# INSIGHT INTO A PHASE MODULATION TECHNIQUE FOR SIGNAL DECORRELATION IN MULTI-CHANNEL ACOUSTIC ECHO CANCELLATION

María Luis Valero and Emanuël A. P. Habets

International Audio Laboratories Erlangen\*, Am Wolfsmantel 33, 91058 Erlangen, Germany {maria.luis.valero,emanuel.habets}@audiolabs-erlangen.de

#### ABSTRACT

High coherence between the loudspeaker signals in a multi-channel communication set-up is known to be detrimental to the performance of a multi-channel acoustic echo cancellation (MC-AEC) system. The MC-AEC performance can be improved by decorrelating the loudspeaker signals prior to their reproduction. The decorrelation process can however degrade the subjective sound quality. A technique which has proven to provide a good trade-off between the MC-AEC performance enhancement and the subjective quality of the decorrelated signals, is applying a time-varying and perceptually motivated phase modulation in the sub-band domain. The aim of this paper is to provide further insight into this technique by analysing the influence on the signals' coherence as well as the MC-AEC performance in terms of its convergence speed.

*Index Terms*— decorrelation, multi-channel acoustic echo cancellation, non-uniqueness problem

# 1. INTRODUCTION

In hands-free communication scenarios, the far-end signals reproduced by the loudspeakers are propagated through the room and acquired by the microphone. In order to reduce the electro-acoustic coupling, an acoustic echo cancellation (AEC) system, [1,2], is commonly used. AEC employs adaptive filtering techniques, [3], to identify the acoustic echo paths that are necessary for the estimation of the echo signals. Afterwards, the estimated echo signals are subtracted from the microphone signal, reducing the coupling. Given a single-input single-output (SISO) communication set-up, the adaptive algorithm is able to uniquely identify the acoustic echo path. This statement does not always hold however for stereophonic (SAEC) and multichannel AEC (MC-AEC), [4–6], where two or more loudspeakers are used for sound reproduction.

It is well-known that if the loudspeaker signals are highly coherent, the adaptive algorithms in MC-AEC converge towards a nonunique solution, [4, 5]. Under these circumstances, the adaptive algorithms converge towards one of the solutions that are able to cancel the echo signal. This solution is usually a linear combination of the true echo paths which depends on the far-end signals reproduced by the loudspeakers, see [6]. Consequently, any change in the far-end room will destabilize the adaptation process. Moreover, in [6] it is also shown that there exists an inverse relation between the loudspeaker signals' coherence and the convergence speed of the adaptive algorithm.

The most-commonly used technique to ensure the convergence and enhance the performance of a MC-AEC system is to decorrelate the loudspeaker signals prior to reproduction, [4–11]. However,



Fig. 1: Block diagramm of SAEC, including processing

decorrelating the signals can lead to severe audio artifacts and to the distortion of the spatial image. Thus, a compromise has to be made between the degradation of the subjective sound quality and the performance of the MC-AEC system. A technique which has proven to provide a good trade-off is the phase modulation decorrelation technique proposed in [8]. This technique applies a time-varying phase modulation in the sub-band domain, whose amplitude is subband dependent. Hereby, a relative phase difference is introduced to pairs of loudspeaker signals, which is optimized to be kept under the threshold of perception for every sub-band.

The aim of this paper is to give further insight into the decorrelation technique proposed in [8]. The coherence of the decorrelated signals is analysed, both theoretically and by means of simulations, for time-varying relative phase difference introduced to pairs of loudspeaker signals. This provides a link between the amplitude of the phase modulation and the performance of a MC-AEC system.

#### 2. NON-UNIQUENESS PROBLEM

Given a multiple-input single-output (MISO) communication set-up with N loudspeakers and one microphone, as depicted in Fig. 1 for N = 2, the signal acquired by the microphone at time instant n is,

$$y(n) = \sum_{i=1}^{N} d_i(n) + s(n) + v(n) = \sum_{i=1}^{N} \mathbf{x}_i^{\mathrm{T}}(n) \mathbf{h}_i(n) + r(n), \quad (1)$$

where, from the perspective of estimating the acoustic echo paths, r(n) = s(n) + v(n) is the interference signal, which comprises the desired near-end speech, s(n), and background noise, v(n). It is assumed that the acoustic echo paths can be modelled as finite impulse response (FIR) filters of length L. Hence, in (1) the acoustic echo paths are defined as,

$$\mathbf{h}_{i}(n) = [h_{i}(n,0), \dots, h_{i}(n,L-1)]^{\mathrm{T}}$$
, with  $i = 1, \dots, N$ .

Further, the acoustic echo signals, denoted by  $d_i(n)$ , are the result of the far-end signals,  $\mathbf{x}_i(n) = [x_i(n), \dots, x_i(n-L+1)]^T$ , being

<sup>\*</sup>A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

reproduced in the near-end room. The error signal after cancellation, denoted as e(n), is the result of subtracting the estimated echo signals,  $\hat{d}_i(n)$ , from the microphone signal, i.e.,

$$e(n) = y(n) - \sum_{i=1}^{N} \hat{d}_i(n) = \sum_{i=1}^{N} (d_i(n) - \hat{d}_i(n)) + r(n).$$
 (2)

In the remainder of this paper, for brevity, only the case N = 2, i.e., stereophonic reproduction set-up, with a SAEC system is considered. The aim of SAEC is to minimize the error signal after cancellation, which is achieved if e(n) = r(n), meaning that,

$$\mathbf{x}_{1}^{\mathrm{T}}(n)(\mathbf{h}_{1}(n) - \hat{\mathbf{h}}_{1}(n)) + \mathbf{x}_{2}^{\mathrm{T}}(n)(\mathbf{h}_{2}(n) - \hat{\mathbf{h}}_{2}(n)) = 0, \quad (3)$$

where  $\hat{\mathbf{h}}_i(n)$  are the estimated echo paths. It can be observed that if there exists a linear relation between the loudspeaker signals, more than one solution will fulfil (3). Assuming that the interference signal, r(n), and the loudspeaker signals,  $\mathbf{x}_i(n)$ , are uncorrelated, the *i*-th channel's update equation of a generalized adaptive filter can be written as,

$$\hat{\mathbf{h}}_{i}(n+1) = \hat{\mathbf{h}}_{i}(n) + \mu_{i}(n)\mathbf{x}_{i}(n)e(n)$$

$$= \hat{\mathbf{h}}_{i}(n) + \mu_{i}(n)(\mathbf{x}_{i}(n)\mathbf{x}_{i}^{\mathrm{T}}(n))(\mathbf{h}_{i}(n) - \hat{\mathbf{h}}_{i}(n))$$

$$\cdots + \mu_{i}(n)(\mathbf{x}_{i}(n)\mathbf{x}_{j}^{\mathrm{T}}(n))(\mathbf{h}_{j}(n) - \hat{\mathbf{h}}_{j}(n)),$$
(4)

with i = 1, 2 and  $j \neq i$ , and where  $\mu_i(n)$  denotes the step-size factor that depends on the adaptive algorithm. Thus, the update of the filter coefficients of the *i*-th channel depends on the cross-correlation between the loudspeaker signals, unless, either the loudspeaker signals were completely uncorrelated or  $\hat{\mathbf{h}}_j(n) = \mathbf{h}_j(n)$ . Consequently, if the loudspeaker signals are highly correlated, the adaptive algorithm usually fails to converge towards the true echo paths. In addition, if the correlation between the loudspeaker signals is time-invariant, the adaptive algorithm converges towards a linear combination of both echo paths that depends on the loudspeaker signals. Moreover, as described in [6], the correlation between the signals also affects the convergence speed of the adaptive algorithm.

The adaptive algorithm can be forced to converge towards the true echo paths by decorrelating the loudspeaker signals prior to reproduction. In addition, a further enhancement in terms of convergence speed can be obtained if the correlation between the loudspeaker signals is time-variant, [12, 13].

#### 3. SUB-BAND DOMAIN PHASE MODULATION

Several decorrelation techniques have been proposed whose aim is to solve the non-uniqueness problem. In this paper we focus on the perceptually motivated decorrelation technique proposed in [8]. The authors of [8] proposed to apply a time-varying phase modulation to the loudspeaker signals in the sub-band domain. The phase modulation was designed to introduce the highest possible phase difference per sub-band, in order to maintain, or only slightly degrade, the perceptual stereo sound quality. The employed phase modulation was designed such that the following requirements are fulfilled:

- (i) The human sensitivity to phase differences decreases with increasing frequency and vanishes at around 4 kHz. Hereafter, the phase difference introduced to the loudspeaker signals has to be kept under the threshold of perception to ensure that the spatial image is kept mostly unaltered.
- (ii) The convergence speed of SAEC is inversely proportional to the correlation between the loudspeaker signals, [6]. Therefore, the maximum possible decorrelation should be applied at every sub-band.

Consequently, the phase modulation is applied as follows. A dedicated filter-bank is used to transform the unprocessed loudspeaker signals,  $\mathbf{z}_i(m)$  in Fig. 1, to the sub-band domain, i.e.,

$$\mathbf{Z}_{i}(m) = \mathcal{FB}\{\mathbf{z}_{i}(m)\},\tag{5}$$

with  $\mathbf{z}_i(m) = [z_i(mR), \dots, z_i(mR - M + 1)]^T$ , where *m* is the frame index, *M* is the transform or input signal frame length, *R* is the shift in samples between subsequent frames, and  $\mathcal{FB}\{\cdot\}$  denotes the filter-bank operation. To manipulate the phase of the loudspeaker signals it is necessary to use a complex-valued filter-bank. In [8] the complex modulated lapped transform (CMLT), [14], was used. The relative phase difference between a pair of loudspeaker signals is modified per sub-band, *k*, by,

$$X_i(m,k) = Z_i(m,k) e^{i(i-1)\alpha(k)\sin(\theta(m))},$$
(6)

where  $i = \sqrt{-1}$ , and  $\theta(m) = 2\pi f_M Rm/f_s$  describes the time variation of the phase modulation. The latter depends on the modulation frequency  $f_M$ , the frame shift R, and the sampling frequency  $f_s$ . Finally,  $\alpha(k)$  is the sub-band dependent amplitude of the phase modulation that was perceptually optimized for each sub-band by means of listening tests in [8]. The resulting  $\alpha(k)$  ranged from  $\alpha(k) = \pi/20$  for the lower sub-bands to  $\alpha(k) = \pi/2$  for frequencies above 2.5 kHz. This technique can be extended to more than one loudspeaker signal pairs by using a different modulation frequency for each pair.

In the following section an analysis of this decorrelation technique is provided. For the analysis, a generalized formulation of (6) is used, i.e.,

$$X_i(m,k) = Z_i(m,k) e^{i\alpha(k)\sin(\theta(m) - (i-1)\varphi)},$$
(7)

where  $\varphi$  denotes a time- and frequency-invariant phase shift. Hereafter, this analysis is valid for different implementations of the phase modulation technique. It is worth mentioning that, the phase modulation in [8], i.e., (6), is obtained using (7) with  $\varphi = \pi$ .

# 4. ANALYSIS OF THE PHASE MODULATION

Let us assume that  $Z_i(m, k)$ , for i = 1, 2, are two stationary random processes with zero-mean and that the inter-band correlations are negligible. Consequently, it is possible to define the cross-spectral density of the unprocessed signals, at frame m and sub-band k, as,

$$S_{ij}^{z}(m,k) = \mathbb{E}\left\{Z_{i}(m,k)Z_{j}^{*}(m,k)\right\}, \ \forall \ m, \ \text{with} \ j \neq i,$$
(8)

where  $E\{\cdot\}$  denotes mathematical expectation. Moreover, let us assume that this cross-spectral density is also stationary, which, as previously mentioned, is detrimental to the SAEC performance. Finally, the spectral variance of the *i*-th unprocessed signal is defined by  $S_{ii}^{z}(m,k) = E\{|Z_i(m,k)|^2\}$ .

#### 4.1. Assumptions on the decorrelated signals

Now, based on the assumptions on the unprocessed loudspeaker signals,  $Z_i(m, k)$ , and given the definition of the phase-modulated signals  $X_i(m, k)$  in (7). First, it can be assumed that the random processes  $X_i(m, k)$  are cyclostationary  $\forall i$ , i.e., their statistical properties are periodic with period T, see [15, 16]. Secondly, it can be assumed that their cross-spectral density is also cyclostationary. Thus,  $S_{is}^x(m, k) = F_i \{X_i(m, k)X_i^*(m, k)\} = S_{is}^x(m + aT, k), \forall m, a$ 

where the integer *a* denotes the period index. The product of the two complex exponential terms in 
$$S_{ii}^{x}(m, k)$$
 using (7) can be written as,

 $e^{i\alpha(k)\sin(\theta(m))}e^{-i\alpha(k)\sin(\theta(m)-\varphi)} = e^{i2\alpha(k)\sin\left(\frac{\varphi}{2}\right)\cos\left(\theta(m)-\frac{\varphi}{2}\right)}.$ 



Fig. 2: Theoretical MSC computed using (12) as a function of  $\alpha$ 

Consequently, it is possible to express the cross-spectral density of the phase-modulated signals as,

$$S_{ij}^{x}(m,k) = S_{ij}^{z}(m,k) e^{i2\alpha(k)\sin\left(\frac{\varphi}{2}\right)\cos\left(\theta(m) - \frac{\varphi}{2}\right)}.$$
 (9)

#### 4.2. Maximum decorrelation analysis

In SAEC, the signals' cross-correlation is commonly evaluated in the frequency domain, as this highly simplifies the analysis, see [5]. Common measures are the normalized cross-spectral density,

$$\rho_{ij}^{x}(m,k) = \frac{S_{ij}^{x}(m,k)}{\sqrt{S_{ii}^{x}(m,k)S_{jj}^{x}(m,k)}},$$
(10)

also denoted as complex coherence, and the magnitude-squared coherence (MSC),  $\gamma_{ij}^x(m,k) = |\rho_{ij}^x(m,k)|^2 \in [0,1]$ . The latter is often used in the analysis and evaluation of the decorrelation techniques, [5, 6, 9, 11]. In this case, the MSC can not be evaluated per frame, as if only one frame is taken into account  $\gamma_{ij}^x(m,k) = 1 \forall m$ . Hereafter, the MSC has to be calculated using time-averages of the cross- and power spectral densities. As a consequence of the periodicity of  $S_{ij}^x(m,k)$ , a proper measure for this analysis is to obtain the MSC of the processed signals by averaging over one period T, i.e,

$$\gamma_{ij}^{x}(m,k) = \frac{\left|\frac{1}{T}\sum_{p=m-T+1}^{m} S_{ij}^{z}(p,k) \mathrm{e}^{i2\alpha(k)\sin\left(\frac{\varphi}{2}\right)\cos\left(\theta(p)-\frac{\varphi}{2}\right)}\right|^{2}}{\frac{1}{T}\sum_{p=m-T+1}^{m} S_{ii}^{z}(p,k)\frac{1}{T}\sum_{p=m-T+1}^{m} S_{jj}^{z}(p,k)}$$

using (9) and the fact that  $S_{ii}^{z}(m,k) = S_{ii}^{x}(m,k)$ . Now, based on the previous assumptions, it is possible to relate the MSC of the processed signals to that of the unprocessed signals,  $\gamma_{ij}^{z}(m,k)$ , by,

$$\gamma_{ij}^x(m,k) = \frac{|S_{ij}^z(m,k)|^2}{S_{ii}^z(m,k)S_{ii}^z(m,k)}\gamma_{\rm pm}(k) = \gamma_{ij}^z(m,k)\gamma_{\rm pm}(k),$$

being  $\gamma_{\rm pm}(k)$  the MSC of the combined modulation function. The latter does not depend on m, as it is averaged over one period, i.e.,

$$\gamma_{\rm pm}(k) = \left| \frac{1}{T} \sum_{p=0}^{T-1} \rho_{\rm pm}(p,k) \right|^2$$
$$= \left| \frac{1}{T} \sum_{p=0}^{T-1} e^{i2\alpha(k)\sin\left(\frac{\varphi}{2}\right)\cos\left(\theta(p) - \frac{\varphi}{2}\right)} \right|^2, \quad (11)$$

where  $\rho_{pm}(p, k)$  denotes the complex coherence of the combined modulation function. Finally, the MSC in (11) reduces to,

$$\gamma_{\rm pm}(k) = \left| \sum_{p=0}^{T-1} \cos\left(2\alpha(k)\sin\left(\frac{\varphi}{2}\right)\cos\left(\theta(p) - \frac{\varphi}{2}\right)\right) \right|^2, \quad (12)$$
  
as 
$$\sum_{p=0}^{T-1} \sin\left(2\alpha(k)\sin\left(\frac{\varphi}{2}\right)\cos\left(\theta(p) - \frac{\varphi}{2}\right)\right) = 0, \forall \varphi, \alpha(k).$$

In Fig. 2 the function  $\gamma_{pm}(k)$  in (12) is depicted as a function of  $\alpha(k)$ . For the evaluation a sub-band invariant amplitude of the phase



Fig. 3: Real value of the complex coherence as a function of  $\alpha$ 

modulation was considered, i.e.,  $\alpha(k) = \alpha \forall k$ . The effective MSC is depicted for three different phase modulation schemes  $\varphi = \pi$ , as in [8], and  $\varphi = \pm \pi/2$ . It can be observed that applying the highest possible phase difference, obtained with  $\varphi = \pi$  and  $\alpha = \pi/2$ , does not provide the lowest possible MSC. This is due to the variation of  $S_{ij}^x(m,k)$  across m. Considering (11) and its simplification in (12), we further assume that the variation of  $S_{ij}^x(m,k)$  over one period can be evaluated using,

$$\operatorname{Re}\{\rho_{\mathrm{pm}}(m,k)\}\approx\cos\left(2\alpha(k)\sin\left(\frac{\varphi}{2}\right)\cos\left(\theta(m)-\frac{\varphi}{2}\right)\right).$$

The minimum complex coherence, i.e.,  $|\rho_{\rm pm}(m,k)|=0$ , is obtained in this case for,

$$\alpha(k)\sin\left(\frac{\varphi}{2}\right)\cos\left(\theta(m) - \frac{\varphi}{2}\right) = \pi/4 + a\pi.$$
(13)

Figure 3 depicts the absolute value of  $\operatorname{Re}\{\rho_{\operatorname{pm}}(m,k)\}\$  for different values of  $\alpha(k)$ , with  $\alpha(k) = \alpha \forall k$ , and  $\varphi$ . It can be observed that depending on  $\alpha(k)$ ,  $|\operatorname{Re}\{\rho_{\operatorname{pm}}(m,k)\}|\$  varies differently between [0, 1]. It can also be observed that if the phase modulation as defined in [8] is used and the maximum possible phase difference is applied, the pre-processed signals are coherent and out of phase over long periods of time. Consequently, the MSC of the pre-processed signals is not reduced to the minimum, as can be observed in Fig. 2.

#### 5. PERFORMANCE EVALUATION

In the previous section, the decorrelation technique proposed in [8] was analysed. The analysis provides insight into the effective and temporal performance of the phase modulation decorrelation technique, which can also be used to further optimize it. As the analysis is based on strict assumptions on the signals' properties, the performance evaluation is provided using both noise and speech signals.

#### 5.1. Validation of the theoretical MSC

In the following evaluation a worst-case scenario was considered in which one monophonic signal was reproduced by both loudspeakers to produce a phantom center. As a result, the loudspeakers signals are fully correlated. The monophonic signal consisted of either Gaussian noise or speech. Afterwards, the phase modulation was applied with a constant amplitude of the phase modulation over frequency, i.e.,  $\alpha(k) = \alpha \forall k$ , at a sampling frequency of 16 kHz. The modulation frequency of the phase modulation was set to  $f_m = 13$  Hz. It is necessary to mention that this rather high value of  $f_m$  may introduce audible artifacts, mostly in the lower frequency range, if the same design of the sub-band dependent amplitude of the phase modulation as in [8] is used. The choice of a high value for  $f_m$  is motivated in Sec. 5.3. The unprocessed loudspeaker signals were



Fig. 4: Measured MSC as a function of  $\alpha$ 

transformed to the frequency domain using the short-time Fourier transform (STFT), with a 256 points Hamming window as low pass prototype. The overlap between subsequent frames was set to 75%, being the frame shift R = 64 samples. The overall MSC of the pre-processed signals was measured for different values of  $\alpha$  and  $\varphi$ ,

$$\widehat{\gamma}^{x} = \frac{1}{K} \sum_{k=0}^{K-1} \frac{\left|\sum_{m=0}^{W-1} X_{1}(m,k) X_{2}(m,k)\right|^{2}}{\sum_{m=0}^{W-1} |X_{1}(m,k)|^{2} \sum_{m=0}^{W-1} |X_{2}(m,k)|^{2}},$$

where K = 256 is the number of sub-bands, and  $W \approx 4000$  is the length of the signals in frames. The results are depicted in Fig. 4, which closely match the theoretical analysis depicted in the Fig. 2.

### 5.2. Evaluation of the AEC performance

The SAEC performance was evaluated using as loudspeaker signals the signals that were pre-processed using  $\varphi = \pi$ . The echo signals were obtained by convolving the loudspeaker signals with two simulated room impulse responses, and background noise was added to the microphone signal to obtain a segmental echo-to-noise ratio (segENR) of 30 dB. The RIRs, of length 4098 taps, were generated using the image method, [17], for a room of dimensions  $5 \times 4 \times 3 m^3$ , and a reverberation time,  $T_{60}$ , of 0.25 s. The adaptive algorithm was designed to estimate 1024 filter coefficients. Thus, to ensure that the only remaining residual echo is the one caused by the mismatch of the estimated filter coefficients, the generated RIRs were truncated to 1024 taps. A non-symmetric set-up was used to guarantee that the generated RIRs already differed on the time-of-arrival of the direct sound and on the early reflections. Hence, the distance between the microphone and the loudspeakers was  $d_L = 1.15$  m and  $d_R = 1.28$ m, and the distance between the loudspeakers was  $d_{LR} = 1.25$  m. The SAEC system was evaluated using the partitioned-block-based NLMS, [18],

$$\hat{H}_{i}^{b}(m+1,k) = \hat{H}_{i}^{b}(m,k) + \mu \hat{S}_{ii}^{x\,b}(m,k)^{-1} X_{i}^{b}(m,k) E^{*}(m,k),$$
  
with  $\mu = 0.0625$ , where b is the block index,  $E(m,k)$  is the sub-  
band domain error signal, and  $\hat{S}_{ii}^{x\,b}(m,k)$  is estimated using a first  
order recursive filter. The SAEC transform length was set to  $M_{\rm af} =$   
256, and the filters were partitioned into  $B = 8$  blocks with 128  
filter coefficients each. The performance was measured using the

$$NMSA(m) = 20 \log_{10}(||\mathbf{h}(m) - \hat{\mathbf{h}}(m)||_2/||\mathbf{h}(m)||_2),$$
  
ERLE(m) = 20 log<sub>10</sub>(||\mathbf{d}(m)||\_2/||\mathbf{y}(m) - \hat{\mathbf{d}}(m))||\_2),

normalized misalignment (NMSA) and the echo return loss enhance-

ment (ERLE), respectively defined as,

where  $|| \cdot ||_2$  denotes the  $l_2$ -norm, and  $\mathbf{d}(m) = \sum_{i=1}^{N} \mathbf{d}_i(m)$  and  $\mathbf{y}(m)$  are the echo and the microphone signal vectors. The results obtained during the initial adaptation stage are depicted in Figs. 5a and 5b, for noise and speech, respectively. It can be observed in both figures that the convergence speed is inversely proportional to the coherence between the pre-processed loudspeaker signals. Hence, the fastest convergence is obtained for  $\alpha = 0.8\pi/2$ .



(c) For noise with  $f_m = 0.75, 13$  Hz and different values of  $\alpha$ Fig. 5: SAEC performance

# **5.3.** Influence of $f_m$ on the SAEC performance

The SAEC performance was evaluated using the pre-processed noise signals with  $f_m = 0.75$  and  $f_m = 13$  Hz. The resulting ERLE and NMSA are depicted in Fig. 5c. It can be observed that the use of a low modulation frequency impairs the SAEC performance, at least if compared to the high  $f_m$  case. This happens due to the facts that i) the correlation between the pre-processed signals varies relatively slow in time and that ii) the loudspeaker signals are coherent during relatively long periods of time. During these periods, the adaptive filter fails to converge towards the true echo paths. On top, the energy of the sum of the echo signals varies inversely proportional to the relative phase difference between them, modulating the error signal in amplitude, as it can be observed in Fig. 5c. Hence, the selection of  $f_m$  defines a trade-off between the subjective quality of the pre-processed signals and that of the resulting error signal.

#### 6. CONCLUSIONS

A theoretical analysis was provided on the effective decorrelation obtained by using a perceptually motivated phase modulation. It was shown how the amplitude of the phase modulation affects both the effective and the instantaneous decorrelation introduced by such a technique. The analysis was empirically validated and the effect of the decorrelation on the SAEC performance was evaluated. It can be concluded that it is not necessary to apply the highest possible phase difference to obtain the maximum decorrelation. In addition, it was also shown that the selection of the modulation frequency affects the output signal of the SAEC system as well as its performance.

#### 7. REFERENCES

- [1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A practical Approach*, Wiley, New Jersey, USA, 2004.
- [2] G. Schmidt, "Applications of acoustic echo control an overview," in *Proc. European Signal Processing Conf. (EU-SIPCO)*, Vienna, Austria, 2004, pp. 9–16.
- [3] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, third edition, 1996.
- [4] M. M. Sondhi, D. R. Morgan and, and J. L. Hall, "Stereophonic acoustic echo cancellation - an overview of the fundamental problem," *IEEE Signal processing letters*, vol. 2, no. 8, pp. 148–151, Aug. 1995.
- [5] 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 transactions* on speech and audio processing, vol. 6, no. 2, pp. 156–165, Mar. 1998.
- [6] D. R. Morgan, J. L. Hall, and J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation," *IEEE transactions on speech and audio processing*, vol. 9, no. 6, pp. 686–696, Sept. 2001.
- [7] T. Gänsler and P. Eneroth, "Influence of audio coding on stereophonic acoustic echo cancellation," in *Proc. IEEE ICASSP*, Seattle, U.S.A., May 1998.
- [8] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *Proc. IEEE ICASSP*, Honolulu, U.S.A., Apr. 2007.
- [9] B. C. Bispo and D. da S. Freitas, "Hybrid pre-processor based on frequency shifting for stereophonic acoustic echo cancellation," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 2012.

- [10] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "Lowcomplexity implementation of a real-time decorrelation algorithm for stereophonic acoustic echo cancellation," *Signal Processing*, vol. 92, no. 11, pp. 2668–2675, 2012.
- [11] L. Romoli, S. Cecchi, and F. Piazza, "A combined approach for channel decorrelation in stereo acoustic echo cancellation exploiting time-varying frequency shifting," vol. 20, no. 7, pp. 717–720, 2013.
- [12] S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," in *Proc. IEEE ICASSP*, Detroit, Michigan, U.S.A., May 1995.
- [13] J. van de Laar and P. C. W. Sommen, "Exploiting nonstationarity in multi-channel acoustic echo cancellation," in *Proc.* 3<sup>rd</sup> *IEEE Benelux signal processing symposium (SPS-*2002), Leuven, Belgium, Mar. 2002.
- [14] H. S. Malvar, "A modulated complex lapped transform and its application to audio processing," in *Proc. IEEE ICASSP*, Phoenix, U.S.A., Mar. 1999.
- [15] W. A. Gardner, "The spectral correlation theory of cyclostationary time-series," *Signal Processing, Elsevier Science Publishers B.V.*, vol. 11, no. 1, pp. 13 – 36, 1986.
- [16] W. A. Gardner, Cyclostationarity in communications and signal processing, IEEE Press, Piscataway, NJ, USA, 1994.
- [17] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
- [18] P. C. W. Sommen, "Partitioned frequency-domain adaptive filters," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, 1989, pp. 677–681.