Dereverb

Speech enhancement with Complex Cepstrum

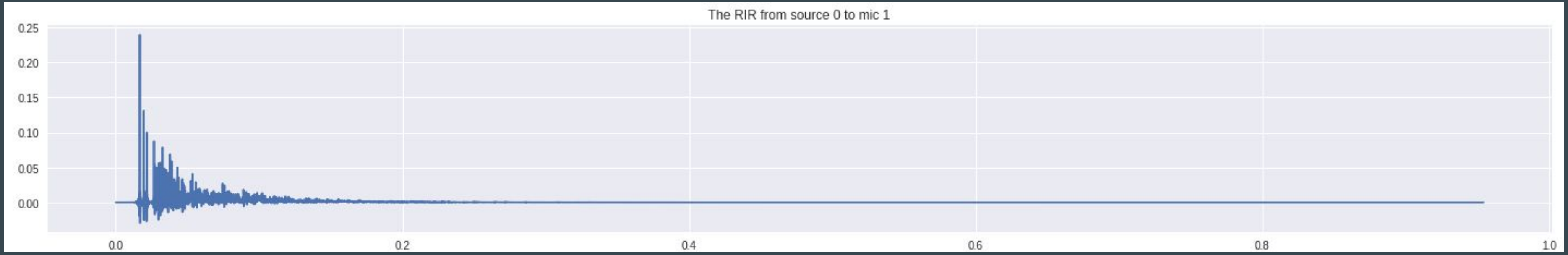
SASP
Rapisarda Claudio, Sani Paolo

Objectives of the project

- Develop a speech enhancement experiment based on short-time segmentation and evaluate it over different RIRs
- Highlight key-points and weaknesses of the Cepstrum-based system commonly reported in literature

Source:

https://github.com/PaoloSani/SASP-Project



The theory behind Cepstrum Deconvolution

$$x(n) = s(n) * h(n) \rightarrow c_x(q) = c_s(q) + c_h(q)$$

$$Fourier\ transform:$$

$$X(\omega_k) = S(\omega_k)H(\omega_k)$$

$$Apply\ the\ logarithm:$$

$$ln[X(\omega_k)] = ln[S(\omega_k)H(\omega_k)] = ln[S(\omega_k)] + ln[H(\omega_k)]$$

$$Compute\ the\ Cepstrum:$$

$$c_x(q) = iDFT\big[ln[X(\omega_k)]\big] = iDFT\big[ln[S(\omega_k)] + ln[H(\omega_k)]\big] = c_s(q) + c_h(q)$$

The theory behind Cepstrum Deconvolution

$$x(n) = s(n) * h(n) \rightarrow c_x(q) = c_s(q) + c_h(q)$$

$$c_w(q) \simeq c_h(q) \longrightarrow c_w(q) = c_x(q)G(q)$$

$$c_x(q) - c_w(q) = c_s(q) + c_h(q) - c_w(q) \simeq c_s(q)$$

The theory behind Cepstrum Deconvolution - Real Cepstrum

$$x(n) = s(n) * h(n)$$

$$c_x(q) = iDFT\left[ln[|X(\omega_k)|]\right]$$

$$c_y(q) = c_x(q)(1 - G(q))$$

$$y(n) = iDFT\left[e^{DFT[c_y(q)] + j\angle X(\omega_k)}\right] = \hat{s}(n)$$

Real Cepstrum

- discard the phase information
- no need for phase unwrapping when computing cepstrum
- phase must be taken when reconstructing the signal

The theory behind Cepstrum Deconvolution - Complex Cepstrum

$$X(\omega_k) = DFT[x(n)] = |X(\omega_k)|e^{j\phi(\omega_k)}$$

$$ln[X(\omega_k)] = ln[|X(\omega_k)|e^{j\phi(\omega_k)}] = ln[|X(\omega_k)| + j\phi(\omega_k)$$

$$\tilde{c}_x(q) = iDFT\left[ln[X(\omega_k)]\right]$$

$$\tilde{c}_y(q) = \tilde{c}_x(q)\left(1 - \tilde{G}(q)\right)$$

$$y(n) = iDFT\left[e^{DFT[\tilde{c}_y(q)]}\right] = \hat{s}(n)$$

Complex Cepstrum

- it retains information about the phase
- need for unwrapping during cepstrum computation

The algorithm

Time domain
- Creation of the reverbered audio file
- Segmentation of the audio file
- Weighting by exponential window

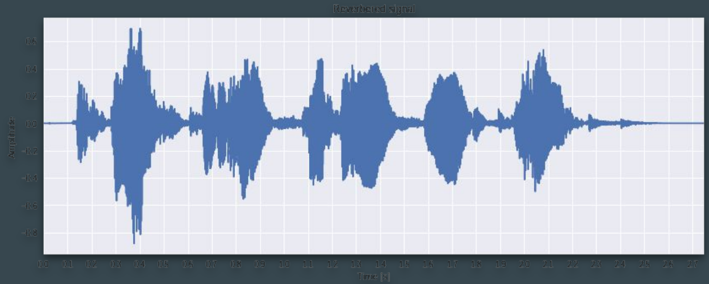
Cepstrum domain
- Cepstrum calculation of the 8 segments
- Mean Cepstrum
- Peaks detection and liftering

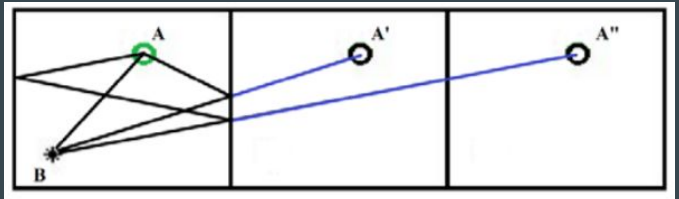
Back to time domain

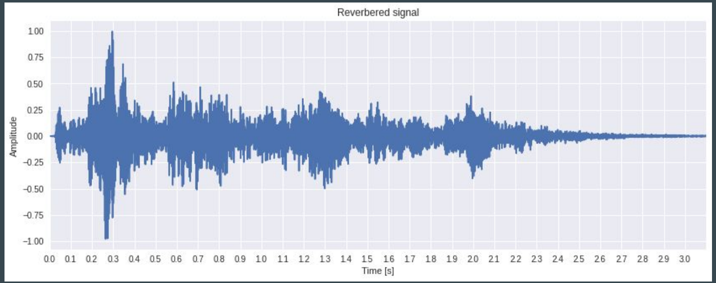
Time Domain Processing

The experiment was carried out using:

- a short RIR, simulated with Image Method, t60 = 0.4 s
- a long RIR, with t60 = 2.55 s (@ 1kHz)







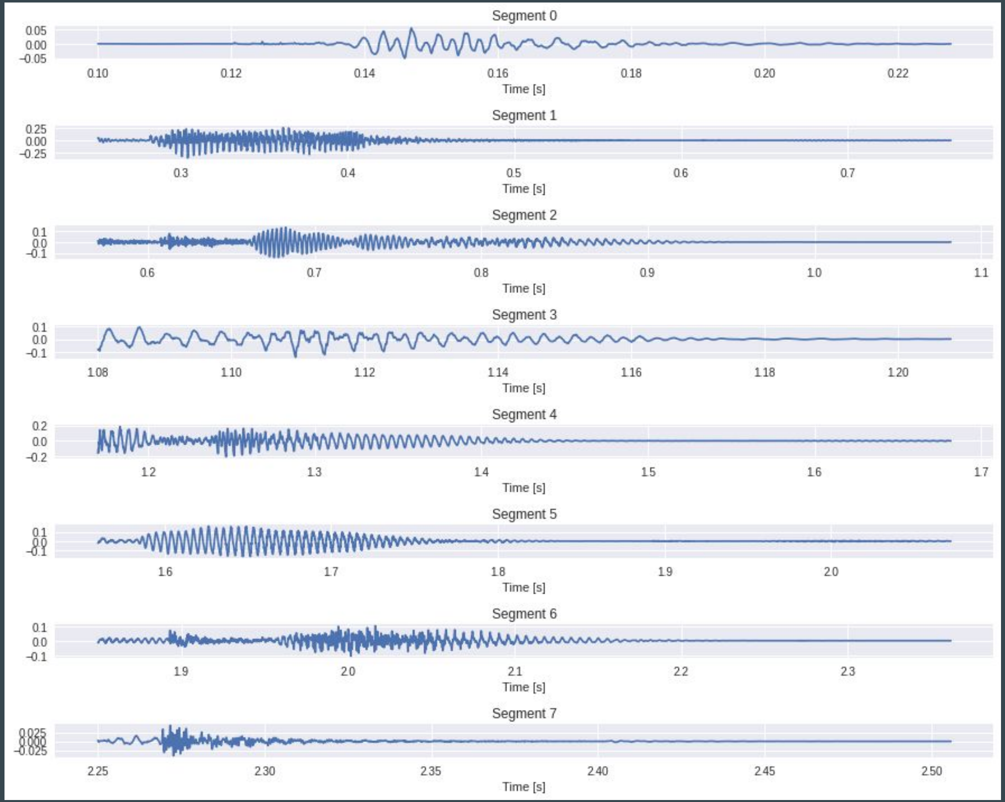
Segmentation and exponential windowing

Segmentation has been done by eye, accounting for speech activity.

Applying an exponential window:

- let us deal with segmentation error
- returns a minimum phase sequence [1]

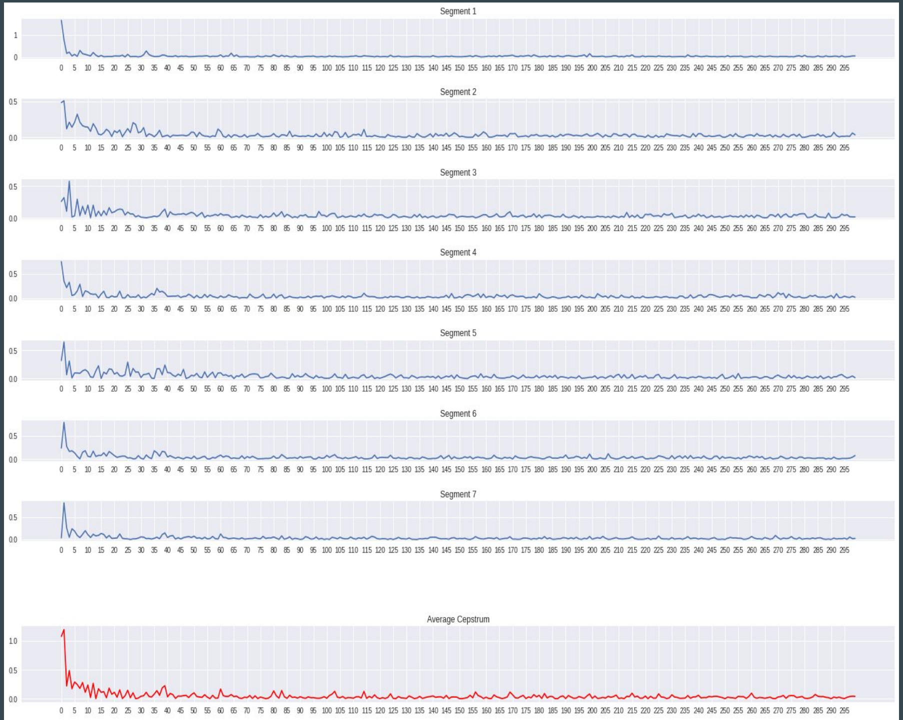
$$w(n) = e^{-\frac{|n - center|}{\tau}}$$



Cepstrum domain processing

In Cepstrum domain, low quefrencies account for slowly varying spectrum components and they are used to identify speech, while high quefrencies refer to the spectrum of the reverb.

Exploiting the statistical properties of speech we extrapolated relevant peaks through the calculation of the average Cepstrum.

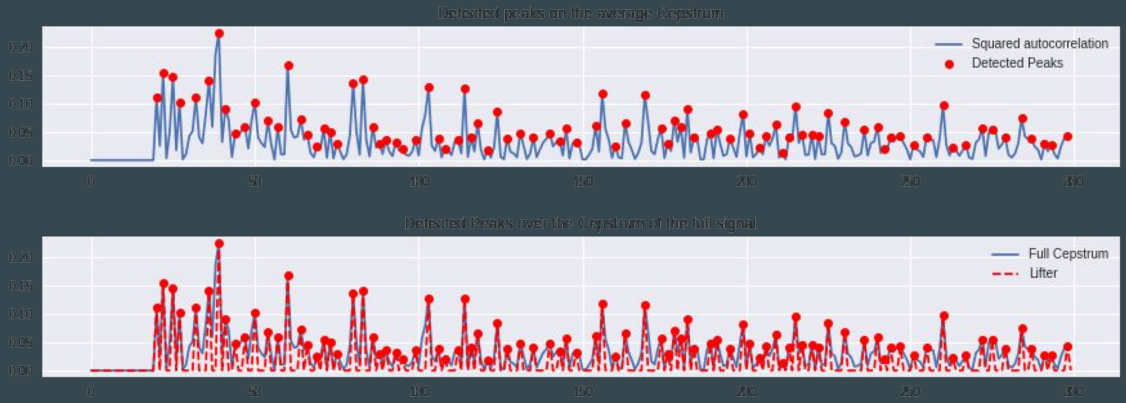


Peak detection and liftering

First 20 ms quefrencies are associated to speech, thus, they are neglected

Parameters for peak detection:

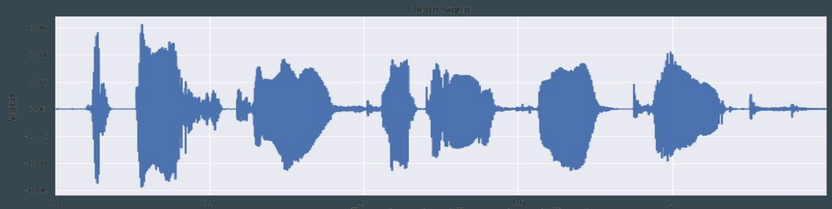
- Neglected quefrencies
- Threshold
- No. of cepstral coefficients taken into account

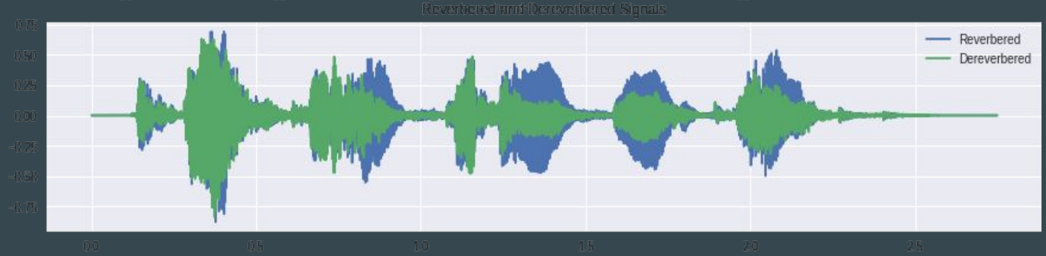


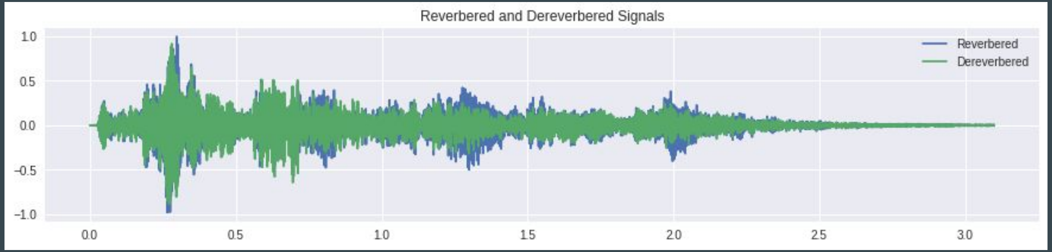
Results

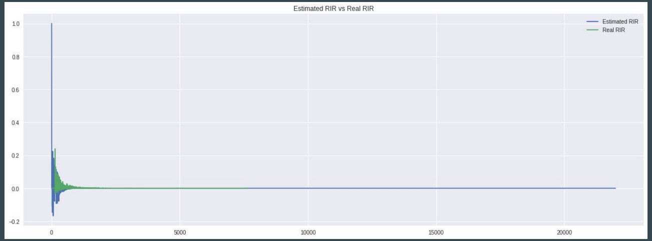
The best result was obtained for the short RIR. As it can be noted, the “boomy” effect of the reverb is greatly reduced.

On the other hand, the longest RIR turned out to be quite uneffected by the dereverberation system.









Discussion

- Segmentation seems to be more important than expected, since a more efficient and automatized system could be implemented in order to minimize overlapping from nearest segments' reverb.
- Thresholding is important as well, more studies could be carried out to find a correct value to choose, also accounting for the mixed speech-reverb peaks.
- Iterative techniques could be used to find the best values for the control parameters (i.e. rate of decay of the exponential window, threshold value for peak detection and number of cepstral coefficients to consider).

Bibliography

Our reference material:

- D. Bees, M. Blostein, P. Kabal, *Reverberant speech enhancement using cepstral processing.* [1]
- S. Subramaniam, A. P. Petropulu, C. Wendt, *Cepstrum-Based Deconvolution for Speech Dereverberation.* [2]
- S. Xizhong, M. Guang, *Complex Cepstrum Based Single Channel Speech Dereverberation.* [3]
- Patrick Ziegler, *Single-Channel Speech Dereverberation.* [4]

Along with other articles collected in our repository.