forum acusticum 2026
8-12 September · Graz, Austria

DUAL PERSONAL SOUND ZONES BASED ON LOW-LATENCY REAL-TIME FREQUENCY-DOMAIN FXLMS ADAPTIVE FILTERING

Philippe Micheau[1], Manuel Melon[2]
[1] CRASH-UdeS, Universite de Sherbrooke, Sherbrooke, Canada
[2] LAUM–UMR 6613 CNRS, Le Mans Universite, Le Mans, France

Abstract

Personal sound zones (PSZ) aims to create bright and dark acoustic zones within a space by filtering a reference signal to drive multiple loudspeakers. Adapting the control filters is challenging due to variations in acoustic conditions or shifts of the zones, since error microphones are assumed to be fixed in the zones and the estimated acoustic transfer functions cannot be updated. Under these constraints, a real-time adaptive filtering approach was developed for two movable sound PSZ zones. Each zone consists of five loudspeakers in front of six microphones mounted on the same frame to ensure a fixed geometry. The ten filters are implemented in the time domain to minimize latency. The filter coefficient are updated using a frequency-domain block Filtered-x Normalized Least Mean Square (FxNLMS) algorithm. The system was deployed on a real-time controller. Performance, evaluated in terms of contrast and error, was tested under varying PZS positions and changing local acoustic conditions. The results show algorithm convergence and acceptable performance. The discussion highlights fundamental advantages and limitations of the presented approach, including sensitivity to uncertain and time-varying transfer paths.

Keywords: *wave field synthesis, Adaptive filter, real-time*

1. Introduction

Personal Sound Zones (PSZ) aim to reproduce an audio content in a defined bright zone (BZ) while minimizing it in a separate dark zone (DZ). This is typically achieved using multiple loudspeakers whose combined acoustic output is controlled through optimized filters. A widely used approach for the filter design is the least-squares method (LSM), also known as pressure matching (PM), which minimizes the error between the reproduced and desired sound fields [1]. The design of filters is often formulated as a dual objective: minimizing reproduction error in the bright zone and reducing acoustic energy in the dark zone. The filter impulse responses can be computed offline using either a frequency-domain or a time-domain formulation [2] [3] [4]. In both cases, all acoustical transfer functions from the loudspeakers to the microphones located within the PSZ are required.

The main limitation of these off-line design approaches is their assumption of time-invariant acoustical transfer functions, which yields fixed control filters. Therefore, any temporal variability in acoustic conditions or displacement of the zones may prevent the off-line computed control filters from achieving optimal error minimization.

While the error microphones are assumed to be fixed within the zones and the identified acoustic transfer functions cannot be updated, the problem is to adapt the control filters online to compensate for variations in acoustic conditions or displacements of the zones.

Under these constraints, the objective is to develop, implement, and evaluate a real-time adaptive filtering approach for movable sound zones.

In the PSZ literature, the use of real-time adaptive filtering can be found in few cases. They usually used the FxLMS algorithm to adapt the control filter in order to update the filters that minimize the error pressure at microphone locations. The FxLMS algorithm is an adaptive method that optimizes the coefficients of an FIR filter using stochastic gradient descent by incorporating the secondary path model into the adaptation process. In 2021, Vindrola *et al.* evaluated the possibility of using the FxLMS algorithm to retrieve the re-calibrated performance without the need to re-characterize the acoustic response of a car cabin [5].

Corresponding author: *Philippe.Micheau@USHerbrooke.ca.*

To the best of the authors' knowledge, the FxLMS has not been used to adapt PSZ filters online.

2. Theory

2.1. Pressure Matching problem

The least-squares method, also known as pressure matching (LSM/PM) has been previously used in Active Noise Control, Sound Field Synthesis and in the generation of PSZ [1]. It offers an expression of the optimal filters that minimize the least squared error between the reproduced field and a "desired" pressure field. The pressure matching (PM) problem of the PSZ multichannel system is to reproduce an audio signal over a bright zone (BZ), whilst minimizing the same audio signal in a dark zone (DZ). Fig. 1 presents the PSZ problem as a multichannel linear time-invariant (LTI) system in discrete time.

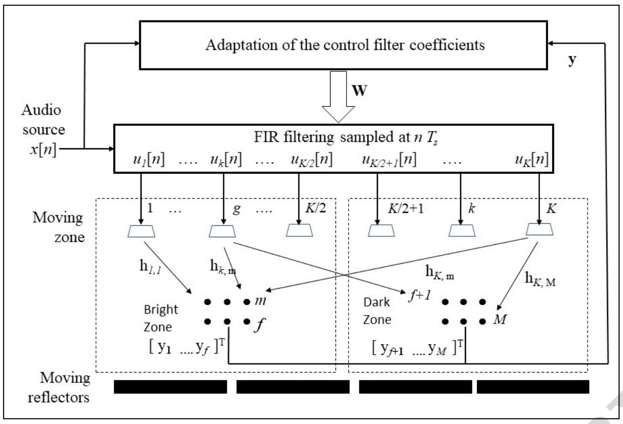

Figure 1: Configuration of PM problem under study.

2.2. Multichannel FIR model of the system

The linear acoustic system response from K loudspeakers to M microphones is modeled with a MIMO filter matrix $\mathbf{h}$

$$\mathbf{y}[n] = \mathbf{h}^T \mathbf{u}[n] \tag{1}$$

where the output vector is

$$\mathbf{y}[n] = [y_1[n] \cdots y_m[n] \cdots y_M[n]]^T \in \mathbb{R}^{M\times 1} \tag{2}$$

with $y_m[n]$ the m^{th} microphone signal at time sample n. The input stacked signal is expressed as

$$\mathbf{u}[n] = [\mathbf{u}_1^T[n] \cdots \mathbf{u}_k^T[n] \cdots \mathbf{u}_K^T[n]]^T \in \mathbb{R}^{N_h K\times 1} \tag{3}$$

with the k^{th} input signal with N_h sample is given by

$$\mathbf{u}_k[n] = [u_k[n]\, u_k[n-1] \cdots u_k[n-N_h+1]\,]^T \in \mathbb{R}^{N_h\times 1} \tag{4}$$

The LTI filter matrix $\mathbf{h} \in \mathbb{R}^{KN_h\times M}$

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_{11} & \mathbf{h}_{21} & \cdots & \mathbf{h}_{M1} \\ \mathbf{h}_{12} & \mathbf{h}_{22} & \cdots & \mathbf{h}_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{1K} & \mathbf{h}_{2K} & \cdots & \mathbf{h}_{MK} \end{bmatrix} \tag{5}$$

where $\mathbf{h}_{mk} \in \mathbb{R}^{N_h\times 1}$ represents the coefficient vector of causal finite impulse response (FIR) filters from the k^{th} input signal (loudspeaker) to the m^{th} output signal (microphone),

$$\mathbf{h}_{mk} = [h_{mk,0}\, h_{mk,1} \cdots h_{mk,N_h-1}\,]^T \in \mathbb{R}^{N_h\times 1} \tag{6}$$

2.3. Definition of desired signals

The desired vector of signals for the M microphones is generated as

$$\mathbf{d}[n] = \boldsymbol{\psi}^T \mathbf{x}[n - N_d] \in \mathbb{R}^{M\times 1} \tag{7}$$

with N_d a time delay and $\mathbf{x}[n] \in \mathbb{R}^{N_h\times 1}$ the mono audio source signal of N_h sample

$$\mathbf{x}[n] = [x[n]\, x[n-1] \cdots u[x - N_h + 1]\,]^T \tag{8}$$

and $\boldsymbol{\psi} \in \mathbb{R}^{N_h\times 1}$. A feasible "desired" delayed impulse response (7) is defined for a given loudspeaker g for the first f microphones

$$\boldsymbol{\psi} = [\mathbf{h}_{1g}(n) \cdots \mathbf{h}_{fg}(n)\; \mathbf{0} \cdots\; \mathbf{0}] \tag{9}$$

where microphones 1 to f are located in the bright zone (BZ), and microphones $f+1$ to M are located in the dark zone. Hence, the desired output audio signals are the time-delayed signals emitted at loudspeaker g for the bright zone. The time delay N_d is used to overcome the causality constraint of the problem.

2.4. Control filters

The input signal driving loudspeaker k is generated by a causal control FIR filter

$$u_k[n] = \mathbf{w}_k^T \phi[n] \tag{10}$$

where $\phi[n] \in \mathbb{R}^{N_w\times 1}$ denotes the signal vector:

$$\phi[n] = [x[n]\, x[n-1] \cdots x[n - N_w + 1]\,]^T \tag{11}$$

where $\mathbf{w}_k \in \mathbb{R}^{N_w\times 1}$ denotes the coefficient vector of the k^{th} control filter:

$$\mathbf{w}_k = [w_{k,0}\; w_{k,1} \cdots w_{k,N_w-1}\,]^T \tag{12}$$

The global coefficient vector collecting the coefficients of the K control filters is

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_K \end{bmatrix} \in \mathbb{R}^{KN_w\times 1}. \tag{13}$$

For a given acoustic configuration χ, the optimal control vector of filters $\mathbf{w}^{(\chi)}$ is the solution of the quadratic criterion minimization problem:

$$\mathbf{w}^{(\chi)} = \arg\min_{\mathbf{w}} \mathbb{E}\{\|\mathbf{d}[n] - \mathbf{y}[\mathbf{w}, n]\|_2^2\} \tag{14}$$

Because the optimal control filter changes when the acoustic configuration changes (χ), it is necessary to adapt $\mathbf{w}$. For this purpose, a multichannel adaptive algorithm can be used [5].

2.5. Multichannel adaptive filtering

Multichannel adaptive filtering is used to adapt control filters on the real system in real time. This section explains that it requires the identified matrix of FIR filters, $\hat{\mathbf{h}}$, to ensure the convergence of the algorithm. Then, the frequency-domain block FxLMS algorithm is introduced to address the computational load issue.

2.5.1. Multichannel FxLMS algorithm

The adaptation of the loudspeaker commands to the real system leads to implement a time-variant control filters

$$\mathbf{w}[n] \in \mathbb{R}^{KN_w \times 1} \tag{15}$$

The time evolution of the control filter matrix is usually obtained with a stochastic gradient descent algorithm (e.g. the MIMO FxLMS algorithm) [6]:

$$\mathbf{w}[n+1] = (1-\lambda)\mathbf{w}[n] - \mu \mathbf{F}^T[n]\mathbf{e}[n] \tag{16}$$

where $\mathbf{e}[n] = \mathbf{d}[n] - \mathbf{y}[n]$ is the error to minimize, λ is the leaky factor and μ is the adaptation coefficient. It is important to note that $F[n]$ is the matrix of the filtered reference $f_{m,k}$ given by

$$f_{m,k}[n] = \widehat{\mathbf{h}}_{m,k}^T \, \mathbf{x}[n] \tag{17}$$

where $\widehat{\mathbf{h}}_{m,k}$ is the identified finite impulse response from loudspeaker k to microphone m. The main interest of this approach is to use the same $\widehat{\mathbf{h}}$ for any configuration χ. Although $\widehat{\mathbf{h}} \neq \mathbf{h}^{(\chi)}$, algorithm (16) updates $\mathbf{w}[n]$ and ensures its convergence toward the optimal filters $\mathbf{w}^{(\chi)}$, provided that $\lambda = 0$ and that the convergence conditions are satisfied. In other words, this approach can update the control filters without any update of the identified acoustic system.

The challenge is that the MIMO (multichannel) FxLMS structure in (16) entails a high computational load due to the large number of parameters [3] and the computation of the filtered reference matrix (17) (obtained by filtering the reference signal x through $\widehat{\mathbf{h}}$). To address the computational load issue, transforming the problem of adaptation into the frequency domain provides an efficient solution.

2.5.2. Frequency Domain Block FxNLMS

The frequency domain approach provides significant computational savings through the use of the Fast Fourier Transform (FFT), which transforms a block of data from the time domain to the frequency domain. In the frequency domain, convolutions performed in the time domain become multiplications, which greatly simplifies the computations. Hence, many researchers in active noise control have focused their attention on developing the frequency domain block LMS (FB-LMS) algorithm, by combining the LMS algorithm with the fast Fourier transform (FFT) and overlap–save method [6], [7], [8], [9], [10].

In this paper, the frequency-domain FxLMS approach originally developed for active noise control is adapted for PSZ. Fig. 2 illustrates the implementation chosen to perform the control filtering in the time domain (10) while performing the reference filtering (17) and the update of the control filters (16) in the frequency domain.

The processing in the frequency domain is performed at the downsampled time index j (with $j = N_B k$) because all time signals are converted in batch of size N_B. Data in batches are concatenated into blocks of size N_F to perform FFT. To avoid time aliasing, the FFT size must respect the following condition: $N_F \geq \max(N_h, N_w) + N_B - 1$.

For each bin l, the FFT of the reference $\widetilde{\mathbf{X}}_l[j]$ is multiplied by the precomputed frequency responses of the filters , which yields MK filtered reference signals $\widetilde{\mathbf{F}}_l[j]$ after application of IFFT, the over-lap and save (OLS) method and FFT. The update of the complex coefficients of the filter $\widetilde{\mathbf{W}}[j]$ in the frequency domain is given by (18)

$$\widetilde{\mathbf{W}}_l[j+1] = (1-\lambda)\widetilde{\mathbf{W}}_l[j] - \mu \sum \frac{\widetilde{\mathbf{G}}_l[j]}{\widetilde{P}_l[j] + \gamma} \tag{18}$$

where $\widetilde{\mathbf{G}}_l[j]$ is the gradient obtained by cross-correlation between the filtered reference $\widetilde{\mathbf{F}}_l[j]$, and the error signal $\widetilde{\mathbf{E}}_l[j]$. For each bin, the gradient is normalized with the calculated power $\widetilde{P}_l[j]$ plus $\gamma > 0$ (to prevent a division by zero). The power per bin is calculated with the following recursive equation:

$$\widetilde{P}_l[j+1] = \alpha \widetilde{P}_l[j] + (1-\alpha)\|\widetilde{\mathbf{F}}_l[j]\|_2^2 \tag{19}$$

where α is the filtering power factor. Normalization ensures that each bin converges at the same rate.

Every nN_B the updated control filters in the frequency domain are loaded into the control filters, which then allow to improve the generated signal in real time according to (10).

3. Experimental setup

The experimental setup is shown in Fig. 3. The speaker array consists of two linear sub-arrays, each 70 cm long and containing 5 speakers (Elipson Planet M powered with HPA D604 amplifiers). The centers of the sub-arrays are 1 m apart and are aligned with the centers of the two zones. Both BZ and DZ are sampled with a 3×2 microphone array (Brüel & Kjaer 1/2" 4192) with a spacing of 6 cm. The shortest distance between a speaker and a microphone is 15 cm. The bright zone and one speaker sub-array are mounted on a plywood panel to allow for their movement. Initially, the two zones are spaced 1 m apart (center-to-center) and are located in an anechoic room without reflectors.

The algorithm was developed in the Simulink environment (Matlab 2025b). It was run on a Speedgoat system (Intel Core i7-3770K 3.5 GHz, 4GB). Analog signal acquisition was performed using a Speedgoat IO106 module featuring 64 simultaneously sampled 16-bit analog input channels with an input range of

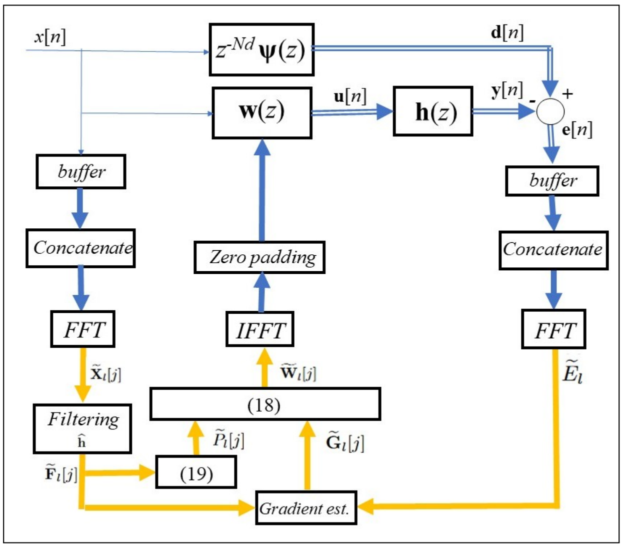

Figure 2: Implementation of the multichannel FD-FxLMS algorithm. Single lines denote single-channel signals, double lines denote multichannel signals, and bold lines denote frame-based signals. Blue indicates real-valued signals, while yellow indicates complex-valued signals.

± 2.5 V. Signal generation was achieved using a Speedgoat IO110 module providing 16 channels of 16-bit analog outputs with an output range of ± 5 V. The source signal $x[n]$ is generated by filtering a white noise signal through two 4th-order Butterworth bandpass shaping filters with a bandwidth ranging from 100 to 1,500 Hz and a gain factor of 0.2.

Four different configurations (see Fig. 4) were tested:

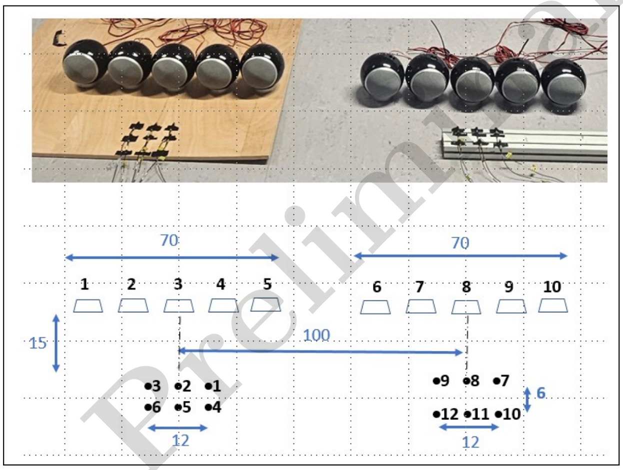

Figure 3: Experimental set-up

(a) Sub-arrays aligned, no reflectors added,
(b) The BZ and its associated sub-array are shifted 30 cm backward, no reflectors added;
(c) Sub-arrays aligned, reflectors added behind SZ;
(d) Sub-arrays aligned, reflectors added behind and above SZ.

For each of the four configurations ($\chi \in \{a, b, c, d\}$), there is one real acoustic response $\mathbf{h}^{(\chi)}$. and one optimal control filter set $\mathbf{w}^{(\chi)}$. For the following, the optimal control filters of the initial configuration, $\mathbf{w}^{(a)}$, will therefore serve as the frozen controller, e.g. the not adapted control filters. The parameters used for

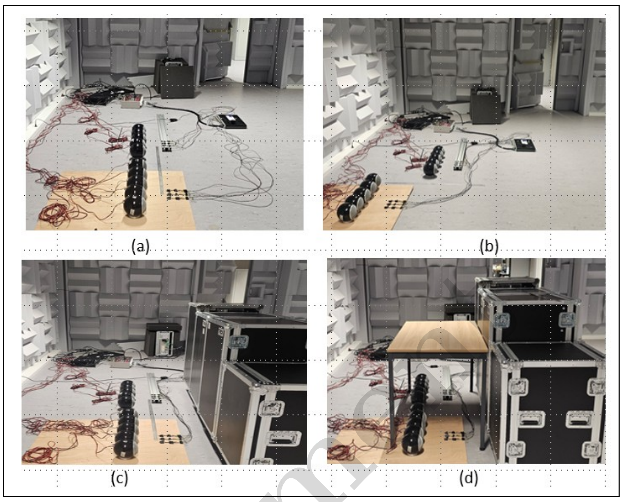

Figure 4: Pictures of the 4 configurations: (a) initial, (b) BZ and its associated sub-array are shifted 30 cm backward, (c) reflectors added, (d) reflectors and table added.

the experiments are summarized in Tab. 1.

Table 1: Parameters of the experiments.

Description	Variable	value
number of loudspeakers	K	10
number of microphones	M	12
sampling frequency	F_s	10 kHz
batch size	N_B	100
FFT points	N_F	1024
transfer function size	N_h	500
delay	N_d	400
control FIR size	N_w	900
leaky factor	λ	0.0
learning factor	μ	0.02
filtering power factor	α	0.99

4. Experimental results

To illustrate the differences between the configurations, two sets of frequency response (FRF) curves are shown in Fig. 5 for configurations (a) and (b). It can be seen that when the microphone is in the near field of the loudspeaker (upper curves in Fig. 5), the differences between the configurations are small (a few dB and a few degrees). At a greater separation distance (lower curves in Fig. 5), the changes are much more pronounced (more than 10 dB for the amplitude and 90° for the phase at certain frequencies). In summary, the transfer functions between a sub-array and the nearest zone do not change significantly when the wooden board is moved or reflectors are added, whereas the cross-transfer functions can change radically.

The performance of the sound zone system is eval-

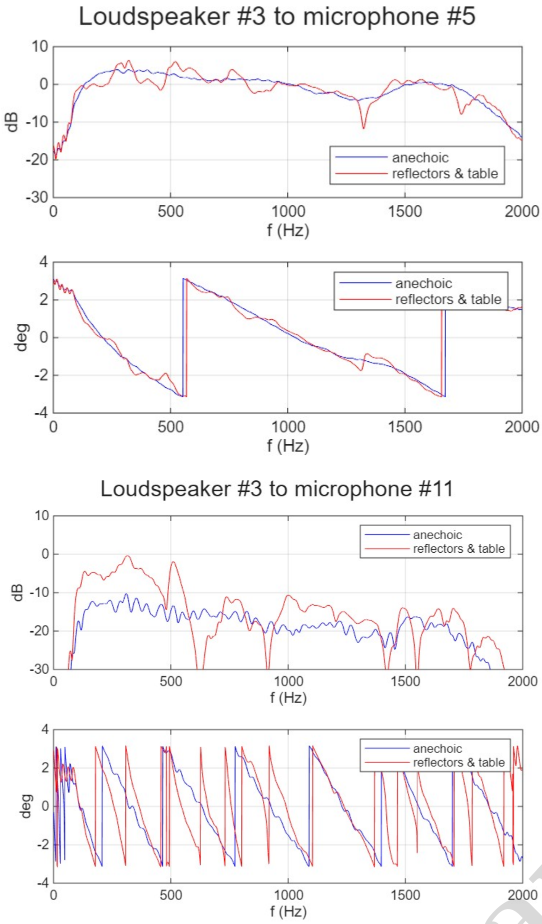

Figure 5: Identifier frequency response $\hat{h}^{(a)}_{3,5}(f)$ and $\hat{h}^{(a)}_{3,11}(f)$ for the anechoic case and $\hat{h}^{(d)}_{3,5}(f)$ and $\hat{h}^{(d)}_{3,11}(f)$ for the reflector-and-table scenario.

uated using two classical metrics, calculated in the frequency domain from the acoustic pressure in the bright zone $\boldsymbol{p_b}(f)$ and the dark zone $\boldsymbol{p_d}(f)$.

The acoustic contrast (AC) between the two zones is defined by

$$\mathrm{AC}(f) = 10 \log \left(\frac{\|\boldsymbol{p_b}(f)\|_2^2}{\|\boldsymbol{p_d}(f)\|_2^2} \right), \tag{20}$$

where $\boldsymbol{p_b}(f) = C_{mic}[y_0(f) \cdots y_f(f)]^T \in \mathbb{C}^f$ and $\boldsymbol{p_d}(f) = C_{mic}[y_{f+1}(f) \cdots y_M(f)]^T \in \mathbb{C}^f$ where $Y_m(f)$ is computed over 1 second using Welch's averaged periodogram method ($N_{\mathrm{FFT}} = 2028$, 50% overlap) and C_{mic} the calibration factor microphone.

The reproduction error (err), relative to the desired pressure $\boldsymbol{p}_t(f)$ in the bright zone, is calculated with

$$\mathrm{err}(f) = \frac{\|\boldsymbol{p_t(f)} - \boldsymbol{p_b(f)}\|_2^2}{\|\boldsymbol{p_t(f)}\|_2^2}. \tag{21}$$

where $\boldsymbol{p_t}(f) = C_{mic}[d_0(f) \cdots d_f(f)]^T \in \mathbb{C}^f$ is calculated with the initial configuration, e.g., with $\hat{\mathbf{h}}^{(a)}$. Fig. 6 represents the acoustic contrast and reproduction error for configuration (a). The acoustic contrast is much more pronounced (23 dB $<$ AC(f) $<$ 39 dB) for f between 100 and 1,200 Hz, and then varies from 18 to 23 dB in the 1,200–1,900 Hz band. The error remains below 3×10^{-2} up to 1,900 Hz.

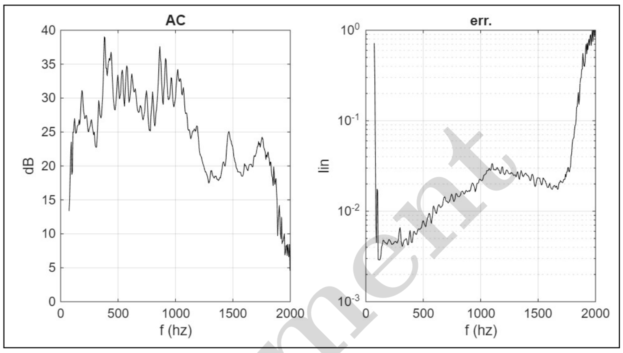

Figure 6: Acoustic contrast (AC) and error (err) for the anechoic configuration, e.g. the initial configuration $\chi = a$.

Fig. 7 represents the metrics for configuration (b). It can be observed that, when using frozen control filters, $\mathbf{w}^{(a)}$, the acoustic contrast is reduced by approximately 10 dB, while the error increases only slightly. This is because, when the wooden board is moved, the sub-array speakers and the microphones move together, thereby maintaining virtually the same transfer functions (the FRFs to the DZ are indeed modified; however, they do not contribute the target pressure). After running the adaptive process for 15 s, the control filters are updated to converge to $\mathbf{w}^{(b)}$. One can see that the metrics recover values quiet similar to those of the initial configuration (a).

For config. (c) (with reflectors added behind BZ and

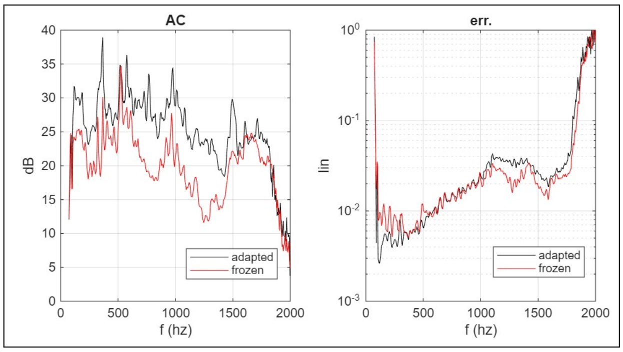

Figure 7: Acoustic contrast (AC) and error (err) for $\chi = b$: red: results with frozen control filters ($\mathbf{w}^{(a)}$), black: adapted filters.

DZ), the results are plotted in Fig. 8. In this case, the reduction in acoustic contrast (up to 18 dB) with the frozen control filters $\mathbf{w}^{(a)}$ is much more pronounced than in config. (b), and the error increases significantly up to 1,500 Hz. After adaptation of the control filters $\mathbf{w}^{(c)}$ has been obtained, the metric values increase even though they are still slightly lower than in the configuration. (a). This decrease, compared to config. (a), is due to the increased complexity of the environment, which makes it more difficult to satisfy the SZ constraints. In this case, all FRFs are modified due to reflectors but not enough to prevent the convergence of the adaptive process.

The results for config. (d) are shown in Fig. 9. Here,

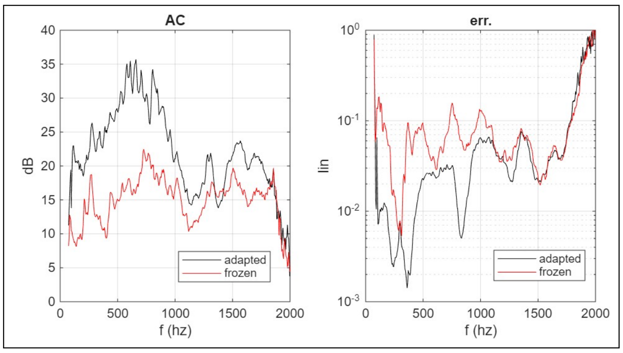

Figure 8: Figure of the configurations (c) reflectors at 35 cm in front of the microphones.

the addition of a reflector surface above the zones and speakers yields major changes in the transfer functions (see Fig. 5). Nevertheless, the algorithm still converges even though the gain in AC after adaptation is restricted to the low frequency band (< 800 Hz). Regarding the error, the gain after adaptation is much greater, meaning that the algorithm successfully canceled out the energy reflected by the panels.

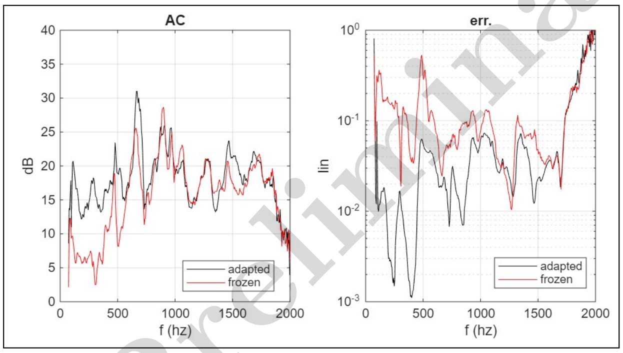

Figure 9: Figure of the configurations (d) reflectors at 35 and table

5. Conclusion

In this paper, a real-time frequency-domain FxNLMS adaptive filtering has been implemented in an SZ configuration. The results showed, for the various cases tested, that the algorithm allows the SZ filters to be adapted after a change, thereby improving the metrics after convergence. These encouraging initial results pave the way for future work in a less academic context, namely online implementation on headrests installed in a car cabin.

6. Acknowledgments

This work has been carried out within the framework of the “International Research Project–Centre Acoustique Jacques Cartier”.

References

[1] M. Poletti, “An investigation of 2-d multizone surround sound systems,” in *125th Convention of the Audio Engineering Society*, Convention paper 7551, pp. 1–9, Oct. 2008.

[2] L. Vindrola, M. Melon, J.-C. Chamard, B. Gazengel, and G. Plantier, “Personal sound zones: A comparison between frequency and time domain formulations in a transportation context,” in *147th Convention of the Audio Engineering Society*, Convention paper 10216, Oct. 2019.

[3] D. Shi, W.-S. Gan, J. He, and B. Lam, “Practical implementation of multichannel filtered-x least mean square algorithm based on the multiple-parallel-branch with folding architecture for large-scale active noise control,” *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 28, no. 4, pp. 940–953, Apr. 2020.

[4] M. F. S. Gálvez, S. J. Elliott, and J. Cheer, “Time domain optimization of filters used in a loudspeaker array for personal audio,” *IEEE-ACM Trans. Audio Speech Lang.*, vol. 23, no. 11, pp. 1869–1878, Nov. 2015.

[5] L. Vindrola, M. Melon, J.-C. Chamard, and B. Gazengel, “Use of the filtered-x least-mean-squares algorithm to adapt personal sound zones in a car cabin,” *J. Acoust. Soc. Am.*, vol. 150, no. 3, p. 1779, 2021.

[6] S. Elliott, *Signal Processing for Active Control*, en. Elsevier, 2000, ISBN: 978-0-08-051713-1.

[7] T. Kosaka, S. J. Elliott, and C. C. Boucher, “A novel frequency domain filtered-x lms algorithm for active noise reduction,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 1, 1997, pp. 403–406.

[8] C. Hansen, S. Snyder, X. Qiu, L. Brooks, and D. Moreau, *Active Control of Noise and Vibration*, 2nd ed. London: CRC Press, Nov. 1, 2012, 1554 pp., ISBN: 9780429162855.

[9] J. Lorente, M. Ferrer, M. de Diego, and A. Gonzalez, “The frequency partitioned block modified filtered-x nlms with orthogonal correction factors for multichannel active noise control,” *Digit. Signal Prog.*, vol. 43, pp. 47–58, 2015, ISSN: 1051-2004.

[10] S. Zhang, Y. S. Wang, H. Guo, C. Yang, X. L. Wang, and N. N. Liu, “A normalized frequency-domain block filtered-x lms algorithm for active vehicle interior noise control,” *Mech. Syst. Sig. Process.*, vol. 120, pp. 150–165, 2019.