

Amplitude and Phase Dereverberation of Monocomponent Signals

A. Belhomme, R. Badeau, Y. Grenier
 LTCI, Télécom ParisTech,
 Université Paris-Saclay,
 75013 Paris, France

E. Humbert
 Invoxia,
 Audio Lab,
 92130 Issy-les-Moulineaux, France

Abstract—While most dereverberation methods focus on how to estimate the amplitude of an anechoic signal, we propose a method which also takes the phase into account. By applying a sinusoidal model to the anechoic signal, we derive a formulation to compute the amplitude and phase of each sinusoid. These parameters are then estimated by our method in the reverberant case. As we jointly estimate the amplitude and phase of the clean signal, we achieve a very strong dereverberation, resulting in a significant improvement of objective dereverberation measures over the state-of-the-art.

I. INTRODUCTION

Due to the different paths that an acoustic wave may follow in an enclosed space, a microphone does not only capture the direct sound of a source: all its reflections are added. This phenomenon is known as reverberation. We know from [1] that if a soft reverberation may be desired to give a feeling of space or to color a sound, strong reverberation is unpleasant. Indeed, it damages speech intelligibility and it reduces automatic speech recognition performance of machines [2].

If the room impulse response (RIR) is known, one can invert it to cancel the reverberation [3], [4]. Such methods require a high computing power and are highly dependent on the speaker position, this is why one prefers to use suppression methods. These methods process the reverberant signal by estimating the magnitude spectrogram of the late reverberation and remove it from the magnitude spectrogram of the input signal. Suppression methods are based on a wide variety of approaches, such as a stochastic model of the RIR [5], a linear prediction model of the speech [6], a non-negative matrix factorization approach [7], or more recently on deep neural networks [8].

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 [8]. In the source separation literature, the idea of modeling the phase has recently been proposed [9] because a similar problem occurs (the phase of the mixture is generally used to synthesize the source signals), but not for dereverberation.

We know from [10] that phase estimation improves the predicted speech quality and we proposed in [11] a dereverberation method that uses phase information to improve the dereverberation performance. However, this method is restricted to linear chirp signals. It assumes that the signal magnitude is known, and it focuses on the estimation of the anechoic phase. In the dereverberation literature, we were not able to find a method that jointly estimates the dereverberated magnitude and phase in a non-supervised way. This is why in this paper we propose a single-microphone dereverberation method for accurately estimating the anechoic amplitude, as well as the anechoic phase, in order to reconstruct a clean signal from its reverberant version.

Section II introduces the amplitude and phase parameters used to model the signals, while Section III derives a method for estimating these parameters when analyzing an anechoic signal. In Section IV we propose a method for estimating these parameters given a reverberant signal. This enables us to synthesize a dereverberated signal, whose dereverberation performance is evaluated in Section V. Finally, in Section VI some conclusions are drawn and future work is presented.

II. MODELS AND NOTATIONS

A. Analysis framework

For all $k \in [0, K - 1]$, $K \in \mathbb{N} \setminus \{0\}$, 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 $\left[-\frac{f_s}{2}, \frac{f_s}{2}\right]$. Since $g_k(t)$ is infinitely differentiable, we denote its time derivatives \dot{g}_k and \ddot{g}_k . 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

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

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 .

B. Anechoic signal

We model a signal $s(t)$ as the sum of $q \in [1, Q]$ complex sinusoids $s_q(t)$, of amplitude $\lambda_q(t)$ and phase $\varphi_q(t)$. For the sake of conciseness, we denote $\lambda_{m,q} = \lambda_q(t_m)$ and $\varphi_{m,q} = \varphi_q(t_m)$ the amplitude and phase of the q -th sinusoid at time t_m , respectively. In the neighborhood of t_m , $s_q(t)$ will be approximated by the following second-order Taylor expansion:

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

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

C. Reverberant signal

The RIR $h(t)$ is modeled by a stochastic model, as in [12], carrying the information of the reverberation time at 60 dB (RT₆₀) [13]. We thus define $h(t) = b(t)p(t)$, 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 $\alpha = \frac{3 \log(10)}{\text{RT}_{60}}$.

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

$$y(t) = (h * s)(t). \quad (3)$$

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

III. ANECHOIC SIGNAL PARAMETERS

In this section we show how to estimate the amplitude and phase parameters from the anechoic signal. From (2), straightforward calculations lead to:

$$\dot{s}_q(t) = \left(\dot{\theta}_{m,q} + \ddot{\theta}_{m,q}(t-t_m) \right) s_q(t), \quad (4)$$

$\forall t$ in the neighborhood of t_m . We now assume that there is only one significant sinusoid q at time-frequency bin $[m, k]$ and thus denote $\dot{\theta}_{m,k} = \dot{\theta}_{m,q}$ and $\ddot{\theta}_{m,k} = \ddot{\theta}_{m,q}$. By noting that $(\dot{g}_k * s) = (g_k * \dot{s})$, (4) shows that $\forall t$ in the neighborhood of t_m we have:

$$(\dot{g}_k * s)(t) = \dot{\theta}_{m,k} (g_k * s)(t) + \ddot{\theta}_{m,k} ((t-t_m)(g_k * s)(t) - (g'_k * s)(t)), \quad (5)$$

with $g'_k(t) = tg_k(t)$. Let $w_{m,k}[m', k']$ be a time-frequency mask measuring whether the same component q is also dominant at time-frequency bin $[m', k']$. Through this mask, $\dot{\theta}_{m,k}$ and $\ddot{\theta}_{m,k}$ are characterized from (5) as the unique minimum of the quadratic function:

$$\sum_{m', k'} w_{m,k}[m', k'] \times \left| S_{\dot{g}}[m', k'] - \left(\dot{\theta}_{m,k} S_g[m', k'] + \ddot{\theta}_{m,k} S_m[m', k'] \right) \right|^2, \quad (6)$$

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

By differentiating (6) with respect to $\dot{\theta}_{m,k}$ and $\ddot{\theta}_{m,k}$ and by zeroing the derivatives, we show that parameters $\dot{\theta}_{m,k}$ and $\ddot{\theta}_{m,k}$ satisfy the linear system¹:

$$A_{m,k} \begin{bmatrix} \dot{\theta}_{m,k} \\ \ddot{\theta}_{m,k} \end{bmatrix} = b_{m,k} \quad (7)$$

with

$$A_{m,k} = \sum w_{m,k} \begin{bmatrix} |S_g|^2 & S_g^* S_m \\ S_g S_m^* & |S_m|^2 \end{bmatrix} \quad (8)$$

and

$$b_{m,k} = \sum w_{m,k} \begin{bmatrix} S_g^* S_{\dot{g}} \\ S_m^* S_{\dot{g}} \end{bmatrix}, \quad (9)$$

where $*$ denotes the complex conjugate.

In other respects, from (2) we derive that $\forall t$ in the neighborhood of t_m :

$$(g_k * s)(t) = a_{m,k} e^{j\varphi_{m,k}} \times \sum_n g_k[n] e^{\dot{\theta}_{m,k}(t-t_m - \frac{n}{f_s}) + \frac{1}{2}\ddot{\theta}_{m,k}(t-t_m - \frac{n}{f_s})^2}, \quad (10)$$

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

$$\sum_{m', k'} w_{m,k}[m', k'] \left(\frac{|S_g[m', k']|^2}{a_{m,k}} + a_{m,k} |G_{m,k}[m', k']|^2 \right), \quad (11)$$

where

$$G_{m,k}[m', k'] = e^{\dot{\theta}_{m,k}(t_{m'} - t_m) + \frac{1}{2}\ddot{\theta}_{m,k}(t_{m'} - t_m)^2} \times \sum_n g_{k'}[n] e^{-\frac{n}{f_s}(\dot{\theta}_{m,k} + \ddot{\theta}_{m,k}(t_{m'} - t_m - \frac{n}{2f_s}))}. \quad (12)$$

By minimizing (11), $a_{m,k}$ is obtained as:

$$a_{m,k}^2 = \frac{\sum w_{m,k} |S_g|^2}{\sum w_{m,k} |G_{m,k}|^2}. \quad (13)$$

Besides, the phase $\varphi_{m,k}$ at time-frequency bin $[m, k]$ is estimated by enforcing phase continuity between successive time frames. Since the instantaneous frequency is time-varying, an accurate implementation of phase unwrapping requires to find the predecessor k_p at time $m-1$ of subband k at time m . We thus propose to implement phase unwrapping as follows:

$$\varphi_{m,k} = \varphi_{m-1, k_p} + \dot{\varphi}_{m-1, k_p} \frac{R}{f_s} + \frac{1}{2} \ddot{\varphi}_{m-1, k_p} \left(\frac{R}{f_s} \right)^2 \quad (14)$$

with

$$k_p = \underset{i \in [0, K-1]}{\text{argmin}} \left| \frac{1}{2\pi} \left(\dot{\varphi}_{m,k} - \ddot{\varphi}_{m,k} \frac{R}{f_s} \right) - f_i \right|,$$

where f_i was defined in Section II-A as the center frequency of filter g_i . Therefore, by analyzing $s(t)$ with windows g_k , \dot{g}_k and g'_k , given a neighborhood $\mathcal{V}_{m,k}$ of time-frequency bins around $[m, k]$, we can compute $a_{m,k}$, $\varphi_{m,k}$, $\dot{\theta}_{m,k}$ and $\ddot{\theta}_{m,k}$.

¹In equations (7) to (9), (13) and (31), indexes m' and k' have been omitted for conciseness.

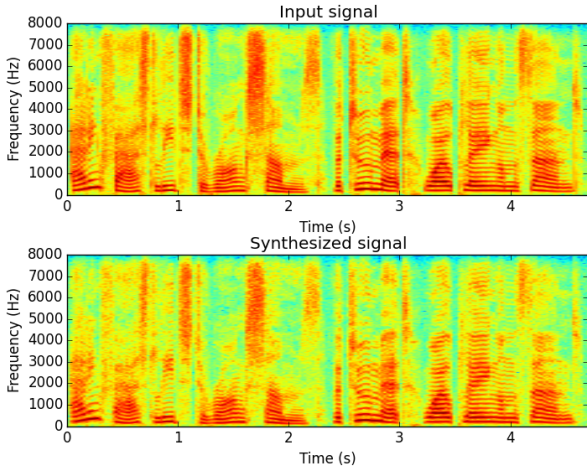


Fig. 1. Magnitude spectrograms of the original and re-synthesized signals

Finally, we estimate the STFT of the anechoic signal from (2):

$$S_g[m, k] = a_{m,k} e^{j\varphi_{m,k}} \sum_n g_k[n] e^{-\frac{n}{T_s} (\hat{\theta}_{m,k} - \hat{\theta}_{m,k} \frac{n}{2T_s})} \quad (15)$$

and we reconstruct the signal $s(t)$ by applying an inverse STFT to (15). As we can see in Figure 1, this framework allows an accurate analysis-synthesis of speech, even in unvoiced sections. Let us see now how to estimate the same parameters from a reverberant signal.

IV. ESTIMATION FROM A REVERBERANT SIGNAL

The goal is to estimate the quadratic terms in (8) and (9) from the reverberant signal $y(t)$ defined in (3), in order to re-synthesize the anechoic signal $s(t)$. To do so, we use the fact that if the RIR $h(t)$ is modeled as in Section II-C, then for any analog signals $x_1(t)$ and $x_2(t)$:

$$\mathbb{E}_b [(h * x_1) \times (h * x_2)] = \sigma^2 p^2 * (x_1 x_2), \quad (16)$$

where \mathbb{E}_b denotes the mathematical expectation *w.r.t.* $b(t)$. This relation can be easily verified by using the fact that $\mathbb{E}_b [b(u)b(v)] = \sigma^2 \delta(u - v)$, 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:

$$\gamma(t) = \frac{1}{\sigma^2} \left(2\alpha \delta(t) + \dot{\delta}(t) \right). \quad (17)$$

Noting that $(g_k * y)(t) = (h * g_k * s)(t)$, (16) leads to:

$$\mathbb{E}_b [|g_k * y|^2] = \sigma^2 p^2 * (|g_k * s|^2). \quad (18)$$

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

$$|g_k * s|^2 = \frac{1}{\sigma^2} \mathbb{E}_b [2\alpha |g_k * y|^2 + 2\Re((g_k * y)^* (\dot{g}_k * y))]. \quad (19)$$

By applying (19) to time t_m , we thus obtain for every time-frequency bin $[m, k]$:

$$|S_g|^2 = \frac{1}{\sigma^2} \mathbb{E}_b [2\alpha |Y_g|^2 + 2\Re(Y_g^* Y_{\dot{g}})]. \quad (20)$$

Likewise, we derive the following expressions:

$$S_g^* S_{\dot{g}} = \frac{1}{\sigma^2} \mathbb{E}_b [2\alpha Y_g^* Y_{\dot{g}} + Y_g^* Y_{\dot{g}} + |Y_{\dot{g}}|^2], \quad (21)$$

$$S_g^* S_{g'} = \frac{1}{\sigma^2} \mathbb{E}_b [2\alpha Y_g^* Y_{g'} + Y_g^* Y_{g'} + Y_g^* Y_{\dot{g}'}], \quad (22)$$

$$|S_{g'}|^2 = \frac{1}{\sigma^2} \mathbb{E}_b [2\alpha |Y_{g'}|^2 + 2\Re(Y_{g'}^* Y_{\dot{g}'})], \quad (23)$$

$$S_{g'}^* S_{\dot{g}} = \frac{1}{\sigma^2} \mathbb{E}_b [2\alpha Y_{g'}^* Y_{\dot{g}} + Y_{g'}^* Y_{\dot{g}} + Y_{g'}^* Y_{\dot{g}'}]. \quad (24)$$

As, in practice, we do not have access to the mathematical expectation, we estimate it with a temporal smoothing by means of a discrete first-order recursive filter, of smoothing parameter $\eta \in [0, 1]$. Hence, for every quadratic term Z in equations (20) to (24), we denote its smoothed version with a bar as:

$$\bar{Z}[m, k] = \eta \bar{Z}[m - 1, k] + (1 - \eta) Z[m, k].$$

We thus estimate $|S_g|^2$, $S_g^* S_{\dot{g}}$, $S_g^* S_{g'}$, $|S_{g'}|^2$ and $S_{g'}^* S_{\dot{g}}$ as:

$$\widehat{|S_g|^2} = \frac{1}{\sigma^2} \left(2\alpha \bar{|Y_g|^2} + 2\Re(\bar{Y_g^* Y_{\dot{g}}}) \right), \quad (25)$$

$$\widehat{S_g^* S_{\dot{g}}} = \frac{1}{\sigma^2} \left(2\alpha \bar{Y_g^* Y_{\dot{g}}} + \bar{Y_g^* Y_{\dot{g}}} + \bar{|Y_{\dot{g}}|^2} \right), \quad (26)$$

$$\widehat{S_g^* S_{g'}} = \frac{1}{\sigma^2} \left(2\alpha \bar{Y_g^* Y_{g'}} + \bar{Y_g^* Y_{g'}} + \bar{Y_g^* Y_{\dot{g}'}} \right), \quad (27)$$

$$\widehat{|S_{g'}|^2} = \frac{1}{\sigma^2} \left(2\alpha \bar{|Y_{g'}|^2} + 2\Re(\bar{Y_{g'}^* Y_{\dot{g}'}}) \right), \quad (28)$$

$$\widehat{S_{g'}^* S_{\dot{g}}} = \frac{1}{\sigma^2} \left(2\alpha \bar{Y_{g'}^* Y_{\dot{g}}} + \bar{Y_{g'}^* Y_{\dot{g}}} + \bar{Y_{g'}^* Y_{\dot{g}'}} \right). \quad (29)$$

From these equations, we can compute matrix $\hat{A}_{m,k}$ and vector $\hat{b}_{m,k}$ as in (8) and (9), in order to estimate $\hat{\theta}_{m,k}$ and $\hat{\theta}_{m,k}$ by following (7):

$$\begin{bmatrix} \hat{\theta}_{m,k} \\ \hat{\theta}_{m,k} \end{bmatrix} = \hat{A}_{m,k}^{-1} \hat{b}_{m,k}. \quad (30)$$

By following (13), the amplitude is then estimated with:

$$\hat{a}_{m,k}^2 = \frac{\sum w_{m,k} |\widehat{S_g|^2}|^2}{\sum w_{m,k} |\widehat{G}_{m,k}|^2}. \quad (31)$$

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

V. PERFORMANCE EVALUATION

Although few time-frequency bins are required in $\mathcal{V}_{m,k}$ to accurately synthesize the signal when it is anechoic, a larger amount is needed in the presence of reverberation. However, expanding the neighborhood when dealing with speech results in cases where there is not only one dominant sinusoid in $\mathcal{V}_{m,k}$, as assumed in Section III. For this reason, the following experimental results focus on monocomponent signals; a solution for multicomponent signals will be exposed in Section VI.

A. Signals and RIRs

In order to evaluate our method, we consider a monocomponent, 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 each frequency, the simulated signal spans the entire frequency range, in 2 seconds. Its spectrogram is plotted in Figure 4-(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 II-C; real RIRs come from the AIR database [14], from which we select regularly spaced RT_{60} s. The spectrogram of a reverberant signal (with $RT_{60} = 2$ s) is plotted in Figure 4-(b).

To assess the performance of our method, we use objective measures from the REVERB challenge toolbox [15]: 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 [16]. We compare our approach with a state-of-the-art suppression method [17], which focuses only on the magnitude of the STFT and ignores the phase information. Our previous work [11] is not included in the benchmark, as it can only deal with linear frequency modulation.

B. 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:

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

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

For each time-frequency bin, the neighborhood $\mathcal{V}_{m,k}$ corresponds to a disc of fixed radius $r = 20$ bins, centered on $[m, k]$. On $\mathcal{V}_{m,k}$, the weights $w_{m,k}[m', k']$ are defined as a function of the distance between $[m, k]$ and $[m', k']$:

$$w_{m,k}[m', k'] = e^{-\frac{1}{2}d_{m,k,r}^2[m', k']},$$

with $d_{m,k,r}[m', k'] = \frac{1}{r}\sqrt{(m - m')^2 + (k - k')^2}$.

C. Results

The scores are plotted in Figures 2 and 3. 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). The gain *w.r.t.* the reverberant signal is more impressive on synthetic RIRs than on real RIRs, but in both cases we achieve a very strong dereverberation.

The spectrogram of a dereverberated signal (with an initial $RT_{60} = 2$ s) is plotted in Figure 4-(c), where we can see that our method completely removes the reverberation. For very

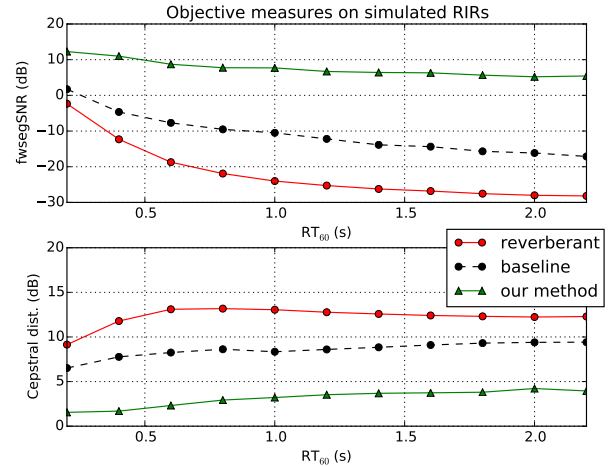


Fig. 2. Objective measures on simulated RIRs

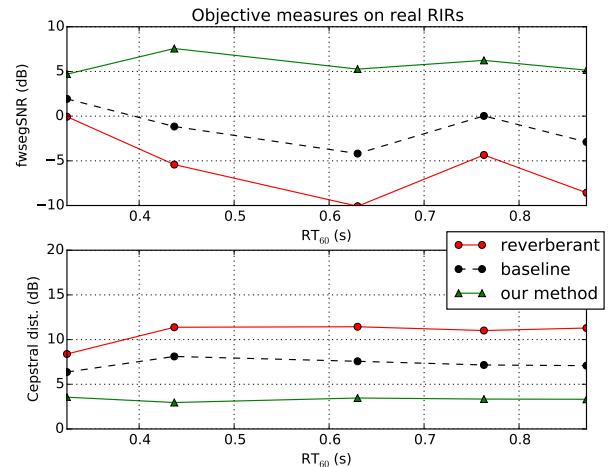


Fig. 3. Objective measures on real RIRs

strong reverberations, some artifacts may appear but can be avoided by choosing a higher value of r . However, a large value of the radius r may not be the optimal solution, as we will see in the next section.

VI. CONCLUSION

Instead of only estimating the STFT magnitude of the anechoic signal to perform dereverberation, we proposed a single-channel method for jointly estimating the magnitude and phase of the STFT. Our estimator almost achieves a perfect reconstruction of monocomponent signals, on both synthetic and real RIRs.

The performance of the method is limited by the size of the neighborhood $\mathcal{V}_{m,k}$, which can be chosen as large as desired for monocomponent signals. However, as soon as we deal with multicomponent signals, we have to constrain $\mathcal{V}_{m,k}$ to include at most one component. For signals such as speech,

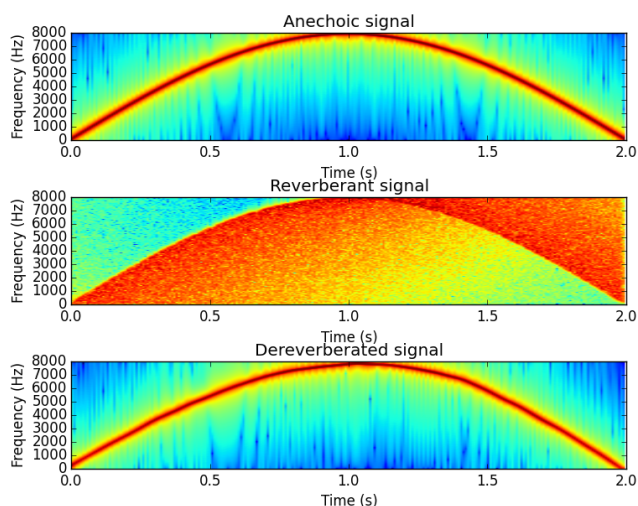


Fig. 4. (a) Anechoic, (b) reverberant and (c) dereverberated signals

it allows only narrow neighborhoods in practice, which is not sufficient for high quality dereverberation.

To overcome this problem, we can enforce a harmonic structure in the signal model by considering $\theta_q(t) = q\theta_1(t)$. Regarding the phase-related terms $\varphi_q(t)$, $\dot{\varphi}_q(t)$ and $\ddot{\varphi}_q(t)$, this assumption holds as speech signals are harmonic. Concerning the amplitude-related terms $\lambda_q(t)$, $\dot{\lambda}_q(t)$ and $\ddot{\lambda}_q(t)$, this assumption is technically necessary but also realistic: in the case of free oscillation, high-frequency harmonics decay faster than low-frequency ones.

This harmonic structure will enable us to replace the local neighborhood $\mathcal{V}_{m,k}$ used in this paper, with a global neighborhood including all the frequency bands, to estimate the amplitude and phase parameters. This higher amount of time-frequency bins will lead to high quality dereverberation. This method will be developed and used to process speech signals in future work.

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] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 2, pp. 145–152, Feb 1988.
- [4] I. Kodrasi, T. Gerkmann, and S. Doclo, "Frequency-domain single-channel inverse filtering for speech dereverberation: Theory and practice," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 5177–5181.
- [5] K. Lebart and J. Boucher, "A new method based on spectral subtraction for speech dereverberation," *ACUSTICA*, vol. 87, no. 3, pp. 359–366, 2001.
- [6] 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.

- [7] N. Mohammadiha and S. Doclo, "Speech dereverberation using non-negative convolutive transfer function and spectro-temporal modeling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 2, pp. 276–289, Feb 2016.
- [8] 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>
- [9] P. Magron, R. Badeau, and B. David, "Complex NMF under phase constraints based on signal modeling: application to audio source separation," in *41st International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Shanghai, China: IEEE, pp. pp. 46–50.
- [10] 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*, Nov 2012, pp. 1–5.
- [11] 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)*, Sept 2016, pp. 1–5.
- [12] J. Polack, "La transmission de l'énergie sonore dans les salles," Ph.D. dissertation, Université du Maine, 1988.
- [13] M. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, no. 3, p. 409, 1965.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] E. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," Ph.D. dissertation, Technische Universiteit Eindhoven, 2007.