Dereverb

Speech enhancement with Complex Cepstrum

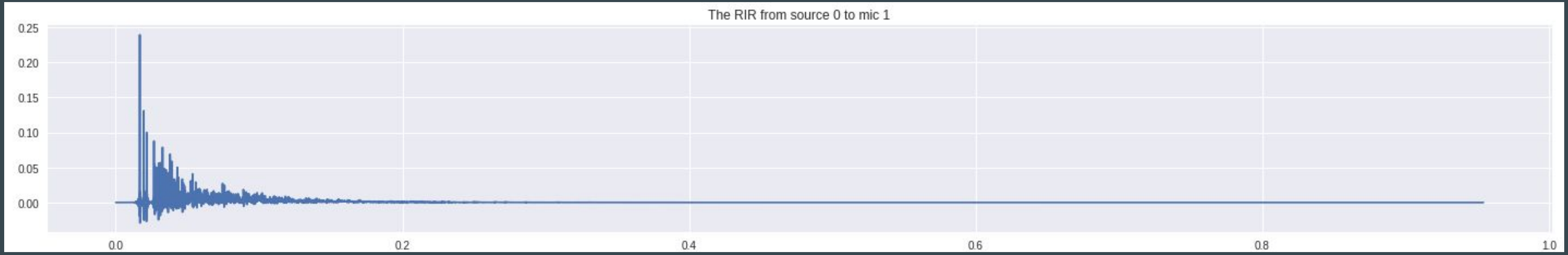
SASP
Rapisarda Claudio, Sani Paolo

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

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

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

$$c_y(q) = c_x(q)\left(1 - G(q)\right)$$

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

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

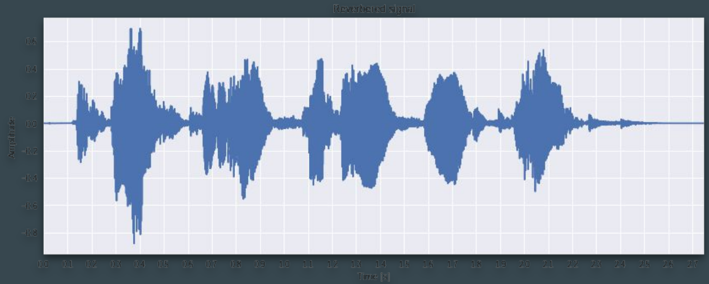
The algorithm

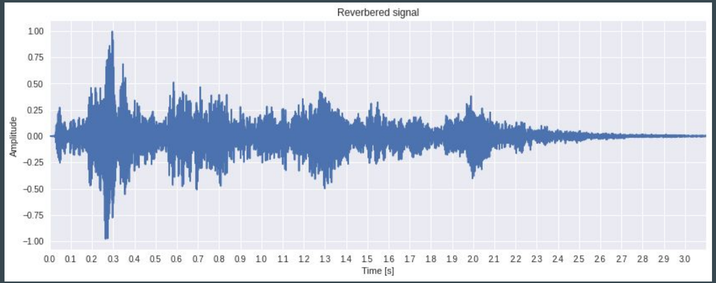
Time domain	Creation of the reverbered audio file	Segmentation of the audio file	Weighting by exponential window
Cepstrum domain	Cepstrum calculation of the 8 segments	Mean Cepstrum	Peaks detection and liftering
		Back to time domain	

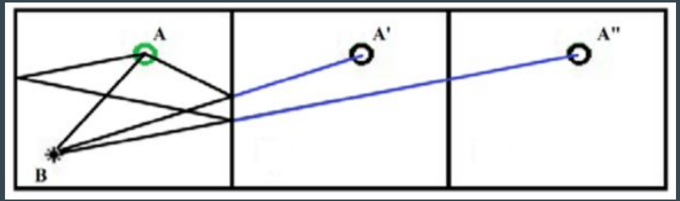
Time Domain Processing

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)







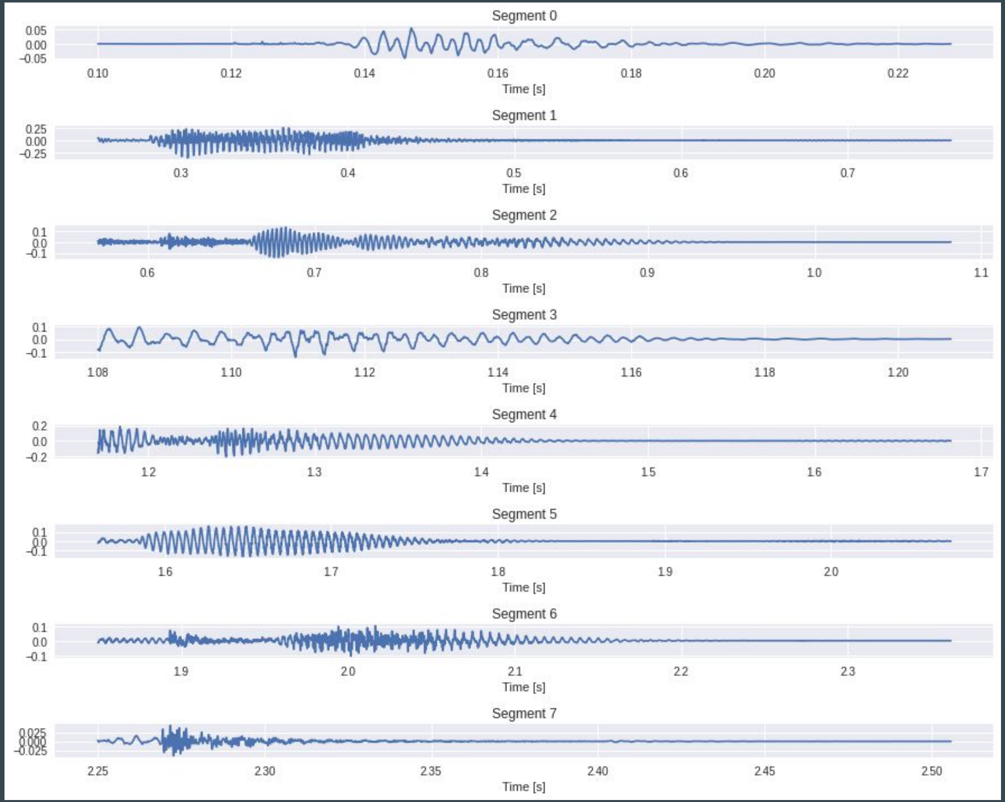
Segmentation and exponential windowing

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]

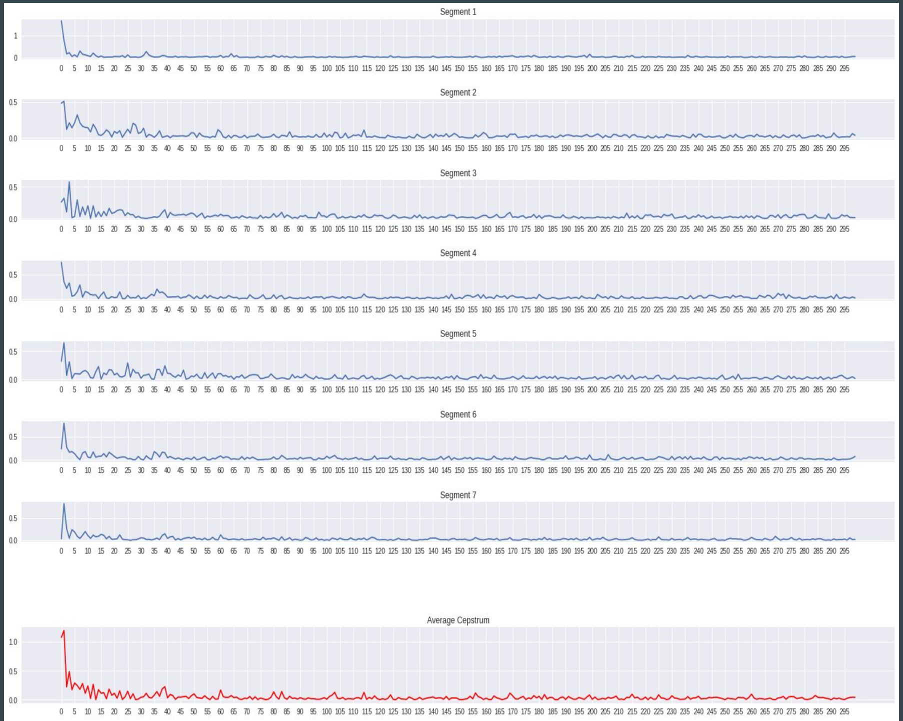
$$w(n) = e^{-\frac{|n-center|}{\tau}}$$



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.

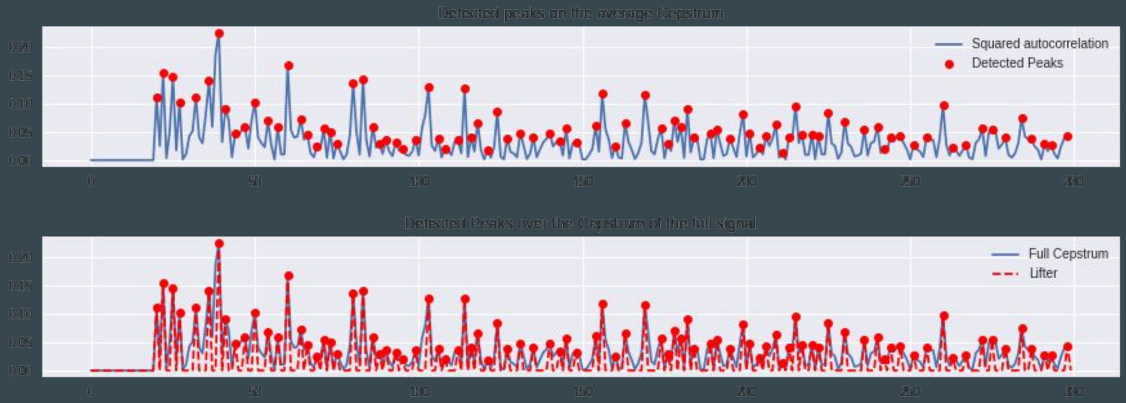


Peak detection and liftering

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

Parameters for peak detection:

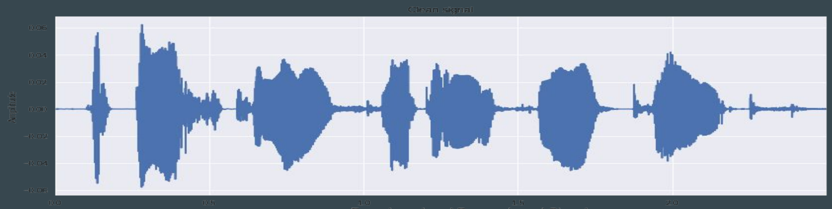
- Neglected quefrencies
- Threshold
- No. of cepstral coefficients taken into account

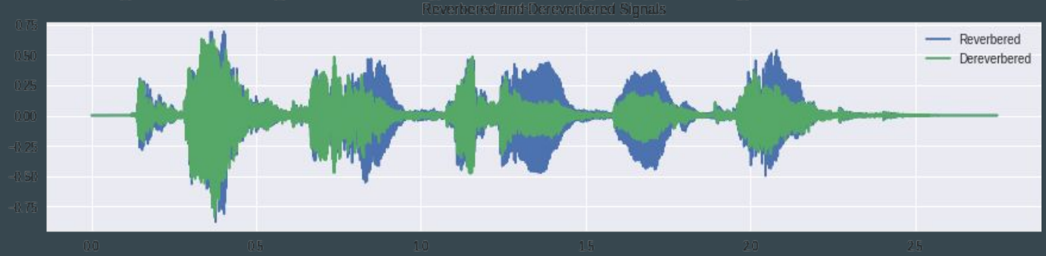


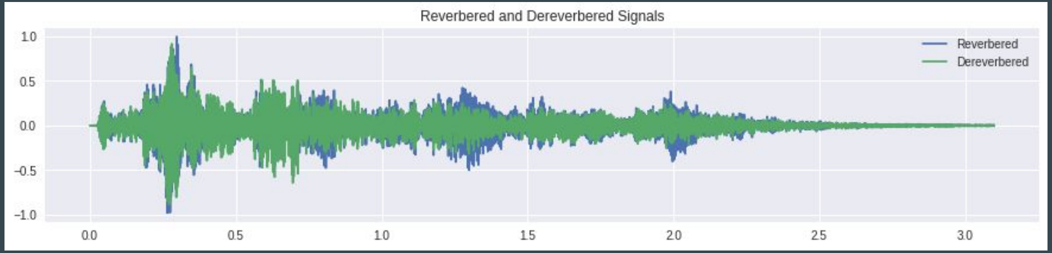
Results

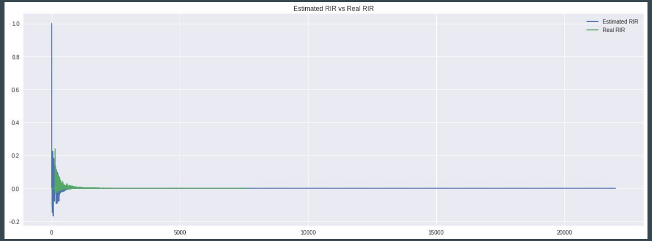
The best result was obtained for the short RIR. As it can be noted, the “boomy” effect of the reverb is greatly reduced.

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

Main references:

- D. Bees, M. Blostein and P. Kabal, "Reverberant speech enhancement using cepstral processing," [Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing, 1991, pp. 977-980 vol.2, doi: 10.1109/ICASSP.1991.150504. [1]
- S. Subramaniam, A. P. Petropulu and C. Wendt, "Cepstrum-based deconvolution for speech dereverberation," in IEEE Transactions on Speech and Audio Processing, vol. 4, no. 5, pp. 392-396, Sept. 1996, doi: 10.1109/89.536934. [2]
- Shen Xizhong and Meng Guang, "Complex cepstrum based single channel speech dereverberation," 2009 4th International Conference on Computer Science & Education, 2009, pp. 7-11, doi: 10.1109/ICCSE.2009.5228535. [3]
- Patrick Ziegler, *Single-Channel Speech Dereverberation,* Master Thesis at Technischen Universität Graz. https://phaidra.kug.ac.at/open/o:43183
[4]

Other references:

- J. Bednar and T. Watt, "Calculating the complex cepstrum without phase unwrapping or integration," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 33, no. 4, pp. 1014-1017, August 1985, doi: 10.1109/TASSP.1985.1164655.
- Cauchi, Benjamin & Kodrasi, Ina & Rehr, Robert & Gerlach, Stephan & Jukic, Ante & Gerkmann, Timo & Doclo, Simon & Goetze, Stefan. (2014). Joint Dereverberation and Noise Reduction using Beamforming and a single-channel speech enhancement scheme. Proceedings of REVERB Challenge Workshop, o1.2.
- S. C. Douglas, H. Sawada and S. Makino, "Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters," in IEEE Transactions on Speech and Audio Processing, vol. 13, no. 1, pp. 92-104, Jan. 2005, doi: 10.1109/TSA.2004.838538.
- D. Naik, "Pole-filtered cepstral mean subtraction," 1995 International Conference on Acoustics, Speech, and Signal Processing, 1995, pp. 157-160 vol.1, doi: 10.1109/ICASSP.1995.479388.
- D. G. Childers, D. P. Skinner and R. C. Kemerait, "The cepstrum: A guide to processing," in Proceedings of the IEEE, vol. 65, no. 10, pp. 1428-1443, Oct. 1977, doi: 10.1109/PROC.1977.10747.