Digital Signal Processing 38 (2015) 22–31

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Denoising sound signals in a bioinspired non-negative spectro-temporal domain

C.E. Martínez [a,c,*], J. Goddard [b], L.E. Di Persia [a,d], D.H. Milone [a,d], H.L. Rufiner [a,c,d]

[a] Research Institute for Signals, Systems and Computational Intelligence, sinc(i), Facultad de Ingeniería, Universidad Nacional del Litoral–CONICET CC217, Ciudad Universitaria, Paraje El Pozo, S3000, Santa Fe, Argentina
[b] Dpto. de Ingeniería Eléctrica, UAM-Iztapalapa, Mexico
[c] Laboratorio de Cibernética, Facultad de Ingeniería-Universidad Nacional de Entre Ríos, Argentina
[d] CONICET, Argentina

ARTICLE INFO

Article history:
Available online 24 December 2014

Keywords:
Approximate auditory cortical representation
Sound denoising
Non-negative sparse coding
Bioinspired signal processing

ABSTRACT

The representation of sound signals at the cochlea and auditory cortical level has been studied as an alternative to classical analysis methods. In this work, we put forward a recently proposed feature extraction method called *approximate auditory cortical representation*, based on an approximation to the statistics of discharge patterns at the primary auditory cortex. The approach here proposed estimates a non-negative sparse coding with a combined dictionary of atoms. These atoms represent the spectro-temporal receptive fields of the auditory cortical neurons, and are calculated from the auditory spectrograms of clean signal and noise. The denoising is carried out on noisy signals by the reconstruction of the signal discarding the atoms corresponding to the noise. Experiments are presented using synthetic (chirps) and real data (speech), in the presence of additive noise. For the evaluation of the new method and its variants, we used two objective measures: the perceptual evaluation of speech quality and the segmental signal-to-noise ratio. Results show that the proposed method improves the quality of the signals, mainly under severe degradation.

1. Introduction

In previous years, several techniques of signal analysis have been applied to audio and speech denoising with relatively good results in controlled conditions [1]. However, it is widely known that the performance of these signal analysis techniques in adverse environments is far from that of a normal human listener [2]. On the other hand, there is an increasing number of new signal processing paradigms that promise to deal with more complex situations. This is the case with sparse coding and compressed sensing [3,4]. Their ability to efficiently solve challenging signal representation problems could be exploited in order to develop new audio and speech processing techniques.

For many years, researchers in the field of signal processing have greatly benefited from the use of methods inspired by human sensory mechanisms. Some examples of this for audio and speech encoding were *mel frequency cepstral coefficients* (MFCC) and *perceptual linear prediction* (PLP) coefficients [5]. Auditory representations of sound at the cochlea have been widely studied. Different mathematical and computational models have been developed that allow the approximate estimation of the so-called *early auditory spectrogram* [6,7]. These investigations have enabled an accurate modeling of the discharge patterns of the auditory nerve [8,9].

Although less known, the underlying mechanisms at the level of the auditory cortex have also been studied and modeled [10]. In experimental conditions – given a sound signal – a pattern of activations can be found at the primary auditory cortex that encodes a series of meaningful cues contained in the signal. This cortical representation seems to use two principles: the need for very few active elements in the representation and the statistical independence between these elements [11]. This behavior of the cortical neurons could be emulated using the fundamentals of *sparse coding* (SC) [12], the *independent component analysis* (ICA) [13] and the notion of *spectro-temporal receptive fields* (STRF). The STRF are defined as the optimal linear filter that convert a time-varying stimulus into the firing rate of an auditory cortical neuron, so that it responds with the largest possible activation [14]. These concepts have led to the development of a number of contemporary auditory models that incorporate different auditory phenomena, for example neural timing information [15], modeling of spectral and

* Corresponding author at: Research Institute for Signals, Systems and Computational Intelligence, sinc(i), Facultad de Ingeniería, Universidad Nacional del Litoral–CONICET CC217, Ciudad Universitaria, Paraje El Pozo, S3000, Santa Fe, Argentina.
E-mail address: cmartinez@fich.unl.edu.ar (C.E. Martínez).

http://dx.doi.org/10.1016/j.dsp.2014.12.008
1051-2004/

temporal content in the cochlear response [9]. A very complete and recent review on biologically-inspired models for speech processing is given in [16].

A number of works have explored the use of auditory models for building robust speech/speaker recognition system. In [17], a model of auditory perception (PEMO [18]) is used to obtain the features in a digit recognition system, after processed with well-established algorithms for speech enhancement (for example, the Ephraim and Malah estimator [19]). In [20], authors proposed the use of the model of Li [21] as a front-end in a hidden Markov model-based speech recognizer. Here, the speech is first pre-processed with state-of-the-art enhancement algorithms ([19,22] and others). More recently, different modifications of the MFCC representation were introduced (noise suppression, temporal masking and others) and compared to standard MFCC and PLP coefficients for speech recognition [23]. As can be seen, these efforts were mainly devoted – differently from our speech enhancement point of view – to build new feature extraction schemes for the recognizers while maintaining standard techniques for the enhancement itself.

In a previous work [24], the *approximate auditory cortical representation* (AACR) which is a set of activations computed using *matching pursuit* (MP) on a discrete dictionary of bidimensional atoms, was presented. These atoms represent the STRF of the auditory cortical neurons. The AACR intends to model the global statistical characteristics of the discharge patterns in the auditory cortex, in a phenomenological rather than a physiological way. This technique provides an approximated representation of the speech signal at the auditory cortical level. It has proved to be beneficial with respect to standard spectro-temporal techniques given the fact that at this higher level in the auditory path, some aspects of the acoustic signal that arrives at the eardrum have been reduced or eliminated [16]. Among these superfluous aspects are the temporal variability of the signal and the relative phase of acoustic waveforms [25]. This approach was then applied to a phoneme classification task in both clean and noisy conditions, showing the advantages of the intrinsic robustness of the sparse coding achieved.

In this work, this approach is adapted to a non-negative matrix factorization (NMF) framework. A non-negative auditory cortical representation is used in order to propose a novel sound denoising algorithm. NMF is a recently developed family of techniques for finding parts-based, linear representations of non-negative data [26–29]. These models deal with the temporal continuity of the signals (which is also found in our auditory spectrograms), such as slow variation of pitch in speech and music through consecutive frames, and were applied to monaural source separation. Regarding the speech processing applications, semi-supervised/supervised approaches were reported [30–33]. In these systems, first statistical models for clean speech/noise are estimated. Then, the input signal is analyzed to obtain the denoised version, which is then applied to the recognition block. In [34] two sparse dictionaries are obtained directly from spectrograms of clean speech and noise. Then, a representation of the noisy speech is obtained by a linear combination of a small number of both type of exemplars, in order to feed a robust speech recognizer.

In the biologically-inspired context, the NMF use data described by using just additive components, e.g. a weighted sum of only positive STRF atoms. This new model still retains its biological analogy, in spite of the fact that positive STRF implies only non-inhibitory behavior. Thus, positive coefficients could be interpreted as firing rates of excitatory cortical neurons. The new proposal of a non-negative auditory cortical denoising algorithm also differs from previous work in the sense that now two STRF dictionary are estimated from clean and noisy signals separately. Then, the dictionaries are combined in a mixed dictionary containing the most representative atoms for each case, obtaining a better representation of the important features of sound and noise for the denoising stage.

The organization of the paper is as follows. Section 2 presents the methods that give the signal representation in the approximate auditory cortical domain. Section 3 outlines the proposed technique to perform the signal denoising in this domain. Section 4 presents the experimental framework and data used in the following experimentation. Section 5 shows the obtained results and the discussions. Finally, Section 6 summarizes the contributions of the paper and outlines future research.

2. Sound signal representation

2.1. Early auditory model

Mesgarani and Shamma [10] proposed a model of sound processing carried out in the auditory system based on psychoacoustic facts found in physiological experiments in mammals. The main idea behind the model is first to obtain a representation of the sound in the auditory system. Then, they further decompose this representation to its spectral and temporal content in the cochlear response.

While the complete model of Shamma consists of two stages, in this work only the first stage was used. This stage produces the *auditory spectrogram* (AS), an internal cochlear representation of the pattern of vibrations along the basilar membrane.

In the following, subscript 'ch' stands for cochlear, 'an' for auditory nerve and 'hc' for hair cell. The first part of the model is implemented by a bank of 128 cochlear filters x_{ch} that process the temporal signal $s(t)$ and yield the outputs

$$x^k_{\text{ch}}(t, f) = s(t) \otimes h^k_{\text{ch}}(t, f), \tag{1}$$

where h^k_{ch} is the impulse response of the k-th cochlear filter [10]. This is a bank of overlapping constant-Q (QERB = 5.88) bandpass filters with center frequencies (CF) that are uniformly distributed along a logarithmic frequency axis, over 5.3 octaves (24 filters/octave, 0–4 kHz). The CF of the filter at location l on the logarithmic frequency axis (in octaves) is defined as

$$f_l = f_0 2^l \,(\text{Hz}), \tag{2}$$

where f_0 is a reference frequency of 1 kHz [10]. The quantity and frequency distribution of the filters proved to be satisfactory for the discrimination of important acoustic clues and for an appropriate reconstruction of speech signals [9].

These 128 filter outputs are transduced into auditory-nerve patterns x_{an} using

$$x^k_{\text{an}}(t, f) = g_{\text{hc}}\big(\partial_t x^k_{\text{ch}}(t, f)\big) \otimes \mu_{\text{hc}}(t), \tag{3}$$

where ∂_t represents the velocity fluid-cilia coupling (highpass filter effect), g_{hc} the nonlinear compression in the ionic channels (sigmoid function of the channel activations) and μ_{hc} the hair-cell membrane leakage modeling the phase-locking decreasing on the auditory nerve (lowpass filter effect) [10]. Finally, the lateral inhibitory network is approximated by a first-order derivative with respect to the tonotopic (frequency) axis, which is then half-wave rectified as

$$x^k_{\text{lin}}(t, f) = \max\big(\partial_f x^k_{\text{an}}(t, f), 0\big) \quad [10]. \tag{4}$$

The AS is then obtained by integrating this signal over a short window, modeling a further loss of phase locking. Fig. 1 shows a scheme of the auditory model as used in this work.

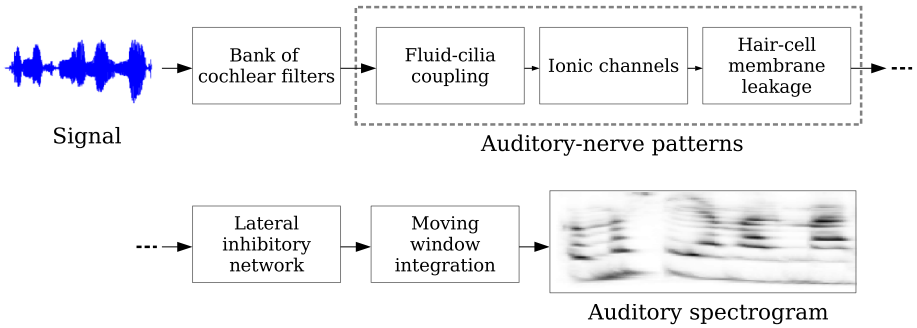

Fig. 1. Early auditory model.

2.2. Sparse coding of auditory spectrogram

We now suppose that the representation of any bidimensional slide signal $\mathbf{x} \in \mathbb{R}^{m \times n}$ obtained from the early auditory model in (4) is given by a linear combination of atoms representing the STRFs, in the form

$$\mathbf{x} = \mathbf{\Phi}\mathbf{a}, \tag{5}$$

where $\mathbf{\Phi} \in \mathbb{R}^{m \times n \times M}$ is the dictionary of M bidimensional atoms and $\mathbf{a} \in \mathbb{R}^M$ is the target representation. The 2-D basis functions of the dictionary are vectorized as $\mathbf{\Phi} = [\vec{\Phi}_1 \dots \vec{\Phi}_M]$ with $\vec{\Phi}_i \in \mathbb{R}^{[mn] \times 1}$. Then, (5) can be alternatively written as $\vec{x} = \sum_{1 \le i \le M} \vec{\Phi}_i a_i$. The desired sparsity is included when the solution is restricted to

$$\min_a \|\mathbf{a}\|_0, \tag{6}$$

where $\|\cdot\|_0$ is the l^0 norm, which counts the number of non-zeroes entries of the vector. This is an NP-complete problem so several approximations were proposed [35].

In order to find the required representation, two problems have to be jointly solved: the estimation of a sparse representation and the inference of a specialized dictionary. The coefficients found with methods based on *basis pursuit* (BP) or MP give both atoms and activations with positive and negative values [36,37]. However, in some applications it could be useful to work only with positive values, thus providing the method with the ability to explain the data from the controlled addition of (only positive) atoms. This is the objective of *non-negative matrix factorization* methods.

2.3. NN-K-SVD algorithm

As it was mentioned in Section 1, there are several approaches to obtain a nonnegative atomic sparse decomposition of data. Among them, in this work the method proposed in [38] is selected given its simplicity, excellent performance in other applications (for example, image classification [39]) and the possibility to explicitly set the number of sparse components to use in the approximation.

Aharon et al. introduced the K-SVD as a generalization of the *k-means* clustering algorithm to solve the sparse representation problem given a set of signals $\mathbf{x}$ to be represented [38]. Moreover, they included a non-negative version of the BP algorithm, named NN-BP, for producing non-negative dictionaries. The method solves the problem

$$\min_a \left\|\mathbf{x} - \mathbf{\Phi}^L \mathbf{a}\right\|_2^2 \quad \text{s.t.} \quad \mathbf{a} \ge 0, \tag{7}$$

where a sub-matrix $\mathbf{\Phi}^L$ that includes only a selection of the L largest coefficients is used. In the dictionary updating, this matrix is forced to be positive by calculating

$$\min_{\vec{\phi}_k, a^k} \left\|\mathbf{E}^k - \vec{\phi}_k a^k\right\|_2^2 \quad \text{s.t.} \quad \vec{\phi}_k, \vec{x}^k \ge 0, \tag{8}$$

for each one of the k selected coefficients. The error matrix $\mathbf{E}^k$ is the residual between the signal and its approximation with the k-th atom $\vec{\phi}_k$ and its respective activation a^k being updated.

The dictionary itself and the activation coefficients are calculated from the SVD of $\mathbf{E}^k = \mathbf{U}\Sigma\mathbf{V}^T$. This decomposition is then truncated to null the negative entries. Finally, the atoms and activations are obtained as the rank-one approximation with the first left and right singular vector as $\phi_k = \mathbf{u}_1$ and $a^k = \mathbf{v}_1$. The complete algorithm, called NN-K-SVD for short [38], is illustrated in Appendix A.

3. Denoising methods

3.1. Non-negative cortical denoising

The main idea of the proposed method is that sound and noise signals can be projected to an approximate auditory cortical space, where the meaningful features of each one could easily be separated. The signals being analyzed could be decomposed into more than one (possibly overcomplete) dictionary containing a rough approach to all the features of interest. More precisely, the method here proposed is based on the decomposition of the signal into two parallel STRF dictionaries, one of them estimated from clean signals and the other one from noise. The estimation of both dictionaries is carried out after obtaining the respective two-dimensional early auditory spectrograms for each type of signals, as was explained before. Given that this type of representation is non-negative, a natural way to obtain both the dictionary and the cortical activations is to use an algorithm that obtains a representation with non-negative constraints. This is especially true in the case of denoising applications, where forcing non-negativity on both the dictionary and the coefficients may help to find the building blocks of the different type of signals [38]. Among the several NMF models reported in literature (some of then summarized in Section 1), we chose for our purposes the above outlined NN-K-SVD.

Before carrying out the denoising, the dictionaries corresponding to clean signals and noise should be estimated. They are produced applying twice the NN-K-SVD algorithm described in Section 2.3, one for each type of signal. The dictionaries are then rearranged according to the activation for the training samples, in descending order. From these two sets, a combined dictionary containing atoms of signal and noise is used in our approach. This new dictionary is composed by the "most representative" atoms of each previous dictionary, by selecting those with greater activation.

Fig. 2 shows a diagram of the method here proposed, which consists of two stages. In the *forward* stage (Fig. 1a), the auditory spectrogram is firstly obtained. Then, using the combined dictionary, the auditory cortical activations that best represent the noisy

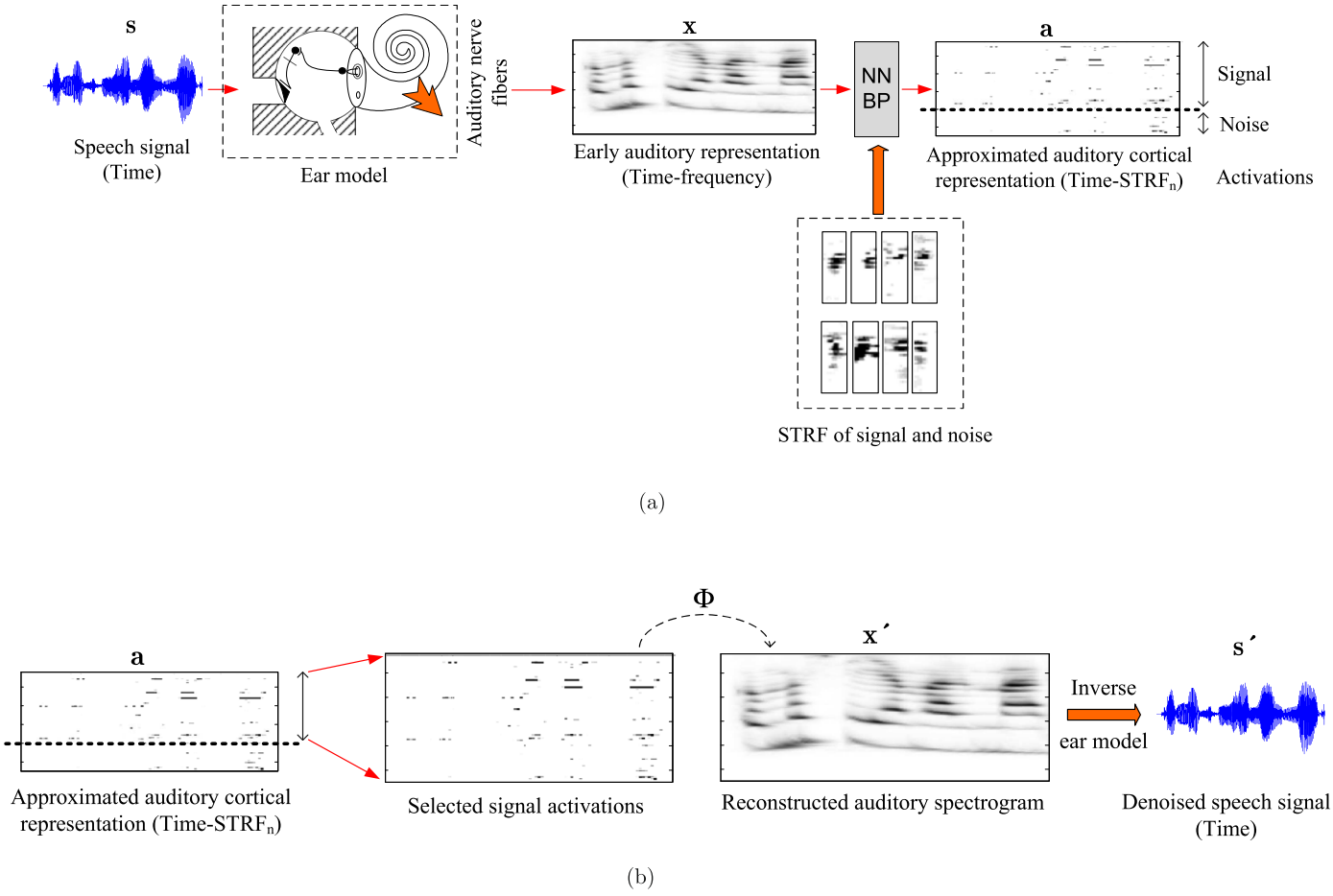

Fig. 2. Diagram of the NNCD method for denoising in the cortical domain. (a) Forward stage: cortical representation. (b) Backward stage: denoised reconstruction.

signal (including both clean and noisy activations) are calculated by means of the non-negative version of the BP algorithm. In the *backward* stage (Fig. 1b), the auditory spectrogram is reconstructed by taking the inverse transform from only the coefficients corresponding to the signal dictionary (synthesis). In this way, the denoising of the signal is carried out in the approximate non-negative auditory cortical domain. Finally, the denoised signal in the temporal domain is obtained by the approximate inverse ear model. The proposed method is named NNCD, which stands for *non-negative cortical denoising*.

The reconstruction of the auditory spectrogram from the cortical response is direct because it only consists of a linear transformation. However, a perfect reconstruction of the original signal from the auditory spectrogram is impossible because of the non-linear operations in the earlier described in Section 2.1. Shamma proposed a method to approximately invert the model and showed through objective and subjective quality tests that the resulting quality of this approximate reconstruction is not degraded [9].

The idea of using a cortical model for sound denoising was also proposed by Shamma in a recent work [10]. The main differences with our approach are that his cortical representation uses the concept of spectro-temporal modulation instead of STRF and non-negative sparse coding, and also the way he incorporates information about signal and noise.

3.2. Speech denoising configurations

We propose applying the NNCD in three different scenarios for denoising speech signals degraded by uncorrelated additive noise:

(a) "NNCD speech": corresponds to the NNCD reconstruction from selected atoms of the speech dictionary, discarding the noise selected atoms.

(b) "Wiener/NNCD noise": applies a Wiener filter to the noisy signal $y(t)$, where the noise estimation $n'(t)$ is given by the NNCD reconstruction from only selected atoms of the noise dictionary.

(c) "NNCD + Wiener": applies a Wiener filter to both previously NNCD estimations of noise $n'(t)$ and speech $s'(t)$.

In cases (b) and (c), the Wiener filter is estimated by means of the Short-Time Fourier Transform (STFT), as $\frac{|S(\omega,\tau)|^2}{|S(\omega,\tau)|^2+|N(\omega,\tau)|^2}$. Here, $S(\omega,\tau)$ and $N(\omega,\tau)$ are the STFT representations of $s(t)$ and $n(t)$ respectively. Note that in case (c), the Wiener filter is estimated from the speech signal $s'(t)$ instead of $s(t)$ [40,41]. Fig. 3 shows the block diagrams of these configurations.

For comparison purposes, different filtering algorithms were also implemented and tested:

- iWiener: the iterative Wiener method [42]. After preliminary experimentation, the number of iterations was fixed at 4.
- apWiener: the speech enhancement based on the use of the *A Priori Signal to Noise ratio* in a minimum mean square error estimation, as given in [43].
- Wavelet: sound denoising using the thresholding of wavelet coefficients. The parameters of this process were: 5 levels of a Daubechies 8 function, soft thresholding using the unbiased

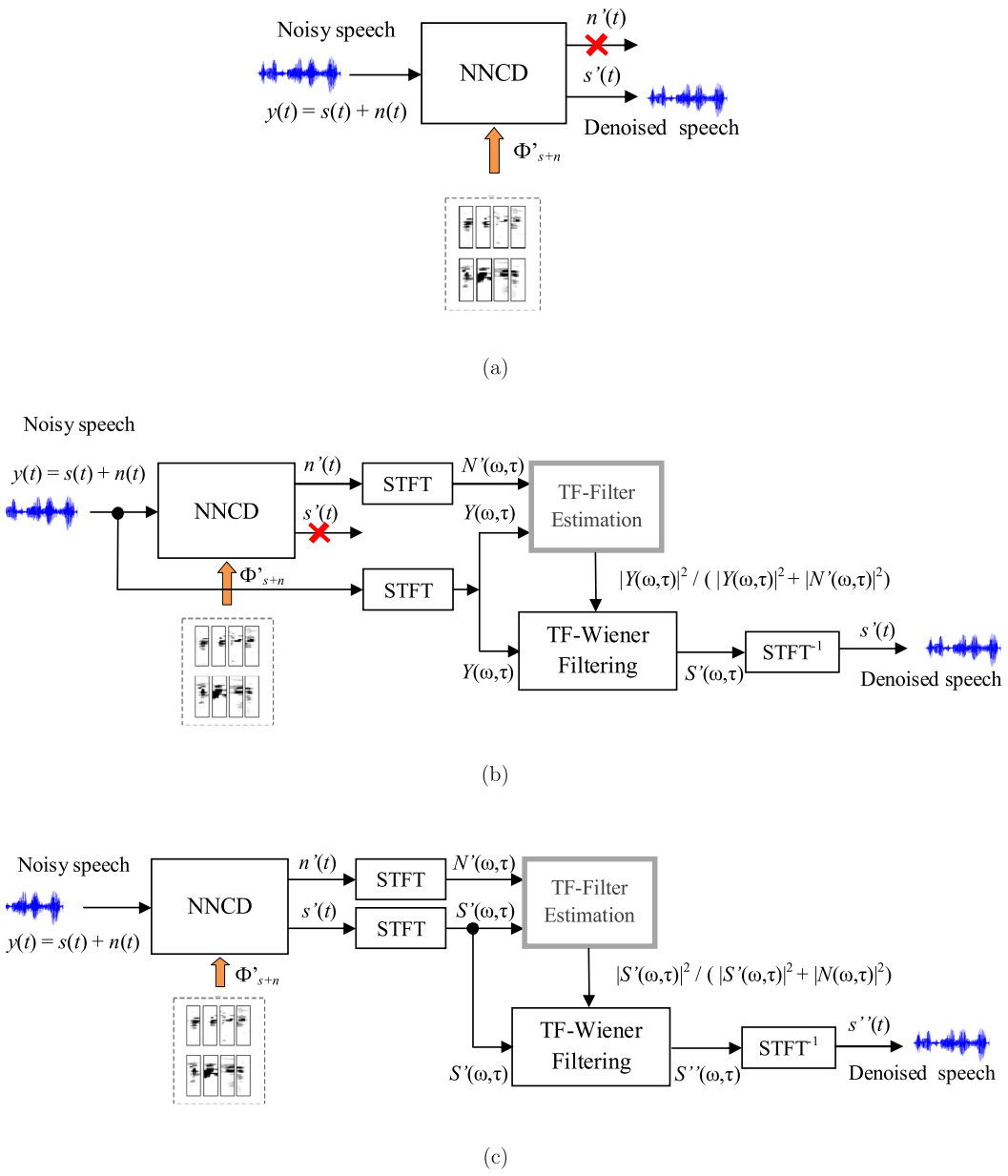

Fig. 3. Schematics of the three configurations proposed to apply the NNCD to speech enhancement: (a): NNCD speech only, (b) Wiener filter with noise estimation given by the NNCD, and (c) Wiener filter calculated with the estimation of signal and noise given by the NNCD.

SURE estimator and rescaling using a single estimation of level noise based on first-level coefficients [44].

- mBand: Multi-band spectral subtraction, a method that takes into account the fact that colored noise affects the speech spectrum differently at various frequencies [45]. The parameters of the algorithm were fixed at 6 frequency bands with a linear spacing between bands.
- BNMF: a recently proposed Bayesian formulation of nonnegative matrix factorization [33]. First, a mean square error estimator for the speech signal is derived, then it learns the NMF noise model online from the noisy signal (unsupervised speech denoising).

Given the nature and characteristics of the artificial/real signals, the Wavelet denoising was used in the experiments with artificial signals, where mBand and BNMF were used in the experiments with speech data.

4. Experimental framework

A series of experiments were carried out to demonstrate the capabilities of the proposed technique. The first of this were carried out on artificial "clean" sound signals constructed by a mixture of chirps and pure tones. Then a second series of experiments were developed to work with real data consisting of speech signals of complete sentences from a single speaker. Noises with different frequency distributions and non-stationary behaviors were additively aggregated to the signals at several signal to noise ratios (SNRs). The proposed technique was then applied to obtain the denoised signals and the performance was evaluated by two objective methods: the *perceptual evaluation of speech quality* (PESQ) score [46] and the classical segmental signal-to-noise ratio (SNRseg) [47].

4.1. Artificial and real signals and noises

A total of 1000 artificial signals were obtained by concatenating 7 different subsignal segments of 64 ms each at a sampling

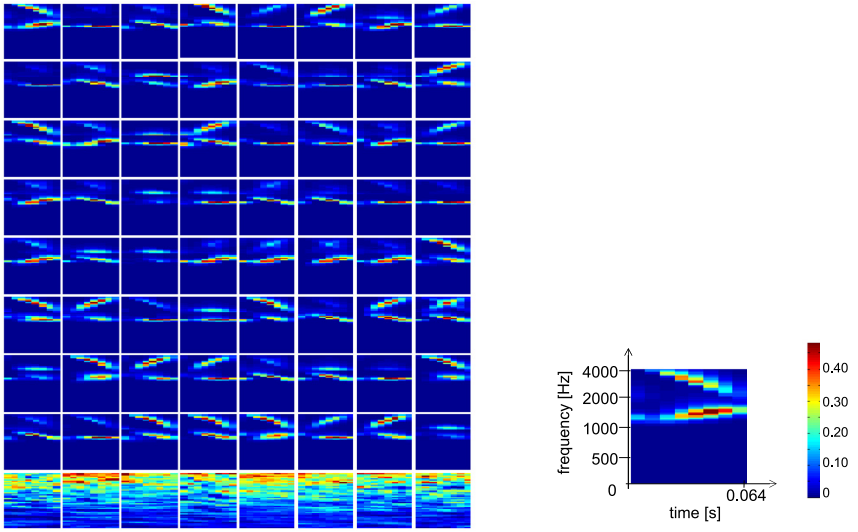

Fig. 4. Example of spectro-temporal receptive fields (STRF) estimated from the early auditory representation of artificial signals and white noise signals, showing the most active atoms of each dictionary (left). A single atom with axis labels and colorbar is also showed (right). The top 8 rows show the 64 most important STRF for clean signals, whereas the last row show the respective STRF for the noise signals. The dimensions of each atom follow the setup outlined in Section 4.2.

frequency of 8 kHz. Each segment consisted of the random combination of up or down chirps and pure tones. In order to restrict all the possible combinations of these features so that a relatively simple dictionary was able to represent them, the spectrogram was divided in two frequency zones, below and above 1200 Hz. Inside each zone only one of the features could occur. Also, the frequency slopes of the chirps are fixed in each zone. Experiments with this type of signals were designed just to illustrate the operation of the method, also for sanity check and to show the feasibility of the method.

The clean speech data was extracted from a widely-used database in the speech recognition field, the TIMIT corpus [48]. The data used in this work corresponds to the set of 10 speech sentences of the speaker FCJF0 in dialectic region number 1. Sentences have a mean length of 5 seconds.

Two kinds of noise with different frequency content were used. On the one hand, the white noise, which exhibits a relatively high frequency content with a non-uniform distribution in the early auditory spectrogram (due to its logarithmic frequency scale), and on the other hand voice babble and street noises with mainly low frequency content in that representation. The white noise was generated by an HF radio channel and the babble noise was recorded in a crowded indoor ambient, both taken from the NOISEX-92 database [49]. The street noise corresponds to an outdoor recording and was taken from the Aurora database [50]. In all the experiments, the noise was first conveniently resampled to the same rate and resolution of the clean signals. The noisy signals were obtained by additively mixing the signals at different SNRs.

4.2. Combined clean-noisy dictionary estimation

First, the auditory spectrograms of clean signals were obtained. Then, the training data for the estimation of the dictionaries was extracted by means of a sliding time-frequency windowing using frames of 64 ms in length with an overlapping of 8 ms.

The dictionaries were generated using complete dictionaries. For the artificial data, 512 atoms of size 64×8 were calculated. Here, the 64 coefficients correspond to a downsampled version of the original 128 coefficients representing the range 0–4 kHz, while the 8 columns correspond each to a window of 8 ms. For speech data, based on preliminary experiments, the number of columns was reduced to 4, given that with 8 windows the dictionary learning process becomes computationally very intensive. Thus, in this case, the dictionaries have 256 atoms of size 64×4.

For the artificial data, 1/10 of the total number of signals was used as training data (100 random selected chirp signals). For the estimation of noise dictionaries, the same ratio of 1/10 was used as the balance of training/test data. For the speech sentences, a 10-fold leave-one-out method was applied, where each partition consisted on 9 sentences for train and 1 sentence for test.

From each dictionary, the most active atoms were collected. Then, they were combined to form new dictionaries with atoms containing both clean and noisy features. The reported results consist of the mean value obtained for the 10 partitions.

4.3. Denoised signals quality estimation

For the speech denoising experiments, two well-known objective speech quality measures were evaluated: the PESQ score and the segmental signal-to-noise ratio (SNRseg).

The PESQ score is an objective quality measure introduced by the International Telecommunication Union (ITU) as a standard for evaluation of speech quality after transmission over communication channels [46]. It uses an auditory representation based on bark scale to compare the original and distorted speech signals. It has been shown to be very well correlated with perceptual tests using *mean opinion score* (MOS) [51] and robust automatic speech recognition results [52]. The measure has an ideal value of 4.5 for clean signals with no distortion, and a minimum of -0.5 for the worst case of distortion.

The segmental signal-to-noise ratio is another quality measure here evaluated. It was obtained as the frame-based average SNR value calculated from the original and the processed signals. Here, short segments of 15–20 ms are used (instead of the whole signals). This time domain measure was computed as in [47], using the MATLAB code provided in [53].

5. Results and discussions

5.1. Non-negative STRF dictionaries

Fig. 4 shows a selection of STRFs from a combined dictionary. Here, the most active (best trained) atoms are presented, 64 atoms for chirp signals and 8 atoms for white noise signals.

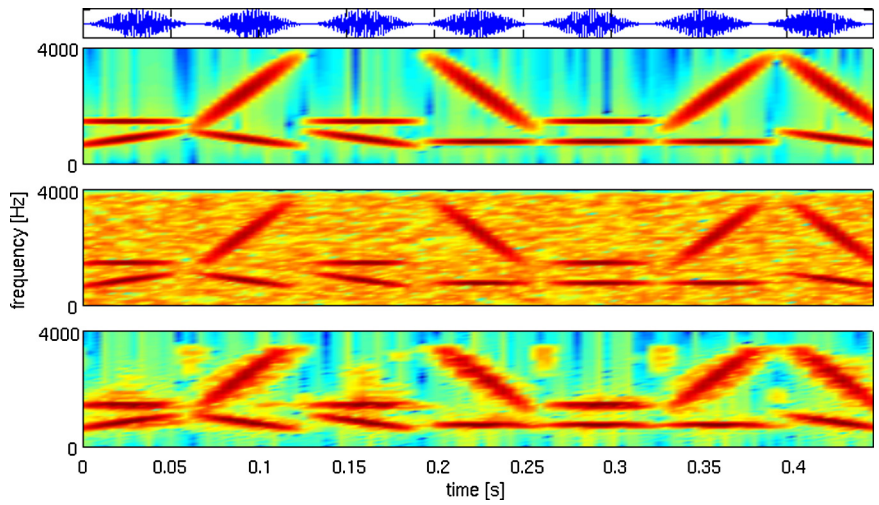

Fig. 5. Example of the denoising of an artificial signal with a combination of 7 windowed segments of random chirps and pure tones. The spectrograms (STFT) of the clean signal (top), a noisy version obtained by the addition of white noise at SNR = 0 dB (middle) and the denoised signal (bottom) are shown. The temporal signal at the top of the figure is given as reference.

Table 1
Raw PESQ scores obtained for artificial signals. The NNCD scheme applied was the scenario (a) given in Section 3.2. In bold face, the best result obtained for each experimental condition.

Noise	SNR (dB)	Signal		
		Noisy	Wavelet	NNCD
White	12	1.93	1.79	**2.16**
	6	1.40	1.43	**2.11**
	0	0.69	0.87	**1.99**
Voice babble	12	1.82	1.72	**2.05**
	6	1.23	1.14	**2.01**
	0	0.56	0.53	**1.91**
Model distortion: 2.11				

It can be clearly seen the features captured by the STRFs in each dictionary are the more prominent ones contained in the training signals. For the first group, some atoms (see, for example, number 2, 3 and 4 in the first row) capture portions of pure tones or chirp signals, while others show the combination of them. For the second group, the atoms show mainly the high energy characteristics of the noise signals. Thus, in the context of sparse coding given in Section 2.2, each segment of the input signal can be represented by a linear combination of selected atoms from these dictionary.

5.2. Artificial signals denoising

Our scheme for denoising was applied using the representation discussed above. The reconstruction of the denoised auditory spectrogram was obtained by selecting only the clean atoms from the 32 greatest activations selected by the NN-BP algorithm. Fig. 5 shows the short-time Fourier transform (STFT) for a clean (top), noisy with white noise at SNR = 0 dB (middle) and denoised signal (bottom), with the temporal signal above the clean spectrogram. In the spectrogram shown at the bottom, the effects of the denoising carried out in the cortical representation by the NNCD can be seen, where the most important features are reconstructed.

Table 1 shows the PESQ scores obtained of denoising the artificial signals. For all cases, there was an increase in the PESQ score when the NNCD was applied to the noisy signals and our method also outperformed the results obtained with the baseline. The improvement was more marked when the noise energy was higher (SNR = 0 dB) and smaller when the signals become cleaner at larger SNR (lower energy of the noise).

The PESQ score for the original (clean) signal after transformation using the auditory model and reconstruction back to the time domain is 2.11. This score measures the distortion from the best quality (PESQ MOS of 4.5) that is introduced by the use of the early auditory model, which is only approximately invertible. Even if the noise is completely removed by the NNCD, there is an intrinsic error introduced by the auditory analysis method. For reference, the PESQ obtained using the NNCD method in the same conditions as in Table 1 but on clean signal (SNR = ∞) was 2.105. The result is almost identical to the one of the auditory model, showing that no additional degradation was introduced. This is because the number of selected coefficients in the NN-K-SVD method is enough to the preserve the quality of the reconstructed signal. In this way, the method not only provides a good enhancement in the noisy case but also preserve the signal when there is no noise. The PESQ values greater than the model distortion (for example, 2.16 for white noise at SNR = 12 dB) are pointing out that small amount of noise are beneficial for the quality of the signal obtained. This effect might be due to the *stochastic resonance*, which concern to non-linear systems (like our proposal) [54].

In order to demonstrate the benefits of using the auditory representation of the signal, an experiment replacing this model with the short-time Fourier transform was carried out. Here, two dictionaries trained with clean chirp signals and white noise were obtained. Then, the NNCD method was applied in the same conditions as in Table 1 for noisy signals at SNR = 0 dB. The PESQ obtained was 1.27, which is better than the wavelet denoising (0.87) but lower than the result obtained using the NNCD method (1.99). This result would be supporting the intrinsic robustness of the sparse representation when using the auditory model.

5.3. Speech denoising

In Fig. 6, a subset of 64 atoms from the dictionary trained with speech data is shown. It can be seen that different particularities of the signals are learned, for example, onset events (see atoms number 1 and 3 in the first row), offset (atom number 5 in the first row), combination of formants (atoms number 2 and 7 in the first row), energy spreading in a wide frequency range possibly given by fricative phonemes (atom number 1 in the last line), etc.

Fig. 7 shows an example of the denoising of real data signals corresponding to speech data. The clean signal corresponds to the sentence /She had your dark suit in greasy wash water all year/ (shown in the top spectrogram). The signal is then contami-

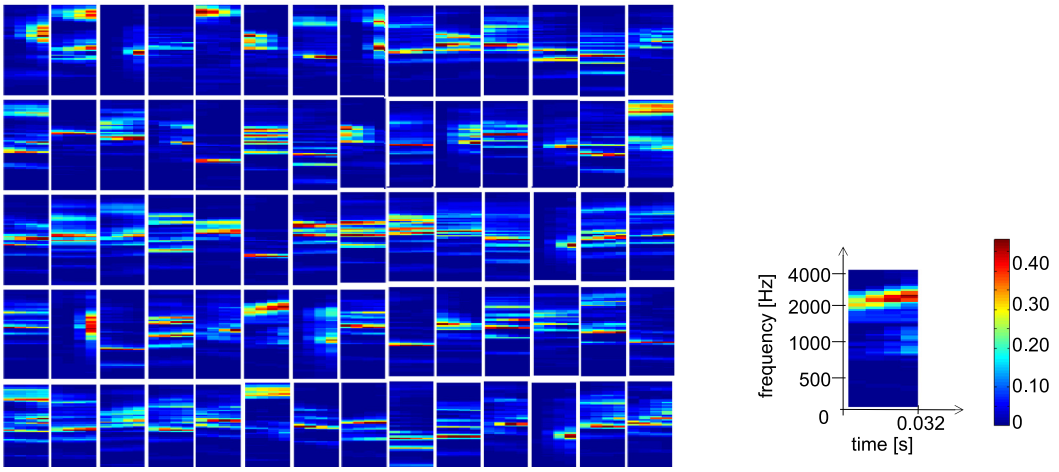

Fig. 6. Examples of spectro-temporal receptive fields (STRF) calculated from the early auditory representation of speech signals (left). A single atom with axis labels and colorbar is also showed (right). The dimensions of each atom follow the setup outlined in Section 4.2.

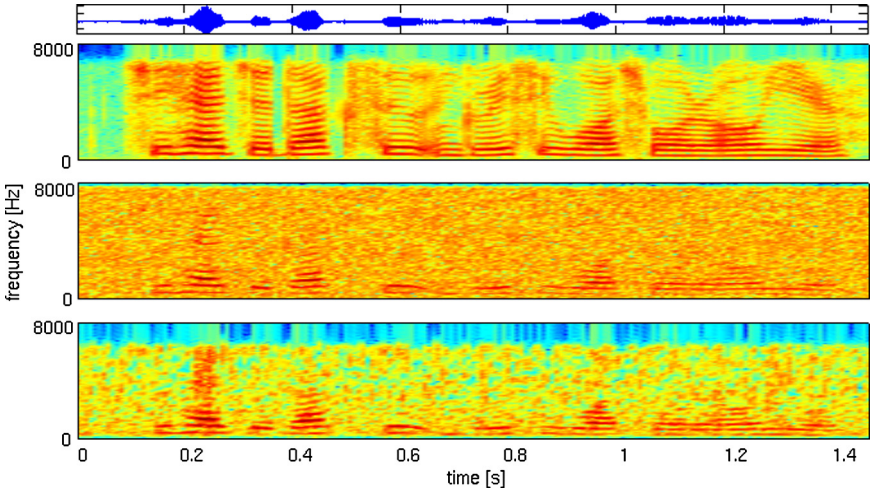

Fig. 7. Example of the auditory cortical denoising result of a speech signal contaminated with white noise at SNR = 0 dB. The spectrograms (STFT) of the clean signal (top), the noisy signal (middle) and the denoised reconstructed signal (bottom) are shown. The acoustic signal at the top of the figure is given as reference.

nated with white noise at SNR = 0 dB. The effects of the noise can be seen in the middle spectrogram, where almost every important speech feature has been masked by the noise. The denoising scheme, however, is able to recover the most prominent formants and to reduce the energy noise as shown in the bottom spectrogram.

For the measures of PESQ and SNRseg, a 10-fold cross validation procedure was applied by training a dictionary with 9 signals and testing with the remaining one. In each case, white and street noise were added with SNR of 12, 6 and 0 dB. The results are summarized in Tables 2 and 3. They show the mean and standard deviation of PESQ and SNRseg scores obtained for the cross validation scheme, being tested on the three different scenarios in the application of NNCD and compared with different baseline methods (see Section 3.2). For each experimental condition, the method that obtained the best denoising quality is emphasized in boldface.

It can be seen that state-of-the-art method performs better only at very high SNR (12 dB), while the NNCD method achieves good results in realistic conditions when the energy noise increases at lower SNR. Here, our method obtains the larger differences in the PESQ and SNRseg scores between the noisy and denoised signals. For example, in the case of white noise at SNR = 0 dB the method improves the PESQ from 1.63 up to 2.12 and SNRseg from −2.77 to 4.56. With respect to the other denoising methods, the NNCD approach performs better for both measures, PESQ and SNRseg, under real and very high non-stationary noise, like the street noise used in these experiments. As an example, it can be seen an improvement in PESQ at SNR = 0 dB from 1.79 up to 2.24 and in SNRSeg from −3.54 up to 3.94. This type of noise presents a more complex structure, which could be captured by our approach.

6. Conclusions

A new denoising method of audio signals was presented, inspired by the biological processing carried out at the primary auditory cortical level. The method obtains a sparse coding of the spectrogram at cochlea level using a non-negative approach. The atoms of the dictionary are calculated from clean signals and noise. Then, the denoising signal is obtained by inverting the model using only the atoms corresponding to the signal, discarding the noise activations.

The performance of the method using synthetic and real signals with additive noise was obtained through two objective quality measures. Results showed that our proposed method and its variants can improve the quality of sound signals, specially under severe conditions.

Future research will be devoted to further improve the performance and also investigate the application of this technique in the preprocessing stage of robust classification systems.

Table 2
Mean raw PESQ scores obtained for speech sentences from the TIMIT corpus. The 'W' and 'S' on the left column stand for White and Street noise. The three scenarios for the NNCD based speech enhancement given in Section 3.2 are denoted as (a), (b) and (c). In bold face, the best quality for each case. For reference, the score for the clean signal after transformation to the cortical domain and reconstruction back to the time domain is 2.15.

	SNR (dB)	Signal					NNCD		
		Noisy	iWiener	apWiener	mBand	BNMF	(a)	(b)	(c)
W	12	2.25 (0.14)	2.59 (0.15)	2.53 (0.15)	**2.66 (0.21)**	2.41 (0.10)	2.46 (0.08)	2.31 (0.14)	2.52 (0.08)
	6	1.92 (0.13)	2.19 (0.08)	2.17 (0.09)	2.18 (0.12)	2.18 (0.10)	2.26 (0.08)	1.97 (0.12)	**2.36 (0.05)**
	0	1.63 (0.18)	1.86 (0.15)	1.84 (0.16)	1.84 (0.18)	1.80 (0.09)	1.99 (0.13)	1.67 (0.17)	**2.12 (0.10)**
S	12	2.57 (0.13)	2.61 (0.13)	2.73 (0.13)	**2.86 (0.11)**	2.30 (0.14)	2.67 (0.11)	2.65 (0.12)	2.71 (0.11)
	6	2.21 (0.10)	2.18 (0.12)	2.39 (0.09)	2.49 (0.11)	2.06 (0.16)	2.45 (0.07)	2.30 (0.09)	**2.51 (0.05)**
	0	1.79 (0.13)	1.76 (0.15)	2.00 (0.10)	2.11 (0.09)	1.82 (0.13)	2.14 (0.08)	1.89 (0.11)	**2.24 (0.06)**

Table 3
Mean SNRseg obtained for speech sentences from the TIMIT corpus. The 'W' and 'S' on the left column stand for White and Street noise. The three scenarios for the NNCD speech enhancement given in Section 3.2 are denoted as (a), (b) and (c). In bold face, the best result for each condition. For reference, the score for the clean signal after transformation to the cortical domain and reconstruction back to the time domain is 5.41.

	SNR (dB)	Signal					NNCD		
		Noisy	iWiener	apWiener	mBand	BNMF	(a)	(b)	(c)
W	12	6.98 (3.42)	8.43 (1.82)	**10.04 (2.95)**	6.91 (1.99)	1.59 (0.30)	5.60 (1.14)	7.63 (3.47)	5.79 (0.90)
	6	1.84 (2.54)	4.50 (1.54)	5.14 (2.12)	5.14 (2.56)	1.62 (0.31)	5.21 (0.62)	2.68 (2.52)	**5.24 (0.70)**
	0	−2.77 (2.00)	2.10 (0.85)	0.04 (1.92)	2.25 (0.23)	1.57 (0.16)	3.84 (0.84)	−2.01 (2.04)	**4.56 (0.79)**
S	12	7.10 (2.31)	6.33 (1.33)	8.67 (2.23)	7.09 (1.31)	1.54 (0.22)	5.75 (0.79)	**8.24 (2.40)**	5.68 (0.48)
	6	1.93 (2.24)	3.79 (1.05)	4.13 (2.40)	4.52 (1.59)	1.69 (0.36)	**5.26 (0.50)**	3.51 (2.15)	4.95 (0.36)
	0	−3.54 (2.27)	1.71 (0.61)	−1.19 (2.55)	2.37 (1.07)	1.57 (0.30)	**3.94 (0.54)**	−1.94 (2.23)	3.89 (0.33)

Initialization: Set the NN random normalized dictionary $\mathbf{\Phi}^{(0)} \in \mathbb{R}^{m\times n\times M}$. Set $J = 1$ and repeat until convergence.
Sparse coding stage: use the NN version of the Basis Pursuit decomposition algorithm to calculate a_i for $i = 1, \ldots, M$.

$$\min_a \|\mathbf{x} - \mathbf{\Phi a}\|_2^2 \quad \text{s.t. } \|\mathbf{a}\|_0 \le L \wedge \mathbf{a} \ge 0.$$

Dictionary update stage: for $k = 1, \ldots, L$

- Define the samples that use $\vec{\phi}_k$: $\omega_k = \{i | 1 \le i \le M, \mathbf{a}_i(k) \ne 0\}$.
- Compute $\mathbf{E}_k = \mathbf{x} - (\mathbf{\Phi a} - \vec{\phi}_k \mathbf{a}(k))$.
- Choose only the columns corresponding to ω_k, and obtain $\mathbf{E}_k^{\omega_k}$.
- Set $A = \mathbf{E}_k^{\omega_k}$,

$$\vec{\phi}_k = \begin{cases} 0, & \mathbf{u}_1(i) < 0 \\ \mathbf{u}_1(i), & \text{otherwise} \end{cases}$$

$$\mathbf{a}(k) = \begin{cases} 0, & \mathbf{v}_1(i) < 0 \\ \mathbf{v}_1(i), & \text{otherwise} \end{cases}$$

where $\mathbf{u}_1$ and $\mathbf{v}_1$ are the first singular vector of A.
Repeat J times:

$$\vec{\phi} = \frac{A\,\mathbf{a}}{\mathbf{a}'\mathbf{a}}. \text{ Project: } \vec{\phi}(i) = \begin{cases} 0, & \vec{\phi}(i) < 0 \\ \vec{\phi}(i), & \text{otherwise} \end{cases}$$

$$\mathbf{a} = \frac{\vec{\phi}'\,A}{\vec{\phi}'\vec{\phi}}. \text{ Project: } \mathbf{a}(i) = \begin{cases} 0, & \mathbf{a}(i) < 0 \\ \mathbf{a}(i), & \text{otherwise} \end{cases}$$

Normalize $\vec{\phi}_k$.

Set $J = J + 1$.

Fig. 8. The NN-K-SVD algorithm.

Acknowledgments

The authors wish to thank: the *Agencia Nacional de Promoción Científica y Tecnológica* (with PICT 2010-1730), the *Universidad Nacional de Litoral* (with CAI+D 2011 #58-511, #58-519, #58-525), the *Universidad Nacional de Entre Ríos* (with PID NOVEL 6121), the *Consejo Nacional de Investigaciones Científicas y Técnicas* (CONICET) from Argentina (with PIP 2011 00284), and the SEP and CONACyT from México (with Program SEP–CONACyT CB-2012-01, No. 182432), for their support.

Appendix A

The pseudocode for the NN-K-SVD method is showed in Fig. 8 [38].

References

[1] Y. Hu, P.C. Loizou, Subjective comparison and evaluation of speech enhancement algorithms, Speech Commun. 49 (7–8) (2007) 588–601.
[2] J.-C. Junqua, J.-P. Haton, Robustness in Automatic Speech Recognition: Fundamentals and Applications, Kluwer Academic Publishers, 1995.
[3] M. Lewicki, T. Sejnowski, Learning overcomplete representations, Neural Comput. 12 (2) (2000) 337–365.
[4] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.
[5] J. Deller, J. Proakis, J. Hansen, Discrete Time Processing of Speech Signals, Macmillan Publishing, New York, 1993.
[6] B. Delgutte, Physiological models for basic auditory percepts, in: H.H. Hawkins, T.A. McMullen, A.N. Popper, R.R. Fay (Eds.), Auditory Computation, Springer, New York, 1996.
[7] H. Rufiner, L. Rocha, J. Goddard, Sparse and independent representations of speech signals based on parametric models, in: Proceedings of the ICSLP'02, 2002, pp. 989–992.
[8] S. Greenberg, The ears have it: the auditory basis of speech perception, in: Proceedings of the International Congress of Phonetic Sciences, vol. 3, 1995, pp. 34–41.
[9] T. Chiu, P. Ru, S. Shamma, Multiresolution spectrotemporal analysis of complex sounds, J. Acoust. Soc. Am. 118 (2) (2005) 897–906.
[10] N. Mesgarani, S. Shamma, Denoising in the domain of spectrotemporal modulations, EURASIP J. Audio Speech Music Process. 2007 (2007) 8.
[11] D. Klein, P. Konig, K. Kording, Sparse spectrotemporal coding of sounds, EURASIP J. Appl. Signal Process. 2003 (7) (2003) 659–667.
[12] B. Olshausen, D. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1?, Vis. Res. 37 (23) (1997) 3311–3325.
[13] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Netw. 13 (4–5) (2000) 411–430.
[14] F. Theunissen, K. Sen, A. Doupe, Spectro-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds, J. Neurosci. 20 (2000) 2315–2331.
[15] D.-S. Kim, S.-Y. Lee, R. Kil, Auditory processing of speech signals for robust speech recognition in real-world noisy environments, IEEE Trans. Speech Audio Process. 7 (1) (1999) 55–69.
[16] R. Stern, N. Morgan, Hearing is believing: biologically-inspired feature extraction for robust automatic speech recognition, IEEE Signal Process. Mag. 29 (6) (2012) 34–43.
[17] M. Kleinschmidt, J. Tchorz, B. Kollmeier, Combining speech enhancement and auditory feature extraction for robust speech recognition, Speech Commun. 34 (1) (2001) 75–91.

[18] T. Dau, D. Püschel, A. Kohlrausch, A quantitative model of the "effective" signal processing in the auditory system, I: model structure, J. Acoust. Soc. Am. 99 (6) (1996) 3615–3622.
[19] Y. Ephraim, D. Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. 32 (6) (1984) 1109–1121.
[20] R. Flynn, E. Jones, Combined speech enhancement and auditory modelling for robust distributed speech recognition, Speech Commun. 50 (10) (2008) 797–809.
[21] Q. Li, F. Soong, O. Siohan, A high-performance auditory feature for robust speech recognition, Interspeech (2000) 51–54.
[22] S. Rangachari, P. Loizou, A noise-estimation algorithm for highly non-stationary environments, Speech Commun. 48 (2) (2006) 220–231.
[23] C. Kim, R. Stern, Power-normalized cepstral coefficients (PNCC) for robust speech recognition, in: Proc. of Acoustics, Speech and Signal Processing, ICASSP, 2012, pp. 4101–4104.
[24] C. Martínez, J. Goddard, D. Milone, H. Rufiner, Bioinspired sparse spectro-temporal representation of speech for robust classification, Comput. Speech Lang. 26 (2012) 336–348.
[25] O.-W. Kwon, T.-W. Lee, Phoneme recognition using ICA-based feature extraction and transformation, Signal Process. 84 (6) (2004) 1005–1019.
[26] P.O. Hoyer, Non-negative matrix factorization with sparseness constraints, J. Mach. Learn. Res. 5 (2004) 1457–1469.
[27] T. Virtanen, Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria, IEEE Trans. Audio Speech Lang. Process. 15 (3) (2007) 1066–1074.
[28] F. Weninger, J. Feliu, B. Schuller, Supervised and semi-supervised suppression of background music in monaural speech recordings, in: Proc. of Acoustics, Speech and Signal Processing, ICASSP, 2012, pp. 61–64.
[29] F. Weninger, J. Feliu, B. Schuller, Source separation using regularized NMF with MMSE estimates under GMM priors with online learning for the uncertainties, Digit. Signal Process. 29 (0) (2014) 20–34.
[30] P. Smaragdis, Convolutive speech bases and their application to supervised speech separation, IEEE Trans. Audio Speech Lang. Process. 15 (1) (2007) 1–12.
[31] K. Wilson, B. Raj, P. Smaragdis, Regularized non-negative matrix factorization with temporal dependencies for speech denoising, Interspeech (2008) 411–414.
[32] R. Vipperla, S. Bozonnet, D. Wang, N. Evans, Robust speech recognition in multi-source noise environments using convolutive non-negative matrix factorization, in: Workshop on Machine Listening in Multisource Environments, CHiME, 2011, pp. 74–79.
[33] N. Mohammadiha, P. Smaragdis, A. Leijon, Supervised and unsupervised speech enhancement using nonnegative matrix factorization, IEEE Trans. Audio Speech Lang. Process. 21 (40) (2013) 2140–2141.
[34] J. Gemmeke, T. Virtanen, A. Hurmalainen, Exemplar-based sparse representations for noise robust automatic speech recognition, IEEE Trans. Audio Speech Lang. Process. 19 (7) (2011) 2067–2080.
[35] B. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 24 (2) (1995) 227–234.
[36] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[37] S.G. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (1993) 3397–3415.
[38] M. Aharon, M. Elad, A.M. Bruckstein, K-SVD and its non-negative variant for dictionary design, in: Proceedings of the SPIE Conference Wavelets, vol. 5914, 2005.
[39] R. Zhang, C. Wang, B. Xiao, A strategy of classification via sparse dictionary learned by non-negative K-SVD, in: 12th IEEE International Conference on Computer Vision Workshops, ICCV Workshops, 2009, pp. 117–122.
[40] Y. Huang, J. Benesty (Eds.), Audio Signal Processing for Next-Generation Multimedia Communication Systems, Kluwer Academic Press, 2004.
[41] D. Milone, L. Di Persia, M.E. Torres, Denoising and recognition using hidden Markov models with observation distributions modeled by hidden Markov trees, Pattern Recognit. 43 (4) (2009) 1577–1589.
[42] J. Lim, A.V. Oppenheim, All-pole modeling of degraded speech, IEEE Trans. Acoust. Speech Signal Process. 26 (3) (1978) 197–210.
[43] P. Scalart, J. Vieira Filho, Speech enhancement based on a priori signal to noise estimation, in: Proc. of Acoustics, Speech and Signal Processing, ICASSP, vol. 2, 1996, pp. 629–632.
[44] D. Donoho, De-noising by soft-thresholding, IEEE Trans. Inf. Theory 41 (3) (1995) 613–627.
[45] S. Kamath, P. Loizou, A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, in: Proc. of Acoustics, Speech and Signal Processing, ICASSP, vol. 4, 2002, pp. 4164–4164.
[46] Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, ITU-T recommendation P.862, 2001.
[47] J. Hansen, B. Pellom, An effective quality evaluation protocol for speech enhancement algorithms, in: Proc. Int. Conf. Spoken Lang. Process, vol. 7, 1998, pp. 2819–2822.
[48] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, DARPA TIMIT acoustic-phonetic continuous speech corpus documentation, Technical report, National Institute of Standards and Technology, 1993.
[49] A. Varga, H. Steeneken, Assessment for automatic speech recognition II: NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems, Speech Commun. 12 (3) (1993) 247–251.
[50] H. Hirsch, D. Pearce, The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions, in: Proceedings of the ISCA ITRW ASR2000, 2000.
[51] A.W. Rix, J.G. Beerends, M.P. Hollier, A.P. Hekstra, Perceptual evaluation of speech quality (PESQ)—a new method for speech quality assessment of telephone networks and codecs, in: Proc. of Acoustics, Speech and Signal Processing, ICASSP, vol. 2, 2001, pp. 749–752.
[52] L. Di Persia, D. Milone, H. Rufiner, M. Yanagida, Perceptual evaluation of blind source separation for robust speech recognition, Signal Process. 88 (10) (2008) 2578–2583.
[53] P. Loizou, Speech Enhancement: Theory and Practice, CRC Press, 2013.
[54] M. McDonnell, D. Abbott, What is stochastic resonance?, Definitions, misconceptions, debates, and its relevance to biology, PLoS Comput. Biol. 5 (5) (2009) e1000348.

Cesar E. Martínez received the Bioengineering degree from National University of Entre Rios (UNER), Argentina, in 1999, and the Ph.D. in Engineering from National University of Litoral (UNL), Argentina, in 2011. During 2000–2001, he was a graduate student at the Universitat Politècnica de València (Spain). He was with the Bioengineering Department (UNER) since 1996 and with the Informatics Department (UNL) since 2003. Currently, he is a Professor in both institutions in the Signal and Image Processing area. His research interests include DSP, machine learning and computational intelligence with applications to robust speech recognition, biometrics, image processing and computer vision.

John C.H. Goddard received a B.Sc. (1st Class Hons) from the University of London in 1972 and a Ph.D. from the University of Cambridge in 1979, both in Mathematics. He is a Professor in the Electrical Engineering Department at the Universidad Autónoma Metropolitana, in Mexico City, where he has lectured since 1993. His current research interests are in the areas of Pattern Recognition, Artificial Intelligence and Artificial Life.

Leandro E. Di Persia was born in Parana, Argentina, in 1977. He received the Bioengineering degree from National University of Entre Rios, Argentina, in 2003, and the Ph.D. in Engineering from National University of Litoral (UNL), Argentina, in 2009. During 2004 and 2005 he was a fellow researcher at Doshisha University, Japan. From 2002 to 2009, he was an auxiliar professor, and since 2009 he is an Assistant Professor in UNL. Since 2011 he is a Researcher at the National Scientific and Technical Research Council. His research interests include speech, audio and biomedical signal processing and artificial intelligence.

Diego H. Milone received the Bioengineering degree (Hons.) from National University of Entre Rios, Argentina, in 1998, and the Ph.D. degree in Microelectronics and Computer Architectures from Granada University, Spain, in 2003. He is a Full Professor in the Department of Informatics at National University of Litoral (UNL). He was Director of the Department of Informatics and Assistant Dean for Science and Technology. Since 2006 he is a Research Scientist at National Scientific and Technical Research Council (CONICET). His research interests include statistical learning, pattern recognition, signal processing, neural and evolutionary computing, with applications to speech recognition, biomedical signals and bioinformatics.

Hugo Leonardo Rufiner was born in Buenos Aires, Argentina, in 1967. He received the Bioengineer degree (Hons.) from National University of Entre Ríos (UNER), in 1992, the M.Eng. degree (Hons.) from the Metropolitan Autonomous University, Mexico, in 1996 and the Dr.Eng. degree from the University of Buenos Aires, in 2005. He is a Full Professor of National University of Litoral and UNER and Research Scientist at the National Scientific and Technical Research Council. He was awarded by the National Academy of Exact, Physical and Natural Sciences of Argentina. His research interests include bioinspired signal processing, artificial intelligence and bioengineering.