A Nonlinear Acoustic Echo Canceller with Improved Tracking Capabilities

Danilo Comminiello, Michele Scarpiniti, Simone Scardapane, Raffaele Parisi and Aurelio Uncini

Abstract This paper introduces the use of a variable step size for a functional link adaptive filter (FLAF). We consider a split FLAF architecture, in which linear and nonlinear filterings are performed in two separate paths, thus resulting well-suited for online filtering applications, like the nonlinear acoustic echo cancellation (NAEC). We focus our attention on the nonlinear path to improve the overall NAEC performance. To this end, we derive a variable step size for the filter on the nonlinear path that shows reliance not only on the nonlinear path, but on the whole split FLAF architecture. The introduction of the variable step size for the nonlinear filter aims at improving the modeling of nonlinear speech signals, thus yielding superior performance in NAEC problems. Experimental results prove the effectiveness of the proposed method with respect to the standard split FLAF involving a fixed step size.

Keywords Functional links · Nonlinear acoustic echo cancellation · Nonlinear adaptive filtering · Nonlinear speech modeling

The work of Danilo Comminiello was partly funded by bdSound.

D. Comminiello (✉) · M. Scarpiniti · S. Scardapane · R. Parisi · A. Uncini
Department of Information Engineering, Electronics and Telecommunications (DIET)
“Sapienza” University of Rome, Via Eudossiana 18, 00184 Rome, Italy
e-mail: danilo.comminiello@uniroma1.it

M. Scarpiniti
e-mail: michele.scarpiniti@uniroma1.it

S. Scardapane
e-mail: simone.scardapane@uniroma1.it

R. Parisi
e-mail: raffaele.parisi@uniroma1.it

A. Uncini
e-mail: aurel@ieee.org

A. Esposito et al. (eds.), *Recent Advances in Nonlinear Speech Processing*,
Smart Innovation, Systems and Technologies 48,
DOI 10.1007/978-3-319-28109-4_24

1 Introduction

Nonlinear acoustic echo cancellation (NAEC) systems are widely used to model nonlinearities rebounding on acoustic echo paths that affect speech signals in hands-free communication systems. Such nonlinearities are mainly caused by loudspeakers and lead to a quality degradation of a speech communication. NAEC systems reduce the effect of nonlinearities, thus improving echo cancellation performance.

In this paper, we focus on a recently proposed NAEC system, which is based on the use of *functional link adaptive filters* (FLAFs) [5]. These filters are characterized by a nonlinear expansion of the input followed by a linear filtering of the expanded signal. In particular, we take into account a split FLAF (SFLAF) architecture [5], which separates the adaptation of linear and nonlinear elements in two parallel paths, each one devoted to a specific task. This structure is particularly significant in NAEC problems [4–6], since the linear path can be exclusively used to estimate the acoustic impulse response, while the nonlinear path can be committed to model any nonlinearity.

Usually, processing a speech signal is made difficult by its nonstationary nature. Moreover, a nonlinearity, like that produced by a loudspeaker, emphasizes the nonstationarity of a signal, such that modeling a distorted speech signal becomes very difficult. In order to improve the modeling of nonlinearities, the tracking performance of the nonlinear filter should be optimized according to the level of nonlinearity that affects a speech signal at each instant [10]. To this end, we propose the use of a variable step size for the adaptive filter on the nonlinear path of the SFLAF. Variable step sizes have been largely used for adaptive filters in linear system identification problems, such as acoustic echo cancellation and adaptive beamforming [1–3, 8, 11–13, 15]. However, in this paper the variable step size is used to provide improved tracking performance in the presence of nonlinear speech.

The rest of the paper is organized as follows: the SFLAF architecture is described in Sect. 2. In Sect. 3, a variable step size is introduced for the adaptive filter on the nonlinear path, and, in Sect. 4, experimental results are shown. Finally, in Sect. 5 our conclusions are presented.

2 The Split Functional Link Adaptive Filter

The *split functional link adaptive filter* (SFLAF) model [5], depicted in Fig. 1, is a parallel architecture including a linear path and a nonlinear path. The former is simply composed of a linear adaptive filter, which completely aims at modeling the linear components of an unknown system; the nonlinear path is composed of a Hammerstein cascade model comprising a functional expansion block and a subsequent adaptive filter.

At nth time instant the SFLAF receives the input sample $x[n]$, which is stored in the linear input buffer $\mathbf{x}_{\mathrm{L},n} \in \mathbb{R}^M = \left[x[n]\ x[n-1]\ \ldots\ x[n-M+1]\right]^T$, where M

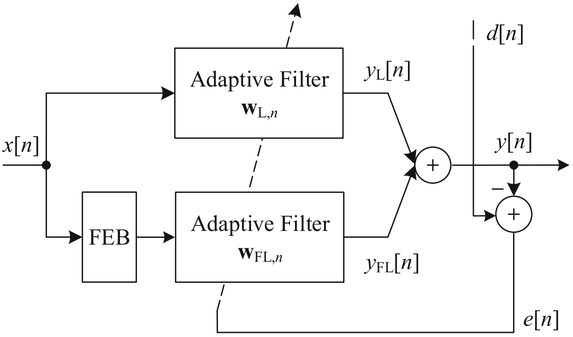


Fig. 1 The split functional link adaptive filter

is the linear input buffer length, i.e., the length of linear adaptive filter $\mathbf{w}_{\mathrm{L},n} \in \mathbb{R}^{M} = \left[\, w_{\mathrm{L},0}\,[n]\; w_{\mathrm{L},1}\,[n] \ldots \; w_{\mathrm{L},M-1}\,[n] \,\right]^{T}$. The adaptive filtering yields the linear output $y_{\mathrm{L}}\,[n] = \mathbf{x}^{T}_{\mathrm{L},n}\mathbf{w}_{\mathrm{L},n-1}$. On the other hand, the nonlinear path receives a subvector of $\mathbf{x}_{\mathrm{L},n}$ as nonlinear input buffer: $\mathbf{x}_{\mathrm{FL},n} \in \mathbb{R}^{M_{\mathrm{i}}} = \left[\, x\,[n]\; x\,[n-1] \ldots x\,[n-M_{\mathrm{i}}+1] \,\right]^{T}$, where $M_{\mathrm{i}} \leq M$ is defined as the nonlinear input buffer length, which can be equal to the linear input buffer length or just a portion of it. The nonlinear path receives the nonlinear buffer $\mathbf{x}_{\mathrm{FL},n}$, which is processed by means of a *functional expansion block* (FEB). The FEB consists of a series of functions, which might be a subset of a complete set of orthonormal basis functions, satisfying universal approximation constraints. The term “functional links” refers to the functions contained in the chosen set $\Phi = \left\{\varphi_0\,(\cdot)\,, \varphi_1\,(\cdot)\,, \ldots, \varphi_{Q_{\mathrm{f}}-1}\,(\cdot)\right\}$, where Q_{f} is the number of functional links. The FEB processes the input buffer by passing each element of the buffer $\mathbf{x}_{\mathrm{FL},n}$ as argument for the chosen functions, each one yielding a subvector $\overline{\mathbf{g}}_{i,n} \in \mathbb{R}^{Q_{\mathrm{f}}}$:

$$\overline{\mathbf{g}}_{i,n} = \left[\, \varphi_0\,(x\,[n-i])\; \varphi_1\,(x\,[n-i]) \ldots \varphi_{Q_{\mathrm{f}}-1}\,(x\,[n-i]) \,\right]. \tag{1}$$

The concatenation of such subvectors yields an *expanded buffer* $\mathbf{g}_n \in \mathbb{R}^{M_{\mathrm{e}}}$:

$$\mathbf{g}_n = \left[\, \overline{\mathbf{g}}^{T}_{0,n}\; \overline{\mathbf{g}}^{T}_{1,n} \cdots \overline{\mathbf{g}}^{T}_{M_{\mathrm{i}}-1,n} \,\right]^{T} \tag{2}$$

where $M_{\mathrm{e}} = Q_{\mathrm{f}} \cdot M_{\mathrm{i}} \geq M_{\mathrm{i}}$ represents the length of the expanded buffer. Note that $M_{\mathrm{e}} = M_{\mathrm{i}}$ only when $Q_{\mathrm{f}} = 1$. The functional expansion chosen for this work is a nonlinear trigonometric series expansion such that:

$$\varphi_j\,(x\,[n-i]) = \begin{cases} \sin\left(p\pi x\,[n-i]\right), & j = 2p-2 \\ \cos\left(p\pi x\,[n-i]\right), & j = 2p-1 \end{cases} \tag{3}$$

where $p = 1, \ldots, P$ is the expansion index, being P the *expansion order*, and $j = 0, \ldots, Q_{\mathrm{f}} - 1$ is the functional link index. Therefore, in the case of trigonometric expansion, the functional link set Φ is composed of $Q_{\mathrm{f}} = 2P$ functional links. Convergence performance of a trigonometric FLAF is shown in [9]. Note that (3) actually refers to a *memoryless* expansion, since it does not involve cross-products of the nth input sample with previous samples. However, the same process holds also for functional expansion with memory. The choice of involving some memory may be decisive when the nonlinearity introduced by the system to be identified is characterized by a dynamic nature, i.e., depends also on the time instant. In our model, we consider the memory of a nonlinearity by taking into account the outer products of the ith input sample with the functional links of the previous K input samples, where K represents the *memory order* (see [5] for a detailed explanation).

The achieved expanded buffer $\mathbf{g}_n$ is then fed into a linear adaptive filter $\mathbf{w}_{\mathrm{FL},n} \in \mathbb{R}^{M_{\mathrm{e}}} = \left[\, w_{\mathrm{FL},0}\left[n\right] \; w_{\mathrm{FL},1}\left[n\right] \; \ldots \; w_{\mathrm{FL},M_{\mathrm{e}}-1}\left[n\right] \right]^T$, thus providing the nonlinear output $y_{\mathrm{FL}}\left[n\right] = \mathbf{g}_n^T \mathbf{w}_{\mathrm{FL},n-1}$. The SFLAF output results from the sum of the two path outputs:

$$y\left[n\right] = y_{\mathrm{L}}\left[n\right] + y_{\mathrm{FL}}\left[n\right] \tag{4}$$

and, thereby, the overall error signal[1] is:

$$e\left[n\right] = d\left[n\right] - y\left[n\right] = d\left[n\right] - \mathbf{x}_{\mathrm{L},n}^T \mathbf{w}_{\mathrm{L},n-1} - \mathbf{g}_n^T \mathbf{w}_{\mathrm{FL},n-1}, \tag{5}$$

which is used for the adaptation of both adaptive filters. In (5), $d\left[n\right]$ represents the desired signal including any near-end additive noise $v\left[n\right]$ and a near-end speech contribution $s\left[n\right]$. We use a standard *normalized least-mean square* (NLMS) algorithm (see for example [14, 16]) to adapt the coefficients of both $\mathbf{w}_{\mathrm{L},n}$ and $\mathbf{w}_{\mathrm{FL},n}$:

$$\mathbf{w}_{\mathrm{L},n} = \mathbf{w}_{\mathrm{L},n-1} + \mu_{\mathrm{L}} \frac{\mathbf{x}_{\mathrm{L},n} e\left[n\right]}{\mathbf{x}_{\mathrm{L},n}^T \mathbf{x}_{\mathrm{L},n} + \delta_{\mathrm{L}}} \tag{6}$$

$$\mathbf{w}_{\mathrm{FL},n} = \mathbf{w}_{\mathrm{FL},n} + \mu_{\mathrm{FL}}\left[n\right] \frac{\mathbf{g}_n e\left[n\right]}{\mathbf{g}_n^T \mathbf{g}_n + \delta_{\mathrm{FL}}} \tag{7}$$

where δ_{L} and δ_{FL} are regularization factors, and μ_{L} is a fixed step size for the filter on the linear path, and $\mu_{\mathrm{FL}}\left[n\right]$ is the variable step size parameter, on which we focus in the next section in order to improve the nonlinear modeling performance.

3 A Variable Step Size for the Nonlinear FLAF

In this section, we derive the *variable step size* $\mu_{\mathrm{FL}}\left[n\right]$ of (7), thus providing a reliable solution to nonlinear speech modeling. In order to yield an algorithm easy to control in practical implementations, similarly to what done in [12] for linear echo cancellation, the derivation is taken considering that no *a priori* information must be required about the nonlinearity to be modeled.

[1] It may also be denoted as *a priori* output estimation error [14] to be distinguished from the *a priori* estimation error, which is defined as $e_{\mathrm{a}}\left[n\right] = e\left[n\right] - v\left[n\right]$, where $v\left[n\right]$ is additive noise.

We start from the consideration that the desired signal $d[n]$ is composed of a signal $\widetilde{x}[n]$, which is generated by the far-end signal convolved with an acoustic impulse response and distorted by any nonlinear process. The desired signal also contains a near-end contribution $s[n]$ and any additive noise $v[n]$, therefore, it can be written as:

$$d[n] = \widetilde{x}[n] + s[n] + v[n] \tag{8}$$

In order to derive the optimal variable step size parameter, we assume that $\widetilde{x}[n]$, $s[n]$ and $v[n]$ are statistically uncorrelated and we take the squares and the expectations of both sides of (8), thus resulting:

$$\mathrm{E}\left\{d^2[n]\right\} = \mathrm{E}\left\{\widetilde{x}^2[n]\right\} + \mathrm{E}\left\{s^2[n]\right\} + \mathrm{E}\left\{v^2[n]\right\} \tag{9}$$

According to the *least perturbation property* [14], at steady state, i.e., for $n \to \infty$, the weights of an adaptive filter no longer change during the adaptation. Therefore, it is reasonable to assume the following approximation:

$$\mathrm{E}\left\{\widetilde{x}^2[n]\right\} \approx \mathrm{E}\left\{y^2[n]\right\} + \mathrm{E}\left\{q^2[n]\right\}. \tag{10}$$

In (10) an irreducible noise term $q[n]$ has been introduced due to the nonlinear approximation. Therefore, Eq. (9) turns into the following one:

$$\mathrm{E}\left\{d^2[n]\right\} - \mathrm{E}\left\{y^2[n]\right\} = \mathrm{E}\left\{s^2[n]\right\} + \mathrm{E}\left\{v^2[n]\right\} + \mathrm{E}\left\{q^2[n]\right\}. \tag{11}$$

The right member of (11) contains the near-end contribution and the irreducible noise, that may be approximated to the *a posteriori* output estimation error $\varepsilon[n]$ at steady state [3, 14]. Therefore, Eq. (11) can be written as:

$$\mathrm{E}\left\{d^2[n]\right\} - \mathrm{E}\left\{y^2[n]\right\} \approx \mathrm{E}\left\{\varepsilon^2[n]\right\}. \tag{12}$$

However, in order to achieve the optimal $\mu_{\mathrm{FL}}[n]$, we need to express $\varepsilon[n]$ in terms of the *a priori* error $e[n]$. A relation between the *a posteriori* and *a priori* error signals may be derived starting from the definition of $\varepsilon[n]$:

$$\varepsilon[n] = d[n] - \mathbf{x}_n^T \mathbf{w}_{\mathrm{L},n} - \mathbf{g}_n^T \mathbf{w}_{\mathrm{FL},n} \tag{13}$$

Replacing the update Eqs. (6) and (7) in (13), and taking into account the *a priori* error signal definition (5), it is possible to achieve the following relation:

$$\varepsilon[n] = \left(1 - \mu_{\mathrm{L}} - \mu_{\mathrm{FL}}[n]\right) e[n] \tag{14}$$

The step size μ_{L} in (14) may be a fixed value or even a variable parameter achieved by any variable step size technique. However, the goal of the paper is to investigate the effects of a variable step size for the nonlinear modeling and thus we consider μ_{L} as a fixed value. Therefore, we can replace Eq. (14) in (12), thus resulting:

$$\mathrm{E}\left\{d^2[n]\right\} - \mathrm{E}\left\{y^2[n]\right\} = \left|1 - \mu_{\mathrm{L}} - \mu_{\mathrm{FL}}[n]\right|^2 \mathrm{E}\left\{e^2[n]\right\} \tag{15}$$

from which we can derive an expression of the variable step size parameter $\mu_{\mathrm{FL}}[n]$:

$$\mu_{\mathrm{FL}}[n] = \left|1 - \mu_{\mathrm{L}} - \sqrt{\frac{\mathrm{E}\left\{d^2[n]\right\} - \mathrm{E}\left\{y^2[n]\right\}}{\mathrm{E}\left\{e^2[n]\right\}}}\right|. \tag{16}$$

From a practical point of view, we evaluate the expectations in terms of power estimates, as done for example in [12], thus achieving:

$$\mu_{\mathrm{FL}}[n] = \left|1 - \mu_{\mathrm{L}} - \frac{\sqrt{\left|\widehat{\sigma}_d^2[n] - \widehat{\sigma}_y^2[n]\right|}}{\widehat{\sigma}_e^2[n] + \xi}\right|. \tag{17}$$

In (17), the general parameter $\widehat{\sigma}_\theta^2[n]$ represents the power estimate of the sequence $\theta[n]$, being $\theta = \{d, y, e\}$, and it can be computed as:

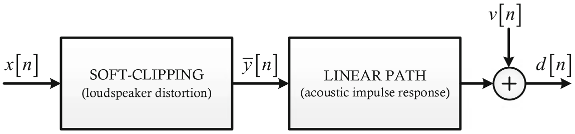
$$\widehat{\sigma}_\theta^2[n] = \beta\widehat{\sigma}_\theta^2[n-1] + (1-\beta)\,\theta^2[n] \tag{18}$$

where β is a forgetting factor, whose value can be chosen as $\beta = 0.99$. A small positive number $\xi = 10^{-4}$ is added in (17) to avoid divisions by zero. Another practical consideration is that, in presence of high dynamic nonlinearities, the power of the estimate of the output signal $\widehat{\sigma}_y^2[n]$ may be larger than the power of the desired signal $\widehat{\sigma}_d^2[n]$; this is the reason why the absolute value of the terms under the square root is considered.

4 Experimental Results

We assess the effectiveness of the proposed FLAF-based architecture in a nonlinear acoustic echo cancellation problem. Experiments take place in a simulated room environment with a reverberation time of $T_{60} \approx 100$ ms measured at 8 kHz sampling frequency. A far-end signal $x[n]$ is reproduced by a simulated loudspeaker and captured by a microphone. In order to have a complete view of the effects of the nonlinearity, we use both a colored noise and a speech signal as far-end input. The colored noise signal is generated by means of a first-order autoregressive model, whose transfer function is $\sqrt{1-\theta^2}/\left(1-\theta z^{-1}\right)$, with $\theta = 0.8$, fed with an independent and identically distributed (i.i.d.) Gaussian random process. In order to simulate a loudspeaker distortion, we apply a symmetrical soft-clipping nonlinearity to the far-end signal [6, 7]:

Fig. 2 Scheme of the NAEC system



$$\bar{y}[n] = \begin{cases} \frac{2}{3\zeta} x[n] & , \; 0 \le |x[n]| \le \zeta \\ \operatorname{sign}(x[n]) \frac{3-(2-|x[n]|/\zeta)^2}{3} & , \; \zeta \le |x[n]| \le 2\zeta \\ \operatorname{sign}(x[n]) & , \; 2\zeta \le |x[n]| \le 1 \end{cases} \tag{19}$$

where $0 < \zeta \le 0.5$ is a nonlinearity threshold. As also described by Fig. 2, the signal $\bar{y}[n]$ is then convolved with an acoustic impulse response related to the simulated room environment, thus achieving the desired signal $d[n]$ acquired by a microphone. The signal $d[n]$ contains also near-end background noise $v[n]$, in the form of additive Gaussian noise, providing 20 dB of *signal-to-noise ratio* (SNR). The length of the acoustic impulse response is $M = 300$.

Performance is evaluated in terms of the *echo return loss enhancement* (ERLE), expressed in dB as: $\text{ERLE}[n] = 10\log_{10}\left(\text{E}\left\{d^2[n]\right\}/\text{E}\left\{e^2[n]\right\}\right)$. We use the following parameter setting: input buffer length $M_i = M$, fixed step-size parameter $\mu_L = 0.2$, regularization parameter $\delta_{FL} = 10^{-2}$ for both the filters of the SFLAF, expansion order $P = 10$, memoryless functional links (i.e., $K = 0$), and distortion threshold $\zeta = 0.15$. We compare the results of the proposed VSS-SFLAF with a SFLAF having the same parameter setting of the VSS-SFLAF, but a fixed step-size value $\mu_{FL} = 0.2$.

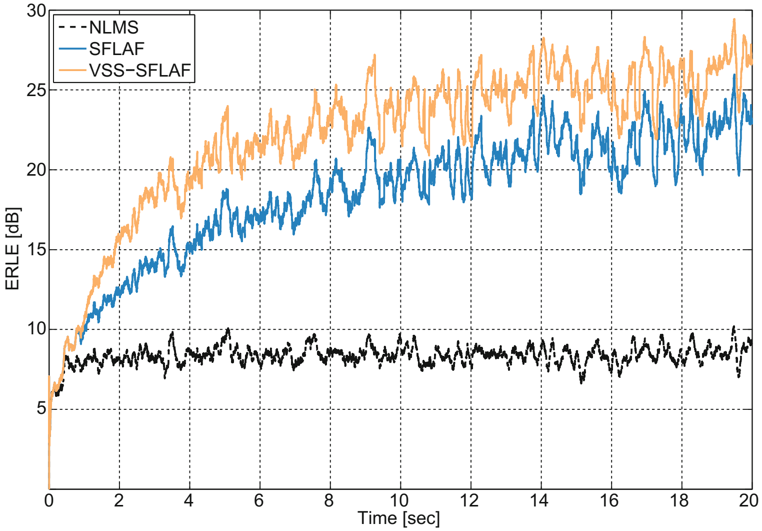


Fig. 3 Performance behavior in terms of ERLE in case of colored noise input

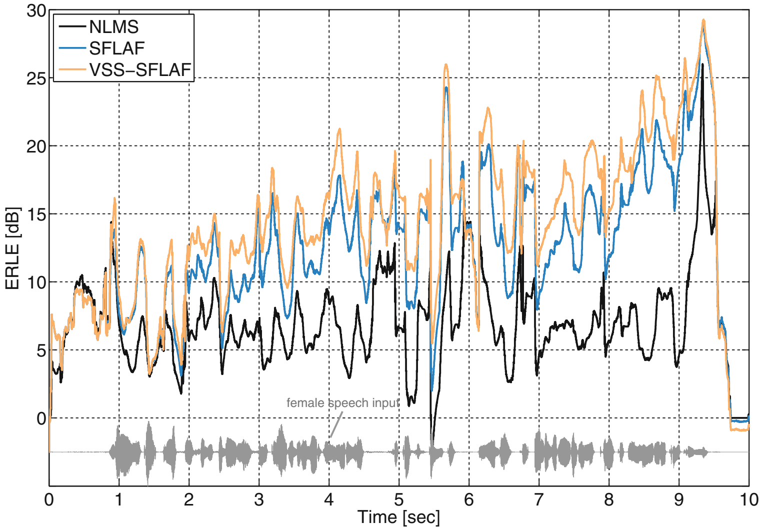


Fig. 4 Performance behavior in terms of ERLE in case of female speech input

In case of colored noise, the VSS-SFLAF achieves a good improvement in terms of tracking performance over the SFLAF, as it is possible to see in Fig. 3, while tending to a similar behavior at steady state. Results becomes more significant when using an input signal with a high nonstationarity, i.e., a speech signal, since performance improvements are more difficult to be obtained in this case. As depicted in Fig. 4, a gain over the SFLAF can be achieved, not only in proximity of the peaks of the speech signal, but for the whole length of the signal.

5 Conclusions

In this paper, a functional link-based nonlinear acoustic echo canceller is proposed involving a variable step size on the nonlinear path of the architecture. The proposed VSS-SFLAF takes advantage from the use of the variable step size, thus improving the tracking performance of nonlinear speech signals. Future research will include the use of a joined variable step size that governs the convergence performance of both filters on the linear and nonlinear paths.

References

1. Aboulnasr, T., Mayyas, K.: A robust variable step-size LMS-type algorithm: analysis and simulations. IEEE Trans. Signal Process. **45**(3), 631–639 (1997)
2. Albu, F., Coltuc, D., Comminiello, D., Scarpiniti, M.: The variable step size regularized block exact affine projection algorithm. In: Proceedings of the IEEE International Symposium on Electronics and Telecommunications (ISETC), pp. 283–286. Timisoara, Romania, Nov 2012
3. Benesty, J., Rey, H., Vega, L.R., Tressens, S.: A nonparametric VSS NLMS algorithm. IEEE Signal Process. Lett. **13**(10), 581–584 (2006)
4. Comminiello, D., Azpicueta-Ruiz, L.A., Scarpiniti, M., Uncini, A., Arenas-García, J.: Functional link based architectures for nonlinear acoustic echo cancellation. In: Proceedings of the IEEE Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA), pp. 180–184. Edinburgh, UK, May 2011
5. Comminiello, D., Scarpiniti, M., Azpicueta-Ruiz, L.A., Arenas-García, J., Uncini, A.: Functional link adaptive filters for nonlinear acoustic echo cancellation. IEEE Trans. Audio Speech Lang. Process. **21**(7), 1502–1512 (2013)
6. Comminiello, D., Scarpiniti, M., Azpicueta-Ruiz, L.A., Arenas-García, J., Uncini, A.: Nonlinear acoustic echo cancellation based on sparse functional link representations. IEEE/ACM Trans. Audio Speech Lang. Process. **7**(22), 1172–1183 (2014)
7. Comminiello, D., Scardapane, S., Scarpiniti, M., Parisi, R., Uncini, A.: Functional Link Expansions for Nonlinear Modeling of Audio and Speech Signals. In: Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1–8. Killarney, Ireland, Jul 2015
8. Comminiello, D., Scarpiniti, M., Parisi, R., Uncini, A.: A novel affine projection algorithm for superdirective microphone array beamforming. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2127–2130. Paris, France, May 2010
9. Comminiello, D., Scarpiniti, M., Parisi, R., Uncini, A.: Convergence properties of nonlinear functional link adaptive filters. IET Electron. Lett. **49**(14), 873–875 (2013)
10. Comminiello, D., Scarpiniti, M., Scardapane, S., Parisi, R., Uncini, A.: Improving nonlinear modeling capabilities of functional link adaptive filters. Neural Networks **69**, 51–59 (2015)
11. Huang, H.C., Lee, J.: A new variable step-size nlms algorithm and its performance analysis. IEEE Trans. Signal Process. **60**(4), 2055–2060 (2012)
12. Paleologu, C., Benesty, J., Ciochină, S.: A variable step-size affine projection algorithm designed for acoustic echo cancellation. IEEE Trans. Audio Speech Lang. Process. **16**(8), 1466–1478 (2008)
13. Vega, Rey: L., Rey, H., Benesty, J., Tressens, S.: A new robust variable step-size nlms algorithm. IEEE Trans. Signal Process. **56**(5), 1878–1893 (2008)
14. Sayed, A.H.: Adaptive Filters. John Wiley & Sons, Hoboken, NJ (2008)
15. Shin, H.C., Sayed, A.H., Song, W.J.: Variable step-size NLMS and adffine projection algorithms. IEEE Signal Process. Lett. **11**(2), 132–135 (2004)
16. Uncini, A.: Fundamentals of Adaptive Signal Processing. Springer International Publishing AG, Cham, Switzerland, Signal and Communication Technology (2015)