Title	Experimental evaluations of TS-BASE/WF in reverberant conditions
Author(s)	Li, Junfeng; Sasaki, Yuuki; Akagi, Masato; Yan, Yonghong
Citation	2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'10): 269-272
Issue Date	2010-03-04
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9970
Rights	This material is posted here with permission of the Research Institute of Signal Processing Japan. Junfeng Li, Yuuki Sasaki, Masato Akagi and Yonghong Yan, 2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'10), 2010, pp.269-272.
Description	

2010 International Workshop on Nonlinear Circuits,
Communication and Signal Processing
NCSP'10, Waikiki, Hawaii, March 3-5, 2010

Experimental evaluations of TS-BASE/WF in reverberant conditions

Junfeng Li†, Yuuki Sasaki†, Masato Akagi† and Yonghong Yan‡

†School of Information Science
Japan Advanced Institute of Science and Technology
1-1, Asahidai, Nomi, Ishikawa, 923-1211, Japan

‡Institute of Acoustics,
Chinese Academy of Science
21, Beisihuan Xilu, Beijing, 100190, China

Abstract

Acoustic signal processing techniques have been widely exploited in modern hearing aids, and played a more important role in dealing with background noise and improving speech quality. We previously proposed a two-stage binaural speech enhancement approach with Wiener filter (TS-BASE/WF), which was proven to be much more effective in suppressing non-stationary multiple-source interfering signals and preserving binaural cues after processing. In addition to background noise, reverberation is another crucial issue in decreasing speech quality under practical conditions. In this paper, we pay main attention to experimental evaluations of the performance of the TS-BASE/WF algorithm in various acoustic conditions with different degrees of reverberation. Experimental results show that the TS-BASE algorithm is able to reduce reverberant components to a limited extent.

1. Introduction

Hearing-impaired listeners have great difficulty in understanding speech in adverse conditions due to the hearing loss and the annoying background noise and reverberation. With the widespread usage of digital hearing aids, advanced signal processing techniques are playing a more important role in modern hearing-aid systems. To deal with these acoustic interferences (background noise and reverberation), a large number of effective speech enhancement techniques have to be integrated into hearing-aid signal processing.

Concerning noise suppression, Nakashima *et al.* presented a frequency domain binaural model (FDBM) for extracting the target signal from the observed binaural signals by exploiting both interaural phase and level differences [1]. More recently, Li *et al.* proposed a two-stage binaural speech enhancement with Wiener filter (TS-BASE/WF) approach based on the psychoacoustic equalization-cancellation theory [2]. Among these algorithms, TS-BASE/WF algorithm has been proven to be much more effective in dealing with non-stationary multiple-source interference and preserving binaural cues after processing. However, no reverberation effect has been considered in its design. Concerning reverberation suppression, Habets *et al.* introduced to remove the late reverberation by spectral subtraction in which the late reverberation is estimated using a statistical model [3]. Gaubitch *et al.* reported a multi-microphone speech dereverberation approach by using the spatial-temporal average technique [4]. More recently, it is noted that Jeub *et al.* collected a binaural impulse response database for evaluating binaural signal processing systems, especially dereverberation techniques [5]. While all these dereverberation algorithms were designed to generate only single-channel output, to our knowledge, only few dereverberation techniques, if available, has been consideration for binaural applications.

The long-term purpose of our research is to design an advanced binaural speech enhancement system for hearing aids, to make hearing aids more robust and intelligent in practical environments, through dealing with both acoustic background noise and reverberation. Towards this purpose, considering the good noise-reduction performance of TS-BASE/WF, in this paper, we investigate its performance in dealing with reverberation effects in various reverberant conditions. In our evaluations, the impulse responses were generated by using the traditional image model [6]. Then, we report the experimental results in terms of objective measures (signal-reverberation ratio and spectral distortion) and subjective measures (waveform and spectrogram). Finally, we give some general remarks on the future research.

2. TS-BASE/WF: Two-stage binaural speech enhancement with Wiener filter

The equalization-cancellation (EC) model was originally developed by Durlach [7] and further improved by Culling and Sumerfiled [8]. Although these EC models could explain many psychoacoustic effects (e.g., BMLD), they function well only in single interference conditions [7] [8]. Inspired by the EC model, the two-stage binaural speech enhancement approach with Wiener filter (TS-BASE/WF) was recently developed [2]. The TS-BASE/WF consists of: (1) interferences estimation by equalizing and cancelling the target signal components, followed by a compensation process; (2) target signal enhancement by a Wiener filter. The block

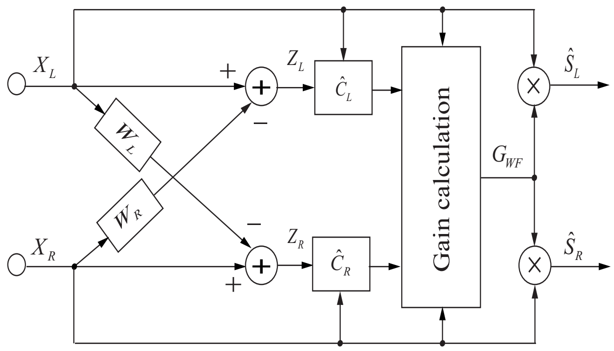

Figure 1: Block diagram of TS-BASE/WF.

diagram of the proposed system is shown in Fig. 1.

2.1. Estimation of interference signal

In binaural applications, HRTFs are normally involved to include the shadowing effects of the head. The cancellation of the target signal is achieved through the equalization and cancellation procedures, yielding the interference-only outputs. It is realized in the following steps [2].

1. In the "equalization" process, two filters are applied to the left and right input signals for equalizing the target signal components in these inputs. Given the binaural inputs, two equalizers can be obtained by using the normalized least mean square (NLMS) algorithm. Based on the assumption that the direction of the target signal is known a priori, two equalizers are pre-learned in the absence of interference signals.

2. In the "cancellation" process, the coefficients of two equalizers are fixed and applied to the observed mixture signals in the presence of interference signals. The target-cancelled signals are derived by subtracting the filter-calibrated inputs at one ear from the input signals at the other ear.

3. In the "compensation" process, a time-variant frequency-dependent compensation factor is exploited to map the target-cancelled signals to the interference components in the input mixture signals. This compensation factor is derived by minimizing the mean square error (MMSE) between the target-cancelled signal and the input mixture signal under the assumption of zero correlation between the target signal and interference signals. Obviously, this zero-correlation assumption cannot be satisfied in reverberant conditions, which further makes this compensation process problematic in reverberant environments.

2.2. Enhancement of target signal

For binaural applications, the system that outputs binaural signals is much preferred. In the proposed TS-BASE/WF, the compensated interference estimates are used to control the gain function of a speech enhancer which is shared in both channels for binaural cue preservation. Specifically, a Wiener filter is used because it is the optimal solution for noise reduction in MMSE sense. Its real gain function contributes to minimize the speech distortion from the frequency-domain filter. The decision-directed adaption mechanism of the a priori SNR helps to reduce the "musical noise" and improve speech quality.

3. Experiments and results

Performance of the TS-BASE/WF algorithm was examined in various acoustic environments with different degrees of reverberation, in terms of objective and subjective measures.

3.1. Data

In experiments, 10 continuous speech sentences uttered by one male speaker were randomly selected from NTT database that has a sampling rate of 44.1 kHz at 16 bit resolution. The room impulse responses (RIRs) with different reverberation times varying from 0.1 s to 1.0 s were simulated using the image method by changing the absorption coefficients of enclosure [6]. These simulated RIRs were first convoluted with the head-related impulse responses (HRIRs) measured at the MIT media lab. [9], which is finally convoluted with the clean speech signals to generate the reverberant binaural signals. The reverberant binaural signals were down-sampled to 16 kHz. In our experiments, the signal source was simulated in the front of the dummy head (i.e., 0 degree). The down-sampled signals that involve reverberation effects were used in our evaluations.

3.2. Objective evaluation measures

The performance of the TS-BASE/WF algorithm was evaluated using the segmental signal-to-reverberation ratio (SEGSRR) [10] and log-spectral distance (LSD) measures [11].

The first, *segmental SRR* (SEGSRR), is used to measure the dereverberation performance in speech enhancement, defined as [10]

$$\mathrm{SEGSRR} = \frac{10}{L}\sum_{\ell=0}^{L-1}\log_{10}\left(\frac{\sum_{k=0}^{K-1}[s_d(\ell K+k)]^2}{\sum_{k=0}^{K-1}[\hat{s}(\ell K+k)-s_d(\ell K+k)]^2}\right), \quad (1)$$

where $s_d(.)$ is the direct-path signal, $\hat{s}(.)$ is the reverberant signal or the enhanced signal processed by the TS-BASE/WF algorithm, and L and K represent the number of frames in the signal and the frame length in samples, respectively. Note that a higher SEGSRR means a higher speech quality of the enhanced signal.

The second measure is *log-spectral distance* (LSD), which is often used to assess the distortion of the processed speech signal, given by [11]

$$\mathrm{LSD} = \frac{10}{L}\sum_{\ell=0}^{L-1}\left(\frac{1}{K}\sum_{k=0}^{K}\left[\log_{10}\mathcal{AS}_d(k,\ell) - \log_{10}\mathcal{A}\hat{S}(k,\ell)\right]^2\right)^{\frac{1}{2}}, \quad (2)$$

where $\mathcal{AS}(k,\ell) \triangleq \max\{|S(k,\ell)|^2, \delta\}$ is the clipped spectral power, such that the log-spectrum dynamic range is confined to about 50 dB (that is, $\delta = 10^{-50/10}\max_{k,\ell}\left\{|S(k,\ell)|^2\right\}$). Note that a lower LSD level indicates less speech distortion.

3.3. Objective evaluation results

The experimental results of SEGSRR and LSD averaged across all sentences in different reverberant conditions are plotted in Fig. 2 and 3.

Fig. 2 shows that the SEGSRRs of the reverberant signals greatly decrease as the reverberation time increases. Specifically, the SEGSRR of the reverberant signal decrease from −5.0 dB in the condition with $\tau_{60} = 0.1$ s to −9.1 dB in the condition with $\tau_{60} = 1.0$ s. Furthermore, the proposed TS-BASE/WF algorithm consistently improves the SEGSRRs in all tested conditions. The improvements in SEGSRR amount to 0.4 dB in the low reverberant condition with $\tau_{60} = 0.1$ s, and 1.0 dB in the high reverberant condition with $\tau_{60} = 1.0$ s. The improved SEGSRR after processing indicates the TS-BASE/WF algorithm is able to improve speech quality in some sense under reverberant conditions. However, the SEGSRRs after TS-BASE/WF's processing is somehow still quite low. The improvements provided by the TS-BASE/WF might attribute to the fact that the TS-BASE/WF algorithm is effective in dealing with the reverberant components coming from the directions other than the direction of the direct-path signal (i.e., 0 degree or the front).

Concerning the results of LSD shown in Fig. 3, it is observed that LSDs of the reverberant signals markedly increases as the reverberation time increases. Moreover, the TS-BASE/WF algorithm consistently reduces LSDs in all tested conditions, especially in high reverberant conditions. The decreased LSDs indicates that the TS-BASE/WF algorithm decreases the speech distortions caused by reverberation effects. However, the LSDs after TS-BASE/WF's processing are still quite high, especially in high reverberant conditions.

3.4. Subjective evaluations

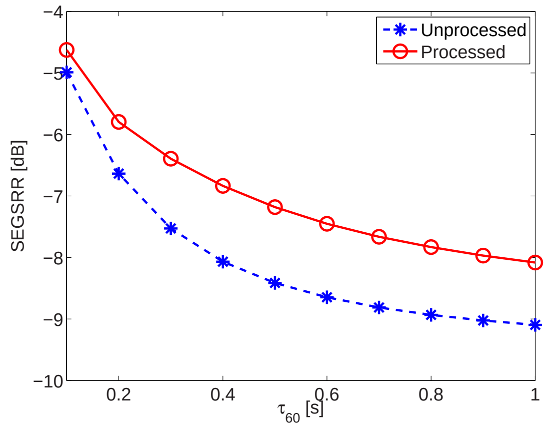

Figure 2: Segmental SRR [dB] of the unprocessed reverberant signal, the processed signal by the TS-BASE/WF algorithm, in different reverberant conditions.

The subjective evaluations of the TS-BASE/WF algorithm were performed using waveforms and spectrograms. Typical examples of waveforms and spectrograms in the reverberant condition with $\tau_{60} = 0.5$ s are plotted in Fig. 4. Fig. 4 (b) shows that the speech signal is greatly smeared by reverberation. Furthermore, as shown in Fig. 4 (c), the TS-BASE/WF algorithm enable remove part of reverberant components from the smeared signal, yielding slightly "brighter" processed signal. However, the processed signal by TS-BASE/WF still involve high degree of reverberation and far from the clean signal. Therefore, it becomes a problem to solve critically and requires immediate attentions.

4. Conclusion

In this paper, we investigated the performance of the TS-BASE/WF algorithm that we previously proposed in dealing with reverberation effects in various reverberant conditions. Experimental results show that the TS-BASE/WF algorithm enable suppress part of reverberant components especially in high reverberant conditions, which is however far from the clean signal that we expected. As a result, it is obviously a future work for us to further evolve the TS-BASE/WF algorithm to make it effective in dealing with reverberant effects in practical conditions.

5. Acknowledgement

The research is partially supported by the China-Japan (NSCF-JSPS) Bilateral Joint Projects, the Grant-in-Aid for Young Scientists (B) (No. 19700156) from the Ministry of Education, Science, Sports and Culture of Japan, and

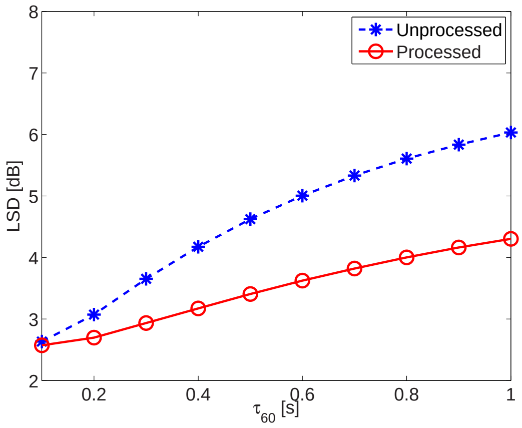

Figure 3: Log-spectral distance (LSD) [dB] of the unprocessed reverberant signal, the processed signal by the TS-BASE/WF algorithm, in different reverberant conditions.

the SCOPE (071705001) of Ministry of Internal Affairs and Communication (MIC), Japan.

References

[1] Nakashima H., Chisaki Y., Usagawa T., Ebata M., (2003). ”Frequency domain binaural model based on interaural phase and level differences,” Acoust. Sci. and Tech., vol. 24, no. 4, pp. 172-178.

[2] Li J., Sakamoto S., Hongo, S., Akagi M., Suzuki Y., (2009), ”Two-stage binaural speech enhancement with Wiener filter based on equalization-cancellation model,” In Proc. WASPAA 2009.

[3] Habets E.A.P., (2004). “Single-channel speech dereverberation based on spectral subtraction,” In Proc. ProRISC, pp. 250-254.

[4] Gaubitch N.D., Naylor P.A. and Ward D. B., (2004). “Mulitmicrophone speech dereverberation using spatio-temporal averaging, in Proc. EUSIPCO, pp. 809812.

[5] Jeub, M., Schfer, M., and Vary, P., (2009).“A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms,” in Proc. DSP, pp. 14.

[6] Allen J.B., Berkley D.A. (1979). “Image method for efficiently simulating small-room acoustics.” JASA, vol. 65, pp. 943-950.

[7] Durlach N.I. (1963). ”Equalization and cancellation theory of binaural masking level differences,” JASA, vol. 35, no. 8, pp. 1206-1218.

[8] Culling J.F. and Summerfield Q. (1995). ”Perceptual segregation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay,” JASA, vol. 98, pp. 785-797.

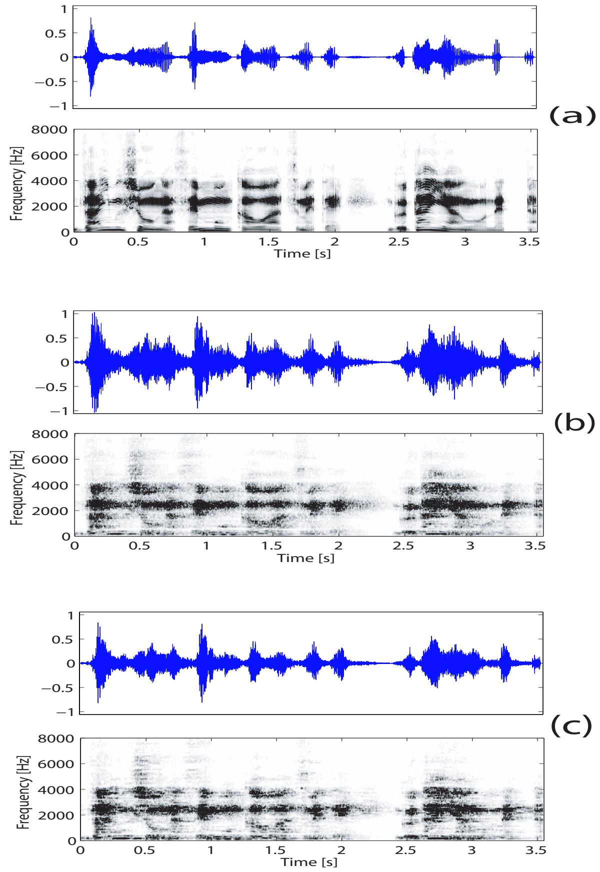

Figure 4: Speech waveforms and spectrograms. (a) Clean signal; (b) Reverberant signal ($\tau_{60} = 0.5$ s); (c) Processed signal with the TS-BASE/WF algorithm.

[9] http://sound.media.mit.edu/resources/KEMAR.html.

[10] Naylor P.A., and Gaubitch N.D. (2005). “Speech dereverberation,” In Proc. IWAENC, p03.

[11] Cohen I. (2004). “Multi-channel post-filtering in nonstationary noise environments,” IEEE Trans. Signal Processing, vol. 52, no. 5, pp. 1149-1160.