NON-LINEAR ACOUSTIC ECHO CANCELLATION USING EMPIRICAL MODE DECOMPOSITION

Leela K. Gudupudi[1]*, Navin Chatlani*[2]*, Christophe Beaugeant*[3] *and Nicholas Evans*[1]

[1]EURECOM, Sophia-Antipolis, France
lastname@eurecom.fr
[2]INTEL, Allentown, U.S.A., [3]INTEL, Sophia-Antipolis, France
firstname.lastname@intel.com

ABSTRACT

The increasing popularity of miniature devices and loudspeakers has fuelled research in non-linear acoustic echo cancellation (NAEC). This paper reports a novel approach to NAEC based on empirical mode decomposition (EMD), a recently developed technique in non-linear and non-stationary signal analysis. EMD decomposes any signal into a finite number of time varying sub-band signals termed intrinsic mode functions (IMFs). The new approach to NAEC presented here incorporates this multi-resolution analysis with conventional power filtering to estimate non-linear echo in each IMF. Comparative experiments with a competitive baseline approach to NAEC based on pure power filtering show that the new EMD approach achieves greater non-linear echo reduction and faster convergence.

Index Terms— Echo cancellation, non-linear modelling, empirical mode decomposition (EMD), intrinsic mode functions (IMF)

1. INTRODUCTION

Acoustic echo cancellation (AEC) plays a vital role in ensuring satisfactory speech quality in many applications, for example hands-free telephony and teleconferencing. The use of miniature transducers in such applications generally introduces non-linearities in the acoustic path which typically degrade the performance of AEC algorithms. Consequently, non-linear acoustic echo cancellation (NAEC) is today an active research area.

Loudspeaker saturation is generally assumed to be the major source of memory-less, non-linearities [1, 2, 3, 4, 5]. Popular, time-domain solutions to NAEC include **cascaded** and **parallel** approaches. In the cascaded approach, non-linearities and the echo path are estimated with a cascaded, adaptive non-linear pre-processor and a finite impulse response (FIR) filter. With the parallel approach, estimation is generally performed with a multi-channel adaptive filter structure.

Both cascaded and parallel approaches have their drawbacks. The cascaded approach requires pre-processor and FIR filter adaptation using a single joint error signal $e\left(n\right)$. As a result the convergence of both filters is interdependent, which leads to possible errors. The parallel approach requires redundant echo path estimation with each sub-filter, which generally leads to slow convergence [3]. Frequency domain solutions were proposed to address these drawbacks. A discrete Fourier transform (DFT) based approach is proposed in [6] and a sub-band domain approach is proposed in [7]. The use of a non-linear transformation based on a raised cosine function for non-linear echo reduction is proposed in [8]. Approaches to NAEC based on kernel adaptive filtering are proposed in [9, 10]. Neural network solutions are also reported in [11]. Despite this considerable volume of research, the current state of the art solutions generally accomplish only modest reductions in non-linear echo. This observation has motivated our pursuit of entirely new solutions to NAEC.

One of the new approaches we are exploring involves empirical mode decomposition (EMD). EMD is a recent adaptive data analysis technique suited to non-linear and non-stationary signals [12]. Applications in speech and audio processing are widely reported and include speech enhancement/noise cancellation [13, 14, 15], source separation [16], voice activity detection [17] and pitch estimation [18]. An EMD-based, sub-band approach to linear AEC is reported in [19]. The suitability of EMD to decompose any signal completely without losing any information has prompted us to investigate its application to NAEC.

This paper reports the first EMD-based approach to NAEC. The work aims to demonstrate the application of EMD in the time domain as a potential solution to NAEC. The approach is based on the decomposition of a full-band microphone signal into a small, finite number of time-varying sub-band signals, termed intrinsic mode functions (IMFs). NAEC is then accomplished through the application of conventional, adaptive power filtering to each IMF using a full-band reference signal. The new approach is shown to outperform a baseline power filter approach in terms of better echo reduction and faster convergence.

2. EMPIRICAL MODE DECOMPOSITION

EMD is a recent approach to non-linear and non-stationary signal analysis [12, 20]. EMD decomposes any signal into a finite number of time varying sub-band signals referred to as intrinsic mode functions (IMFs). IMFs are not predefined, as is the case with the Fourier and wavelet transforms, but are adaptively extracted from the input data and accordingly serve as adaptive basis functions.

The EMD algorithm examines the input signal between two consecutive extrema and iteratively extracts the highest frequency components between these two points [20]. The remaining local, low frequency components can then be extracted by consecutive iterations. This procedure identifies the different oscillatory modes in the input. IMFs are symmetric with respect to a local zero-mean and the number of zero crossings and extrema differ at most by one [12].

As described in [12, 14], a signal $y\left(n\right)$ is decomposed into a set of M IMFs according to the following procedure known as *sifting*:

1. Identify all extrema (local maxima and minima) of the signal, $y\left(n\right)$.
2. Obtain the upper envelope $e_{max}\left(n\right)$ and the lower envelope $e_{min}\left(n\right)$ by interpolating the local maxima and minima, respectively.
3. Compute the local mean $m\left(n\right) = \frac{e_{min}(n)+e_{max}(n)}{2}$.
4. Extract the detail signal $d\left(n\right) = y\left(n\right) - m\left(n\right)$.
5. $d\left(n\right)$ can be considered as an IMF if it has zero mean and all its local maxima and minima are positive and negative respectively. If not, steps 1–4 are repeated with $d\left(n\right)$ in place of $y\left(n\right)$.
6. For the next IMF, the entire process is applied to the residual $r_1\left(n\right) = y\left(n\right) - d\left(n\right)$.
7. Iterate on the residual until the number of extrema in the residual is smaller than two or until a maximum number of iterations is reached. Assign the last residual as $r\left(n\right)$.

The above *sifting* process decomposes any signal $y(n)$ into a set of frequency ordered IMF components $y_j\left(n\right);\ j = 1, \cdots, M$. Each successive IMF contains successively lower frequency components. Together they represent $y(n)$ according to:

$$y\left(n\right) = \sum_{j=1}^{M} y_j\left(n\right) + r\left(n\right) \tag{1}$$

Full details of EMD are available in [12, 20]. While there is an on-line EMD algorithm [21], the work reported here was performed with an 'off-line' implementation, i.e. by application of EMD to entire signals. This was deliberate in order to demonstrate the application of EMD to non-linear echo cancellation while avoiding additional problems inherent to on-line processing [21].

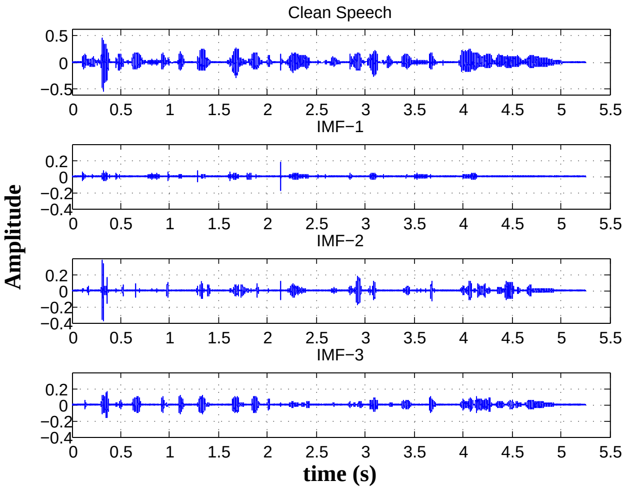

Fig. 1. An illustration of EMD. Illustrated is a clean speech signal (top) and the first 3 IMFs.

An example of EMD is illustrated in Fig. 1. Illustrated is a clean speech signal sampled at 8kHz and the first three IMFs. The spectral content of consecutive IMFs corresponds to decreasing frequency. The first few IMFs correspond to a bandwidth of approximately 1kHz to 4kHz, the bandwidth which typically contains the majority of the higher-order non-linear echo components. Other IMFs are predominant with other echo components. This data-adaptive technique thus decomposes a non-linear input into a set of IMFs which can further be characterised as either non-linear-dominant or linear-dominant. By taking advantage of this principle, we propose a novel approach to NAEC based on EMD.

3. EMD FOR NAEC

The new EMD-based NAEC scheme illustrated in Fig. 2 is essentially standard except for EMD decomposition, resynthesis and the use of multiple filter chambers. The downlink/reference signal is denoted by $x\left(n\right)$, the loudspeaker output signal by $x_{out}\left(n\right)$ and the uplink/microphone output signal by $y\left(n\right)$. In this first attempt to employ EMD for NAEC we suppose no near-end speech and no background noise. The uplink signal thus contains echo alone.

The microphone output $y\left(n\right)$ is decomposed by EMD into M IMFs according to the approach described in Section 2. Each IMF is then adaptively estimated from the downlink/reference signal $x\left(n\right)$ by one of M filter chambers (FCs). Each FC contains the P^{th} order conventional power filter model [1] illustrated in Fig. 3. The power filter model

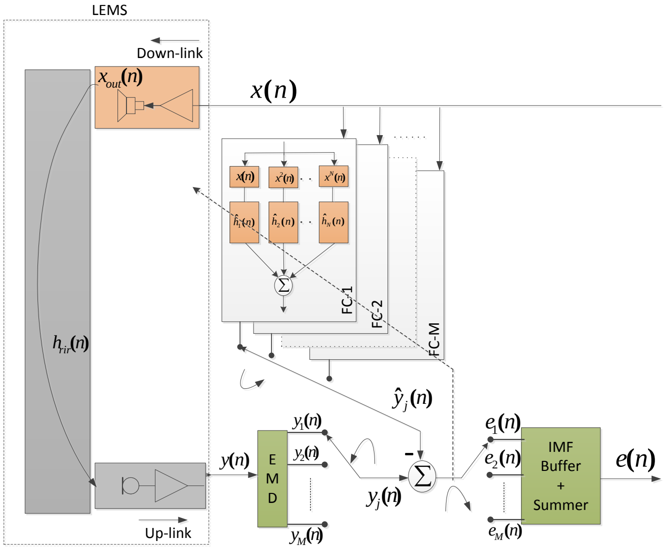

Fig. 2. Structure of EMD based NAEC.

is an efficient approach to the identification of non-linear acoustic echo paths. The sub-filters adaptively estimate the acoustic channel and loudspeaker impulse response, collectively referred to as the loudspeaker enclosure microphone system (LEMS) illustrated in Fig. 2.

Decomposition of the microphone signal $y\,(n)$ produces M IMF signals $y_j; j = 1, \cdots, M$ where each IMF represents a distinct frequency range. Accordingly, each corresponding FC requires fewer filter taps than would otherwise be required in the case of a full-band signal. The output of each FC $\hat{y}_j(n)$ is subtracted from the corresponding IMF $y_j(n)$ thereby generating individual error signals $e_j\,(n)$. Each error signal is used in the conventional manner to update FC sub-filter coefficients $h_p\,(n)\,; p = 1, \cdots, P$. Finally, the individual error signals are summed together to reconstruct the full-band error signal:

$$e(n) = \sum_{j=1}^{M} e_j\,(n) + r\,(n) \tag{2}$$

3.1. Echo generation

In this work, it is assumed that the loudspeaker is a memoryless non-linear system and that it is the only source of non-linearity. For experimentation purposes, microphone output signals with non-linear echo are generated artificially according to:

$$y\,(n) = \sum_{p=1}^{P}\sum_{i=0}^{L-1} x^p(n-i)h_p(i) \tag{3}$$

where $x\,(n)$ is the downlink/reference signal and $h_p(n)$ is the L-tap linear filter in the P^{th} channel. It represents the combination of the loudspeaker response $h_{p_L}\,(n)$ and the room impulse response $h_{rir}\,(n)$:

$$h_p\,(n) = h_{p_L}\,(n) * h_{rir}\,(n) \tag{4}$$

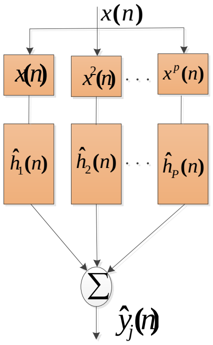

Fig. 3. Block diagram of a multi-channel, P^{th} order power filter.

Here, $h_{p_L}\,(n)$ is the P^{th} order simplified, diagonal (one-dimensional) loudspeaker Volterra kernel. It is measured empirically as explained in our previous work [22, 23] using the non-linear system identification method reported in [24, 25].

3.2. Adaptive filtering

EMD produces a total of M IMF signals $y_j; j = 1, \cdots, M$. Corresponding error signals $e_j; j = 1, \cdots, M$ are thus expressed by:

$$\begin{aligned} e_j\,(n) &= y_j\,(n) - \hat{y}_j\,(n) \\ e_j\,(n) &= y_j\,(n) - \sum_{p=1}^{P} \hat{\mathbf{h}}_p^T\,(n)\,\mathbf{x}^p(n) \end{aligned} \tag{5}$$

where $\hat{\mathbf{h}}_p\,(n)$ is the estimated sub-filter vector of length N_p, $\mathbf{x}^p(n) = [x^p(n), \ldots, x^p(n - N_p + 1)]^T$, and $\hat{y}_j\,(n) = \sum_{p=1}^{P} \hat{\mathbf{h}}_p^T\,(n)\,\mathbf{x}^p(n)$ is the output of the j^{th} FC. Due to its simplicity we used a normalised least mean square (NLMS) adaptive filtering algorithm within each FC. The NLMS algorithm for sub-filter $\hat{\mathbf{h}}_p\,(n)$ is derived using an approach similar to that given in [5]. Updates are applied in the usual manner according to:

$$\hat{\mathbf{h}}_p\,(n+1) = \hat{\mathbf{h}}_p\,(n) + \frac{\mu_p}{\|\mathbf{x}^p\|_2^2}\mathbf{x}^p e_j\,(n) \tag{6}$$

4. EXPERIMENTAL WORK

The following reports a performance comparison of the new EMD-based approach to NAEC to a baseline power filtering approach. All experiments were conducted with speech

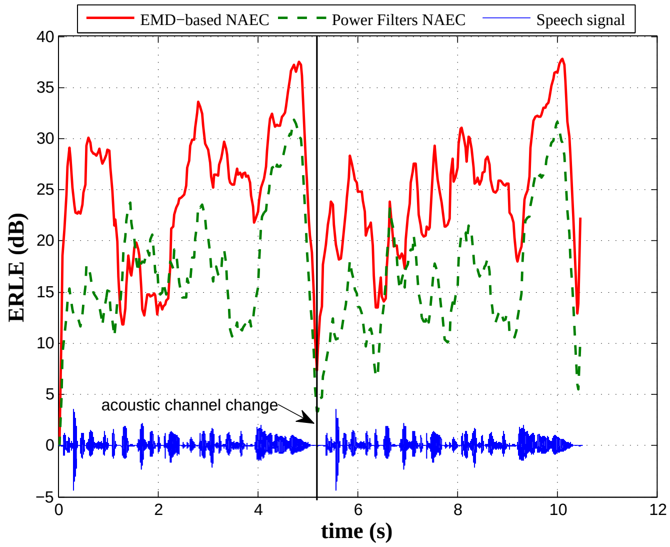

Fig. 4. A performance comparison in terms of ERLE for the new EMD-based approach to NAEC and a baseline power filter approach.

signals contaminated artificially according to an empirically measured loudspeaker non-linearity function. Performance is assessed in terms of the echo return loss enhancement (ERLE).

4.1. Experimental setup

Experiments were performed with diagonal (one-dimensional) loudspeaker Volterra kernels $h_{p_L}(n)$ for values of $p \leq 5$ and with 32 taps in all cases. The acoustic channel was modelled with a fixed 256-tap room impulse response (RIR) $h_{rir}(n)$ selected from the Aachen RIR database [26]. All experiments were performed with a clean speech downlink/reference signal $x(n)$ of approximately 10 seconds duration with a sampling frequency of 8kHz. A change in the acoustic channel is introduced after approximately 5 seconds simply by delaying the RIR by 2.5 ms. This is done to compare the dynamic re-convergence performance of each algorithm. The $P = 5$ sub-filters, each of 287 taps, are generated according to Eq. 4 before the microphone output/non-linear echo signal $y(n)$ is then generated according to Eq. 3.

We used the EMD routines available in [27] for decomposition into $M = 10$ IMFs[1]. The order of the power filters can be adjusted individually in each FC according to the spectral properties of the corresponding IMF. Over-modelling the order of power-filters in the FCs increases computational complexity and the unnecessary degrees of freedom lead to noisy estimates $\hat{y}_j(n)$. For the test whose results are illustrated in Fig. 4, the first 4 FCs each contain 5 adaptive sub-filters, the 5th FC has only 4 adaptive sub-filters whereas the 6th and 7th FCs have only 3. FCs 8–10 consist of a single, 287-tap linear transversal filter. For all multi-channel FCs, the first sub-filter, which corresponds to the linear system response, has 128 taps. All other sub-filters have 32 taps.

Finally, the baseline power filter approach has $P = 5$ sub-filters, each with 287 taps. Neither the EMD-based nor power filter approach uses orthogonalization since; with the number of sub-filter taps used in these experiments, it does not improve performance [3].

4.2. Experimental results

ERLE results for the EMD and the baseline power filter approaches to NAEC are illustrated in Fig. 4 for a common test speech signal. The EMD approach is shown to outperform the baseline system; it attains a higher level of ERLE, around 8-10 dB more than the baseline. The use of different orders of power filters provides a convenient means of improving NAEC performance, thus minimising gradient noise due to over-modelling.

Fig. 4 also illustrates the response of each approach upon initialisation and to a discrete change in the acoustic echo path which occurs at approximately 5 seconds. In both cases the EMD approach is shown to converge more rapidly than the baseline system. This is due to the lower spectral dynamic range in each IMF compared to the full-band signal in the baseline approach.

While the proposed EMD-based NAEC not only delivers greater average echo attenuation, faster convergence and thus better performance in the case of a dynamically changing acoustic path, it is not without cost. This entails increased computational complexity, principally due to the EMD decomposition and the use of multiple FCs. While there is scope to reduce the computational load via further optimisation, the current system is approximately 1.8-times more demanding in terms of computation.

5. CONCLUSIONS

This paper reports the first application of empirical mode decomposition (EMD) to non-linear acoustic echo cancellation (NAEC). The EMD solution entails the decomposition of the microphone signal into intrinsic mode functions and their utilisation in otherwise-conventional echo cancellation using adaptive filtering. When compared to the power filter baseline system, experimental results demonstrate improved NAEC performance in terms of greater echo reduction and faster convergence. The proposed structure is also more robust to dynamic changes in the acoustic channel. While a modest increase in computational complexity is a drawback, there is scope to reduce this through further optimisation.

[1] M varies for each speech signal; it depends on the stopping criteria used in the process outlined in Section 2. It is not the purpose of this paper to address such issues which have been analysed in detail elsewhere [12, 27]. Accordingly, the 10th IMF is equivalent to the sum of $r(n)$ and all higher-order IMFs.

6. REFERENCES

[1] F. Kuech, A. Mitnacht, and W. Kellermann, “Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters,” in *Proc. ICASSP*, Mar. 2005.

[2] M. Z. Ikram, “Non-linear acoustic echo cancellation using cascaded Kalman filtering,” in *Proc. ICASSP*, May 2014.

[3] M. I. Mossi, C. Yemdji, N. W. D. Evans, C. Beaugeant, and P. Degry, “Robust and low-cost cascaded non-linear acoustic echo cancellation,” in *Proc. ICASSP*, May 2011.

[4] B. S. Nollett and D. L. Jones, “Nonlinear echo cancellation for hands-free speakerphones,” in *Proc. NSIP*, Sept. 1997.

[5] A. Stenger and W. Kellermann, “Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling,” *Elsevier Signal Processing*, vol. 80, pp. 1747–1760, Sept. 2000.

[6] S. Malik and G. Enzner, “Fourier expansion of hammerstein models for nonlinear acoustic system identification,” in *Proc. ICASSP*, May 2011.

[7] D. Zhou, V. DeBrunner, Y. Zhai, and M. Yeary, “Efficient adaptive nonlinear echo cancellation, using subband implementation of the adaptive volterra filter,” in *Proc. ICASSP*, May 2006, vol. 5.

[8] H. Dai and W. Zhu, “Compensation of loudspeaker nonlinearity in acoustic echo cancellation using raised-cosine function,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, no. 11, Nov 2006.

[9] J. M. Gil-Cacho, M. Signoretto, T. van Waterschoot, M. Moonen, and S. H. Jensen, “Nonlinear acoustic echo cancellation based on a sliding-window leaky kernel affine projection algorithm,” *IEEE Transactions on Audio, Speech, and Language Processing*, Sept. 2013.

[10] J. M. Gil-Cacho, T. Van Waterschoot, M. Moonen, and S.H. Jensen, “Nonlinear acoustic echo cancellation based on a parallel-cascade kernel affine projection algorithm,” in *Proc. ICASSP*, Mar. 2012.

[11] A. N. Birkett and R. A. Goubran, “Acoustic echo cancellation using nlms-neural network structures,” in *Proc. ICASSP*, May 1995.

[12] N. E. Huang, Z. Shen, et al., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, Mar. 1998.

[13] N. Chatlani and J. J. Soraghan, “Speech enhancement using adaptive empirical mode decomposition,” in *Proc. DSP*, July 2009.

[14] L. Zao, R. Coelho, and P. Flandrin, “Speech Enhancement with EMD and Hurst-Based Mode Selection,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 5, pp. 899–911, May 2014.

[15] N. Chatlani and J. J. Soraghan, “EMD-Based Filtering (EMDF) of Low-Frequency Noise for Speech Enhancement,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, May 2012.

[16] B. Mijovic, M. De Vos, I Gligorijevic, J. Taelman, and S. Van Huffel, “Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, Sept. 2010.

[17] Md. K. I Molla, K. Hirose, S. K. Roy, and S. Ahmad, “Adaptive thresholding approach for robust voiced/unvoiced classification,” in *Proc. ISCAS*, May 2011.

[18] S. K. Roy and W. P. Zhu, “Pitch estimation of noisy speech using ensemble empirical mode decomposition and dominant harmonic modification,” in *Proc. CCECE*, May 2014.

[19] X. He, R. Goubran, and P. Liu, “A novel sub-band adaptive filtering for acoustic echo cancellation based on empirical mode decomposition algorithm,” *International Journal of Speech Technology*, vol. 17, no. 1, 2014.

[20] G. Rilling, P. Flandrin, and P. Gonçalves, “On empirical mode decomposition and its algorithms,” in *Proc. NSIP*, June 2003.

[21] P. Flandrin, P. Gonalves, and G. Rilling, *Hilbert-Huang Transform and its Applications*, chapter 4, World Scientific, 2005.

[22] L. K. Gudupudi, C. Beaugeant, N. W. D. Evans, M. I. Mossi, and L. Lepauloux, “A comparison of different loudspeaker models to empirically estimated nonlinerities,” in *Proc. HSCMA*, May 2014.

[23] L. K. Gudupudi, C. Beaugeant, and N. W. D Evans, “Characterization and modelling of non-linear loudspeakers,” in *Proc. IWAENC*, Sept. 2014.

[24] A. Farina, A. Bellini, and E. Armelloni, “Non-linear convolution: A new approach for the auralization of distorting systems,” in *Audio Engineering Society Convention 110*, May 2001.

[25] A. Novak, L. Simon, F. Kadlec, and P. Lotton, “Nonlinear system identification using exponential swept-sine signal,” *IEEE Transactions on Instrumentation and Measurement*, Aug. 2010.

[26] M. Jeub et al., “Aachen Impulse Response (AIR) database,” 2009, [Online].

[27] P. Flandrin et al., “Matlab/C codes for EMD and EEMD with examples,” 2007, [Online].