This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Time-varying allpass filters using spectral-shaped noise for signal decorrelation in stereophonic acoustic echo cancellation

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© IEEE

VERSION

VoR (Version of Record)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Tangsangiumvisai, N., Jonathon Chambers, and A.G. Constantinides. 2019. "Time-varying Allpass Filters Using Spectral-shaped Noise for Signal Decorrelation in Stereophonic Acoustic Echo Cancellation". figshare. <https://hdl.handle.net/2134/5767>.

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

TIME-VARYING ALLPASS FILTERS USING SPECTRAL-SHAPED NOISE FOR SIGNAL DECORRELATION IN STEREOPHONIC ACOUSTIC ECHO CANCELLATION

N. Tangsangiumvisai¹, J. A. Chambers², and A. G. Constantinides³

¹Department of Electrical Engineering, Chulalongkorn University, Bangkok, 10330, Thailand.

²Centre for Digital Signal Processing Research, King's College, London, WC2R 2LS, U.K.

³Department of Electrical and Electronic Engineering, Imperial College, London, SW7 2BT, U.K.

E-mail : ntang@ee.eng.chula.ac.th, jonathon.chambers@kcl.ac.uk, a.constantinides@ic.ac.uk

Abstract

In this paper, a signal decorrelation technique is presented that aims to minimise the degradation on stereo signal perception, whilst alleviating the non-uniqueness problem in Stereophonic Acoustic Echo Cancellation (SAEC). The proposed technique is based on higher-order time-varying allpass filters (HO-TV-APFs) utilising spectral-shaped noise, together with the interpolation method. The test results on real speech signals indicate improved performance in both objective and subjective manners.

1 Introduction

The *non-uniqueness problem* of Stereophonic Acoustic Echo Cancellation (SAEC) in a stereo teleconferencing system can be tackled by trying to decorrelate artificially the stereo signals, provided that their quality is minimally affected. A pre-processing approach, employing higher-order time-varying allpass filters (HO-TV-APFs), is proposed in [1] to decorrelate the stereo signals while maintaining stereo perception. This approach is underpinned by the constant magnitude response of allpass filters, and in particular, the *near-to-linear phase response* characteristics of HO-TV-APFs. Therefore, the stereo perception of the signals is minimally degraded.

The effect of employing HO-TV-APFs is empirically shown to be related to the introduction of a random noise signal to each channel of the system. This is equivalent to the signal decorrelation approach in [2], which uses the addition of independent random noise directly to each of the stereo input signals. However, the sound quality of the speech input signals is degraded unless the added noise level is at least 13 – 15dB below each of the speech signals [3]. On the other hand, if insufficient noise is added, the signals cannot be decorrelated adequately. In contrast to the random noise addition approach, the level of random noise added to the stereo signals in each channel, when employing the HO-TV-APFs technique, can be adjusted by the variation of the HO-TV-APFs parameters so that sufficient signal decorrelation can be obtained and the stereo perception is preserved.

Since the dominant cues of the speech signals are

located in the frequency range below 1kHz [4], the noise introduced by the HO-TV-APFs is therefore suggested to be minimum where such baseband signal frequencies are present, so as not to degrade the sound quality. In addition, the power spectral density of the random noise that is introduced to the system by the HO-TV-APFs method should be emphasised at frequencies far beyond 1kHz, the range to which human ears are less sensitive, in order to decorrelate the stereo signals more efficiently. This can be achieved by exploiting the first-order differentiator, which shapes the power spectral density of its input signal emphasising the high frequencies, provided that the stereo signals are interpolated before employing the pre-processing so that the noise signal in the high-frequency region can be removed by the use of the lowpass filter and the decimation process. The proposed technique should result in improvement in signal decorrelation capability and preservation of the quality of stereo signals after employing the pre-processing approach.

This paper is divided into five sections as follows. Section II describes the HO-TV-APFs signal decorrelation method. In Section III, an approach of filtered noise, namely spectral-shaped noise, to be used with the HO-TV-APFs is given. Section IV includes the results when applying the proposed signal decorrelation technique to the SAEC system. Finally, the conclusions are given in Section V.

2 Higher-order time-varying allpass filters for signal decorrelation

2.1 Structure of HO-TV-APFs

Figure 1: A schematic diagram of SAEC

The transmitted stereo signals $x_i(n)$, for $i = 1, 2$, are passed through two TV-APFs $\mathbf{a}_i(n)$ of any order, as illustrated in Fig. 1 to decorrelate the stereo signals. Higher-order time-varying allpass filters can approximate linearity in the phase response and hence, a desirable approximate *constant group delay* can be obtained. The higher the order of allpass filters, the better the constancy of the group delay. This is achieved by representing the allpass filter in the form of

$$\mathbf{A}_i(z, n) = \frac{\sum_{l=1}^N \alpha_{i,l}(n) z^{-(N-l)} + z^{-N}}{1 + \sum_{l=1}^N \alpha_{i,l}(n) z^{-l}} \quad (1)$$

where N is the order of the allpass filters, $l = 1, \dots, N$ and $\alpha_{i,l}(n)$ are the coefficients of the allpass filter in each channel, for $i = 1, 2$, that vary with time. The update equation of the HO-TV-APFs parameter is given by

$$\alpha_{i,l}(n) = \alpha_{i,l,int} + r_{i,l}(n), \quad i = 1, 2, \quad l = 1, \dots, N \quad (2)$$

This means that each TV-APF parameter $\alpha_{i,l}(n)$ deviates around its initial value. From observations, it is noteworthy that, for any order N , the time-varying parameter $\alpha_{i,N}(n)$ has the highest impact on the deviation of the inter-aural delay [5]; whereas the other $\alpha_{i,l}(n)$ parameters, for $l = 1, \dots, N - 1$, do not drastically affect the inter-aural delay, particularly in the low-frequency range.

2.2 Computational complexity

The use of HO-TV-APFs does not drastically increase the system computational complexity. In terms of the number of real multiplications (RMPs) per input sample, the TV-APFs operation requires $2N + 1$ RMPs per input sample, where N is the order of the TV-APFs. This is based on canonical form of the realisation of the system.

3 On improvement of the signal decorrelation method

3.1 The modified decorrelated SAEC system

A modified decorrelated SAEC system is thus introduced by the use of the interpolation process, together with the HO-TV-APFs decorrelation method, in order to decorrelate the stereo signals efficiently and to preserve stereo perception. The random noise that is used in the update equation of each HO-TV-APF parameter is *spectral-shaped*, based upon the first-order differentiator, so that its power level is reduced at low frequencies but increases at high frequencies.

The stereo input signals, $x_i(n)$ for $i = 1, 2$, are interpolated by a factor of $I_h = 2$, and then used as the inputs of the adaptive filters. The interpolated stereo inputs accommodate the emphasised high-frequency noise. This also helps to create a buffer band for the baseband signals which further avoids the sound quality degradation from the noise addition effect when employing the HO-TV-APFs method. The modified schematic diagram

for an SAEC system employing the HO-TV-APFs signal decorrelation method is illustrated in Fig. 2.

The sampling frequency, f'_s , operated in the near-end room is I_h times the sampling frequency, f_s , in the far-end room. Then the interpolated signal $xx_i(n)$ in each channel is passed through the so-called *anti-imaging filter*, which is necessary to recover the baseband signal of interest and eliminate the unwanted image components. The anti-imaging filters in both channels are designed to be half-band lowpass filters, seen as LPF_1 in Fig. 2, whose impulse responses are given by $lp_1(n)$, expressed by its frequency response of

$$\text{LPF}_1(\omega) = \begin{cases} I_h, & 0 \leq |\omega| \leq \frac{\pi}{I_h} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where ω is the normalised angular frequency, with the order of $N_L = 55$. By employing the time-varying allpass filters $\mathbf{a}_i$, $i = 1, 2$ on each channel, the decorrelated signals, $xall_i(n)$, are fed to the adaptive filters $\hat{\mathbf{h}}_i(n)$ for identifying the unknown systems $\mathbf{h}_i$. The high-frequency components introduced in the decorrelated stereo signals, $xall_i(n)$, can be removed before transmitting to the near-end room by the use of another set of identical lowpass filter LPF_2 of order $N_L = 55$ in each channel, given by

$$\text{LPF}_2(\omega) = \begin{cases} 1, & |\omega| \leq \frac{\pi}{I_h} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

whose impulse responses are given by $lp_2(n)$. Hence, the adaptive filters in this system will converge to the convolution of $lp_2(n)$ and the unknown systems $\mathbf{h}_i$, which are assumed to be time-invariant, i.e.

$$\hat{\mathbf{h}}_i(n) = lp_2(n) * \mathbf{h}_i \quad (5)$$

The error signal, $e(n)$, is therefore given by

$$e(n) = d(n) - \sum_{i=1}^2 \hat{\mathbf{h}}_i^T(n) \mathbf{x}all_i(n) \quad (6)$$

where the desired signal is defined as

$$d(n) = \sum_{i=1}^2 \hat{\mathbf{h}}_i^T \mathbf{x}lp_i(n) \quad (7)$$

The decorrelated input vector and its lowpass version vector are $\mathbf{x}all_i(n) = [xall_i(n), xall_i(n-1), \dots, xall_i(n-L+1)]^T$ and $\mathbf{x}lp_i(n) = [xlp_i(n), xlp_i(n-1), \dots, xlp_i(n-L+1)]^T$ respectively, when L is the length of the adaptive filters $\hat{\mathbf{h}}_i(n)$ and the unknown systems $\mathbf{h}_i$.

For this modified system, the power level of the interpolated signals is reduced by half. Thus, the variance of the random noise signal that is used for the time-variation of the HO-TV-APFs is also reduced by one half. The values of the HO-TV-APFs parameters are given in Table 1 when $f'_s = 16\text{kHz}$, where $\alpha_{i,N,int}$ represents the

Figure 2: A schematic diagram for an SAEC system employing time-varying allpass filtering signal decorrelation technique using spectral-shaped random noise.

N	for 50 – 200 μ s inter-aural delay					
	$\alpha_{i,N,int}$	$r_{i,N}$	$\alpha_{i,N-1,int}$	$r_{i,N-1}$	$\alpha_{i,N-2,int}$	$r_{i,N-2}$
1	± 0.2	0.35	-	-	-	-
2	± 0.08	0.27	0	0.12	-	-
3	± 0.06	0.18	0	0.2	0	0.2

Table 1: Initialised parameters and deviation of the allpass filters for different orders, N , when $f_s = 16kHz$. ($i=1,2$)

initial value of the parameter, $\alpha_{i,N}$, $r_{i,N}$ denotes the deviation and the positive and negative values are assigned to $\alpha_{1,N,int}$ and $\alpha_{2,N,int}$, respectively. For each order N , these parameters are initialised to guarantee a minimum inter-aural delay of 50 μ s and with maximum inter-aural delay in the low frequencies of 200 μ s. Note that, the inter-aural delay can exceed 200 μ s in the high frequencies for increased level of signal decorrelation and thus, misalignment performance.

The error signal $e(n)$ is decimated by a factor of $I_h = 2$. Another lowpass filter LPF₃ of order N_L , is required to eliminate the spectrum of the error signal and noise introduced within the high-frequency range of $\frac{\pi}{I_h} < \omega < \pi$. Its frequency response satisfies the following condition

$$LP_3(\omega) = \begin{cases} 1, & |\omega| \leq \frac{\pi}{I_h} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

The decimated error signal, $e'(n)$, is transmitted back to the far-end room, with the sampling frequency of f_s , to satisfy the transmission channel bandwidth requirements.

3.2 Spectral-shaped random noise

The required characteristics of the noise signal that is used to model the time-variation of HO-TV-APFs may be achieved by utilising the *first-order differentiator*. Note that only a first-order differentiator is considered here for low complexity. For this spectral-shaped noise, the generated random noise is shaped out of the baseband of interest, without reducing the magnitude of the original noise.

If a white noise sequence, $r(n)$, is the input to the first-order differentiator, the output signal, $rs(n)$, is given by

$$rs(n) = r(n) - r(n-1) \quad (9)$$

For HO-TV-APFs employing spectral-shaped random noise, the update equation of the time-varying parameter is given by

$$\alpha_i(n) = \alpha_{i,int} + rs_i(n) \quad (10)$$

for $i = 1, 2$.

The reduction of the noise level in the low-frequency region is expected to decrease the sound degradation of the stereo signals. Although the noise level in the high-frequency region rises and can exceed that of the white noise, the sound quality will not be disturbed due to the employment of the lowpass filter and the decimation process. In fact, these high-frequency noise components should potentially yield higher degrees of signal decorrelation.

4 Experimental results

4.1 Simulation results

The experiments were carried out under a statistically stationary assumption for the near-end room in Fig. 2, i.e. the coefficients of h_i were fixed, and modelled with exponentially decaying envelopes of length $L = 256$. Stereo speech signals, sampled at 8kHz, were normalised to have zero mean and unity variance and used as the input signals of the system with SNR at the microphone in the near-end room of 30dB. The adaptive filters employed in

this system were also of length $L = 256$ and controlled by the two-channel normalised least mean square (NLMS2) algorithm with $\mu = 0.3$. The TV-APFs were investigated for the first- and third-order cases, whereas two different types of noise signal, i.e. white noise (conventional case as in [1]) and spectral-shaped noise, were used within the TV-APFs. For comparison, the Weight Error Vector Norm (WEVN) performance, as given by

$$\text{WEVN}(n) = 10 \times \log_{10} \left(\frac{\sum_{i=1}^2 \|\mathbf{h}_i * \mathbf{lp}_2 - \hat{\mathbf{h}}_i(n)\|_2^2}{\sum_{i=1}^2 \|\mathbf{h}_i * \mathbf{lp}_2\|_2^2} \right) \quad (11)$$

was evaluated.

(a) First-order TV-APFs ($N=1$)

(b) Third-order TV-APFs ($N=3$)

Figure 3: Weight error vector norm performance using HO-TV-APFs with different noise sources for signal decorrelation in SAEC.

By employing the HO-TV-APFs decorrelation method with wideband noise, $r(n)$, and spectral-shaped noise, $rs(n)$, the final WEVN for the $N = 1$ case were improved by 5dB and 7dB, respectively, as compared to that without any decorrelation methods. Similarly, for the

$N = 3$ case, the improvement of the final WEVN of 5.5dB and 7dB were also shown. These are illustrated in Fig. 3(a) and Fig. 3(b). The WEVN of the spectral-shaped noise technique shows 2dB improvement on the conventional HO-TV-APFs decorrelation method for both cases.

4.2 Subjective listening tests

Based on the ITU-R five-category impairment scale in [6], subjective listening tests were undertaken with 20 listeners. The original signal pair was compared with different versions of those signals obtained by the proposed HO-TV-APFs decorrelation method, which were based on the wideband random noise, $r(n)$, and the spectral-shaped random noise, $rs(n)$, for both cases of $N = 1, 3$. The MOS values of the original signals, the first-order TV-APFs method using $r(n)$ and $rs(n)$, and the third-order TV-APFs method using $r(n)$ and $rs(n)$ were respectively graded as 4.45, 3.20, 3.65, 3.10 and 3.20. It can be seen that the HO-TV-APFs method employing the spectral-shaped random noise $rs(n)$ can preserve the stereo perception of the speech signals much better than that using the wideband random noise $r(n)$, for the first-order case. However, the tests show that approximate stereo signal quality is obtained.

5 Conclusions

The HO-TV-APFs decorrelation technique for SAEC using spectral-shaped random noise and the interpolation method has been presented. Further improvement in signal decorrelation capability, as compared with the conventional HO-TV-APFs approach, has been demonstrated with minimum degradation on stereo signal perception and insignificant additional computational cost.

6 References

- [1] N. Tangsangumvisai, J.A. Chambers and A.G. Constantinides, "Higher-order time-varying allpass filters for signal decorrelation in stereophonic acoustic echo cancellation", *IEE Electronics Letters*, vol.35, no.1, Jan. 1999, pp. 88–90.
- [2] M.M. Sondhi, D.R. Morgan and J.L. Hall, "Stereophonic Acoustic Echo Cancellation-An Overview of the Fundamental Problem", *IEEE Signal Processing Letters*, vol.2, no.8, Aug. 1995, pp. 148–151.
- [3] J.D. Johnston and K. Brandenburg, "Advances in Speech Signal Processing : Wideband Coding-Perceptual Considerations for Speech and Music", Marcel Dekker, 1992.
- [4] F.L. Wightman and D.J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization", *Journal on Acoustical Society of America*, vol.91, no.3, Mar. 1992, pp. 1648–1661.
- [5] E. Zwicker and H. Fastl, "Psychoacoustics : Facts and Models". Springer-Verlag, 1990.
- [6] "CCITT Recommendation E.432", *International Telecommunication Union (ITU)*, 1992.