Amplitude and Phase Dereverberation of Harmonic Signals

Arthur Belhomme, Roland Badeau, Yves Grenier, Eric Humbert
arthur.belhomme@telecom-paristech.fr

LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

Outline

Outline

Introduction

Reverberation, a natural process

- Results from a direct sound... and all its reflections,
- Spreads the signal in the time-frequency domain.

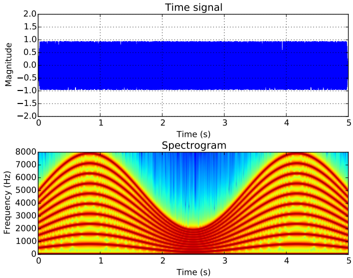


Figure: From anechoic to reverberant signal

Introduction

Dereverberation, a speech enhancement task

- Cancellation methods: estimate the room impulse response (RIR),
- Suppression methods: estimate the late reverberation.

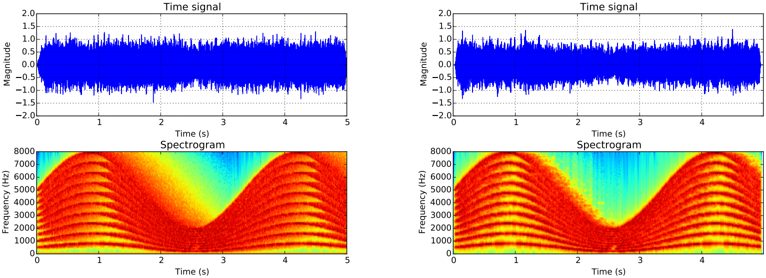


Figure: From reverberant to dereverberated signal

Introduction

Phase, a forsaken issue

- For cancellation methods, there is no problem of phase,
- For suppression methods, there are many ways of estimating the dereverberated magnitude.

Suppression methods estimate a **dereverberated amplitude** but synthesize with the **reverberant phase** ⇒ reintroduces reverberation.

Previous contribution

We proposed a method which jointly estimates the **dereverberated amplitude** and the **dereverbated phase** of the signal.

Our goal and main contribution

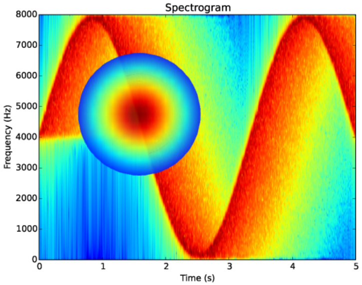


Figure: Example of how our previous contribution works

- Previous method assumed at most one component in different regions of the time-frequency plane,
- Need large neighborhoods to perform strong dereverberation,
- **We now alleviate this condition by considering a harmonic model of signals.**

Outline

Model and notation

▷ We model the anechoic signal $s(t)$ by a sum of Q complex sinusoids $s_q(t)$ of log-amplitude $\lambda_q(t)$ and phase $\varphi_q(t)$:

$$s(t) = \sum_{q=1}^{Q} s_q(t) = \sum_{q=1}^{Q} e^{\lambda_q(t)+j\varphi_q(t)},$$

▷ $\varphi_q(t)$ is related to the instantaneous frequency $f_q(t)$ by:

$$f_q(t) = \frac{1}{2\pi}\dot{\varphi}_q(t).$$

We assume the signal to be harmonic, which implies:

$$f_q(t) = q\, f_1(t), \forall q \in [1, Q].$$

Harmonic model

▷ Hence, $\forall q \in [1, Q]$ we have:

$$\dot{\varphi}_q(t) = q\,\dot{\varphi}(t), \qquad \ddot{\varphi}_q(t) = q\,\ddot{\varphi}(t),$$

with $\dot{\varphi}(t) = \dot{\varphi}_1(t)$ and $\ddot{\varphi}(t) = \ddot{\varphi}_1(t)$.

▷ Our method also needs to assume harmonic ratios between the log-amplitude derivatives. Hence, $\forall q \in [1, Q]$ we have:

$$\dot{\lambda}_q(t) = q\,\dot{\lambda}(t), \qquad \ddot{\lambda}_q(t) = q\,\ddot{\lambda}(t),$$

with $\dot{\lambda}(t) = \dot{\lambda}_1(t)$ and $\ddot{\lambda}(t) = \ddot{\lambda}_1(t)$.

Second-order approximation

▷ Let $\theta_q(t) = \lambda_q(t) + j\varphi_q(t)$ and $\theta(t) = \theta_1(t)$, then:

$$\theta_q(t) = q\,\theta(t),\ \forall q \in [1, Q].$$

▷ Considering a sampling frequency f_s and a time-shift of R samples, the time t_m of frame m is defined by $t_m = m\frac{R}{f_s}$.

▷ We approximate each complex sinusoid by its 2nd order Taylor expansion around time t_m:

$$s_q(t) = a_{m,q} e^{j\varphi_{m,q}} e^{q\left(\dot{\theta}_m(t-t_m)+\frac{1}{2}\ddot{\theta}_m(t-t_m)^2\right)}, \tag{1}$$

with $a_{m,q} = e^{\lambda_q(t_m)}$, $\varphi_{m,q} = \varphi_q(t_m)$ and $\theta_m = \theta(t_m)$.

Key equation - 1

▷ We work with the odd-frequency Short Time Fourier Transform (oSTFT), with K band-pass filters $g_k(t),\ k \in [0, K-1]$:

$$S_g[m,k] = (g_k * s)(t_m).$$

▷ By differentiating (1), we have:

$$\dot{s}_q(t) = q\left(\dot{\theta}_m + \ddot{\theta}_m (t - t_m)\right) s_q(t). \quad (2)$$

▷ By convolving (2) with g_k we have

$$(\dot{g}_k * s_q)(t) = q\,\dot{\theta}_m (g_k * s_q)(t) + q\,\ddot{\theta}_m \left((t - t_m)(g_k * s_q)(t) - \left(g'_k * s_q\right)(t)\right), \quad (3)$$

where $g'_k(t) = t g_k(t)$.

Key equation - 1

- ▷ We assume at most one harmonic q at $[m, k]$ and consider a mask $w_{m,q}[m', k'] \in [0, 1]$ measuring whether the same harmonic is also dominant at $[m', k']$.
- ▷ From (3), we can show that $\begin{bmatrix} \dot{\theta}_m \\ \ddot{\theta}_m \end{bmatrix}$ is the unique solution of the system

$A_m \begin{bmatrix} \dot{\theta}_m \\ \ddot{\theta}_m \end{bmatrix} = b_m$ with:

$$A_{m,k} = \sum_{q=1}^{Q} q^2 \sum_{m',k'} w_{m,q} \begin{bmatrix} |S_g|^2 & S_g^* S_m \\ S_g S_m^* & |S_m|^2 \end{bmatrix}, \text{ and}$$

$$b_{m,k} = \sum_{q=1}^{Q} q \sum_{m',k'} w_{m,q} \begin{bmatrix} S_g^* S_{\dot{g}} \\ S_m^* S_{\dot{g}} \end{bmatrix},$$

Key equation - 2

▷ Again, by convolving (1) with g_k and considering the same mask $w_{m,q}[m', k']$, we can show that:

$$a_{m,q}^2 = \frac{\sum_{m',k'} w_{m,q}|S_g|^2}{\sum_{m',k'} w_{m,q}|G_{m,q}|^2}, \quad (4)$$

with:

$$G_{m,q}[m', k'] = e^{q(t_{m'}-t_m)\left(\dot{\theta}_m+\frac{1}{2}\ddot{\theta}_m(t_{m'}-t_m)\right)} \sum_n g_{k'}[n]e^{-q\frac{n}{f_s}\left(\dot{\theta}_m+\ddot{\theta}_m\left(t_{m'}-t_m-\frac{n}{2f_s}\right)\right)}.$$

▷ We thus need to solve the linear system before computing $a_{m,q}$.

Back to the time domain

- ▷ Once $\dot{\theta}_m$ and $\ddot{\theta}_m$ are obtained, we can compute the amplitude $a_{m,q}$.
- ▷ For $\varphi_{m,q}$, we perform phase unwrapping, from $\dot{\varphi}_m$ and $\ddot{\varphi}_m$.
- ▷ The oSTFT is then obtained with:

$$S_g[m,k] = \sum_{q=1}^{Q} a_{m,q} e^{j\varphi_{m,q}} \sum_{n} g_k[n] e^{-q\frac{n}{f_s}\left(\dot{\theta}_m - \ddot{\theta}_m \frac{n}{2f_s}\right)}. \quad (5)$$

- ▷ The time signal is finally obtained by applying an inverse oSTFT to (5).

Outline

Introduction

Parameters estimation

Estimation in presence of reverberation

Performance evaluation

Conclusion

Model and notation

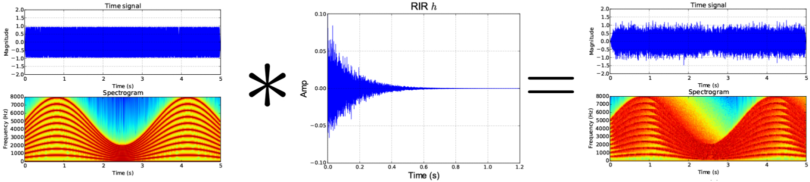
▷ We model the RIR with the stochastic model:

$$h(t) = b(t)p(t),$$

with $b(t) \sim \mathcal{N}(0, \sigma^2)$ i.i.d., $p(t) = e^{-\alpha t}\mathbb{1}_{t \geq 0}$ and $\alpha = \frac{3\log(10)}{\mathrm{RT}_{60}}$.

▷ The reverberant signal $y(t)$ is obtained as the convolution:

$$y(t) = (h * s)(t).$$



Key equation - 3

▷ For any real analog signals x_1 and x_2, we can show that:

$$\mathbb{E}_b\left[(h * x_1)(h * x_2)\right] = \sigma^2 p^2 * (x_1 x_2),$$

where $\mathbb{E}_b$ denotes the mathematical expectation w.r.t. $b(t)$.

▷ it can be easily proved that the inverse filter of $\sigma^2 p^2$ is:

$$\gamma(t) = \frac{1}{\sigma^2}\left(2\alpha\delta(t) + \dot{\delta}(t)\right).$$

We replace x_1, x_2 by $(u_k * s)$ and apply γ to obtain the quadratic terms of A_m and b_m, forming the linear system $A_m \begin{bmatrix} \dot{\theta}_m \\ \ddot{\theta}_m \end{bmatrix} = b_m$. Depending on the entry of the matrix, u_k can be g_k, $\dot{g}_k$ or g'_k.

TELECOM ParisTech

Back in the time domain

▷ Once we have estimated $\widehat{A}_m$ and $\widehat{b}_m$, the amplitude and phase parameters are estimated as follows:

$$\begin{bmatrix} \widehat{\dot{\theta}}_m \\ \widehat{\ddot{\theta}}_m \end{bmatrix} = \widehat{A}_m^{-1} \widehat{b}_m.$$

▷ $\widehat{a}_{m,q}$ is estimated as before, from the estimated $\widehat{|S_g|^2}$ and $\widehat{G}_{m,q}$

▷ The estimated $\widehat{S}_g$ is obtained as before, and the time signal by an inverse oSTFT.

Outline

Introduction

Parameters estimation

Estimation in presence of reverberation

Performance evaluation

Conclusion

Dataset

To evaluate our method, we used:

- A harmonic signal of instantaneous frequency ranging from 0 Hz to 8 kHz in 2 seconds.
- RIRs simulated according to the stochastic model, with a reverberation time ranging from 0.2 s to 2.2 s,
- RIRs recorded in real conditions, from the AIR database (Jeub et al., 2009).

Objective measures

We use the REVERB challenge toolbox (Kinoshita et al., 2016):

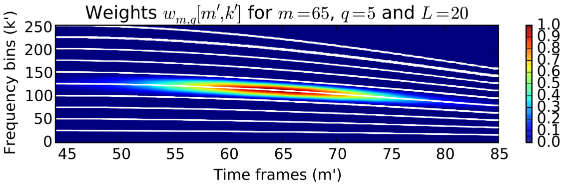
- The frequency-weighted segmental SNR (fwsegSNR), in dB, to evaluate the level of reverberation.
  $\Rightarrow$ The higher the better.
- The Cepstral Distance, in dB, to evaluate the level of distortion.
  $\Rightarrow$ The lower the better.

Both indexes are defined in *Evaluation of Objective Quality Measures for Speech Enhancement* (Hu and Loizou, 2008).

Evaluation

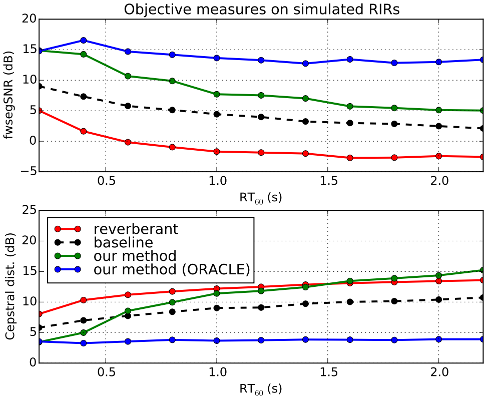
▷ Convolving the anechoic signal with all the RIRs led in a wide variety of reverberant signals,

▷ We processed them with our method, with or without the "ORACLE" localization of harmonics:



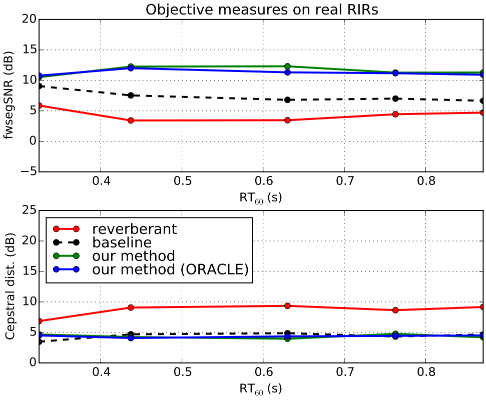
▷ We also processed them with a standard suppression method *Late Reverberant Spectral Variance Estimation based on a Statistical Model* (Habets et al., 2009) to compare our results.

Simulated RIRs



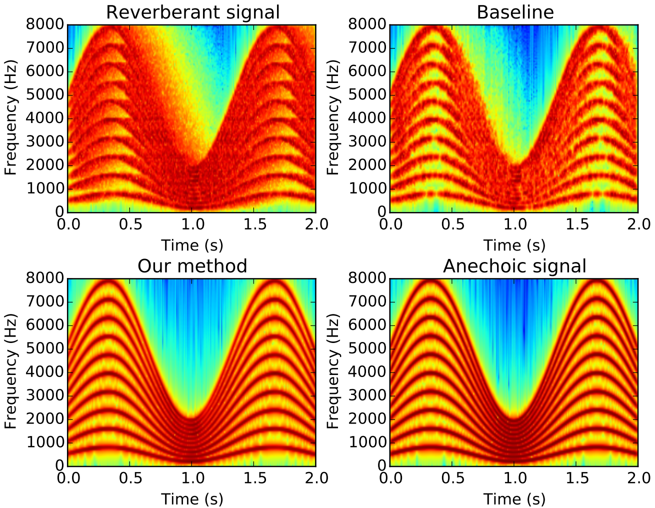
Great in term of fwsegSNR, some distortion with blind localization.

Real RIRs



Lower improvement, but RIRs are less reverberant.

Example of spectrograms



Outline

Introduction

Parameters estimation

Estimation in presence of reverberation

Performance evaluation

Conclusion

Conclusion

Problem encountered

Our previous work was restricted to monocomponent signals to perform high-quality dereverberation.

Our new method

Based on a harmonic signal model, we can compute averages over the full reverberant spectrogram to estimate the anechoic signal.

Performance of our method

Very good results in terms of dereverberation, but some distortion is introduced in case of inaccurate harmonic location. Future work will use a model of noise to process speech as harmonics + noise.

Time for questions

THANK YOU !