Efficient Affine Projection Tanh Algorithm for Acoustic Feedback Cancellation

Felix Albu
Department of Electronics
Valahia University of Targoviste
Targoviste, Romania
felix.albu@valahia.ro

Marius Ionita
Department of Electronics
Valahia University of Targoviste
Targoviste, Romania
marius.ionita@valahia.ro

***Abstract*—In this paper, a low-complexity implementation of the recently proposed affine projection tanh algorithm in conjunction with a frequency shifting performed on speech segments for acoustic feedback cancellation (AFC) is presented. Dichotomous coordinate descent (DCD) iterations having a variable input parameter are used to reduce its numerical complexity. The simulation results show that the proposed approach can achieve better performance than competing methods for both incoming speech and music signals.**

Keywords—dichotomous coordinate descent, affine projection tanh algorithm, acoustic feedback cancellation, prediction error method, frequency shifting, voice activity detection.

I. INTRODUCTION

In hearing aid (HA) devices, the inherent acoustic feedback loop between amplified microphone signals played through loudspeakers leads to reverberation echoes and howling [1]-[4]. The feedback problem is caused by the small-sized and open-fitting HAs [2]. One suboptimal solution is feedforward suppression using notch filtering [3]. A better solution is to use feedback cancellation adaptive algorithms [1],[2]. The bias introduced by the adaptive filter can be reduced using various approaches: frequency shifting [4] – [5], probe noise injection [6], two microphone solutions [7]-[8], the prediction-error-method AFC (PEM-AFC) [1], etc. The PEM-AFC involves using the inverse of the estimated all-pole filter to prefilter the loudspeaker and microphone signals before adapting the AFC. A robust AFC method based on the affine projection algorithm (APA) using the improved practical variable step size (IPVSS) from [9], the prediction error method (PEM-IPVSS-APA) and frequency shifting (FS) was proposed in [10]. Hybrid implementations reduce the duration of the howling (e.g. the hybrid normalized least mean square (H-NLMS) algorithm [11], the combination of the PEM with the simplified Kalman filter (SKF) [12], hybrid approach using the maximum Versoria-criterion (MVC)-based algorithm [13] and the switched PEM with a soft-clipping (swPEMSC) algorithm [14]. A low complexity of the later algorithm using dichotomous coordinate descent iterations proposed in [15] has been presented in [16] and called swPEMSC-DCD.

In this paper, we propose an improved version of the algorithm of [16]. Three novel contributions of this study are presented. First, the affine projection tanh algorithm (APTA) [17] was used instead of the APA as in [16]. In [18] it was shown that, for identification purposes, it is better to use a controlled tanh non-linearity on the loudspeaker output signal. In [11], the tanh non-linearity is applied on the error signal. The novelty and advantage of using APTA is that the tanh non-linearity is applied on the preprocessed error signals with the pre-whitening filter. Secondly, its low complexity is achieved by using a new variable input parameter for the dichotomous coordinate descent (DCD) iterations [15] in APTA (DCD-APTA). Combining both the APTA and modified DCD have not yet been used for AFC. Thirdly, the frequency shifting is used instead of the probe noise as in [20] and is applied only when a voice activity detector (VAD) detects speech, not all the time as in previous studies [10]. The acronym used for the proposed algorithm is swPEMSC-DCD-APTA. The remainder of this paper is organized as follows. Section II presents the swPEMSC-DCD-APTA, and Section III presents the simulation results. Finally, the conclusions and future work are presented.

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS - UEFISCDI, project number PN-III-P4-PCE-2021-0780, within PNCDI.

II. THE PROPOSED ALGORITHM

Figure 1 illustrates the scheme of the proposed method. It is based on the scheme of [14], with important changes related to the used swPEMSC-DCD-APTA and the FS + VAD block.

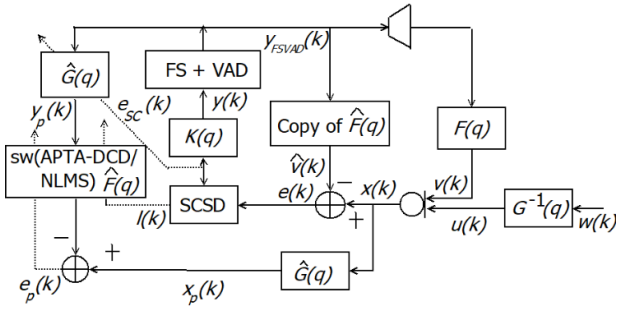

Fig. 1. The proposed AFC scheme

The microphone signal is the sum of the incoming signal, $u(k)$, and the feedback signal, $v(k)$:

$$x(k) = u(k) + F(q)\,y_{FSVAD}(k), \tag{1}$$

where k is the time index, $y_{FSVAD}(k)$ is the loudspeaker signal, $v(k) = F(q)\,y_{FSVAD}(k)$, $F(q) = \mathbf{f}^T\mathbf{q}$, $\mathbf{f}$ is the impulse response of the feedback path having L_f coefficients, $\mathbf{q} = \left[1\ q^{-1}\ \ldots\ q^{-L_f+1}\right]^T$, q^{-1} is the time-shift operator, $y(k) = K(q)e(k)$, where $K(q) = |K|q^{-d_k}$, d_k and $|K|$ are the delay and the gain of the forward path respectively. The error signal given by

$$e(k) = x(k) - \hat{F}(q)\,y_{FSVAD}(k), \tag{2}$$

where $\hat{F}(q)$ is the estimated feedback path having $L_{\hat{f}}$ coefficients. Also, we have $u(k) = G^{-1}(q)w(k)$, (common

assumption for PEM [1], $w(k)$ is a white Gaussian noise) $x_p(k) = \hat{G}(q)x(k)$, $y_{FSVAD}(k)$ is the loudspeaker signal and output of the block that performs a frequency shifting taking into account the output of the VAD.

The pre-whitened signals from Fig. 1 are $e_p(k) = x_p(k) - \hat{F}(q)y_p(k)$, $y_p(k) = \hat{G}(q)y_{FSVAD}(k)$, and $x_p(k) = \hat{G}(q)x(k)$. The AR coefficients are found using the Levinson-Durbin method [11]. The soft clipping stability detector (SCSD) is used to switch between the prediction error method using soft clipping (PEMSC) and NLMS (PEMSC-NLMS) and PEMSC using APTA-DCD (PEMSC-APTA-DCD). As in [11], [12] and [14] the approach allows us to deal with both situations when the system is close to instability or has converged. It is using a nonlinear function of the error signal [11]:

$$l(k) = L\left\{\left|e(k) - e_{SC}(k)\right| < \theta\right\} \tag{3}$$

$$e_{SC}(k) = \rho \tanh\left(e(k)/\rho\right) \tag{4}$$

where L is a binary function (1 for a positive value and 0 otherwise), θ is a threshold and ρ is a parameter [11]. The signal $y(k) = K(q)e_{SC}(k)$ passes through a frequency-shifting block only during the speech segments detected by the VAD, otherwise it is left unchanged. When the VAD detects silence $y_{FSVAD}(k) = y(k)$, otherwise, if detects speech we have:

$$y_{FSVAD}(k) = y(k)\cos(2\pi f_0 k) - y_H(k)\sin(2\pi f_o k), \tag{5}$$

where $y_H(k)$ is the Hilbert transform of $y(k)$ [10]. Because the FS adds roughness for high f_o values, a small value of f_o is recommended [5]. We used $f_o = 3$ in our simulations as in [16].

By replacing the APA part with APTA [17], the update rule for the swPEMSC-APTA is the following:

$$\hat{\mathbf{f}}(k) = \hat{\mathbf{f}}(k-1) + \frac{\mu_1 l(k)\mathbf{y}_p(k)e_p(k)}{\mathbf{y}_p^T(k)\mathbf{y}_p(k) + \delta_{NLMS}} + \mu_2\left[1 - l(k)\right]$$
$$\cdot\mathbf{Y}_p(k)\left[\mathbf{Y}_p^T(k)\mathbf{Y}_p(k) + \delta_{APTA}\mathbf{I}_N\right]^{-1}\tanh\left(\beta\mathbf{e}_p(k)\right), \tag{6}$$

where N is the projection order, $\mathbf{I}_N$ is the identity matrix, $\mathbf{Y}_p(k)$ is a $L_{\hat{f}} \times N$ matrix collecting N recent pre-whitened loudspeaker signal vectors of length $L_{\hat{f}}$ $\mathbf{y}_p(k) = \left[y_p(k),...,y_p\left(k - L_{\hat{f}} + 1\right)\right]^T$ vectors, μ_1 and μ_2 are step-sizes ($\mu_2 >> \mu_1$), β is the APTA parameter, δ_{NLMS} and δ_{APTA} are the regularization parameters for the NLMS and APTA parts, respectively.

The update rule of Eq. 6 for high projection orders is very complex, because it involves the inversion of a $N \times N$ matrix. The numerical complexity of this part is proportional with N^3, the number of multiplications and divisions being $\left(N^2 + 2N\right)L_{\hat{f}} + N^3 + N$ [14]. Therefore, the overall complexity of the update rule must be reduced for high-projection-order values. DCD iterations are proposed to solve the linear system of Eq. 6:

$$\left[\mathbf{Y}_p^T(k)\mathbf{Y}_p(k) + \delta_{APTA}\right]\mathbf{z}(k) = \tanh\left(\beta\mathbf{e}_p(k)\right). \tag{7}$$

We then use the $\mathbf{z}(k)$ solution to obtain the following swPEMSC-DCD-APTA update rule:

$$\hat{\mathbf{f}}(k) = \hat{\mathbf{f}}(k-1) + \frac{\mu_1 l(k)\mathbf{y}_p(k)e_p(k)}{\mathbf{y}_p^T(k)\mathbf{y}_p(k) + \delta_{NLMS}} +$$
$$\mu_2\left[1 - l(k)\right]\mathbf{Y}_p(k)\mathbf{z}(k). \tag{8}$$

A. Dichotomous coordinate descent

The equations of the multiplier-less and division-less dichotomous coordinate descent algorithm version from Table IX in [19] are used to find the $\mathbf{z}(k)$ vector. The accuracy of the DCD solution of the linear system depends on the maximum expected amplitude H of the solution (chosen as a power of two), the number of bits M_b and the number of updates, N_u. The DCD iterations are 'successful' if the solution and the residual vector are updated otherwise, they are 'unsuccessful' [19]. In the worst-case scenario, the maximum number of additions in our case is $N\left(2N_u + M_b - 1\right) + N_u$ [19]. The use of DCD provides an important complexity reduction in terms of multiplications and divisions, especially if the N value is high. The value of H is fixed at each iteration, it should be high enough, otherwise, the convergence could be too slow [19]. We observed that the maximum value of the solution $\mathbf{z}(k)$ can vary a lot during disturbances (howling, impulsive noises, ill-conditioned matrices). The ceiling of the closest value to the log2 of the absolute maximum elements of the previous solution is computed. This exponent value is allowed to be within a minimum and a maximum number depending on the desired range. The H value for the current algorithm iteration is the power of two of the computed exponent value. These computations add $N+4$ comparisons and a logarithmic operation to the overall complexity.

III. SIMULATION RESULTS

The performance of the considered algorithms was investigated using the feedback path characteristics shown in Fig. 2. They were also used in [10], [14] and [21]. Speech (male and female voices from NOIZEUS database [22]) and music signals of 50 seconds were used as incoming signals. The tracking behavior was investigated by switching after 25 seconds from the normal path (F_1) to the closest feedback path (F_2) [14]. The equations of misalignment (MIS) and added stable gain (ASG) from [10] were used to evaluate the performance of the AFC system.

The perceptual evaluation of speech quality (PESQ) measure [23] was computed for speech signals. The order of the prediction error filter was 20, the sampling frequency was $f_s = 16$ kHz, the forward path gain was 30 dB, $L_f = 100$, $L_{\hat{f}} = 64$, $d_K = 96$, $d_{f_b} = 1$, $M_b = 16$, $\mu_U = 0.8$, $\mu_L = 0.0008$, the projection order was $N = 4$, and the regularization factor values are those from Table I of [10].

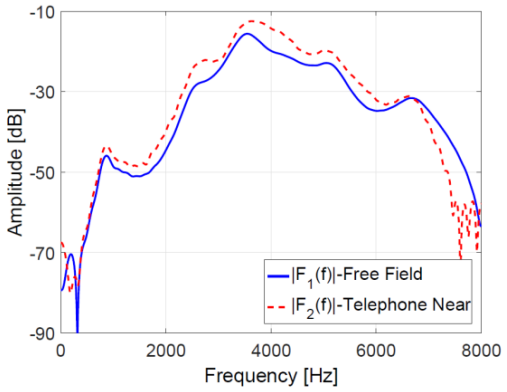

Fig. 2. The amplitude responses of the measured feedback paths [10].

The remaining parameters were $\gamma = 0.9999$, $\theta = 0.15$, $\rho = 2$, β was 0.125 for simulations shown in Figs. 4-6, and $\delta_{APTA} = \delta_{NLMS} = 10^{-6}$. The low complexity VAD from [24] was used. We assured that at least the first 6 frames were silence frames as proposed in [24] and the same parameters from [24] were used (overlapping frames of 20 ms by 50%, -3-frame neighborhood, offset of 5 dB, $\alpha = 0.95$, $\gamma_0 = 6$ dB, $\gamma_1 = 2.5$ dB, $E_0 = 30$ dB, $E_1 = 50$ dB, NFFT=256).

In Fig. 3, the norm of the difference between the ideal solution for the implicit system of equations (Eq. 7) and the solution using various numbers of updates is plotted ($N_u = 1$ in Fig. 3a and $N_u = 4$, and $N_u = 8$ in Fig. 3b).

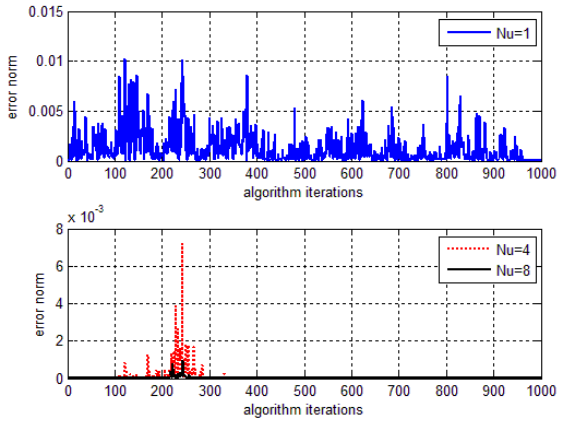

Fig. 3. The error norm for various N_u values; a) $N_u = 1$ b) $N_u = 4$ and $N_u = 8$

As expected from other DCD implementations (e.g. [19], [25]-[26]) using more DCD iterations leads to a performance closer to that obtained using the exact solution for the linear system of equations. It can be noticed that the smallest errors are obtained for $N_u = 8$. The attained precision level is sufficient for the proposed algorithm and this value was used for the DCD iterations in further simulations.

In Fig. 4 the influence of using a variable H value for the DCD iterations is investigated. In Fig. 4a the MIS performance difference between the MIS performances of the proposed algorithm obtained when using a fixed H value of $2^{-1} = 0.5$ and a variable H value respectively. The incoming signal was a speech signal. The exponent of H was allowed to vary around -2 (i.e. between -2 and 1). Therefore, H takes one of the the following values depending on the maximum vector solution absolut value: 0.25, 0.5, 1, 2 and 4. The power of two exponents of H are shown in Fig. 4b.

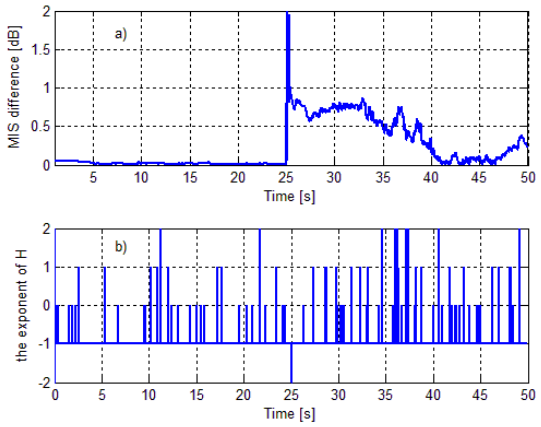

Fig. 4. The MIS performance using music input signal with a sudden change of the feedback path after 25 seconds b) The exponent values of H.

It can be noticed that while the MIS performance improvement is rather small for F_1, it is higher for F_2 (about 0.5 dB on average). It can be concluded that a variable H value could lead to performance improvement due to a better adaptation of the DCD approach since H initialize its step value [19].

In Fig. 5, the influence of the FS for an incoming music signal is examined by comparing the MIS and ASG performances of the proposed algorithm with and without the FS. It can be seen that the FS block has a positive effect most of the time; both the average MIS and ASG values are better with about 3 dB.

Similar conclusions can be obtained when using speech signals and the FS + VAD block is used. The reason to use a VAD is that the howling was less annoying when frequency shifting was not performed on silence frames. It can be seen from Fig. 6b that the ASG improvement is happening for most of the signal. The average improvement for both feedback paths over the case when the FS + VAD block is bypassed is about 5 dB and the average PESQ value is increased by 0.8 (from 3.3 to 4.1). In Fig. 6a, the input to the FS +VAD block and the VAD output are shown. The influence of the VAD output on the overall algorithm performance depends on the extension of the silence regions in the incoming signal and it accounts for a PESQ score improvement of about 0.2 for the shown example. The output of the VAD was almost always 1 for the used music signals, therefore, for the incoming music signals, the ASG and MIS curves were very close regardless if the VAD was used or not.

The influence of β is investigated in Fig. 7. Four values were used: 0.125, 0.5, 1 and 2.

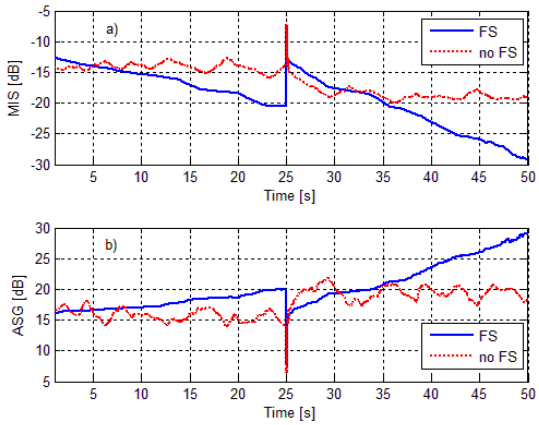

Fig. 5. Performance using music input signal with a sudden change of the feedback path after 25 seconds with and without frequency shifting, respectively a) MIS b) ASG.

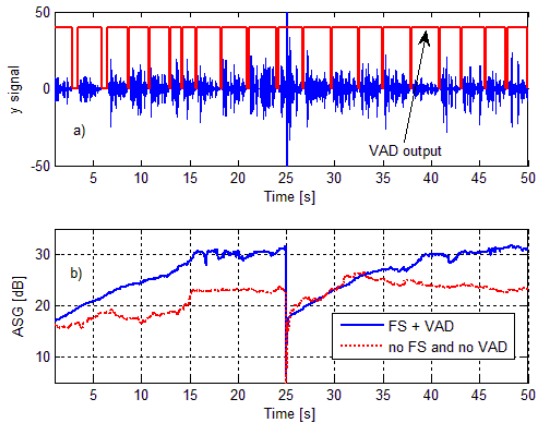

Fig. 6. a) The input signal to the FS + VAD block and the VAD output b) The ASG Performance using speech input signal with a sudden change of the feedback path after 25 seconds

It can be noticed that $\beta = 0.5$ lead to the best MIS and ASG values most of the time and brings about 0.5 dB average improvement over the next performing β value. This value was used for the next simulations.

The performances of the proposed algorithm, swPEMSC-DCD [16], and H-NLMS [11] with a music signal input and feedback path change from F_1 to F_2 after 25 seconds are shown in Fig. 8. The proposed algorithm has an improved ASG performance of about 2.5 dB over the swPEMSC-DCD and about 6.3 dB over the HNLMS algorithm. Also, if using the swPEMSC-DCD-APTA, the duration of the howling period is reduced. In Fig. 9 a zoom of the 0.3 seconds loudspeaker signal after the 25 seconds is shown. Howling is happening at this time due to the change of the feedback path. It can be noticed from Fig. 9 that the swPEMSC-DCD and swPEMSC-DCD-APTA generates shorter annoying higher amplitude sounds than the HNLMS algorithm. One advantage of the swPEMSC-DCD-APTA is its low numerical complexity, although it is about 15% higher than that of the swPEMSC-DCD algorithm for the same used parameters due to performing a VAD on 20 ms frames and tanh operations.

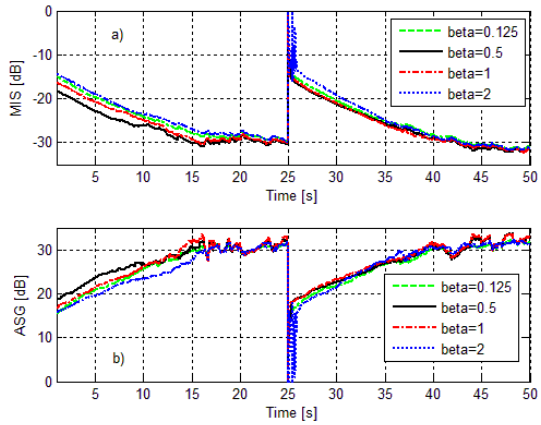

Fig. 7. Performance using speech input signal with a sudden change of the feedback path after 25 seconds and various β values a) MIS b) ASG.

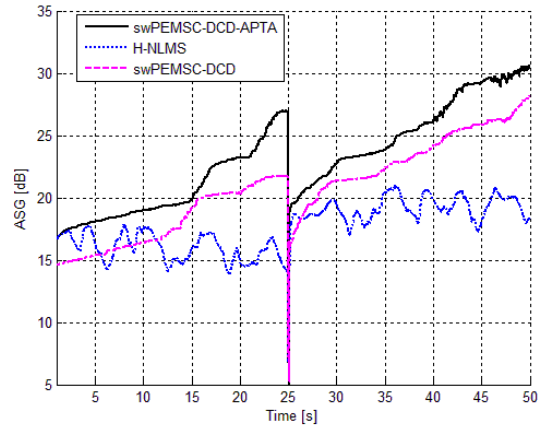

Fig. 8. ASG performance using music input signal with a sudden change of the feedback path after 25 seconds

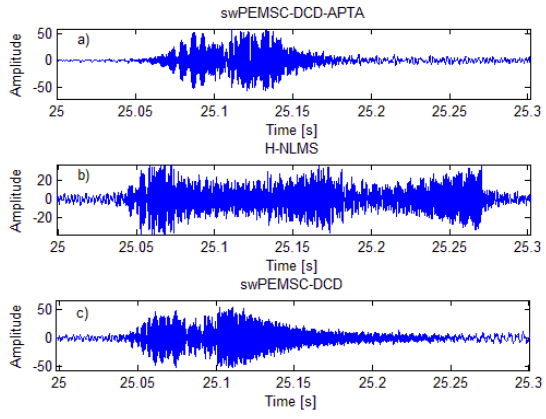

Fig. 9. Zoom on the howling signal after the sudden change of the feedback path after 25 seconds; a) swPEMSC-DCD-APTA b) H-NLMS c) swPEMSC-DCD

The increase of complexity due to the use of a variable H for the DCD iterations, and VAD method is rather small and acceptable if taking ito account the overall complexity of the adaptive filtering method and the performance

improvements. The overall reduction if the numer of additions is taken into account for using the DCD depends on the projection value, the number of updates and the number of bits [19]. For the considered parameters it was around 30% reduction in the number of additions, but if some minor performance losses are acceptable, it could be higher for higher projection orders, lower number of updates and lower number of bits. It was shown in [16] that the swPEMSC-DCD algorithm complexity is several times lower than that of the swPEMSC [14] and about 40% lower than the H-NLMS [11] in terms of multiplications. Therefore, the complexity of the proposed algorithm is also smaller that that of the swPEMSC and H-NLMS algorithms in terms of multiplications for the considered parameters.

IV. Conclusions and Future Work

A low-complexity acoustic feedback cancellation method combining the affine projection tanh algorithm, a variable input parameter DCD approach, frequency shifting, and voice activity detection is proposed. The performance of swPEMSC-DCD-APTA is compared with that of competing methods for AFC systems for hearing aids and its advantages are proved through simulations. Future work will be focused on reducing the howling by using variable steps size versions [27], advanced adaptive filtering [28] and howling detection methods [29].

References

[1] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," in *Proc. IEEE*, pp. 1–40, 2011.

[2] H. Dillon, *Hearing aids*, 2nd edition, Boomerang Press, 2008.

[3] T. van Waterschoot and M. Moonen, "Comparative evaluation of howling detection criteria in notch-filter-based howling suppression," *J. Audio Eng. Soc*. 58(11), pp. 923–940, 2010.

[4] M. Guo and B. Kuenzle, "On the periodically time-varying bias in adaptive feedback cancellation systems with frequency shifting," in *Proc. of ICASSP 2016*, Shanghai, China, pp. 539–543.

[5] F. Strasser and H. Puder, "Adaptive feedback cancellation for realistic hearing aid applications," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2322–2333, 2015.

[6] M. Guo, S. H. Jensen, and J. Jensen, "Novel acoustic feedback cancellation approaches in hearing aid applications using probe noise and probe noise enhancement*," IEEE Trans. Audio, Speech Lang. Process.* 20(9), pp. 2549–2563, 2012.

[7] F. Albu, R. Nakagawa, S. Nordholm, "Proportionate Algorithms for Two Microphones Active Feedback Cancellation", in *Proc. of EUSIPCO 2015*, Nice, France, pp. 290-294

[8] C.R.C. Nakagawa, S. Nordholm, F. Albu and W.-Y. Yan, "Closed-loop feedback cancellation utilizing two microphones and transform domain processing," in *Proc. of ICASSP 2014*, Firenze, Italy, pp. 3645–3649.

[9] C. Paleologu, J. Benesty, S. Ciochina, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Trans. Audio Speech Lang. Process.* (16), pp. 1466–1478, 2008.

[10] L.T.T. Tran, H. Schepker, S. Doclo, H.H. Dam, and S.E Nordholm, "Adaptive feedback control using improved variable step-size affine projection algorithm for hearing aids," in *Proc. of APSIPA-ASC 2017,* Kuala Lumpur, Malaysia.

[11] S. Nordholm, H. Schepker, L.T.T. Tran, S. Doclo, "Stability-controlled hybrid adaptive feedback cancellation scheme for hearing aids", *The Journal of the Acoustical Society of America,* 143 (1), 150-166, 2018.

[12] F. Albu, S. Nordholm and L.T.T. Tran, "The Hybrid Simplified Kalman Filter for Adaptive Feedback Cancellation" in *Proc. of COMM 2018*, Bucharest Romania, pp. 45-50.

[13] M. T. Akhtar, F. Albu, and A. Nishihara, "Maximum Versoria-criterion (MVC)-based adaptive filtering method for mitigating acoustic feedback in hearing-aid devices," *Appl. Acoust*., vol. 181, Oct. 2021.

[14] L.T.T. Tran, and S.E. Nordholm, "A Switched Algorithm for Adaptive Feedback Cancellation Using Pre-Filters in Hearing Aids", *Audiol. Res.* (11), pp. 389-409, 2021.

[15] Y.V. Zakharov and T.C. Tozer, "Multiplication-free iterative algorithm for LS problem," *Electron. Lett*., vol. 40, no. 9, pp. 567–569, Apr. 2004

[16] F. Albu, H.G. Coanda, D. Coltuc, "Low Complexity Acoustic Feedback Cancellation Algorithm with Frequency Shifting and Dichotomous Coordinate Descent Iterations", in *Proc. of ISETC 2022*, Timisoara, Romania, Nov. 2022.

[17] Q. Zhang, S. Wang, D. Lin, and S. Chen, "Robust Affine Projection Tanh Algorithm and Its Performance Analysis," *Signal Processing*, vol 202, 2023.

[18] D. J. Freed, "Adaptive feedback cancellation in hearing aids with clipping in the feedback path," *J. Acoust. Soc. Am.* 123(3), pp. 1618–1626, 2008.

[19] Y. Zakharov, G. White, and J. Liu, "Low complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3150–3161, July 2008.

[20] K. Patel and I.M.S. Panahi, "Efficient Real Time Acoustic Feedback Cancellation using Adaptive Noise Injection Algorithm", in *Proc. of EMBC 2020*, Montreal, Canada, pp. 972–975.

[21] T. Sankowsky-Rothe, M. Blau, H. Schepker, and S. Doclo, "Reciprocal measurement of acoustic feedback paths in hearing aids," *J. Acoust. Soc. Am*., vol. 138, no. 4, pp. EL399–EL404, 2015.

[22] Y. Hu and P. C. Loizou, "Subjective comparison of speech enhancement algorithms," in ICASSP, 2006, pp. 153–156.

[23] ITU-R BS-1387, "Method for objective measurements of perceived audio quality," International Telecommunications Union, Geneva, Switzerland, 100 pp., 1998.

[24] J. Ramirez, J. C. Segura, C. Benitez, A. de la Torre, and A. Rubio, "Efficient voice activity detection algorithms using long-term speech information," *Speech Commun*., vol. 42, pp. 271–287, Apr. 2004.

[25] R. Arablouei, K. Dogancay, S. Werner, "Recursive total least-squares algorithm based on inverse power method and dichotomous coordinate descent iterations," *IEEE Trans. Signal Process*., 63 (8), pp. 1941-1949, 2015

[26] Y. Yu, L. Lu, Z. Zheng, W. Wang, Y. Zakharov, R.C. de Lamare, "DCD-based recursive adaptive algorithms robust against impulsive noise," *IEEE Trans. Circuits Syst*. 67(7), 2020

[27] F. Albu, C. Paleologu and S. Ciochina, "New variable step size affine projection algorithms", in Proc. of IEEE COMM 2012, Bucharest, Romania, pp. 63-66, 2012.

[28] M. T. Akhtar, F. Albu, and A. Nishihara, "Prediction Error Method (PEM)-Based Howling Cancellation in Hearing Aids: Can We Do Better?", *IEEE Access*, vol. 11, pp. 337-364, Dec. 2022.

[29] Y. Alkaher and I. Cohen, "Temporal Howling Detector for Speech Reinforcement Systems", *Acoustics* 2022, 4(4), pp. 967-995.