not difficult to remove the inversion in (13), by normalizing the elements of $\mathbf{z}_{i,k}(t)$ to have unity variance. In this case, $\mathbf{R} = \mathbf{I}$, thus, $\mathbf{c}_{i,k} = E[x_i(t)\mathbf{z}_{i,k}(t)]$.

### 3.3. Independent Component Analysis

In this section we study the third step of the algorithm. Now that we had available the filtered and mixed subband signals obtained from the bank of band-pass filters. In other words, we have split wide-band signals into sub-band (narrow-band) signals. An important property of a narrow band signal is that they have less effects of convolution. In fact, the convolutive mixture turns approximately into an instantaneous mixture, as the bandwidth diminishes.

In this work, we use a simple batch fixed-point algorithm which carries out blind source separation, called RICA (robust independent component analysis) [9]. Its derivation is as follows.

We propose to extract the independent components sequentially. Notice also that the inputs of the ICA blocks are the outputs of the bank of band pass filters. (see Fig. 1). In this case, the ICA inputs for the $k$-th subband are signals $x_{i,k}$, $\forall i$ For this purpose we use $m$ single processing units for each subband. In order to simplify notation, let us consider a single (say *i-th*) processing unit described as:

$$y_i(t) = \mathbf{w}_i^T \mathbf{x}_i(t) = \sum_{j=1}^{m} w_{1j} x_j(t), \quad (14)$$

$$\varepsilon_i(t) = y_i(t) - \sum_{p=1}^{L} b_{1p} y_i(t-\Delta_p) = \mathbf{w}_i^T \mathbf{x}_i(t) - \mathbf{b}_i^T \tilde{\mathbf{y}}_i, \quad (15)$$

where $\Delta_p$ is a time delay, $\mathbf{w}_i = [w_{11}, w_{12}, \ldots w_{im}]^T$, $\mathbf{x}_i = [x_{11}, x_{12}, \ldots x_{1m}]^T$, $\tilde{\mathbf{y}}_i = [y_i(t-1), y_i(t-2), \ldots, y_i(t-L)]^T$, and $\mathbf{b}_i = [b_{11}, b_{12}, \ldots, b_{1L}]^T$. It should be noted that each block $\mathrm{ICA}_i$ has two such processing units. The cost function can be evaluated as follows:

$$J(\mathbf{w}_i, \mathbf{b}_i) = E\{\varepsilon_i^2\} = \mathbf{w}_i^T \hat{\mathbf{R}}_{\mathbf{x}_i\mathbf{x}_i} \mathbf{w}_i - 2\mathbf{w}_i^T \hat{\mathbf{R}}_{\mathbf{x}_i\tilde{\mathbf{y}}_i} \mathbf{b}_i + \mathbf{b}_i^T \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i} \mathbf{b}_i, \quad (16)$$

where $\hat{\mathbf{R}}_{\mathbf{x}_i\mathbf{x}_i} = E\{\mathbf{x}_i\mathbf{x}_i^T\}$, $\hat{\mathbf{R}}_{\mathbf{x}_i\tilde{\mathbf{y}}_i} = E\{\mathbf{x}\tilde{\mathbf{y}}_i^T\}$ and $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i} = E\{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T\}$, mean estimation of true values of covariance and cross-correlation matrices: $\mathbf{R}_{\mathbf{x}_i\mathbf{x}_i}, \mathbf{R}_{\mathbf{x}_i\tilde{\mathbf{y}}_i}, \mathbf{R}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}$, respectively. In order to estimate vectors $\mathbf{w}_i$ and $\mathbf{b}_i$ we evaluate the gradients of the cost function and equalize them to zero as follows:

$$\frac{\partial J_i(\mathbf{w}_i, \mathbf{b}_i)}{\partial \mathbf{w}_i} = 2\hat{\mathbf{R}}_{\mathbf{x}_i\mathbf{x}_i}\mathbf{w}_i - 2\hat{\mathbf{R}}_{\mathbf{x}_i\tilde{\mathbf{y}}_i}\mathbf{b}_i = \mathbf{0}, \quad (17)$$

$$\frac{\partial J_i(\mathbf{w}_i, \mathbf{b}_i)}{\partial \mathbf{b}_i} = 2\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}\mathbf{b}_i - 2\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\mathbf{x}_i}\mathbf{w}_i = \mathbf{0}, \quad (18)$$

Solving the above matrix equations we obtain a new iterative algorithm

$$\mathbf{w}_i = \hat{\mathbf{R}}_{\mathbf{x}_i\mathbf{x}_i}^{-1} \hat{\mathbf{R}}_{\mathbf{x}\tilde{\mathbf{y}}_i} \mathbf{b}_i \quad (19)$$

$$\mathbf{b}_i = \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}^{-1} \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\mathbf{x}_i} \mathbf{w}_i = \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}^{-1} \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i y_i}, \quad (20)$$

where the matrices $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}$ and $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i y_i}$ are estimated on basis of parameters $\mathbf{w}_i$ obtained in previous iteration step.

Remark: It should be emphasized here that in our derivation we have assumed that $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}$ and $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i y_i}$ are independent of actual evaluated vector $\mathbf{w}_i$, i.e. they are estimated on value of $\mathbf{w}_i(t-\Delta_1)$ in the previous iteration step. This two phase procedure is similar to EM schemes: (i) freeze the covariance and cross-correlation matrices and learn the parameters of the processing unit $(\mathbf{w}_i, \mathbf{b}_i)$; (ii) freeze $\mathbf{w}_i$ and $\mathbf{b}_i$ and learn new statistics (i.e. matrices $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_i y_i}$ and $\mathbf{R}_{\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i}$) of estimated source signal, then go back to (i) and repeat. Hence, in phase (i) our algorithm extract a source signal, whereas in phase (ii) it learns the statistics of the source.

In order to avoid the trivial solution $\mathbf{w}_i = \mathbf{0}$ we can perform the normalization of the vector $\mathbf{w}_i$ to unit length in each iteration step as $\mathbf{w}_{1*} = \mathbf{w}_i / \|\mathbf{w}_i\|$ (what ensures that $E\{y_i^2\} = 1$).

The above algorithm can be considerably simplified. It should be noted that in order to avoid the inversion of autocorrelation matrix $\mathbf{R}_{\mathbf{x}_i\mathbf{x}_i}$ in each iteration step we can apply as preprocessing the standard pre-whitening or standard PCA (principal component analysis) and next normalization of sensor signals to unit variance. In such case $\hat{\mathbf{R}}_{\mathbf{x}_i\mathbf{x}_i} = \mathbf{I}_n$ and the algorithm simplifies to

$$\mathbf{w}_i = \hat{\mathbf{R}}_{\mathbf{x}_i\tilde{\mathbf{y}}_i}\mathbf{b}_i = \hat{\mathbf{R}}_{\mathbf{x}_i\overline{\mathbf{y}}_i}, \quad (21)$$

where $\hat{\mathbf{R}}_{\mathbf{x}_i\overline{y}} \cong \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_i(t)\,\overline{y}_i(t)$ and $\overline{y}_i = \mathbf{b}_i^T\,\tilde{\mathbf{y}}_i = \sum_{p=0}^{L} b_{1p} y_i(t-\Delta_p)$. In the case of discrete signals, for a sampling frequency $f_s$, it will be $\Delta_p = \frac{p}{2\pi}$.

Furthermore, it is worth to note that the learning rule (20) can be simplified as

$$b_{1p} = \frac{E\{y_i(t-\Delta_p)y_i(t)\}}{E\{y_i^2(t-\Delta_p)\}} \quad (22)$$

if we use single delay units with time delay $p$ instead of FIR filter with lengths $L$.

The above formulas (21) and (20) builds up our basic simplified learning algorithm. A length of FIR filter should be chosen sufficiently large but a value of $L \approx 10$ was enough in our experiments. However, as shown by our extensive computer simulations, in practice it is sufficient to use only a single delay unit with suitably chosen delay $p$ if some a priori information about source signals is available. The suitable choice

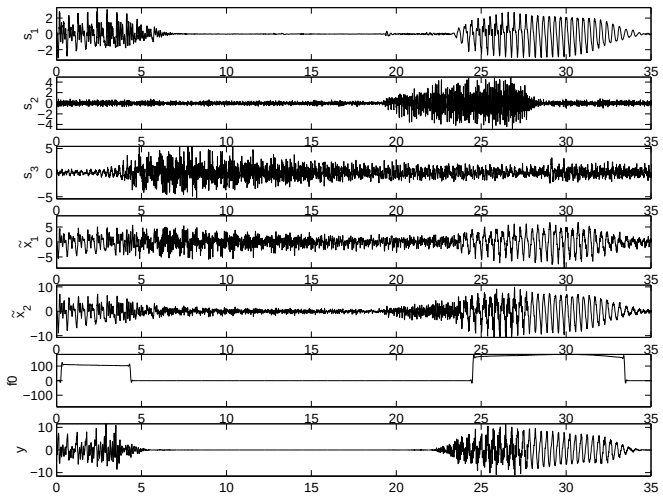

Figure 2: Example of the original speech signals $[s_1 s_2 s_3]^T$, the mixed/convolved ones $[\tilde{x}_1 \tilde{x}_2]$, and the resulting signal $y$ along with the extracted fundamental frequency $f0$.

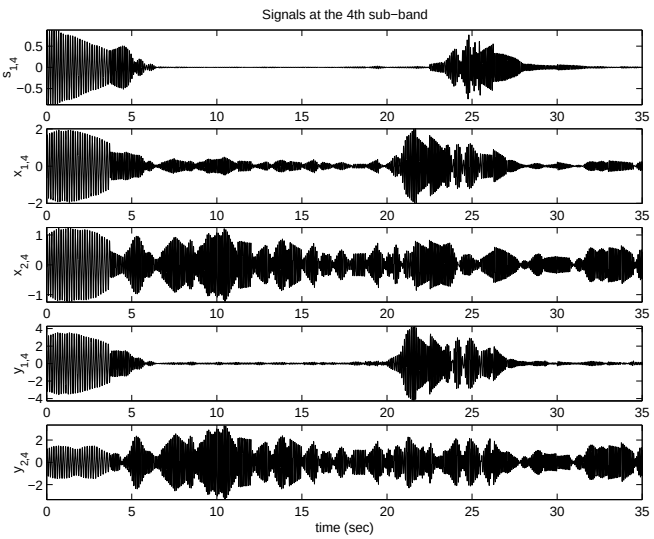


Figure 3: Intermediary signals (the fourth sub-band), just before and after the bank of ICA algorithms, along with the *desired* one.

of single delay $p$ depends on autocorrelation function of extracted source [4].

## 4. SIMULATION RESULTS

We have carried out simulations where we mixed and convolved three independent speech signals into two mixtures, as modeled by (2), where $n = 3$ and $m = 2$. The *desired* signal was a Japanese male speech and the interferences were a male singing plus the sound of a laugh. The task was to find the signal with the highest energy. This simulation aimed to mimic the case when one speaker is close to the listener, but there is some background interference, possibly caused by other speakers or music.

Fig. 2 shows an example of the original speech signals, the mixed/convolved ones, and the resulting signal along with the extracted fundamental frequency. On the other hand, Fig. 3 shows intermediary signals, just before and after the bank of ICA algorithms. Notice that this processing enhanced in fact the quality of the signals.

## 5. DISCUSSIONS

In this work, our objective was not to extract the original source from a convolution mixture. Rather, we aimed to enhance a higher energy speech signal which would be more ineligible. The task was particularly difficult due to the mixture be an over-complete one: there was more signals than sensors (microphones). Moreover, there was the reverbation effect which is usually present in real environment.

Thus, we have divided the signals in the frequency space. The advantage of this is twofold: 1) we are diminishing the probability of finding two signals in the same frequency band and; 2) diminishing the convolutive effect, which, gradually turns into a instantaneous mixture as the frequency band diminishes.

Preliminary results shown here and tested with other signals have shown that this technique may be promising towards the enhancement of signals embedded in noisy environment, and modeling the auditory system.

## 6. REFERENCES

[1] S. Amari and A. Cichocki: "Adaptive blind signal processing - neural network approaches," *Proceedings IEEE* (invited paper), Vol.86, No.10, Oct. 1998, pp. 2026-2048.

[2] S. Amari, "ICA of temporally correlated signals - learning algorithm," *Proc. ICA '99*, Aussois, France, pp. 13-18, Jan. 1999.

[3] A. K. Barros and N. Ohnishi, "Removal of quasi-periodic sources from physiological measurements" *Proc. ICA '99*, Aussois, France, pp. 185-189, Jan. 1999.

[4] A. K. Barros and A. Cichocki, "RICA - Reliable and robust program for Independent Component Analysis",
Report and MATLAB program of Riken, web page

http://www.riken.nagoya.jp/sensor/allan/RICA or http://go.to/RICA

[5] A. K. Barros and N. Ohnishi, "Amplitude estimation of quasi-periodic physiological signals by wavelets", submitted to a journal.

[6] A. Belouchrani, K. Meraim, J.-F. Cardoso and E. Moulines, "A blind source separation technique based on second order statistics". *IEEE Trans. on Signal Processing*, 45, pp. 434-444, 1997.

[7] Berthommier, F., and Meyer, G. (1995), "Source separation by a functional model of amplitude demodulation", *Proc. Eurospeech*, pp. 135-138

[8] A.Cichocki, R. Thawonmas and S. Amari. "Sequential blind signal extraction in order specified by stochastic properties", *Electronics Letters*, vol. 33, No. 1, pp. 64-65, Jan. 1997.

[9] A.Cichocki, A. K. Barros. "Robust batch algorithm for sequential blind extraction of noisy biomedical signals", *Proc. ISSPA '99*, Australia, 1999.

[10] P. Comon, (1994) "Independent component analysis, a new concept?" *Signal Processing*, 24, pp. 287 - 314.

[11] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach", *Signal Processing*, vol. 45, pp. 59 - 83, 1995.

[12] S. Haykin, *Adaptive filter theory.* Englewood Cliffs, NJ: Prentice-Hall, 1991.

[13] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis". *Neural Computation* (9), 1483 - 1492, 1997.

[14] C. Jutten and J. Hérault "Independent component analysis versus PCA," *Proc. EUSIPCO*, pp. 643 - 646, 1988.

[15] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigne, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds", *Speech Communication*, 27, pp.187-207 1999.

[16] T-W Lee, *Independent component analysis.* Kluwer Academic Publishers, 1998.

[17] J. Luo, B. Hu, X-T. Ling and R-W. Liu, "Principal independent component analysis". *IEEE Trans. on Neural Networks* 10, 4, pp. 912-917, 1999.

[18] L. Molgedey, H.g. Schuster, "Separation of a mixture of independent signals using time-delayed correlations, *Phys. Rev. Lett.*, vol. 72(23), pp. 3634-3637, 1994.

[19] S. Ikeda and N. Murata, "A method of ICA in time frequency domain," *Proc. ICA '99*, Aussois, France, pp. 365-370, Jan. 1999.

[20] S. Orfanidis, *Optimum Signal Processing* McGraw-Hill, N. York 1990.

[21] A. Papoulis. *Probability, random variables, and stochastic processes.* McGraw-Hill, 1991.

[22] Weintraub, M. "A theory and computational model of auditory monaural sound separation", Doctoral dissertation, Stanford University.1985

[23] Zissmann, M. A., and Weinstein, C. J. (1990). "Automatic talker activity labeling for co-channel talker interference suppression.", Proc. IEEE-ICASSP, 813-816.

[24] Virag, N., "Single channel speech enhancement based on masking properties of the human auditory system". *IEEE Trans. on Signal Processing*, Vol. 7, No. 2, pp. 126-137, 1999.