A bioinspired spectro-temporal domain for sound denoising

C. Martínez[1,2,⋆], J. Goddard[4], L. Di Persia[1,3], D. Milone[1,3] and H. Rufiner[1,2,3]

[1] sinc(i): Research Institute for Signals, Systems and Computational Intelligence
Dpto. Informática, Facultad de Ingeniería - Universidad Nacional del Litoral
CC217, Ciudad Universitaria, Paraje El Pozo, S3000, Santa Fe, Argentina
[2] Laboratorio de Cibernética, Fac. de Ingeniería-Universidad Nacional de Entre Ríos
[3] CONICET, Argentina
[4] Dpto. de Ingeniería Eléctrica, UAM-Iztapalapa, México

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 calculated from 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. Results on synthetic and real data show that the proposed method improves the quality of the signals, mainly under severe degradation. This communication corresponds to a journal paper published in 2015 in DSP (Elsevier) [1].

1 Introduction

For many years, researchers in the field of signal processing have greatly benefited from the use of methods inspired by human sensory mechanisms. In a previous work [2], the *approximate auditory cortical representation* (AACR) was presented. This technique provides an approximated representation of the speech signal at the auditory cortical level by means of a set of activations computed using *matching pursuit* (MP) on a discrete dictionary of bidimensional atoms. In this work, this approach is adapted to a non-negative matrix factorization (NMF) framework. In the biologically-inspired context, the NMF use data described by using just additive components, e.g. a

⋆ Corresponding author: cmartinez@fich.unl.edu.ar

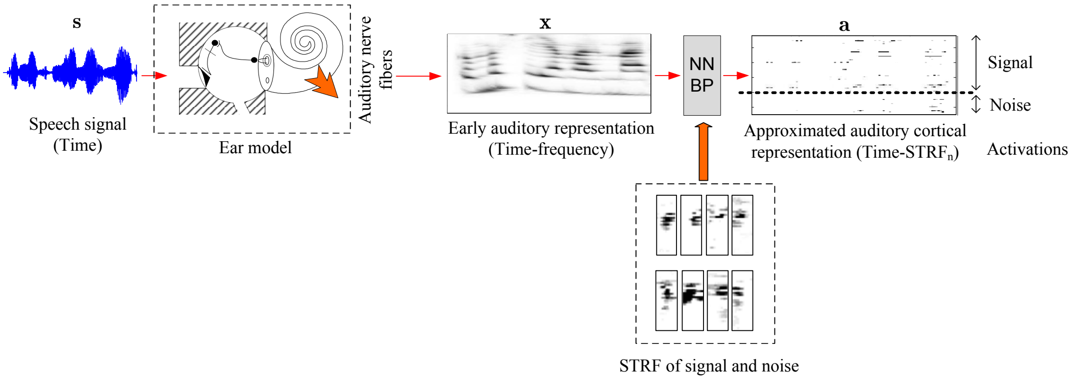

Fig. 1. Forward stage: cortical representation.

weighted sum of only positive atoms. 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 dictionaries 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.

2 Non negative cortical denoising

The main idea is that sound and noise signals can be projected to an approximate auditory cortical space using NN-K-SVD [3], where the meaningful features of each one could easily be separated. In the *forward* stage (Fig. 1), the auditory spectrogram is obtained [4]. Then, using a combined dictionary, the activations that best represent the noisy signal are calculated. In the *backward* stage (Fig. 2), the auditory spectrogram is reconstructed by taking the inverse transform from only the coefficients corresponding to the signal dictionary –denoising–. The proposed method is named NNCD, which stands for *non-negative cortical denoising*.

3 Experimental and results

Experiments were carried out on artificial data (chirps and pure tones) and real data (speech signals of complete sentences). Noises with different frequency distributions and non stationary behaviours

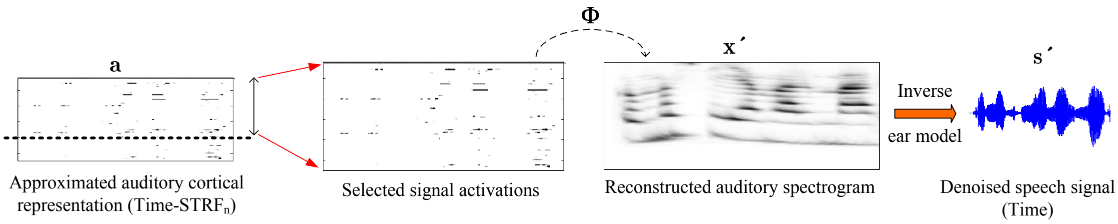

Fig. 2. Backward stage: denoised reconstruction.

Table 1. 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. In bold face, the best quality for each case.

	SNR (dB)	Signal			NNCD
		Noisy	mBand	BNMF	
W	12	2.25	**2.66**	2.41	2.52
W	6	1.92	2.18	2.18	**2.36**
W	0	1.63	1.84	1.80	**2.12**
S	12	2.57	**2.86**	2.30	2.71
S	6	2.21	2.49	2.06	**2.51**
S	0	1.79	2.11	1.82	**2.24**

were additively aggregated. The proposed technique was then applied to obtain the denoised signals and the performance was evaluated by two objective methods: the PESQ score and the classical segmental signal-to-noise ratio (SNRseg) against recent methods such as mBand (multi-band spectral substraction) and BNMF (Bayesian non-negative matrix factorization). Results showed that our method can improve the quality of sound signals, specially under severe conditions.

References

1. C. Martínez, J. Goddard, L. E. Di Persia, D. Milone and H. L. Rufiner. Denoising sound signals in a bioinspired non-negative spectro-temporal domain. *Digital Signal Processing*, 38:22–31, 2015.
2. C. Martínez, J. Goddard, D. Milone and H. Rufiner. Bioinspired sparse spectro-temporal representation of speech for robust classification. *Computer Speech and Language*, 26:336–348, 2012.
3. M. Aharon, M. Elad and A.M. Bruckstein. K-SVD and its non-negative variant for dictionary design. In *Proceedings of the SPIE conference wavelets*, volume 5914, 2005.
4. N. Mesgarani and S. Shamma. Denoising in the domain of spectrotemporal modulations. *EURASIP Journal on Audio, Speech and Music Processing*, 2007:8 pages, 2007.