**Proceedings of the Second International Conference
BIOMEDICAL ENGINEERING
February 16-18, 2004, Innsbruck, Austria**

On the Extraction of the Snore Acoustic Signal by Independent Component Analysis

F. Vrins[1], J. Deswert, D. Bouvy, V. Bouillon, J. A. Lee[1], C. Eugène[2] and M. Verleysen[1]
[1] Microelectronics Laboratory, Machine Learning Group
[2] Electrotechnics and Instrumentation Laboratory
Department of Electricity
Université catholique de Louvain (UCL),
Place du Levant 3, 1348 Louvain-la-Neuve, Belgium
{vrins, lee, verleysen}@dice.ucl.ac.be, eugene@lei.ucl.ac.be

Abstract

Physicians are interested in the acoustic signal of snore, because it allows them to diagnose the patient and eventually to avoid several dangerous accidents. Today, its measure is not satisfactory for various reasons. In this paper, we explore a new way to measure this signal: Blind Source Separation (BSS). We give encouraging results of source separation in this application but also stress the obstacles which prevent a perfect separation when BSS is achieved by the classic linear, instantaneous and unnoiseless model of Independent Component Analysis, an undoubtfully very promising signal processing method in the biomedical world.

Key Words
Snoring, non-invasive measurement, independent component analysis.

1 Introduction

Recently, new signal processing techniques appeared in the biomedical application's world (*a.o.* electroencephalogram [1], fetal-electrocardiogram [2, 3], ...), like Blind Sources Separation (BSS). BSS consists in recovering statistically independent sources by analyzing only mixtures of them (sources are -supposed to be- unknown). Under certain assumptions, Independent Component Analysis (ICA) is able to achieve BSS. However, several 'real-world' problems prevent ICA to recover perfectly each source. Indeed, the theoretical model does not exactly correspond to reality. Nevertheless, in many cases, ICA could give interesting results to the physicians.

A new biomedical application is considered here: the extraction of an acoustic snore signal. This is an important application: snoring has many consequences on the patient's life, and the acoustic signal could really help the physicians to diagnose breathing problems and to evaluate the risk incurred by the patient [4]. Today, it is difficult to record this signal. A brief analysis of the situation shows that this problem corresponds to the framework of BSS, and an attempt of solution by ICA seems natural. In this paper, we explore a first tentative to extract the acoustic snore signal by Independent Component Analysis, and we stress the obstacles of the BSS of acoustic signal measured in real environment in general which make the problem complex.

The paper is organized as follows: section 2 describes the snoring phenomenon, and stresses its interest. Section 3 presents the framework of BSS. Considerations are given in section 4 about the feasibility of the snore signal extraction by ICA. We present the signal processing in order to achieve the extraction of the snore signal in section 5. We show results for mixtures including snore, narrow-spectrum and wide-spectrum signals in section 6. Discussions and conclusions are given in section 7 and 8.

2 The snoring

The snoring is a well-known physical phenomenon, appearing (most often on male patients) during sleep. The acoustic signal of snore is caused by the vibrations of the pharynx slack tissues. Several snore levels exist, classified with respect to their power spectral content, position of the patient during snoring, etc. In addition to the acoustic nuisances, snoring can cause others physical problems for people which are directly suffering from, like tiredness during the day. It also increases the probability of cardiovascular and cerebral accidents.

2.1 Specificity of the signal

The acoustic signal of snore allows diagnosing the patient. Its measure could be associated with other physical data, like electroencephalograms, electro-oculograms, etc. The power is concentrated under 5kHz. The time structure of a typical snore signal is shown in Figure 1 (f_{samp}, the sample frequency, is 44.1 kHz). The snore signal is almost periodic and stationary on windows larger than its period.

2.2 Measurement: state-of-the-art

Currently, two methods exist to record the snore signals: microphones and piezoelectric sensors. Among the micro-

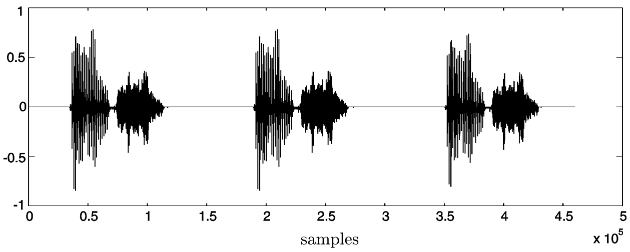

Figure 1. Time structure of a snore signal (total time-window $\simeq$ 10 seconds).

phones used in practice [5, 6], the *ambiant* ones are corrupted by ambiant noise, the *thorax* ones distort the signal (low-frequency filter) and those placed *under the nose* are not convenient for the patient. The *piezoelectric sensors* (placed under the mattress) are very convenient for snore detection, but not for spectral analysis, because of distorsions induced by the sensor at higher frequencies. It could be interesting to explore new ways to enhance the signal recovery (*i.e.* with a high signal-to-noise ratio).

3 Blind Source Separation

Blind Source Separation (BSS) consists in recovering m statistically independent sources ($\mathbf{s} = [s_1, \ \ldots, s_m]^T$), from n mixtures of them ($\mathbf{x} = [x_1, \ldots, x_n]^T$), with $n = m$ (if $n > m$, a dimension reduction of the observations $\mathbf{x}'$ has to be done by selection or projection to an m-dimensional subspace). In most of the algorithms which are able to achieve BSS, the mixtures are supposed to be whitened (zero-mean and covariance matrix equal to the unit matrix): $\mu_{\mathbf{x}} = E\{\mathbf{x}\} = 0$, $C_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{I}$ where E denotes the expectation. If they are not, a linear transformation of the x_i's (based on the EVD of $C_{\mathbf{x}}$) must be done. Statistical independence between the sources s_i expresses that nothing could be known about s_k even if one knows $s_{l \neq k}$ ($1 \leq i, k, l \leq m$). Formally, this implies that the product of their marginal f_{s_i} densities (PDF) equals the joint density (JPDF) $f_{\mathbf{s}}$:

$$\prod_{i=1}^{m} f_{s_i}(s_i) = f_{\mathbf{s}}(\mathbf{s}) \ . \tag{1}$$

A particular case of BSS is when the mixture is linear without additive noise:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \ , \tag{2}$$

where $\mathbf{A}$ is an $m \times m$ mixing matrix. This problem can be solved by Independent Component Analysis (ICA). ICA allows recovering estimations $\hat{\mathbf{s}}$ of the original sources $\mathbf{s}$, up to a permutation and a scale factor ($\hat{s}_i = \mathbf{P}_{ii} s_k$, $\mathbf{P}_{ii} \in \mathbb{R}$) by finding an unmixing matrix $\mathbf{W}$ such that:

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{A}\mathbf{s} \simeq \mathbf{P}\mathbf{D}\mathbf{s} \ , \tag{3}$$

where $\mathbf{P}$ and $\mathbf{D}$ are $m \times m$ permutation and diagonal matrices, respectively. The assumptions on the model are multiple: the sources, in addition to their independence, must be stationary and the mixtures are supposed to be linear, instantaneous (*source-sensor* delays are equal for all x_i) and noise-free. However, many other algorithms exist to deal with particular cases (non-linearities [7], noise [8], ...).

4 Feasibility of ICA applied to snore

The extraction of the acoustic snore signal seems to correspond with the general problem of sources separation. Indeed, signals recorded by microphones are mixtures of independent sources: the snore signal itself and other perturbation signals (electromagnetic noise from instruments or parasitic acoustic sources, like ambiant noise). Furthermore, this method circumvents some of the drawbacks detailed in section 2.2 and is non-invasive (the microphones are not in contact with the patient).

Simple de-noising (using classical pass-band filters or wavelets) is not satisfying, because the *noise* (*i.e.* all contributions in the records which are not properly *snoring*, coming from the snorer or from its environment) present in the snore's frequency band will not be removed !

We have detailed in the previous section the conditions that allow ICA to solve a BSS problem. Before applying this method on snore signals, we have to check that:

1. the signals recovered by ICA include the interesting information;
2. the assumptions of the model are valid.

4.1 Output of ICA

First, we have to remind that ICA recovers sources only up to a scale factor and permutation (given by $\mathbf{D}$ and $\mathbf{P}$, respectively), meaning that only the waveform of the recovered sources will be relevant after identification of the snore signal among the outputs (this identification is done by analyzing the frequency content of the signals).

Many ICA algorithms try to maximize a *contrast function* Φ, which has the drawback of growing monotonically with the magnitude (*i.e.* with the variance) of the output signals, leading to find a separating system $\mathbf{W}$ arbitrarily large. In order to circumvent this problem, the outputs are normalized at each step of the algorithm, thus all of the estimated sources $\hat{\mathbf{s}}$ will be unit variance. This implies that the sum of the squared elements of a row of the unmixing matrix $\mathbf{W}$ is equal to one:

$$\begin{aligned} \sigma^2_{\hat{s}_j} &= \sum_{i=1}^{m} \left(\mathbf{WA}\right)^2_{ji} \sigma^2_{s_i} = \sum_{i=1}^{n} \mathbf{W}^2_{ji} \sigma^2_{x_i} \\ &= \sum_{i=1}^{n} \mathbf{W}^2_{ji} = 1 \ . \end{aligned} \tag{4}$$

This magnitude indetermination can be avoided by a measure of acoustic power in parallel (with an adequate probe), as suggested by Figure 2. In this paper, we focus on the snore extraction problem, other steps (signal identification

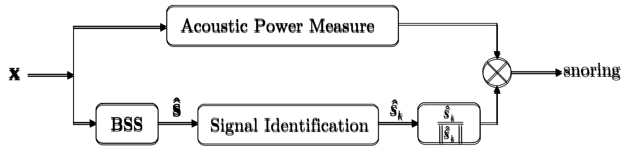

Figure 2. Snoring measurement process: wave-form captured by Blind Source Separation, power level measured by an appropriate probe.

and power measurement) could be easily achieved in parallel by other means, as previously explained.

4.2 ICA assumptions

We process two mixtures of two sources ($n = m = 2$). The assumptions of stationarity (mean, covariance and auto-correlations constant in time) and independence of the sources intuitively correspond to reality, but dealing with linearity and instantaneity of the mixtures is less trivial. In a standard room (reverberant environment), the power of the recorded signals is attenuated. These signals also suffer from multipath propagation effects due to multiple reflections, implying various delays that depend on the distances between the sources and sensors [9]. The measured signals are actually linear mixtures of sources, but also of time-delayed versions of themselves. This corresponds to the mathematical model of convolution, which is more realistic but also more complex. Nevertheless, in this first study of the extraction of a snore signal, we try to solve this problem with the classical model of ICA, but we compensate the simplicity of the model by some measurement precautions (*e.g.* signals are recorded in an echo-free room, with professional recording material, etc.). We test in the following subsections the validity of assumptions of *linearity* and *instantaneity*.

4.2.1 Linearity

The assumption of the linearity of acoustic mixtures in the air is verified by the following experience. Consider one microphone and two sources s_1 and s_2. Record a signal x_1=s_1 (turn *off* s_2) and then record x_2=s_2 (turn *off* s_1). Next, record $x_3 = s_1 + s_2$ (both sources *on*). Then linearity assumption holds if the covariance between x_3 and $\alpha x_1 + \beta x_2$ is one for α=β, which is effectively the case (in echo-free room), as shown in Figure 3 (a).

4.2.2 Instantaneity

The difference between delays could have multiple physical origins. Indeed, the sound speed is finite and the distances between sensors and sources are in general not equal. Delays could also be induced by the transfer function of the instrumentation recording each acoustic signal, which could be different for each measurement chain even with professional material. Some authors have tried to include in the BSS algorithm a delay correction, but they were confronted to the problem of local ‘attractors’ (local minima) [10].

Int this work, we propose to chose the geometrical configuration of sources and sensors shown in Figure 4 in order to limit the delays to a single number (*i.e.* τ^*, the number of sample associated to the time-delay). Indeed, if τ_i denotes the time-propagation of the sound for a distance i, we have:

$$x_1(t) = a_{11} s_1(t - \tau_a) + a_{12} s_2(t - \tau_b) \quad (5)$$
$$x_2(t) = a_{21} s_1(t - \tau_{a+c}) + a_{22} s_2(t - \tau_{b+c}) \quad (6)$$

Denoting $s'_1(t) = s_1(t - \tau_a)$, $s'_2(t) = s_2(t - \tau_b)$, and according that $\tau_{i+j} = \tau_i + \tau_j$, we find:

$$x_1(t) = a_{11} s'_1(t) + a_{12} s'_2(t) \quad (7)$$
$$x_2(t) = a_{21} s'_1(t - \tau_c) + a_{22} s'_2(t - \tau_c) \quad (8)$$

Next, it will be possible to correct the error due to the delay τ_c if the environment is non-dispersive. The sample delay $\tau^* = f_{samp} \tau_c$ is measured by optimizing a relevant criterion before applying a standard ICA algorithm. This method of delay correction is valid for $m = n = 2$. Note that $\hat{\mathbf{s}}$ (the estimated sources extract from the observed signals $\mathbf{x}$) will be actually the estimated of $\mathbf{s}'$, a slightly delayed versions of the original sources $\mathbf{s}$. The first solution

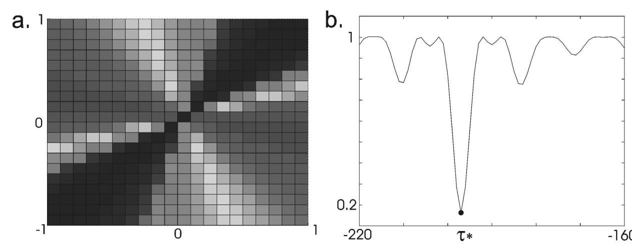

Figure 3. a) Covariance between $\alpha x_1 + \beta x_2$ and $x_1 + x_2$ *vs* α (x axis) and β (y axis). The covariance is maximum (darkest grey) for $\alpha = \beta$, implying that the linearity assumption of the mixture of acoustic signals in echo-free room holds; b) Determinant of covariance between x_1 and x_2 ($E\{x_1(t)x_2(t - \tau)\}$) *vs* τ ($\tau^* \simeq -197$).

to estimated τ^* is a linear method: we compute several time-delayed covariance matrices between mixtures:

$$\Sigma_\tau = E\{x_1(t)x_2(t - \tau)\} \ , \quad (9)$$

and we search the one (say Σ_{τ^*}) that maximizes the sum of its squared extra-diagonal entries, *i.e.* which has the lowest determinant (see Figure 3 (b)):

This is equivalent to finding the sample delay τ^* for which the observations x_i are the closest from $x_{j \neq i}$: although x_i and $x_{j \neq i}$ are sums of independent sources weighted in a different way, their correlation must be maximum when they are synchronized. A second (nonlinear) possible way is to maximize the mutual information between the x_i, but it requires the estimation of the probability density functions and a numerical integration, which is not convenient and very time-consuming. The first method is thus used here, even if the processing is not a 'real-time' one.

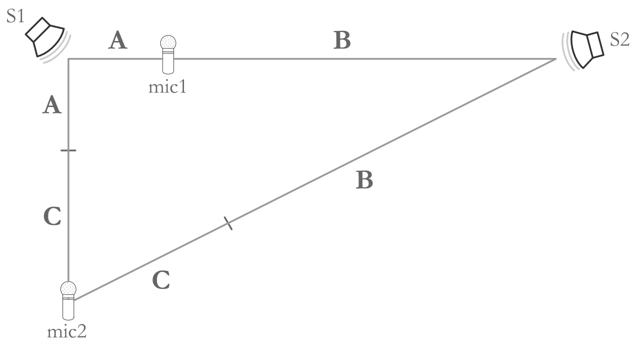

Figure 4. Geometrical configuration of sources and microphones.

5 Processing and extraction of the snore

In section 4, we have explained that it is theoretically possible to extract a snore signal by ICA, and that the power and permutation indeterminations could be easily circumvented by other means. In this section, we detail the processing of the acoustic mixtures $\mathbf{x}$ from the emission of the sources $\mathbf{s}$ to their (delayed) estimations $\hat{\mathbf{s}}$.

5.1 Recording

As we use the simple superimposition model, we have to respect it as well as possible: the difference between delays is compensated by choosing a particular geometric configuration (see Figure 4) in echo-free room of the sources and sensors and finding τ^*, the delay which minimizes the determinant of the covariance matrix Σ_τ, as explained in the previous section. The sources are emitted by two speakers and recorded through two non-saturated microphones (in order to avoid post-nonlinearities in the recorded mixtures).

5.2 Extraction using FastICA

We analyzed mixtures of two sources: snore and signals with different spectral contents; we suggest to use the FastICA algorithm [11] to perform ICA on these mixtures. The function Φ optimized by FastICA to recover independent source is based on nongaussianity [12], measured through the negentropy. This approach lays on two principles: *the central limit theorem* and the *maximum differential entropy of a Gaussian variable*.

The central limit theorem The central limit theorem [13] says that if u is a sum of *n* random iid variables ($n \to \infty$), then $f_u(u)$ tends to a Gaussian function. In other words, it means that the PDF of a mixture of independent variables is *closer from a Gaussian* than the PDF of each variable involved in the mixture.

Maximum differential entropy of a Gaussian variable The differential entropy [13] $h(x)$ is a measure of randomness of a variable x:

$$h(x) \doteq -\int_u f_x(u) \log\big(f_x(u)\big) du \ , \qquad (10)$$

One of the most important results of information theory [13, 14] is that a variable x_G which has a Gaussian PDF has the highest differential entropy (among all unbounded variables x) for a given variance ($\sigma_x^2 = \sigma_{x_G}^2$):

$$h(x) \leqslant h(x_G) = \frac{1}{2}\log(2\pi e)\sigma_x^2 \ , \qquad (11)$$

with equality if and only if x is Gaussian.

Negentropy: a measure of nongaussianity Melting those two principles, we can say that to find one original source, we have to find an output $\hat{s}_k$ which is very different from a Gaussian one [1], *i.e.* which maximizes the nongaussianity, measured through the negentropy $J(\hat{s}_k)$ defined as:

$$J(\hat{s}_k) \doteq h(x_G) - h(\hat{s}_k) \geq 0 \ . \qquad (12)$$

In order to avoid the estimation of the PDF $f_{\hat{s}_k}$, a good approximation is derived, which is actually the measure of nongaussianity used in FastICA:

$$\Phi = \hat{J}(\hat{s}_k) \propto \left(E\{g(\hat{s}_k)\} - E\{g(x_G)\}\right)^2 \ , \qquad (13)$$

where g is a non-linear function, chosen according to the input signals $\mathbf{x}$ [11].

6 Test on snore signals

In a real-world context, applications that fully respect all ICA assumptions proves rare. However, the results can be satisfactory in many situations. This is the reason why we try to apply ICA, keeping in mind that the instantaneity of the mixture is not guaranteed and could raise multiple difficulties. In this section, the snoring source consists in a simulated snore emitted by a speaker. We give in Figures 5 and 6 the observed mixtures $\mathbf{x}$ and the waveforms of the estimated sources $\hat{\mathbf{s}}$ for two experiences. In the first one, the snore is mixed with a pure sine wave; in the second one, it was mixed with a large-band spectrum signal (music track). Two microphones are used, positioned with regard to the acoustic sources as in Figure 4.

[1]this is one of the reason why at most one source could be Gaussian in any ICA problem.

6.1 Snore and pure sine mixtures

We mixed the snore signal with narrow-band signal: a pure sine wave at 1076 Hz. The mixtures recorded by the microphones and the estimated sources after applying FastICA are given in Figure 5 (the figure scale prevents to distinguish the sine wave otherwise than the black rectangle, but allows identifying the snore signal). The estimated sources plotted in the frequency space (after Fourier transform) show that the sine is almost perfectly recovered (actually, as we can see it on the third graph of Figure 5, the estimated sine is slightly polluted, by the snore), and so is the snore.

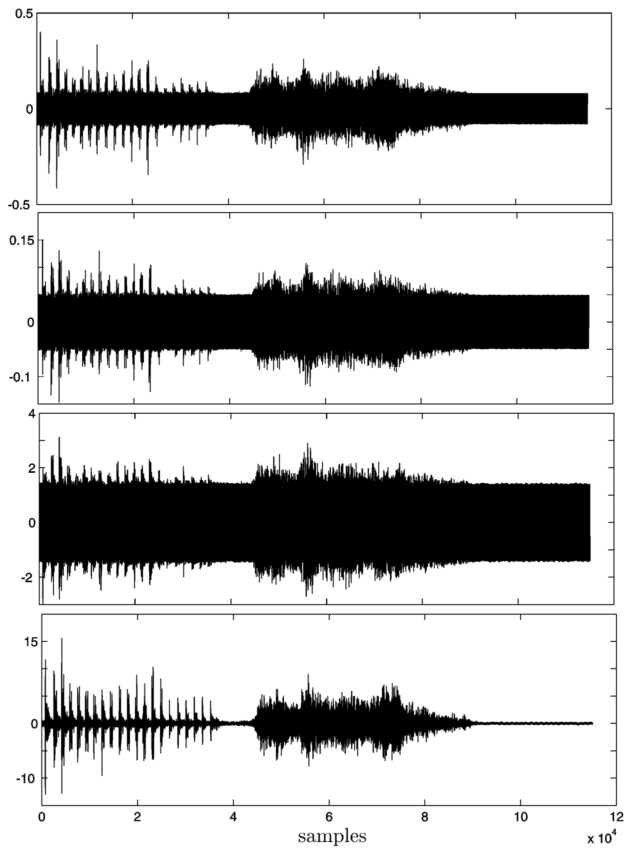

Figure 5. From top to bottom: 2 mixtures (x_1, x_2) of a sine wave and a snore signal; recovered sine wave $\hat{s}_1$; and recovered snore signal $\hat{s}_2$ (a frequency spectrum analysis confirms the quality of the extraction).

6.2 Snore and large-band spectrum mixtures

We mixed then the snore signal with a large-band one (music track). The mixtures measured by the sensors and the estimated sources after FastICA are given in Figure 6. The estimated sources are very far away from the original ones (indeed, the temporal structure of the estimated ones are similar, contrarily to the true ones). The algorithm of separation fails to separate the snore from the music track.

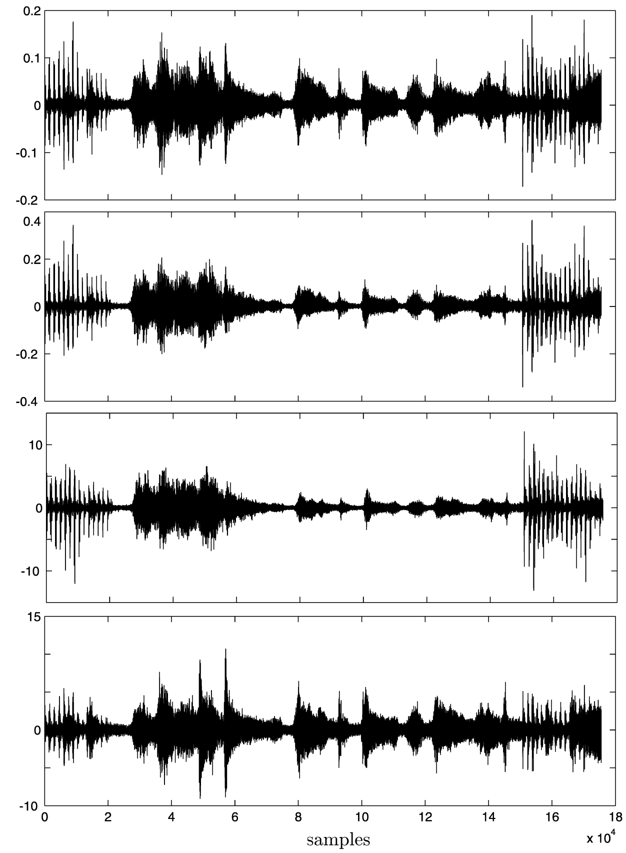

Figure 6. From top to bottom: 2 mixtures (x_1, x_2) of large band and snore; recovered large band signal $\hat{s}_1$; and recovered snore $\hat{s}_2$.

7 Discussion

The separation results on a mixture of snore and narrow-band signals (separation well performed) and, on the other hand, snore and large-band signals (not well performed) are very different. This is due to the delays: even if those due to the propagation time could vanish (by choosing an appropriate geometrical configuration and applying a time-shifting), those induced by the transfer functions of the measuring instruments are very thorny, because they depend on the frequency content of the signal : the compensation should have to be effective for each frequency in the case of large-band signals ! Indeed, as the delay depends on the frequency - even if the air is a non-dispersive medium - we must correct the mixtures by shifting each frequency component present in the mix in a different way; on the contrary, for the unique-frequency signal (here the sine), a simple shift suffices. This problem (occurring at various levels in all real-acoustic signal applications) can drastically decrease the performances of separation. Experiences with professional instruments lead to better results, but the separation is not perfect yet.

8 Conclusion

The extraction of a snore acoustic signal by BSS is promising. Indeed, microphones catch acoustic signals which are mixtures of different independent sources. In this paper, which is a first attempt to extract a snore signal only from linear mixtures of several sources, we applied a simple method to achieve BSS : Independent Component Analysis. In this work, we have used the simple instantaneous superimposition model, but actually, the most realistic model (which is also much more complex) is the multichannel convolution with post-nonlinearities. Even if all assumptions of the ICA model are not fully respected (as in most real-world applications), we tried to solve the problem with a basic ICA model. Among all imperfections, the instantaneity seems to be the most critical one: the presence of noise or nonlinearities does not have so dramatic consequences on the convergence of the ICA algorithm. That is the reason why the simplicity of the linear superimposition model has to be compensated by several precautions (like particular geometrical configurations of sources and sensors, limitations of the reverberations of the acoustic signals, etc.); despite these precautions, the extraction is still not perfect. Troubles come from the hardware equipment! The transfer functions of two measurement chains (even with professional hardware) are different, generating delays between mixtures. In this application, we face all problems appearing in the BSS of acoustic signals in real environment; for instance it appears that the delays depending on the frequency (induced by the measurement chain and unavoidably in real application) have a bad influence on the signals separation (hardware imperfections still decrease the performances of BSS in 'real-world' applications).

Many software improvements can be considered: dealing with propagation delays in more general geometrical situations (leading to a better real-time approach), with those induced by the different transfer functions of the recording instrumentation (probably the most difficult point), with the nonlinearities due to the microphones, multichannel blind deconvolution (for the recording in a standard hospital room), etc.

9 Acknowledgment

The authors would like to thank Laborelec for the material and experimentation facilities used in this work and Dr G. Liistro (UCL Pneumology dpt) for his clinical advises. Michel Verleysen is a Senior Research Associate of the the Belgian National Funds for the Scientific Research (FNRS).

References

[1] R. Vigario and E. Oja, Independence, a new criterion for the analysis of the electromagnetic fields in the global brain?, *Neural Networks*,13 (8-9), 2000, 891-907.

[2] L. De Lathauwer, B. De Moor and J. Vandewalle, Fetal electrocardiogram extraction by blind source subspace separation, *IEEE Trans. Biomedical Engineering*,47(5), 2000, 567-572.

[3] F. Vrins, J. A. Lee, M. Verleysen, V. Vigneron and C. Jutten, Improving independent component analysis performances by variable selection, *Proceedings of the 13^{th} IEEE workshop on Neural Networks for Signal Processing (NNSP 2003)*, Toulouse, France, 2003, 359-368.

[4] J. Rogelio Perez-Padilla, E. Slawinsli, L.M. Difrancesco, R. R. Feige, J. E. Remmers and W. A. Whitelaw, Characteristics of the snoring noise in patients with and without occlusive sleep apnea, *American Review of Respiratory Disease*, 147, 1993, 635-644.

[5] F. Dalmasso and R. Prota, Snoring: analysis, measurement, clinical implications and applications, *European Respiratory Journal*, 9, 1996, 146-159.

[6] V. Hoffstein, Snoring, *Chest*,109, 1996, 201-222.

[7] A. Taleb and C. Jutten, Source separation in post-nonlinear mixtures, *IEEE Trans. on Signal Processing*, 47(10), 1999, 2807-2820.

[8] A. Hyvärinen, Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood, *Neurocomputing*, 22, 1998, 49-67.

[9] S. Choi, Acoustic source separation: Fundamental issues, *Proceedings of the International Conference on Speech Processing*, Seoul, Korea, 1999, 505-510.

[10] K. Torkkola, Blind separation of delayed sources based on information maximization, *Proceedings of the International Conference on Acoustics, Speech & Signal Processing (ICASSP)*, Atlanta, USA, 1996, 3509-3512.

[11] A. Hyvärinen, Fast and robust fixed-point algorithms for Independent Component Analysis, *IEEE Trans. on Neural Networks*, 10(3), 1999, 626-634.

[12] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis* (New York, John Wiley and Sons, Inc. 2001).

[13] T. M. Cover and J. A. Thomas, *Elements of information theory* (New York, John Wiley and Sons, Inc. 1991).

[14] C.E. Shannon, *The mathematical theory of Communication* (University of Illinois Press, Urbana, IL, 1948).