A low-latency single channel blind source separation algorithm for cochlear implants

W. Nogueira, J. Hidalgo, R. Marxer, J. Janer
Music Technology Group, Pompeu Fabra University

Motivation

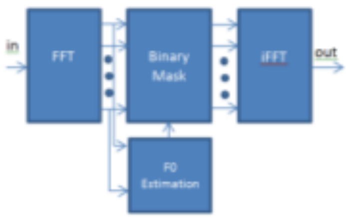
- Speech understanding in noise conditions is more difficult for cochlear implant users than for normal hearing listeners
- Noise Reduction techniques (multichannel or single channel) provide with improvements in speech understanding for CI users
  - Multichannel: Improvement if direction of arrival of the target and noise differ (not always available)
  - Single Channel: Improvement if noise is more stationary than speech

→ Look for single channel noise reduction techniques that can improve speech understanding under non-stationary conditions

Methods

Low-Latency Blind Source Separation (LL-IS) [2]

- Assumption: Vocal component localized around the partials
- Optimal mask: Zeros around the partials and ones elsewhere
- Many parts of vocal components (consonants, fricatives or breath) are not localized in the harmonic region.



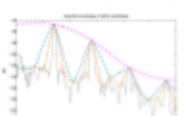
- F0 of the source must be estimated (3 steps):

(1) Pitch likelihood estimation

Linear decomposition (similar to NMF)
Assumption: Spectrum is a linear combination of elementary spectra (basis components)
Tikhonov regularization used to estimate components (simpler implementation than NMF but some gains can be negative)

(2) Timbre classification

- To estimate pitch → need to select right values from pitch likelihood estimation
- Pitch candidates → Classified using SVM
- Envelopes calculated using interpolation on the magnitude of the spectrum and at the harmonic frequency bins (variant of MFCCs)



Spectrum magnitude (solid line) and the harmonic spectral envelopes (colored dashed lines) of three pitch candidates

- Workflow supervised training method:

Based on the pitch information envelopes are created and timbre features extracted. Finally classification is made comparing with a test dataset and the model is generated

(3) Pitch tracking

- Dynamic programming algorithm (2 steps)
- (1) Viterbi determines optimal pitch track
- (2) Determines voiced/unvoiced frames

- Runs in real-time on a standard PC (latency below 250ms)

Instantaneous mixing model (IMM) [1] (M mixtures, N sources)

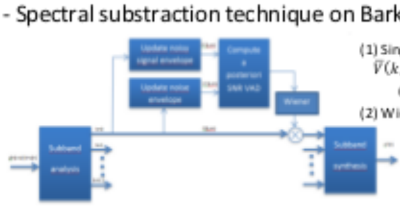
$$x_m(t) = \sum_{n=1}^{N} a_{mn} s_n(t), \qquad m = 1, \cdots, M$$

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{pmatrix} = \begin{pmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1N}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2N}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}(t) & a_{M2}(t) & \cdots & a_{MN}(t) \end{pmatrix} \cdot \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{pmatrix}$$

→ Assumption: Source signals modified by amplitude scalings
→ Uses NMF with kullback-Leibler dist. as cost function

Single channel noise reduction (Spectral Substraction,NR)

- Spectral substraction technique on Bark Domain



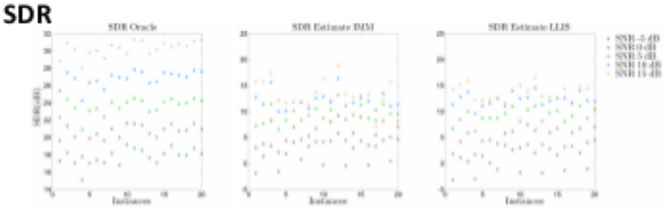
(1) Single Pole recursive method
$$\bar{V}(k,m) = \alpha \bar{V}(k,m-1) + (1-\alpha)|V(k,m)|,$$
$$0 < \alpha < 1$$
(2) Wiener Filter

Evaluation Setup

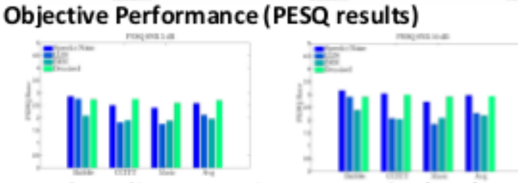
- LL-BSS method was compared to a state of the art spectral substraction algorithm
- Algorithms designed as front-end pre-processing techniques (input an audio waveform mixed with noise and delivering a "cleaned" version of the audio waveform through loudspeakers).
  - All experiments here presented were performed in a sound treated room delivering the cleaned signals through loudspeakers.
- Speech Test [3]: Cardenas bysillabic word test [1]) mixed with different noises (CCITT noise, music noise and babble noise). 2 Lists of 20 words

Results

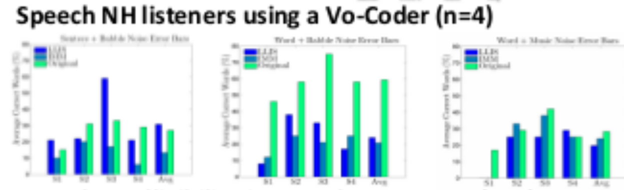
SDR



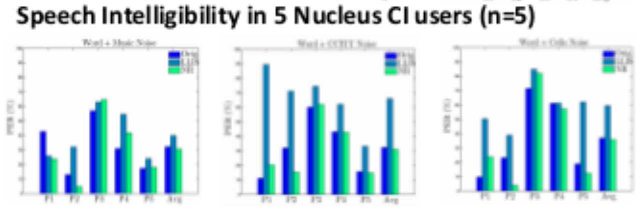
Objective Performance (PESQ results)



Speech NH listeners using a Vo-Coder (n=4)



Speech Intelligibility in 5 Nucleus CI users (n=5)



Conclusions

- Design of a low-latency source separation algo for speech enhancement
- Comparison to 2 well known algorithms (Spectral Substraction and NMF)
- Objective results based on SIR, SDR show improvements for LL-IS
- No improvement in PESQ
- No improvement in normal hearing listeners and CI users
- Reason: Consonants missing when doing the separation speech/noise

References

[1] J.-l Durrieu, , B. David, S. Member, "A musically motivated mi-level representation for pitch estimation and musical audio source separation", IEEE Journal on Selected Topics on Signal Processing, 2011
[2] R. Marxer, J. Janer and J. Bonada, Low-Latency Instrument Separation in Polyphonic Audio Using Timbre Models, Latent Variable Analysis and Signal Separation - 10th International Conference, LVA/ICA, Israel, 2012.
[3] María Rosa Cárdenas, Victoria Marrero, Cuaderno de logoaudometría, Guía de referencia rápida, Vol. 129, Uned Series, ISBN: 8436230116, 9788436230116 , 1994