A NOVEL DECORRELATION APPROACH FOR MULTICHANNEL SYSTEM IDENTIFICATION

Laura Romoli, Stefania Cecchi and Francesco Piazza

A3Lab - DII - Università Politecnica delle Marche
Via Brecce Bianche, 60131 Ancona Italy
l.romoli@univpm.it

ABSTRACT

Multichannel sound reproduction systems aim at providing an optimal acoustical sensation to the listener by enhancing the listening experience and improving the spatial sound impression of the virtual scene. In this context, many audio applications requiring adaptive processing schemes have been developed in order to improve the performance of the audio reproduction. However, online multichannel systems identification requires the introduction of suitable solutions for overcoming the well-known non-uniqueness problem related to the correlation existing among the loudspeaker signals. This issue has been deeply investigated considering stereophonic systems but its extension to multichannel systems may not be so straightforward. In this paper, a novel solution for speech and audio signals is presented based on the "missing-fundamental" phenomenon. Experimental results proved the effectiveness of the approach also making comparisons with the existing state of the art.

***Index Terms*—** Channel decorrelation, missing-fundamental phenomenon, multichannel system identification

1. INTRODUCTION

The main objective of a multichannel sound reproduction system is to give an optimal acoustical sensation to the listener by enhancing the listening experience and improving the spatial sound impression of the virtual scene [1, 2, 3]. In this context, multichannel systems identification has become essential for several audio applications requiring adaptive processing schemes (e.g., acoustic echo cancellation [4, 5, 6] and room equalization [7, 8]). Unfortunately, when a linear relationship exists among the loudspeaker signals, the correlation matrix results ill-conditioned and there is not a unique solution of the normal equation [4]. This is commonly known as the "non-uniqueness problem". Indeed, when the modeling filters have the same length as the actual impulse responses (IRs), the normal equation of the adaptive algorithm has no unique solution. Actually, in the real scenario, the infinite length of the IRs ensures a unique solution, but the covariance matrix is very ill-conditioned [4]. Therefore, a method to reduce interchannel coherence must be introduced in order to obtain good performance in the system identification.

Several efforts have been done in the field of acoustic echo cancellation mainly focusing on stereophonic signals as summarized in [9]. These approaches can be divided into two main categories: methods based on the direct alteration of the stereo signal (e.g., half-wave rectifier [10], time-varying all-pass filters [11][12], phase modulation [13], frequency shifting [14], "missing-fundamental" theory [4, 9]) and other techniques based on the introduction of an external signal to both channels, exploiting psychoacoustic phenomena (e.g., introduction of masked noise considering the human auditory system [15][16]).

The application of these solutions to multichannel reproduction systems may not be much straightforward. Some of the cited approaches have been generalized to the multichannel scenario, e.g, the phase modulation approach [13] and the half-wave rectifier distortion [10]. The former is designed for surround sound systems like 5.1 while the latter has a high impact on the audio quality of the input signal. The goal of this work is that of providing a novel solution for multichannel reproduction systems that results suitable for speech and teleconferencing applications and also for high quality audio systems. Starting from the results obtained for stereophonic signals [4, 5, 6, 9, 17], a novel approach for multichannel audio systems decorrelation is presented. The approach is still based on the missing fundamental theory but novel aspects are introduced for dealing with more than two channels. In particular, taking into consideration the techniques previously proposed by the same authors for stereophonic signals, the missing-fundamental based approach has been combined with the introduction of time-varying allpass filters since they result suitable for a multichannel extension.

The paper is organized as follows. The novel solution for multichannel decorrelation is presented in Section 2 discussing the main original issues to deal with a multichannel scenario. Then, the validation by means of several simulations is reported in Section 3 making also comparisons with another well known approach existing in the literature. Finally, conclusions are drawn in Section 4.

2. PROPOSED APPROACH

A novel channel decorrelation solution for multichannel reproduction systems is discussed starting from the results reported in our previous works focused on stereophonic acoustic echo cancellation [4, 9]. Psychoacoustic criteria are taken into consideration in order to avoid annoying audible distortion. In particular, the missing-fundamental based approach has been combined with the introduction of time-varying allpass filters as described in the following.

A typical scenario of multichannel system identification is taken into consideration with L loudspeakers and M microphones. For the sake of clarity, the scenario is reported in Fig. 1 with only one of the M microphones since the proposed approach can be discussed independently for each microphone. Each input channel $x_l(n)$, being $l = 1, \cdots, L$, is divided into two subbands using the low-pass filter $H_{\mathrm{lp}}(z)$ and the high-pass filter $H_{\mathrm{hp}}(z)$, respectively. Then, an adaptive notch filter $H_l(z, n)$ is applied in the low-frequency range for estimating and removing the fundamental frequency while in the high-frequency range the signal phase is altered through a second-order time-varying allpass filter $F_l(z, n)$ varying with the estimated

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

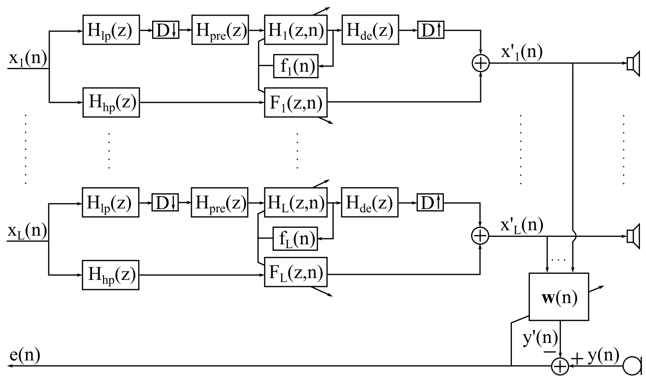


Fig. 1. Multichannel system identification with decorrelation.

fundamental frequency $f_l(n)$. In this way, each processed channel $x'_l(n)$ is obtained and applied to the multichannel adaptive system represented by the matrix $\mathbf{w}(n) = [\mathbf{w}_1(n) \cdots \mathbf{w}_L(n)]$ composed of $L \cdot N$ filters coefficients. The error signal at each time instant n between the output of the multichannel adaptive filter $y'(n)$ and the desired output signal $y(n)$ is then computed as follows [18]:

$$e(n) = y(n) - y'(n) = y(n) - \sum_{l=0}^{L} \sum_{k=0}^{N} x_l(n-k) w_l(k). \quad (1)$$

Finally, the filters coefficients are updated by minimizing the power of the error signal in Eq. (1). Considering the techniques previously proposed by the same authors for stereophonic signals, the missing-fundamental based approach and the introduction of time-varying allpass filters have been chosen since their extension to the multichannel scenario is quite straightforward. In the following, the two techniques applied in the low and high-frequency range are described in details.

2.1. Missing-fundamental approach

In the low-frequency band, the "missing-fundamental" phenomenon is exploited for lowering the interchannel coherence [4]. This is a well known psychoacoustic phenomenon related to the perception of pitch (i.e., the fundamental frequency) without the corresponding frequency actually being contained in the signal. For a complex tone consisting of different frequency harmonics, the pitch is related to the fundamental frequency presence. When it is removed from the set of harmonics, the perceived sound is almost unchanged: the pitch perception does not change but there is a slight alteration of the sound timbre due to the number of reproduced harmonics. This phenomenon has been explained as a human brain capability of processing the information present in the overtones to calculate the "missing-fundamental" [19]. This effect is usually employed to enhance low-frequency sound reproduction, especially in the case of small loudspeakers [20] and inexpensive audio systems, whereby a set of harmonics above the loudspeaker cut-off frequency is created in order to increase the bass experience. In our approach, the "missing-fundamental" effect is exploited to reduce the coherence among the L input channels. While in the two-channel scenario the adaptive notch filter was applied only on one channel of the stereo signal, in the multichannel scenario the fundamental frequency is estimated and removed from all the L input channels applying L second-order lattice form notch filters. The l-th notch filter is described by the following transfer function:

$$H_l(z,n) = \frac{1 + 2k_l(n)z^{-1} + z^{-2}}{1 + k_l(n)[1 + \alpha_l(n)]z^{-1} + \alpha_l(n)z^{-2}}, \quad (2)$$

being $l = 1, \cdots, L$ the channel index. This function is described by the adaptive coefficient $k_l(n)$, related to the tracked frequency $f_l(n)$, and the pole-zero contraction factor $\alpha_l(n)$ controlling the bandwidth of the filter [21]. Besides removing the fundamental frequency, the notch filter can change its cut-off frequency at each new sample to track the time-varying fundamental frequency. In this way, decorrelation is provided in the whole low-frequency band acting as happens with time-varying all-pass filters [11].

The notch filter related to each channel is charcterized by a different value of the contraction factor $\alpha_l(n)$ that performs a slow variation. In this way, the disparity among channels is guaranteed also when the channels are characterized by the same fundamental frequency. In particular, the vector $\boldsymbol{\alpha}_l(n) = [\alpha_1(n), \cdots, \alpha_L(n)]$ is updated as follows:

$$\boldsymbol{\alpha}_l(n) = \begin{cases} s[\boldsymbol{\alpha}_l(n-1), 1] & \text{if } \left(\mathrm{n} - \mathrm{Q}\left\lfloor \frac{\mathrm{n}}{\mathrm{Q}} \right\rfloor\right) = 0 \\ \boldsymbol{\alpha}_l(n-1) & \text{otherwise,} \end{cases} \quad (3)$$

where $s(\cdot, 1)$ is a right circular shift of one sample performed every Q samples, being Q the block length. Assuming the scenario with $L = 3$ channels, Fig. 2 shows the magnitude response of the notch filter of each channel with the vector $\boldsymbol{\alpha}(0) = [0.95, 0.75, 0.55]$ at time instant 0 and the same notch at 200 Hz proving that different bandwidths and depths of the notch can be obtained with different values α_l.

The coefficient $k_l(n)$ is bounded in the range $(-1, 1)$ to prevent the filter from diverging using the following sigmoid function:

$$k_l(n) = \frac{2}{1 + e^{-g_l(n)}} - 1, \quad (4)$$

where $g_l(n) \in R$ is the parameter to be minimized in the cost function computed as fully described in [9]. The estimated frequency $f_l(n)$ can be computed at every step from the knowledge of $k_l(n)$ since

$$f_l(n) = \frac{f_s}{D} \cdot \frac{1}{2\pi} \cos^{-1}[-k_l(n)], \quad (5)$$

being f_s the sampling frequency and D the down-sampling factor.

In order to improve the estimation and tracking performance of the approach, pre-emphasis and de-emphasis stages are introduced. Pre-emphasis is typically used in the field of automatic speech recognition for boosting the amount of energy in the high-frequency band where the glottal pulse produces less energy than

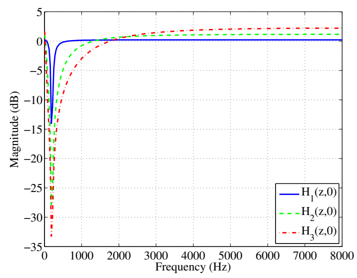


Fig. 2. Magnitude response of three notch filters considering $\boldsymbol{\alpha}(0) = [0.95, 0.75, 0.55]$ and the same notch at 200 Hz.

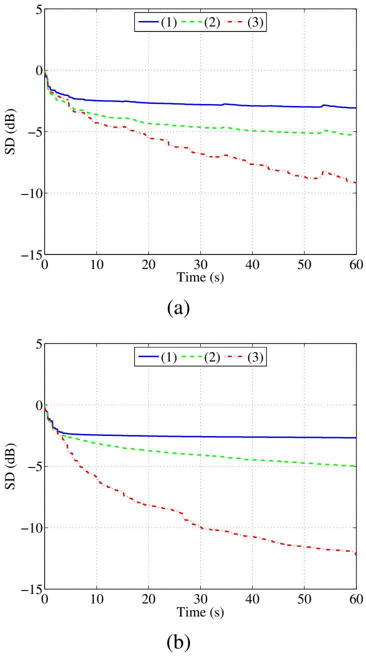


Fig. 3. SD obtained (1) without decorrelation, (2) with the half-wave rectifier techinque [10], and (3) with the proposed approach. (a) Speech signal. (b) Music signal.

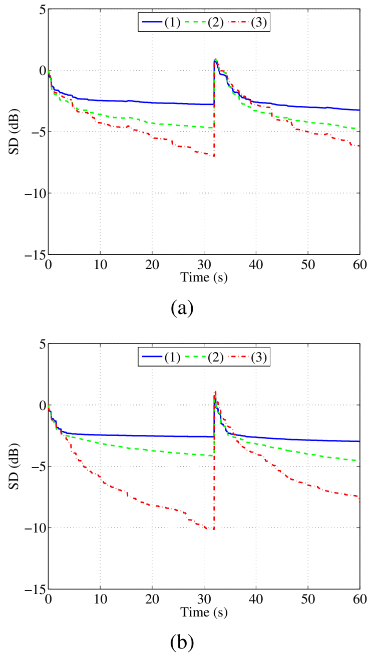


Fig. 4. SD obtained (1) without decorrelation, (2) with the half-wave rectifier techinque [10], and (3) with the proposed approach considering an abrupt change. (a) Speech signal. (b) Music signal.

in the low-frequency band, especially for voiced segments [22, 23]. Differently, in this approach, pre-emphasis is used to emphasize the low-frequency band for better tracking the fundamental frequency and its effect is undo by de-emphasis. Thus, they are characterized by the following transfer functions:

$$H_{\text{pre}}(z) = \frac{1}{1-\nu z^{-1}} \tag{6}$$

$$H_{\text{de}}(z) = 1-\nu z^{-1}, \tag{7}$$

where $0 < \nu < 1$. The pre-emphasis stage offers the advantage of improving the tracking performance, especially when the low-frequency range includes also some harmonics.

2.2. Second-order time-varying all-pass filtering

Regarding the high-frequency range, L second-order time-varying all-pass filters are applied in order to alter the phase of the input channels (Fig. 1). The time variation of the all-pass filters has to be chosen in such a way that does not alter the spatial perception of the speech [11]. Thus, allpass filters characterized by a pole with multiplicity 2 related to the adaptive coefficient $k_l(n)$ of Eq. (4) have been considered. The l-th allpass filter has the following transfer function [24]:

$$F_l(z,n) = \frac{k_l^2(n) - 2k_l(n)z^{-1} + z^{-2}}{1-2k_l(n)z^{-1}+k_l^2(n)z^{-2}}, \tag{8}$$

where the restriction $|k_l(n)| < 1$ described in Section 2.1 ensures the causality and stability of the filter. Furthermore, the restriction derived from the known "just noticeable inter-aural delay" [25] is also satisfied. More specifically, this delay represents the minimum change in the inter-aural time delay between the two ears that causes a noticeable change in the perception of the direction of sound. The introduced delay must vary between $30\,\mu$s and $200\,\mu$s [25]. As demonstrated in [11, 26], considering the group delay of Eq. (8) and $|k_l(n)| < 1$, the maximum change in the time of arrival of each frequency results limited to within about $40\,\mu$s.

3. ALGORITHM VALIDATION

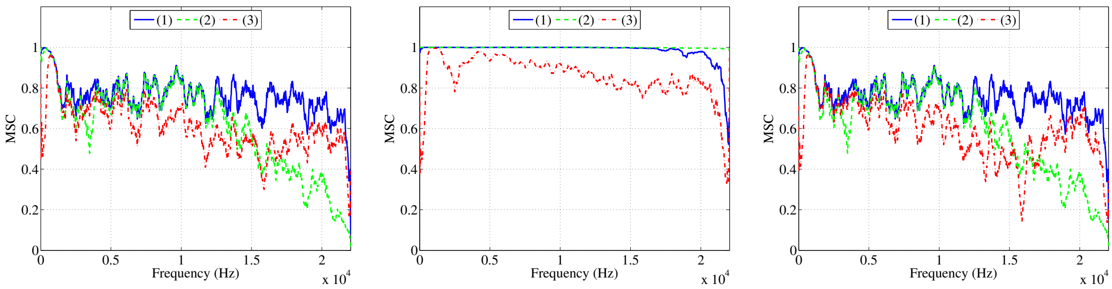
In this section, the effectiveness of the proposed approach in improving the performance achievable for online system identification is presented showing how the interchannel coherence affects the correct estimation of the actual IRs. Tests have been carried out considering the online system identification without any decorrelation algorithm as the reference scenario and comparing the obtained improvement also taking into consideration a well-known approach still existing in the literature, i.e., the half-wave rectifier nonlinearity with nonlinear parameter $\gamma = 0.5$ [10].

Results are presented in terms of the spectral distance $SD_m(i)$ at microphone m between the spectral representations $G_l(k,i)$ and $W_l(k,i)$ of the actual and estimated IRs, respectively, as follows:

$$SD_m(i) = \frac{\sum_{l=0}^{L}\sum_{k=0}^{K}|G_l(k,i) - W_l(k,i)|^2}{\sum_{l=0}^{L}\sum_{k=0}^{K}|G_l(k,i)|^2}, \tag{9}$$

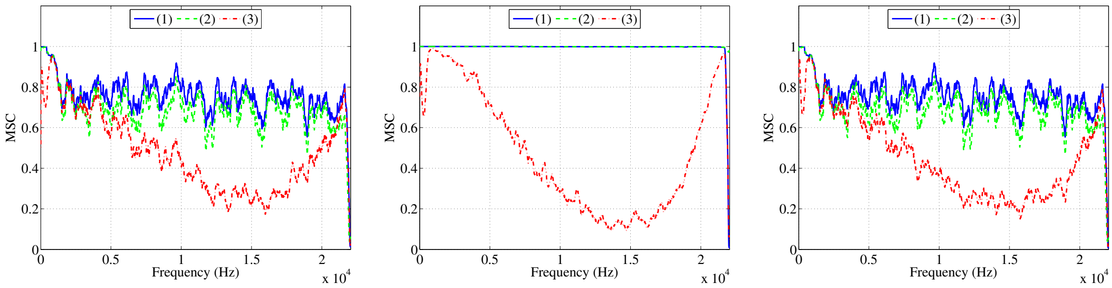
being i the frame index, $k = 1,\cdots,K$ the discrete frequency, and l the loudspeaker index. Moreover, the influence of the correlation among the loudspeaker signals on the convergence performance is underlined showing the interchannel coherence in terms of magnitude-squared coherence (MSC). MSC is described by the following equation as a function of the discrete frequency k:

$$\text{MSC}(k) = \frac{\left|P_{x_j x_h}(k)\right|^2}{P_{x_j x_j}(k)P_{x_h x_h}(k)}, \tag{10}$$



(a) (b) (c)

Fig. 5. MSC obtained considering a speech signal (1) without decorrelation, (2) with the half-wave rectifier techinque [10], and (3) with the proposed approach. (a) Channel 1 vs channel 2. (b) Channel 1 vs channel 3. (c) Channel 2 vs channel 3.



(a) (b) (c)

Fig. 6. MSC obtained considering a music signal (1) without decorrelation, (2) with the half-wave rectifier techinque [10], and (3) with the proposed approach. (a) Channel 1 vs channel 2. (b) Channel 1 vs channel 3. (c) Channel 2 vs channel 3.

where $P_{x_j x_j}$ and $P_{x_j x_h}$ are the auto power spectra and cross power spectra between channels j and h ($j \neq h$), respectively.

All the presented tests have been done considering both a speech signal (SQAM disk [27]) and a music signal sampled at 44.1 kHz and IRs of length 2048 samples simulated using the image method [28]. The room has fixed dimensions of $3\,\text{m}\times 2\,\text{m}\times 2.5\,\text{m}$ with a reverberation time of about 0.2 s. Adaptive filters of length 2048 samples have been used considering a fixed step size $\mu = 0.025$ for system identification using a frequency-domain adaptive filtering algorithm as described in [7, 29]. Two test sessions were carried out assuming the same scenario with $M = 3$ microphones and $L = 3$ loudspeakers but considering a change in the position of the microphones at $t \approx 32$ s in the second test in order to evaluate the response of the algorithm to an abrupt change.

Regarding the implementation of the proposed method, a decimation factor of 32 is applied to the low-frequency channel to increase spectral resolution and $\nu = 0.5$ is used for the pre/de-emphasis factor. The vector $\boldsymbol{\alpha}$ has been defined considering uniformly distributed values in the range from 0.55 to 0.95 in order to provide a time-varying bandwidth and depth of the notch filter and thus to obtain the disparity among channels. Therefore, the vector $\boldsymbol{\alpha} = [0.95, 0.75, 0.55]$ has been defined applying a frame-by-frame shifting every 0.1 s. All these parameters have been fixed to minimize the average SD but optimizing the trade-off between channel decorrelation and audio quality preservation. Informal listening tests were carried out to obtain preliminary subjective results.

Fig. 3 reports the obtained SD averaged over the 3 microphones for a speech and a music signal, respectively. It is evident that the proposed approach improves both the accuracy and speed of the convergence towards the actual IRs with respect to the nonlinear preprocessing discussed in [10]. Analogous comments can be provided considering Fig. 4 where it is evident that the proposed approach shows a good response to the abrupt change still improving both the accuracy and the speed of the convergence in estimating the actual IRs. Moreover, considering the results obtained with the reference scenario without decorrelation (Figs. 3 and 4), it is evident that the correlation among channels affects the performance as also confirmed by Figs. 5 and 6 where the interchannel coherence among channels is reported. Indeed, the better convergence performance provided by the proposed approach corresponds to lower values of the MSC.

4. CONCLUSION

A novel solution for multichannel reproduction systems that results suitable for both speech and audio signals has been presented in this paper. Starting from previous work by the same authors especially focused on stereophonic acoustic echo cancellation and introducing novel aspects to deal with more than two channels, a novel decorrelation approach for preprocessing the input signals to multichannel adaptive filtering algorithm has been discussed based on the "missing-fundamental" phenomenon and second-order time-varying allpass filters. Experimental results show that the approach results effective in the improvement of adaptive filtering algorithm performance for system identification also making comparison with the half-wave rectifier nonlinearity still proposed in the literature. Future work will be oriented to the introduction of the proposed method for improving performance of multichannel audio reproduction systems.

5. REFERENCES

[1] P. Peretti, S. Cecchi, L. Romoli, and F. Piazza, "Performance Evaluation of Adaptive Algorithms for Wave Field Analysis/Synthesis using Sound Field Simulations," in *Computational Simulations and Applications*, 2011.

[2] P. Thüne and G. Enzner, "Assessment of Multichannel Acoustic System Identification Using a Spectral-Importance Weighted Misalignment," in *Proc. Workshop on Acoustic Echo and Noise Control*, Aachen, Germany, Sep. 2012.

[3] ——, "Trends in Adaptive MISO System Identification for Multichannel Audio Reproduction and Speech Communication," in *Proc. 8th Int'l Symposium on Image and Signal Processing and Analysis*, Trieste, Italy, Sep. 2013.

[4] L. Romoli, S. Cecchi, and F. Piazza, "A Combined Approach for Channel Decorrelation in Stereo Acoustic Echo Cancellation Exploiting Time-Varying Frequency Shifting," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 717–720, Jul. 2013, DOI = 10.1109/LSP.2013.2264936.

[5] L. Romoli, S. Cecchi, P. Peretti, and F. Piazza, "A Mixed Decorrelation Approach for Stereo Acoustic Echo Cancellation based on the estimation of the fundamental frequency," *IEEE Trans. Audio, Speech and Language Processing*, vol. 20, no. 2, pp. 690–698, Feb. 2012.

[6] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "A Combined Psychoacoustic Approach for Stereo Acoustic Echo Cancellation," *IEEE Trans. Audio, Speech and Language Processing*, vol. 19, no. 6, pp. 1530–1539, Nov. 2011.

[7] S. Cecchi, L. Romoli, F. Piazza, and A. Carini, "A Multichannel and Multiple Position Adaptive Room Response Equalizer in Warped Domain," in *Proc. 8th Int'l Symposium on Image and Signal Processing and Analysis*, Trieste, Italy, Sep. 2013.

[8] L. Fuster, M. de Diego, M. Ferrer, A. Gonzalez, and G. Pinero, "A biased multichannel adaptive algorithm for room equalization," in *Proc. 20th European Signal Processing Conference*, Bucharest, Romania, Aug. 2012, pp. 1344–1348.

[9] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "Low-complexity implementation of a real-time decorrelation algorithm for stereophonic acoustic echo cancellation," *Signal Processing*, vol. 92, no. 11, pp. 2668–2675, Nov. 2012.

[10] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 156–165, Mar. 1998.

[11] M. Ali, "Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-pass Filtering for Signal Decorrelation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, Seattle, WA, USA, May 1998, pp. 3689–3692.

[12] A. Hirano, K. Nakayama, and K. Takebe, "Stereophonic Acoustic Echo Canceler with Pre-Processing - Second-Order Pre-Processing Filter and its Convergence," in *Proc. Workshop on Acoustic Echo and Noise Control*, Kyoto, Japan, Sep. 2003, pp. 63–66.

[13] J. Herre, H. Buchner, and W. Kellermann, "Acoustic Echo Cancellation for Surround Sound using Perceptually Motivated Convergence Enhancement," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, Honolulu, HI, USA, Apr. 2007, pp. 17–20.

[14] B. C. Bispo and D. D. S. Freitas, "Hybrid pre-processor based on frequency shifting for stereophonic acoustic echo cancellation," in *Proc. 20th European Signal Processing Conference*, Bucharest, Romania, Aug. 2012, pp. 2447 – 2451.

[15] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem," *IEEE Signal Process. Lett.*, vol. 2, pp. 148–151, Aug. 1995.

[16] A. Gilloire and V. Turbin, "Using Auditory Properties to Improve the Behaviour of Stereophonic Acoustic Echo Cancellers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, Seattle, WA, USA, May 1998, pp. 3681–3684.

[17] L. Romoli, S. Cecchi, L. Palestini, P. Peretti, and F. Piazza, "A Novel Approach to Channel Decorrelation for Stereo Acoustic Echo Cancellation based on Missing Fundamental Theory," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA, Mar. 2010, pp. 329–332.

[18] J. Benesty and Y. Huang, *Adaptive Signal Processing: Application to Real-World Problems*. Springer, 2003.

[19] E. Larsen and R. M. Aarts, *Audio Bandwidth Extension*. J. Wiley & Sons, 2004.

[20] S. Cecchi, E. Moretti, and F. Piazza, "A New Approach to Bass Enhancement Based on Prony's Method," in *Proc. 15th International Conference on Digital Signal Processing*, Cardiff, UK, Jul. 2007, pp. 535–538.

[21] J. Lee, E. Song, Y. Park, and D. Youn, "Effective Bass Enhancement Using Second-Order Adaptive Notch Filter," *IEEE Trans. Consum. Electron.*, vol. 54, pp. 663–668, May 2008.

[22] A. M. Peinado and J. C. Segura, *Speech Recognition over digital channels: Robustness and Standards*. J. Wiley & Sons, 2006, pp. 10-11.

[23] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Halls, 2009, pp. 330.

[24] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. Prentice Hall International Inc., 1999, pp. 274-279.

[25] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Springer-Verlag, 1990.

[26] N. Tangsangiumvisai, J. A. Chambers, and A. G. Constantinides, "Higher-order time-varying allpass filters for signal decorrelation in stereophonic acoustic echo cancellation," *Electronics Lett.*, vol. 35, no. 1, pp. 88–90, Jan. 1999.

[27] "Sound Quality Assessment Material, recordings for subjective tests," Technical Review 3253-E, European Broadcasting Union, Apr. 1988.

[28] J. B. Allen and D. A. Berkeley, "Image method for efficiently simulating small - room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.

[29] L. Romoli, S. Squartini, and F. Piazza, "A Variable Step-Size Frequency-Domain Adaptive Filtering Algorithm for Stereophonic Acoustic Echo Cancellation," in *Proc. 18th European Signal Processing Conference*, Aalborg, Denmark, Aug. 2010, pp. 26–30.