Amplitude and Phase Dereverberation of Harmonic Signals

Arthur Belhomme, Roland Badeau, Yves Grenier, Eric Humbert

AMPLITUDE AND PHASE DEREVERBERATION OF HARMONIC SIGNALS

Arthur Belhomme,[1,2] *Roland Badeau,*[1] *Yves Grenier,*[1] *Éric Humbert*[2]

[1] LTCI, Télécom ParisTech,Université Paris-Saclay, 75013 Paris, France
{*arthur.belhomme, roland.badeau, yves.grenier*}*@telecom-paristech.fr*
[2] Invoxia, Audio Lab, 92130 Issy-les-Moulineaux, France
eric.humbert@invoxia.com

ABSTRACT

While most dereverberation methods focus on how to estimate the magnitude of an anechoic signal in the time-frequency domain, we propose a method which also takes the phase into account. By applying a harmonic model to the anechoic signal, we derive a formulation to compute the amplitude and phase of each harmonic. These parameters are then estimated by our method in presence of reverberation. As we jointly estimate the amplitude and phase of the clean signal, we achieve a very strong dereverberation on synthetic harmonic signals, resulting in a significant improvement of standard dereverberation objective measures over the state-of-the-art.

Index Terms— dereverberation, phase, sinusoidal modeling,

1. INTRODUCTION

A sound emitted in an enclosed space reacts with the different surfaces of the room, which produces reverberation. If a soft reverberation may be desired to give a feeling of space or to color a sound [1], strong reverberation damages speech intelligibility and reduces automatic speech recognition performance of machines [2]. Hence, one can perform dereverberation to enhance the sound by estimating the anechoic signal. Most existing methods, known as suppression methods, process the reverberant signal by estimating the magnitude spectrogram of the late reverberation and removing it from the magnitude spectrogram of the input signal. The estimation of the late reverberation can be based on a wide variety of approaches, such as a stochastic model of the room impulse response (RIR) [3], a linear prediction model of speech [4], or more recently on deep neural networks [5].

However, once the dereverberated magnitude spectrogram is computed, suppression methods use the reverberant phase to synthesize the dereverberated signal. This is the main drawback of these methods, because using this corrupted phase reintroduces reverberation and distortion in the signal, as shown in [5]. We know from [6] that phase estimation improves the predicted speech quality and we proposed in [7] a dereverberation method that uses phase information to improve the dereverberation performance. However, this method was restricted to linear chirp signals, assuming the signal amplitude to be known, and it focused on the estimation of the anechoic phase. We then proposed in [8] a method that jointly estimates the amplitude and phase of any kind of signals in a non-supervised way, resulting in high quality analysis/synthesis. This method performs dereverberation by computing local averages of time-frequency data, within areas where only one sinusoidal component is predominant. However, as soon as the components of the signal are too close in the time-frequency domain, which is generally the case for speech, the amount of time-frequency bins to be averaged is not sufficient and the dereverberation performance collapses. In this paper, we apply a harmonic model to the signal in order to get rid of this constraint and compute averages in the full frequency band, which allows us to process multicomponent signals. Previously, a dereverberation method exploiting the harmonic structure of signals had been introduced in [9], but contrary to our method, [9] requires a substantial training stage and does not derive a theoretical expression of the dereverberated amplitude and phase.

Section 2 introduces the amplitude and phase parameters used to model the signals, while Section 3 derives a method for estimating these parameters when analyzing an anechoic signal. In Section 4 we propose a method for estimating these parameters in presence of reverberation. This enables us to synthesize a dereverberated signal, whose dereverberation quality is evaluated in Section 5. Finally, in Section 6 some conclusions are drawn and ideas for future work are presented.

2. MODELS AND NOTATIONS

2.1. Analysis framework

For all $k \in [0, K-1]$, let $g_k(t),\ t \in \mathbb{R}$, be the complex impulse response of an analog band-pass filter, centered at frequency $f_k > 0$. We consider a sampling frequency $f_s > 0$ and we assume that the support of the frequency response of g_k is included in $\left[-\frac{f_s}{2}, \frac{f_s}{2}\right]$. We choose $g_k(t)$ infinitely differentiable and denote its time derivatives $\dot{g}_k = \dfrac{dg_k}{dt}$ and $\ddot{g}_k = \dfrac{d^2 g_k}{dt^2}$. For any analog signal $s(t)$ whose frequency support is also included in $\left[-\frac{f_s}{2}, \frac{f_s}{2}\right]$, we define $\forall m \in \mathbb{Z}, k \in [0, K-1]$,

$$S_g[m,k] = (g_k * s)(t_m), \tag{1}$$

where $*$ denotes the convolution operator, $t_m = m\frac{R}{f_s}$ and R is called the *hop size*. In the same way, we define $S_{\dot{g}}[m,k] = (\dot{g}_k * s)(t_m)$ and $S_{\ddot{g}}[m,k] = (\ddot{g}_k * s)(t_m)$. From now on, we consider that filters g_k are designed so that $S_g[m,k]$ forms a short-term Fourier transform (STFT) of signal s.

2.2. Signal model

The anechoic signal $s(t)$ is modeled as the sum of Q complex harmonic sinusoids $s_q(t)$, of log-amplitude $\lambda_q(t)$ and phase $\varphi_q(t)$:

$$s(t) = \sum_{q=1}^{Q} s_q(t) = \sum_{q=1}^{Q} e^{\lambda_q(t) + j\varphi_q(t)}\ . \tag{2}$$

For the sake of conciseness, we denote $\lambda_{m,q} = \lambda_q(t_m)$ and $\varphi_{m,q} = \varphi_q(t_m)$ the log-amplitude and phase of the q-th harmonic at time t_m, respectively, and we use the same notation for the derivatives of $\lambda_q(t)$ and $\varphi_q(t)$.

As we enforce a harmonic model, we obtain harmonic ratios of the frequency parameters: $\dot{\varphi}_{m,q} = q\dot{\varphi}_{m,1}$ and $\ddot{\varphi}_{m,q} = q\ddot{\varphi}_{m,1}$. Furthermore, we also assume harmonic ratios of the log-amplitude parameters: $\dot{\lambda}_{m,q} = q\dot{\lambda}_{m,1}$ and $\ddot{\lambda}_{m,q} = q\ddot{\lambda}_{m,1}$. This assumption is technically necessary for our method, but it is also realistic in the case of free oscillation, where high-frequency harmonics decay faster than low-frequency ones.

In the neighborhood of time t_m, $s_q(t)$ can then be approximated by the following second-order Taylor expansion:

$$s_q(t) = a_{m,q} e^{j\varphi_{m,q}} e^{q\left(\dot{\theta}_m (t-t_m) + \frac{1}{2}\ddot{\theta}_m (t-t_m)^2\right)} \tag{3}$$

where $a_{m,q} = e^{\lambda_{m,q}}$. Parameters $\dot{\theta}_m$ and $\ddot{\theta}_m$ are the first and second-order time derivatives of $\theta(t) = \lambda_1(t) + j\varphi_1(t)$ at time t_m, respectively.

2.3. Model of RIR

To model reverberation, we adopt the stochastic model of RIR proposed in [10] which carries the information of the reverberation time at 60 dB (RT_{60}) [11]. We thus define the RIR $h(t) = b(t)p(t)$, where $b(t)$ is a centered real-valued white noise of variance σ^2, damped by a decreasing envelope $p(t) = e^{-\alpha t}\mathbf{1}_{t\geq 0}$ of decay rate $\alpha = \frac{3\log(10)}{\mathrm{RT}_{60}}$.

The reverberant signal $y(t)$ is obtained by the convolution of $h(t)$ and $s(t)$:

$$y(t) = (h * s)(t). \tag{4}$$

As in (1), we denote $Y_g[m,k]$ the STFT of $y(t)$ at time-frequency bin $[m,k]$.

3. SIGNAL PARAMETERS ESTIMATION

We show in this section how the amplitude and phase parameters can be estimated from the anechoic signal. From (3), straightforward calculations lead to:

$$\dot{s}_q(t) = q\left(\dot{\theta}_m + \ddot{\theta}_m (t-t_m)\right) s_q(t), \tag{5}$$

$\forall t$ in the neighborhood of t_m. We now assume that there is only one significant harmonic q at time-frequency bin $[m,k]$.

By noting that $(\dot{g}_k * s) = (g_k * \dot{s})$, (5) shows that $\forall t$ in the neighborhood of t_m we have:

$$\begin{aligned}(\dot{g}_k * s)(t) = q\dot{\theta}_m (g_k * s)(t) + \\ q\ddot{\theta}_m \left((t-t_m)(g_k * s)(t) - (g'_k * s)(t)\right),\end{aligned} \tag{6}$$

with $g'_k(t) = t g_k(t)$. Let $w_{m,q}[m',k'] \geq 0$ be a time-frequency mask measuring whether the same harmonic q is also dominant at time-frequency bin $[m',k']$. Through this mask, $\dot{\theta}_m$ and $\ddot{\theta}_m$ are characterized from (6) as the unique minimum of the quadratic function:

$$\begin{aligned}\sum_{q,m',k'} w_{m,q}[m',k'] \times \\ \left| S_{\dot{g}}[m',k'] - q\left(\dot{\theta}_m S_g[m',k'] + \ddot{\theta}_m S_m[m',k']\right)\right|^2,\end{aligned} \tag{7}$$

where $S_m[m',k'] = (t_{m'} - t_m) S_g[m',k'] - S_{g'}[m',k']$.

By differentiating (7) with respect to $\dot{\theta}_m$ and $\ddot{\theta}_m$ and by zeroing the derivatives, we show that parameters $\dot{\theta}_m$ and $\ddot{\theta}_m$ satisfy the linear system[1]:

$$A_m \begin{bmatrix}\dot{\theta}_m \\ \ddot{\theta}_m\end{bmatrix} = b_m \tag{8}$$

with

$$A_m = \sum_{q=1}^{Q} q^2 \sum w_{m,q} \begin{bmatrix}|S_g|^2 & S_g^* S_m \\ S_g S_m^* & |S_m|^2\end{bmatrix} \tag{9}$$

and

$$b_m = \sum_{q=1}^{Q} q \sum w_{m,q} \begin{bmatrix}S_g^* S_{\dot{g}} \\ S_m^* S_{\dot{g}}\end{bmatrix}, \tag{10}$$

where * denotes the complex conjugate.

In other respects, from (3) we derive that $\forall t$ in the neighborhood of t_m:

$$\begin{aligned}(g_k * s)(t) = a_{m,q} e^{j\varphi_{m,q}} \times \\ \sum_n g_k[n] e^{q\left(\dot{\theta}_m \left(t - t_m - \frac{n}{f_s}\right) + \frac{1}{2}\ddot{\theta}_m \left(t - t_m - \frac{n}{f_s}\right)^2\right)},\end{aligned} \tag{11}$$

where $g_k[n] = g_k(\frac{n}{f_s})$. Hence $a_{m,q}$ is characterized as the unique minimum of the function

$$\sum_{m',k'} w_{m,q}[m',k'] \left(\frac{|S_g[m',k']|^2}{a_{m,q}} + a_{m,q} \left|G_{m,q}[m',k']\right|^2\right), \tag{12}$$

where

$$\begin{aligned}G_{m,q}[m',k'] = e^{q\left(t_{m'} - t_m\right)\left(\dot{\theta}_m + \frac{1}{2}\ddot{\theta}_m \left(t_{m'} - t_m\right)\right)} \times \\ \sum_n g_{k'}[n] e^{-q\frac{n}{f_s}\left(\dot{\theta}_m + \ddot{\theta}_m \left(t_{m'} - t_m - \frac{n}{2f_s}\right)\right)}.\end{aligned} \tag{13}$$

By minimizing (12), $a_{m,q}$ is obtained as:

$$a_{m,q} = \sqrt{\frac{\sum w_{m,q}|S_g|^2}{\sum w_{m,q}|G_{m,q}|^2}}. \tag{14}$$

Besides, the phase $\varphi_{m,q}$ of the q-th harmonic at time frame m is estimated by enforcing phase continuity between successive time frames:

$$\varphi_{m,q} = \varphi_{m-1,q} + q\dot{\varphi}_{m-1}\frac{R}{f_s} + \frac{q}{2}\ddot{\varphi}_{m-1}\left(\frac{R}{f_s}\right)^2,\ m > 0, \tag{15}$$

with a random initial phase $\varphi_{0,q}$. In conclusion, by analyzing $s(t)$ with windows g_k, $\dot{g}_k$ and g'_k, given a neighborhood $\mathcal{V}_m$ of time-frequency bins around frame m, we can estimate $a_{m,q}$, $\varphi_{m,q}$, $\dot{\theta}_m$ and $\ddot{\theta}_m$.

Finally, we estimate the STFT of the anechoic signal from (3):

$$S_g[m,k] = \sum_{q=1}^{Q} a_{m,q} e^{j\varphi_{m,q}} \sum_n g_k[n] e^{-q\frac{n}{f_s}\left(\dot{\theta}_m - \ddot{\theta}_m \frac{n}{2f_s}\right)} \tag{16}$$

[1] In equations (8) to (10), (14) and (27), indexes m' and k' have been omitted for conciseness.

and we reconstruct the signal $s(t)$ by applying an inverse STFT to (16). This method gives an accurate estimation of the anechoic signal parameters. Let us see now how to take reverberation into account, in order to estimate the same parameters from a reverberant signal, and thus obtain a dereverberated signal.

4. ESTIMATION IN PRESENCE OF REVERBERATION

The goal is to estimate the quadratic terms in (9) and (10) from the reverberant signal $y(t)$ defined in (4), instead of $s(t)$. To do so, we use the fact that if the RIR $h(t)$ is modeled as in Section 2.3, then for any analog real signals $x_1(t)$ and $x_2(t)$:

$$\mathbb{E}_b\left[(h * x_1) \times (h * x_2)\right] = \sigma^2 p^2 * (x_1 \times x_2), \tag{17}$$

where $\mathbb{E}_b$ denotes the mathematical expectation *w.r.t.* $b(t)$. This relation can be easily verified by using the fact that $\mathbb{E}_b\left[b(u)b(v)\right] = \sigma^2\delta(u-v)$, where $\delta(t)$ denotes the Dirac distribution. Moreover, it can be easily proved that the impulse response of the inverse filter of $\sigma^2 p^2$ is:

$$\gamma(t) = \frac{1}{\sigma^2}\left(2\alpha\delta(t) + \dot{\delta}(t)\right). \tag{18}$$

By noting that $(g_k * y) = (h * g_k * s)$, (17) leads to:

$$\mathbb{E}_b\left[|g_k * y|^2\right] = \sigma^2 p^2 * \left(|g_k * s|^2\right). \tag{19}$$

Applying the inverse filter $\gamma(t)$ in (18) to (19) results in:

$$|g_k * s|^2 = \frac{1}{\sigma^2}\mathbb{E}_b\left[2\alpha|g_k * y|^2 + 2\Re\left((g_k * y)^*\,(\dot{g}_k * y)\right)\right]. \tag{20}$$

By applying (20) to time t_m, we thus obtain for every time-frequency bin $[m,k]$:

$$|S_g|^2 = \frac{1}{\sigma^2}\mathbb{E}_b\left[2\alpha|Y_g|^2 + 2\Re\left(Y_g^* Y_{\dot{g}}\right)\right]. \tag{21}$$

Likewise, we derive the following expressions:

$$S_g^* S_{\dot{g}} = \frac{1}{\sigma^2}\mathbb{E}_b\left[2\alpha Y_g^* Y_{\dot{g}} + Y_g^* Y_{\ddot{g}} + |Y_{\dot{g}}|^2\right], \tag{22}$$

$$S_g^* S_{g'} = \frac{1}{\sigma^2}\mathbb{E}_b\left[2\alpha Y_g^* Y_{g'} + Y_{\dot{g}}^* Y_{g'} + Y_g^* Y_{\dot{g}'}\right], \tag{23}$$

$$|S_{g'}|^2 = \frac{1}{\sigma^2}\mathbb{E}_b\left[2\alpha|Y_{g'}|^2 + 2\Re\left(Y_{g'}^* Y_{\dot{g}'}\right)\right], \tag{24}$$

$$S_{g'}^* S_{\dot{g}} = \frac{1}{\sigma^2}\mathbb{E}_b\left[2\alpha Y_{g'}^* Y_{\dot{g}} + Y_{\dot{g}'}^* Y_{\dot{g}} + Y_{g'}^* Y_{\ddot{g}}\right]. \tag{25}$$

As, in practice, we do not have access to the mathematical expectation, we estimate it with a temporal smoothing by means of a first-order autoregressive filter of transfer function defined by the $\mathcal{Z}$-transform $\frac{1-\eta}{1-\eta z^{-1}}$, with smoothing parameter $\eta = 0.7$. We thus obtain the estimation of the quadratic terms $\widehat{|S_g|^2}$, $\widehat{S_g^* S_{\dot{g}}}$, $\widehat{S_g^* S_{g'}}$, $\widehat{|S_{g'}|^2}$ and $\widehat{S_{g'}^* S_{\dot{g}}}$. From these estimations, we can compute matrix $\widehat{A}_m$ and vector $\widehat{b}_m$ as in (9) and (10), in order to estimate $\widehat{\theta}_m$ and $\widehat{\ddot{\theta}}_m$ by following (8):

$$\begin{bmatrix}\widehat{\theta}_m \\ \widehat{\ddot{\theta}}_m\end{bmatrix} = \widehat{A}_m^{-1}\widehat{b}_m. \tag{26}$$

By following (14), the amplitude is then estimated with:

$$\widehat{a}_{m,q} = \sqrt{\frac{\sum w_{m,q}\widehat{|S_g|^2}}{\sum w_{m,q}\widehat{|G_{m,q}|^2}}}. \tag{27}$$

The phase $\widehat{\varphi}_{m,q}$ is then estimated by phase unwrapping as in Section 3, and the signal is reconstructed in the same way.

5. PERFORMANCE EVALUATION

In our previous work, the evaluation part was restricted to mono-component signals because we did not introduce the harmonic model exploited in this paper. Now, we can apply our dereverberation method to harmonic signals. However, as speech signals are not always harmonic, we only deal with synthetic harmonic signals and propose in Section 6 a solution to process realistic speech signals.

5.1. Dataset and evaluation

In order to evaluate our method, we consider a multicomponent, harmonic, frequency-modulated signal. The sampling frequency f_s is set to 16 kHz, allowing a maximum instantaneous frequency of 8 kHz. To ensure that the estimator performs well at every frequency, the simulated signal spans the entire frequency range, in 2 seconds. Its spectrogram is plotted in Figure 2-(a).

The anechoic signal is then convolved with simulated and real RIRs, of various RT_{60}s. Simulated RIRs are generated according to the model presented in Section 2.3; real RIRs come from the AIR database [12], from which we select regularly spaced RT_{60}s. At a same RT_{60}, the real RIRs are much less reverberant than the simulated ones, which are modeled to emulate a diffuse-field reverberation. The spectrogram of a reverberant signal (with $\mathrm{RT}_{60} = 2.2$ s) is plotted in Figure 2-(b). For this example we used a synthetic RIR to simulate an especially strong and diffuse reverberation.

To assess the performance of our method, we use objective measures from the REVERB challenge toolbox [13]: the *fwsegSNR* to assess the level of reverberation (the higher the better) and the *cepstral distance* to assess the level of distortion (the lower the better); both are defined in [14]. We compare our approach with a state-of-the-art suppression method [15], which focuses only on the magnitude of the STFT and ignores the phase information. Our previous work is not included in the benchmark, as it cannot deal with such multicomponent signals.

5.2. Estimator settings

We split the frequency axis in $K = 256$ bins, centered on the reduced frequencies $\nu_k = \frac{k+0.5}{2K}$ for $k \in [0, K-1]$. The analysis/synthesis window of length $2K-1$ is defined as:

$$g_k[n] = \cos^3\left(\pi\tfrac{n}{2K}\right) e^{2j\pi\nu_k n},\ \forall n \in [-K+1, K-1]$$

and we choose a hop size of $R = \frac{K}{2}$ samples (75% overlap), which can be proved to guarantee perfect reconstruction.

For each frame m, the neighborhood $\mathcal{V}_m$ corresponds to the time-frequency bins $[m', k']$ where $|m - m'| \leq L$. We choose a width of $L = 10$ time frames. On $\mathcal{V}_m$, the weights $w_{m,q}[m', k']$ are defined as the product of a temporal mask $w_m[m']$ and a harmonic mask $w_q[m', k']$. The temporal mask is directly a function of the distance between time frames m and m':

$$w_m[m'] = e^{-\frac{1}{2}\left(\frac{|m-m'|}{L}\right)^2},$$

while the harmonic mask is defined as:

$$w_q[m', k'] = \begin{cases} \frac{1}{2} - \frac{1}{2}\cos\left(\pi\frac{k'-k_q+d_{q+1}}{d_{q+1}}\right) & \text{if } k_{q+1} \geq k' \geq k_q \\ \frac{1}{2} - \frac{1}{2}\cos\left(\pi\frac{d_{q-1}+k'-k_q}{d_{q-1}}\right) & \text{if } k_{q-1} \leq k' < k_q \end{cases}$$

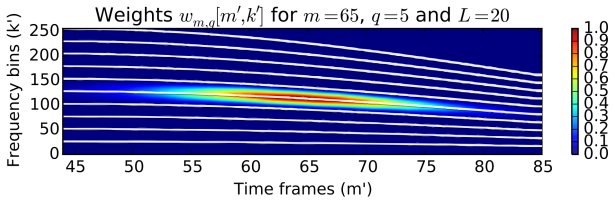

Figure 1: Superposition of harmonics trajectories (white) and the corresponding mask $w_{m,q}[m', k']$ for the 5-th harmonic

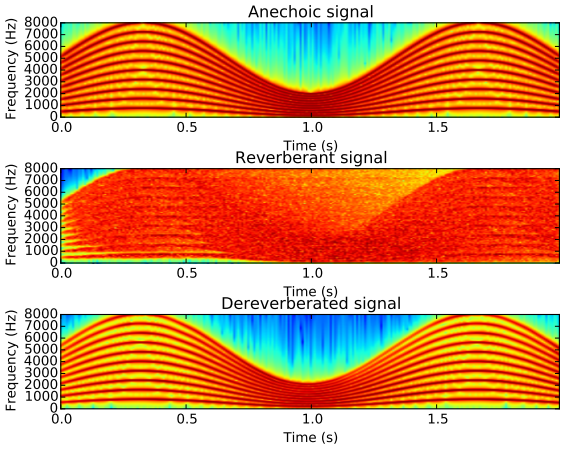

Figure 2: (a) Anechoic, (b) reverberant and (c) dereverberated

for k in $[k_{q-1}, k_{q+1}]$, where k_q matches the peak of the magnitude spectrogram at frame m, corresponding to the q-th harmonic. The distances d_{q-1} and d_{q+1} are defined as $d_{q-1} = k_q - k_{q-1}$ and $d_{q+1} = k_{q+1} - k_q$; we set $k_0 = 0$ and $k_{Q+1} = K$.

The weights $w_{m,q}[m', k'] = w_m[m']w_q[m', k']$ are thus computed from the magnitude spectrogram of either $s(t)$ (ORACLE performance, as in Figure 1) or $y(t)$. If a simple peak detection is sufficient to estimate the k_q for anechoic signals, it is not robust enough for reverberant signals. As reverberation smears the magnitude spectrogram, the q-th peak at frame m may correspond to another harmonic at a previous frame. This results in a noisy fluctuation of the k_q, that we smooth over time with a Savistky-Golay filter [16] in order to recover a more continuous variation.

5.3. Results

The scores are plotted in Figures 3 and 4. We see that dereverberation improves both the *fwsegSNR* and the *cepstral distance*. Moreover, the scores of the dereverberated signal obtained with our method (green) always show a significant improvement *w.r.t.* the baseline method (black) regarding the *fwsegSNR*. However, an error on k_q at frame m results in a slight shift of the frequency content, which degrades the *cepstral distance* (in Figure 3 but not in Figure 4, where the RIRs are less reverberant).

If we accurately locate the harmonics (ORACLE) the proposed method almost achieves a perfect reconstruction and the scores are nearly independent of the reverberation time (blue lines); an example with an initial $\text{RT}_{60} = 2.2$ s is plotted in Figure 2-(c).

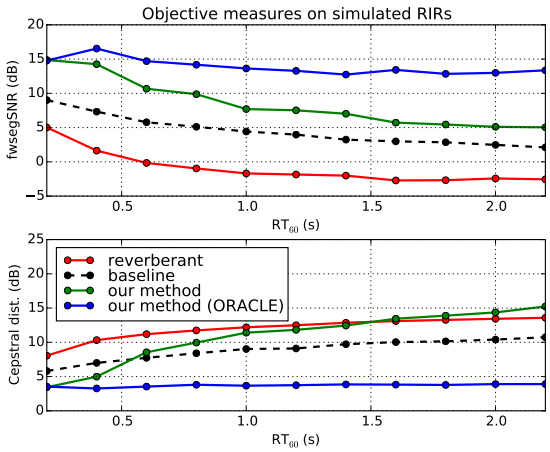

Figure 3: Objective measures on simulated RIRs

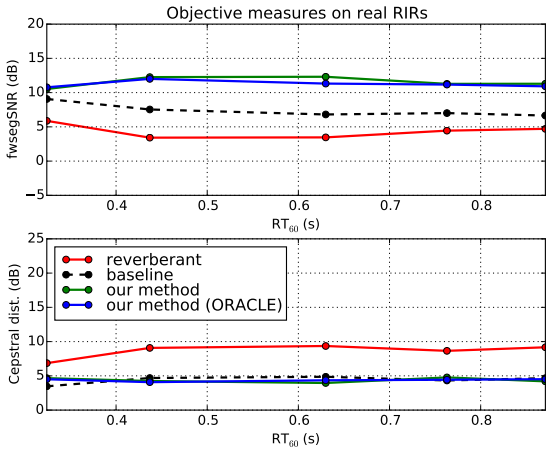

Figure 4: Objective measures on real RIRs

6. CONCLUSION

Instead of only estimating the STFT magnitude of the anechoic signal to perform dereverberation, we proposed a method for jointly estimating the magnitude and phase of the STFT. If the harmonics of the reverberant signal are well located, our estimator almost achieves a perfect reconstruction, on both synthetic and real RIRs.

The performance of the proposed method is limited by the harmonics localization. For monocomponent signals, a peak detection is sufficient to estimate the time-frequency bins $[m, k]$ where the component is predominant, but when dealing with multiple components smeared by reverberation, it is not robust enough. This is why we need a more robust estimation of the k_q to remove the artifacts (that sound like vibratos) in strong reverberant conditions ($\text{RT}_{60} \geq 1$ s).

Future work will tackle speech signals, by applying either the presented method or another method based on a noisy signal model, depending on whether $\mathcal{V}_m$ includes voiced sections or fricative and plosive sounds.

7. REFERENCES

[1] J. Wen and P. Naylor, "An evaluation measure for reverberant speech using decay tail modelling," in *European Signal Processing Conference (EUSIPCO)*, Florence, Italy, September 2006.

[2] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, 2012.

[3] K. Lebart and J. Boucher, "A new method based on spectral subtraction for speech dereverberation," *ACUSTICA*, vol. 87, no. 3, pp. 359–366, 2001.

[4] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Suppression of late reverberation effect on speech signal using long-term multiple-step linear prediction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 534–545, May 2009.

[5] X. Xiao, S. Zhao, D. H. Nguyen, X. Zhong, D. Jones, E. Chang, and H. Li, "Speech dereverberation for enhancement and recognition using dynamic features constrained deep neural networks and feature adaptation," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–18, 2016. [Online]. Available: http://dx.doi.org/10.1186/s13634-015-0300-4

[6] T. Gerkmann, M. Krawczyk, and R. Rehr, "Phase estimation in speech enhancement; unimportant, important, or impossible?" in *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, Nov 2012, pp. 1–5.

[7] A. Belhomme, Y. Grenier, R. Badeau, and E. Humbert, "Anechoic phase estimation from reverberant signals," in *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Xi'an, China, Sept 2016, pp. 1–5.

[8] A. Belhomme, R. Badeau, Y. Grenier, and E. Humbert, "Amplitude and phase dereverberation of monocomponent signals," in *2017 European Signal Processing Conference (EUSIPCO)*, Sept 2017, submitted for publication.

[9] T. Nakatani and M. Miyoshi, "Blind dereverberation of single channel speech signal based on harmonic structure," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 1, April 2003, pp. I–92–5 vol.1.

[10] J. Polack, "La transmission de l'énergie sonore dans les salles," Ph.D. dissertation, Université du Maine, Le Mans, France, 1988.

[11] M. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, no. 3, p. 409, 1965.

[12] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proceedings of the 16th International Conference on Digital Signal Processing*, ser. DSP'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 550–554.

[13] K. Kinoshita, M. Delcroix, S. Gannot, E. Habets, R. Haeb-Umbach, W. Kellermann, V. Leutnant, R. Maas, T. Nakatani, B. Raj, A. Sehr, and T. Yoshioka, "A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–19, January 2016.

[14] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 229–238, Jan 2008.

[15] E. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," Ph.D. dissertation, Technische Universiteit Eindhoven, Netherlands, 2007.

[16] A. Savitzky and M. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.