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 and ones elsewhere
- 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
     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)
  3. Pitch tracking
     - Dynamic programming algorithm (2 steps)
       - Viterbi determines optimal pitch track
       - Determines voiced/unvoiced frames

→ Workflow supervised training method:

Instantaneous mixing model (IMM) [1] (M mixtures, N sources)
\[ x_n(t) = \sum_{m=1}^{M} a_{nm} s_m(t), \quad m = 1, \ldots, M \]
\[ a_{nm} = \begin{bmatrix} a_{n0} & a_{n1} & \cdots & a_{nM-1} \end{bmatrix} \]
\[ s_m(t) = \begin{bmatrix} s_1(t) & s_2(t) & \cdots & s_M(t) \end{bmatrix} \]

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

Single channel noise reduction (Spectral Subtraction, NR)
- Spectral subtraction technique on Bark Domain

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 Subtraction 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