World Journal of Engineering Research
Print ISSN: 2959-9865
Online ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3083

VISION-GUIDED SECONDARY-PATH MODEL UPDATING FOR PLAYBACK-REFERENCED ACTIVE HEADREST NOISE CONTROL

XuDong Wu, Cong Zhang*, JiaXing Luo, ZeXiong Zhang
College of Automotive and Energy Engineering, Tongji University, Shanghai 201804, China.
**Corresponding Author: Cong Zhang*

Abstract: Time-varying secondary paths induced by head motion can introduce filtered-x mismatch and degrade the stability margin and usable attenuation bandwidth of active headrest noise control. This article proposes a vision-guided secondary-path model updating method for playback-referenced active headrest noise control. A monocular camera provides latency-compensated three-dimensional ear positions, from which ear–loudspeaker distances are estimated to update the internal secondary-path models through an equivalent delay shift and amplitude scaling of nominal impulse responses. The updated models are injected into filtered-x generation and normalised adaptation, while the underlying controller structure remains unchanged. To evaluate the proposed method under controlled dynamic conditions, a synchronised offline replay protocol is adopted, in which the electrical playback reference and binaural error signals are logged under a common hardware clock and replayed for repeatable comparisons. Experiments on a semi-anechoic active headrest platform in the 80–800 Hz band with a 3 dB criterion show that the proposed update enlarges the maximum stable step size by 67%–300% and improves attenuation continuity, yielding a median threshold-bandwidth gain of approximately 10–90 Hz depending on the condition.
Keywords: Active headrest; Active noise control; Playback-referenced control; Secondary-path model updating; Vision-guided ear tracking

1 INTRODUCTION

Active headrest noise control (AHR-ANC) is an important local active sound control technique for improving acoustic comfort in compact listening zones near the listener's ears. By embedding secondary loudspeakers in the headrest, the system can generate anti-noise in the immediate vicinity of the ears without requiring global sound-field control in the entire cabin. This local-control characteristic makes AHR-ANC attractive for headrest- or seat-integrated implementations where packaging space is limited and the target zone is tightly constrained. However, this same localization also makes the system more sensitive to geometric changes between the listener and the secondary sources. In practice, even moderate head motion alters the relative positions among the ears, loudspeakers, and surrounding reflective structures, causing the effective secondary paths to vary over time. When the controller continues to use fixed secondary-path models under such conditions, filtered-x mismatch arises and may manifest as a reduced stable step-size range, intermittent attenuation loss, and degraded continuity of effective control bandwidth. These effects are consistent with the well-established sensitivity of FxLMS-type algorithms to secondary-path modelling errors [1-3].

A large body of work has addressed secondary-path uncertainty in active noise control through online modelling, auxiliary-noise injection, variable-step adaptation, and other robustification strategies. Classical online identification methods can improve model tracking, but they often rely on additional excitation or structural modifications that are not always desirable in listener-centred systems [4-7]. For active headrest scenarios in particular, practical deployment is constrained by limited sensing budgets, strict real-time requirements, and the need to avoid intrusive auxiliary signals near the listener. These considerations make it difficult to directly transfer more elaborate online-modelling schemes into compact multichannel headrest systems. At the same time, prior studies on head-tracking-assisted local control have shown that geometry awareness can substantially improve robustness to listener motion and extend usable control towards higher frequencies by mitigating model mismatch as the listening geometry evolves [8-11]. Recent active headrest studies have further highlighted the importance of evaluating not only nominal attenuation, but also the continuity and usability of control performance under dynamic multichannel conditions [12-15].

Despite these advances, two issues remain insufficiently addressed for playback-referenced multichannel active headrest control. First, many existing approaches either require extra excitation or modify the controller architecture, whereas a practically deployable headrest implementation benefits from a minimal-intervention strategy that preserves a standard FxLMS signal path. Second, under dynamic ear motion, performance should not be assessed solely by average attenuation or static frequency responses; rather, it is necessary to examine whether the controller maintains a usable and continuous attenuation region over time, especially when the secondary path varies continuously. For playback-referenced feedforward systems, this issue is particularly relevant because the electrical playback signal is known and repeatable, making it possible to isolate the effect of internal model updating from other confounding factors if an appropriate evaluation protocol is used.

Motivated by these considerations, this article proposes a vision-guided secondary-path model updating scheme for a playback-referenced feedforward 2×2 MIMO FxLMS active headrest controller. A monocular camera provides latency-compensated 3D ear states, from which ear–loudspeaker distances are estimated and converted into an equivalent per-path delay shift and amplitude scaling of nominal secondary-path impulse responses. The resulting update is injected only into the internal models used for filtered-x generation and normalised adaptation, while the control-output structure itself remains unchanged. In this sense, the proposed method targets the dominant geometry-induced component of secondary-path drift using a low-dimensional and implementation-friendly correction, instead of performing full online secondary-path re-identification.

To evaluate this strategy under controlled dynamic conditions, we adopt a synchronized offline replay protocol in which the electrical playback reference and binaural error channels are logged under a common hardware clock and replayed for repeatable comparisons. This design allows the influence of secondary-path updating to be examined while holding the reference, replay segment, controller form, and evaluation settings fixed. Dynamic usability is then quantified in the 80–800 Hz band using stability-margin and attenuation-continuity metrics rather than relying on a single averaged attenuation index.

The contributions of this article are twofold. First, we develop a geometry-driven equivalent secondary-path update that is directly compatible with a conventional playback-referenced 2×2 MIMO FxLMS pipeline and reduces the dominant filtered-x mismatch caused by head motion. Second, we establish a synchronized offline replay evaluation framework that enables controlled quantification of both stability-margin expansion and effective-bandwidth continuity improvement under time-varying secondary paths.

2 PLAYBACK-REFERENCED FEEDFORWARD 2×2 MIMO FXLMS

2.1 System Definition

Consider a feedforward 2×2 AHR-ANC system with two headrest loudspeakers ($j\in\{1,2\}$) and two near-ear error microphones ($i\in\{L,R\}$) located close to the left and right ears [2, 12, 13]. Let the reference be $x(n)$, the control outputs be $u_1(n)$ and $u_2(n)$, and the binaural errors be $e_L(n)$ and $e_R(n)$. In our implementation, $x(n)$ denotes the electrical playback (excitation) signal driving the primary-disturbance loudspeaker; no upstream reference microphone is used. For an arbitrary error channel $i\in\{L,R\}$, the error signal is

$$e_i(n)=d_i(n)-\sum_{j=1}^{2}\sum_{m=0}^{M-1}\mathrm{h}_{ij}(m,t)u_j(n-m), \tag{1}$$

where $d_i(n)$ denotes the disturbance component at ear i and $\mathrm{h}_{ij}(m,t)$ denotes the (time-varying) secondary-path impulse response from loudspeaker j to error microphone i. The controller outputs are generated by filtering the exogenous reference

$$u_j(n)=\sum_{k=0}^{L_w-1} w_j(k,n)x(n-k),\ j=1,2, \tag{2}$$

where $w_j(k,n)$ are adaptive coefficients and L_w is the control-filter length. Define $\mathbf{w}_j(n)=[w_j(0,n),\ldots,w_j(L_w-1,n)]^T$.

2.2 Filtered-x Mismatch as a Signal-Processing Bottleneck

MIMO FxLMS adaptation forms filtered-x signals using an internal secondary-path estimate $\hat{\mathrm{h}}_{ij}(m,t)$ [2, 16]. When head motion causes $\mathrm{h}_{ij}(m,t)$ to drift while $\hat{\mathrm{h}}_{ij}(m,t)$ remains fixed, the filtered-x signals become mismatched. A convenient view is to express the mismatch through an additive modelling error

$$\mathrm{h}_{ij}(m,t)=\hat{\mathrm{h}}_{ij}(m,t)+\delta\mathrm{h}_{ij}(m,t), \tag{3}$$

which induces a bias between the true and estimated filtered-x regressors. The filtered reference associated with path (i,j) is

$$\hat{x}_{ij}(n,t)=\sum_{m=0}^{M-1}\hat{\mathrm{h}}_{ij}(m,t)x(n-m),\ \ \hat{\mathbf{x}}_{ij}(n,t)=\left[\hat{x}_{ij}(n,t),\ldots,\hat{x}_{ij}(n-L_w+1,t)\right]^T \tag{4}$$

Using a normalised multiple-error form, the 2×2 MIMO FxLMS update is

$$\mathbf{w}_j(n+1)=\mathbf{w}_j(n)+\mu\frac{\sum_{i\in\{L,R\}}e_i(n)\hat{\mathbf{x}}_{ij}(n,t)}{\epsilon+\sum_{i\in\{L,R\}}\|\hat{\mathbf{x}}_{ij}(n,t)\|_2^2},j=1,2, \tag{5}$$

where μ is the step size and $\epsilon>0$ is a small regularization constant. Therefore, as $//\,\delta\mathrm{h}_{ij}(m,t)\,//$ increases, the expected update direction deviates from the ideal descent direction associated with the true secondary paths, reducing the admissible step-size range and potentially causing divergence or intermittent degradation—especially in coupled multichannel settings.

3 VISION-GUIDED PARAMETRIC SECONDARY-PATH UPDATING

3.1 Parametric Model for Ear-Motion-Induced Secondary-Path Variations

Let $h_{ij}^0(m)$ denote nominal secondary-path impulse responses measured at a reference ear position, with nominal ear–loudspeaker distances r_{ij}^0 for each ear $i \in \{L,R\}$ and loudspeaker $j \in \{1,2\}$. Monocular vision provides 3D ear positions, which are temporally robustified and latency-compensated to yield $\hat{\mathbf{p}}_i(t-\tau)$, where τ is an end-to-end sensing/processing latency [14, 17]. From $\hat{\mathbf{p}}_i(t-\tau)$, the ear–loudspeaker distances $r_{ij}(t-\tau)$ are computed. We model the dominant effect of moderate geometry change by a low-dimensional parametric transformation of $h_{ij}^0(m)$:

$$h_{ij}(m,t) \approx g_{ij}(t) h_{ij}^0(m-\Delta n_{ij}(t)) + \varepsilon_{ij}(m,t), \tag{6}$$

where $\Delta n_{ij}(t)$ represents the propagation-delay change (in samples), $g_{ij}(t)$ captures distance-related amplitude scaling, and $\varepsilon_{ij}(m,t)$ aggregates unmodeled effects (e.g., scattering-induced spectral-shape changes and reflections). This formulation makes the secondary-path drift tractable: the time variation is primarily captured by two scalar parameters per channel, $\Delta n_{ij}(t)$ and $g_{ij}(t)$.

3.2 Geometry-Driven Update Rule and Controller Integration

Using the distance change, the delay shift is computed as

$$\Delta n_{ij}(t) = \text{round}\left(\frac{f_s}{c}\left[r_{ij}(t-\tau) - r_{ij}^0\right]\right), \tag{7}$$

where f_s is the sampling rate and c is the sound speed. The internal estimate is updated by

$$\hat{h}_{ij}(m,t) = g_{ij}(t) h_{ij}^0(m-\Delta n_{ij}(t)), \tag{8}$$

with consistent zero padding/truncation for out-of-range indices. The scaling $g_{ij}(t)$ can be implemented by a physically motivated mapping (e.g., inverse-distance scaling) or by a calibrated distance–gain curve. The updated $\hat{h}_{ij}(m,t)$ is injected into filtered-x generation (4) and normalised MIMO FxLMS adaptation (5); critically, the control output structure (2) remains unchanged.

Baselines. NV (no update): $\hat{h}_{ij}(m,t) \equiv h_{ij}^0(m)$. VG (vision-guided): $\hat{h}_{ij}(m,t)$ updated by (7)–(8) using $\hat{\mathbf{p}}_i(t-\tau)$.

3.3 Minimal Rationale for Stability-Margin Expansion

Under the mismatch model (3)–(4), updating $\hat{h}_{ij}(m,t)$ by (7)–(8) reduces the dominant component of $\delta h_{ij}(m,t)$ attributable to geometry-induced delay and amplitude drift, thereby reducing filtered-x regressor mismatch. As a result, the expected update direction in the normalised adaptation (5) becomes closer to the ideal descent direction associated with the true secondary paths. This decreases the effective gradient bias and mitigates step-size-induced divergence, which is consistent with the observed enlargement of the maximum stable step size μ_{stab} in Sec. V, see Figure 1.

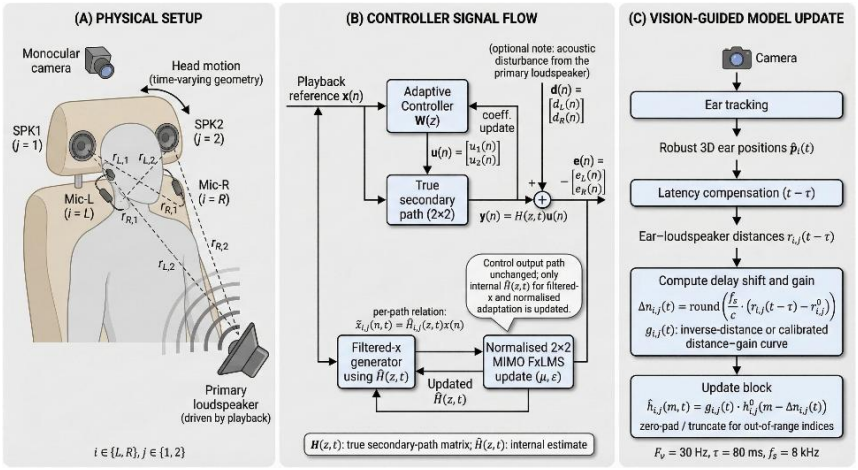

Figure 1 Vision-guided Secondary-path Model Updating Integrated with a Playback-referenced 2×2 Multichannel FxLMS Active Headrest Controller

4 EXPERIMENTAL PROTOCOL AND DYNAMIC USABILITY METRICS

4.1 Platform and Parameters

Experiments were conducted on a semi-anechoic active headrest (AHR) platform equipped with two headrest-mounted secondary loudspeakers and two near-ear error microphones. The primary disturbance was reproduced by a fixed-position loudspeaker driven by a known broadband playback signal. During data acquisition, the electrical playback reference $x(n)$ and the binaural error-microphone signals were synchronously logged under a common hardware clock, so that the reference and response channels remained temporally aligned throughout the subsequent replay analysis.

The controller was implemented as a playback-referenced feedforward 2×2 MIMO FxLMS system. The sampling rate was set to f_s=8000 Hz. Each control filter had length L_w=512, and each secondary-path impulse response was represented by an FIR model of length M=512. These settings were kept fixed in all comparisons to ensure that the evaluated differences were attributable to the internal secondary-path model update rather than to changes in controller complexity or numerical resolution. The vision-guided update rate was F_v=30 Hz, which matches the temporal refresh of the ear-state sequence used by the updating module. The effective sensing-to-update latency was τ=80 ms, and this latency was incorporated explicitly by shifting the ear-state sequence before model updating.

Six dynamic test conditions, denoted Case 1–Case 6, were constructed by combining illumination variation and occlusion state (unoccluded or occluded), with Case 6 representing the most challenging condition. This design makes it possible to test the proposed updating strategy under progressively more difficult visual conditions while retaining the same underlying control architecture and replay procedure. As a result, the influence of vision-assisted secondary-path compensation can be examined over a range of practically relevant sensing qualities rather than under a single nominal scenario.

It should be noted that the purpose of this section is not to optimize controller hyperparameters independently for each condition. Instead, a single parameter set is retained across all runs so that the comparison between the no-update baseline (NV) and the vision-guided update mode (VG) remains controlled and interpretable. In this way, any improvement in stability or usable attenuation region can be attributed to the proposed model-updating mechanism itself, rather than to case-specific retuning.

4.2 Synchronized Offline Replay Protocol

To isolate the effect of secondary-path model updating from other sources of variability, we adopt a synchronized offline replay protocol. In each replay run, the recorded electrical reference $x(n)$ is held fixed, while the control outputs $\{u_1(n),u_2(n)\}$ are re-computed by the same 2×2 MIMO FxLMS algorithm using the selected internal secondary-path estimate $\hat{\mathrm{h}}_{ij}(m,t)$. This setup provides a controlled comparison framework in which the excitation, replay segment, controller topology, and numerical settings are identical across runs.

Within this framework, the effective latency τ is implemented as a time shift applied to the ear-state sequence, such that the update at time ttt uses the delayed state vector $\hat{\mathbf{p}}_i(t-\tau)$. The internal estimate $\hat{\mathrm{h}}_{ij}(m,t)$ is refreshed at the vision rate F_v and is held constant between successive updates. Accordingly, the system captures the practical situation in which geometric information is available only at discrete visual refresh instants and must be used with non-zero delay. This treatment is important because the proposed method is intended as an implementation-oriented correction strategy rather than an idealized zero-latency oracle update.

In all replay runs, the evaluation band, attenuation threshold, control-filter length, secondary-path model length, normalized adaptation form, and numerical safeguards are kept identical. If regularization ϵ\epsilonϵ or output limiting is applied, these settings are also fixed across all conditions. Therefore, the only intentional difference between NV and VG is whether the internal secondary-path model is kept fixed or updated according to the latency-compensated visual estimate. This controlled design is essential for interpreting the experimental outcomes, because it suppresses confounding effects that would otherwise arise from changes in excitation content, controller form, or implementation details.

A further advantage of the offline replay design is repeatability. Since the same recorded reference and error channels can be replayed under matched settings, the comparison does not depend on trial-to-trial differences in disturbance realization or manual re-execution. This is particularly useful for dynamic AHR evaluation, where even small differences in replay conditions can obscure the effect of the internal model update. The replay-based protocol thus provides a practical bridge between real-system recording and controlled algorithmic comparison.

4.3 Metrics for Attenuation Continuity

All metrics are computed within the 80–800 Hz band using a 3 dB attenuation threshold. Rather than relying only on a single average attenuation value, we evaluate dynamic usability from three complementary aspects: stability margin, threshold-satisfying time-frequency coverage, and threshold bandwidth continuity. This choice reflects the fact that, under time-varying secondary paths, two controllers may produce similar mean attenuation while differing substantially in whether useful attenuation is maintained continuously across time and frequency.

4.3.1 Maximum stable step size

We first quantify the stability margin through the maximum stable step size, denoted μ_{stab}. The step size μ is scanned over a discrete grid from 1×10^{-3} to 2×10^{-2}, with a resolution of 1×10^{-3} up to 1.2×10^{-2} and 2×10^{-3} beyond that point. The exact scanned set is reported in Fig. 2. For a given condition and update mode, μ_{stab} is defined as the largest μ that remains stable under the adopted instability criterion.

Instability is declared if any of the following occurs:

(a) the pre-limiter saturation ratio exceeds 0.5;
(b) the maximum controller coefficient norm exceeds 100; or
(c) the in-band attenuation over 80–800 Hz drops below 0.5 dB.

This definition intentionally combines numerical and performance-based failure symptoms. The first two criteria capture clear divergence tendencies in the controller output or coefficient growth, while the third excludes operating

points that may remain numerically bounded but no longer provide meaningful in-band control. Since μ_{stab} is identified by discrete scanning, it should be interpreted at the present scan resolution and under the adopted replay segment and failure criterion, rather than as a continuous theoretical stability bound.

4.3.2 Threshold time–frequency coverage

To characterize whether attenuation remains available over time and frequency, we compute a time-frequency attenuation map $A(f,t)$ from the STFT power spectra of the disturbance-only signals $\{d_L(n),d_R(n)\}$ and the residual errors $\{e_L(n),e_R(n)\}$ over the same replay segment. The STFT uses a 32 ms window (256 samples at 8 kHz), 75% overlap, and $NFFT$=256. Let

$$P_d(f,t)=\frac{|S_{dL}(f,t)|^2+|S_{dR}(f,t)|^2}{2},\ P_e(f,t)=\frac{|S_{eL}(f,t)|^2+|S_{eR}(f,t)|^2}{2},$$

and define the attenuation map as

$$A(f,t)=10\log_{10}\frac{P_d(f,t)+10^{-12}}{P_e(f,t)+10^{-12}}.$$

We then form the binary threshold mask

$$M(f,t)=\mathrm{I}\{A(f,t)\geq 3\ \mathrm{dB}\}. \tag{9}$$

The threshold coverage ratio is defined as the fraction of time-frequency bins within 80–800 Hz for which the attenuation reaches or exceeds 3 dB. This metric reflects the overall occupancy of useful attenuation in the analysis plane. Compared with a frame-averaged spectrum, it is more sensitive to fragmentation, intermittent collapse, and patchy recovery of the attenuation region, all of which are relevant when the secondary path varies dynamically.

4.3.3 Threshold bandwidth

Coverage alone does not indicate whether the threshold-satisfying region remains contiguous from low to higher frequencies. To quantify this aspect, we define the threshold bandwidth for each time frame as the highest frequency within 80–800 Hz such that the attenuation remains above the threshold for all frequencies from 80 Hz up to that frequency:

$$f_{\mathrm{thr}}(t)=max\{f\in[80,800]:A(f',t)\geq 3\ \mathrm{dB},\forall f'\leq f\}. \tag{10}$$

This definition is intentionally conservative. It measures not merely the presence of isolated threshold-satisfying bins, but the upper edge of a continuous usable attenuation band extending upward from 80 Hz. Consequently, it provides a more direct description of effective bandwidth continuity under dynamic operation.

We summarize $f_{\mathrm{thr}}(t)$ by its distribution over time and report the median difference

$$\Delta_{\mathrm{median}}=\mathrm{median}\left(f_{\mathrm{thr}}^{\mathrm{VG}}\right)-\mathrm{median}\left(f_{\mathrm{thr}}^{\mathrm{NV}}\right),$$

together with 95% confidence intervals obtained via bootstrap resampling over time frames with B=800 repetitions using the 2.5th and 97.5th percentiles of the bootstrapped medians. For visualization only, the displayed $f_{\mathrm{thr}}(t)$ curve is smoothed using a 0.2 s moving median. The smoothing is not used in the statistical summary itself; it is introduced solely to improve readability of the plotted trajectories.

Taken together, the three metrics provide complementary information. The maximum stable step size quantifies admissible adaptation strength, the threshold coverage ratio quantifies how much of the time-frequency plane remains useful, and the threshold bandwidth describes how continuously that usefulness extends toward higher frequencies. This combination is therefore better suited to dynamic active headrest evaluation than a single average attenuation measure.

5 RESULTS

5.1 Stability-Margin Expansion

Figure 2 summarizes the step-size scan results obtained under identical replay segments, controller topology, and numerical settings for the no-update baseline (NV) and the vision-guided updating mode (VG). Across all six tested conditions, the stability boundary is shifted consistently toward larger step sizes when vision-guided secondary-path updating is enabled. The resulting increase in the maximum stable step size, μ_{stab} , ranges from 67% to 300% relative to NV, demonstrating that the proposed update enlarges the admissible adaptation range without altering the underlying 2×2 MIMO FxLMS structure.

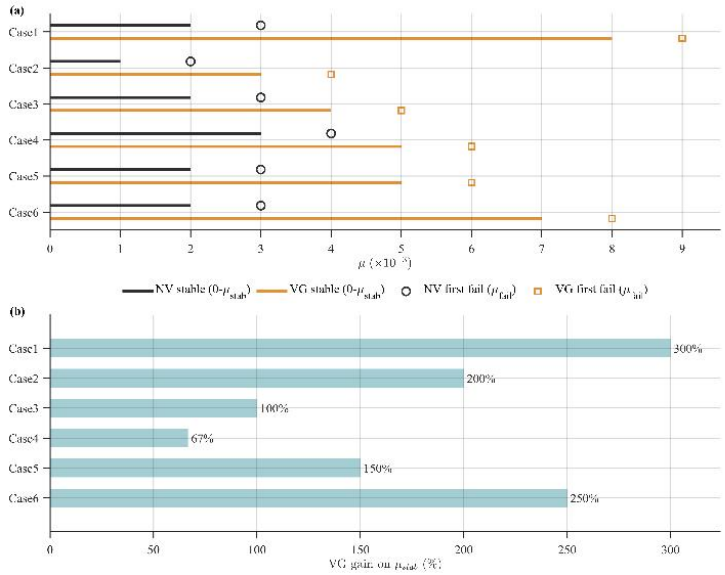

Figure 2 Step-size Scan Showing the Stability Boundary μ_{stab} and First Failure for NV and VG across Six Conditions

This behaviour is consistent with the intended role of the proposed method. In the fixed-model baseline, head-motion-induced secondary-path drift causes the filtered-x signals to be formed with increasingly mismatched internal models. As this mismatch grows, the effective adaptation direction becomes more biased relative to the ideal descent direction associated with the true time-varying secondary paths, and the controller becomes less tolerant to aggressive step sizes. By contrast, VG reduces the dominant geometry-induced component of this mismatch through delay and gain correction of the nominal impulse responses. Although the update remains first-order and does not fully reconstruct all spectral-shape changes, it is sufficient to move the internal model closer to the true path trajectory, thereby reducing gradient bias and postponing instability.

From a practical perspective, the observed increase in μ_{stab} is important because it provides additional adaptation headroom under dynamic conditions. In multichannel active headrest control, usable performance is often limited not by the nominal controller form itself, but by how conservatively the step size must be selected in the presence of secondary -path uncertainty. A larger stable step-size region means that the controller can operate with faster effective adaptation while still remaining within a bounded and practically useful regime. This is especially relevant in compact local-control systems, where the geometric relation between the ears and secondary loudspeakers varies continuously and can otherwise force the algorithm into a narrow operating window.

It is also worth emphasizing that the reported stability boundaries should be interpreted in the context of the adopted replay protocol and discrete scan resolution. Here, μ_{stab} is not a theoretical closed-form bound, but an experimentally identified boundary under fixed replay content, the stated instability criterion, and the scanned μ grid used in Fig. 2. Nevertheless, because the replay segment, controller settings, and failure criteria are kept identical for NV and VG, the relative shift in the stability boundary provides a controlled and meaningful comparison. Under this matched protocol, the consistent expansion across all six cases supports the conclusion that geometry-driven secondary-path updating improves robustness to time-varying path mismatch rather than yielding an isolated gain in only one nominal condition.

5.2 Attenuation Continuity and Effective Bandwidth

Figure 3 provides complementary system-level evidence that the proposed update improves dynamic usability within 80 –800 Hz under the 3 dB criterion. Whereas Fig. 2 evaluates how far the adaptation strength can be increased before instability occurs, Fig. 3 focuses on how effectively useful attenuation is maintained over time and frequency during dynamic operation. This distinction is important because a controller may remain numerically stable yet still exhibit fragmented or intermittent attenuation, which reduces its practical value in listener-centred noise control.

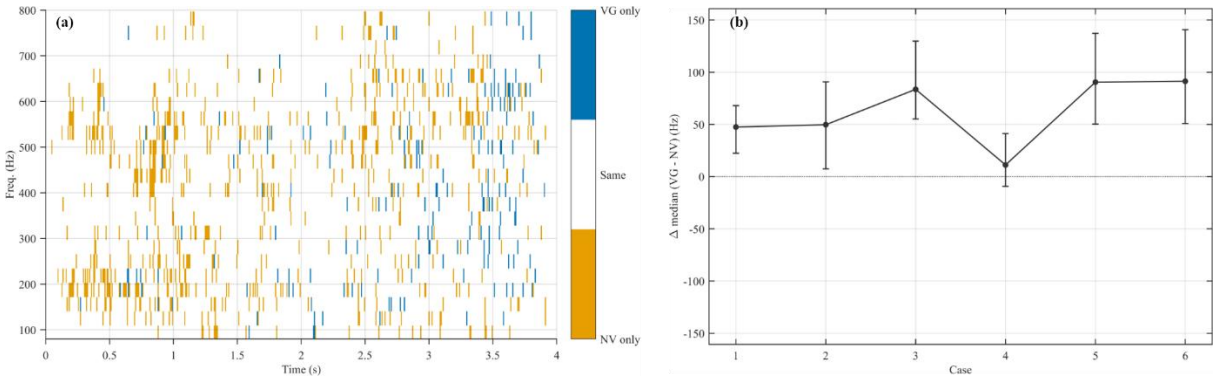

Figure 3 Continuity and Effective Bandwidth under Dynamic Conditions: (a) Categorical Difference Map of Threshold -satisfying Time–frequency Bins (Case 6); (b) Median Threshold-bandwidth Gain (Δ_{median}) with 95% Confidence Intervals across Six Conditions

The categorical difference map in Fig. 3(a), shown for the most challenging condition (Case 6), indicates that the dominant improvement arises from VG-only regions rather than from a wholesale redistribution of all threshold-satisfying bins. In other words, the proposed update does not simply shift the attenuation pattern arbitrarily; instead, it mainly recovers time-frequency regions in which the NV baseline fails to preserve a continuous threshold-satisfying band. These VG-only regions are particularly visible during intervals where NV exhibits fragmentation and shrinkage of the attenuation region. This observation is consistent with the expected effect of reduced filtered-x mismatch: when the internal secondary-path model tracks the dominant delay and amplitude drift more closely, the controller is better able to maintain a coherent attenuation region instead of breaking into scattered, disconnected patches.

The distributional result in Fig. 3(b) further supports this interpretation. Aggregated over time, VG shifts the threshold-bandwidth distribution upward and yields a median gain Δ_{median} of approximately 10–90 Hz depending on the condition, with 95% confidence intervals reported for all six cases. Because the threshold bandwidth is defined conservatively as the highest frequency up to which the 3 dB criterion is satisfied continuously from 80 Hz, this gain should not be interpreted merely as the appearance of isolated higher-frequency bins. Rather, it indicates that the usable attenuation band more often extends continuously toward higher frequencies when vision-guided updating is applied. This continuity-oriented interpretation is central to dynamic active headrest evaluation, since perceptually and operationally useful control depends not only on whether attenuation appears somewhere in the spectrum, but also on whether it remains stably available as a contiguous band over time.

Another notable aspect of Fig. 3 is that the improvement is most informative under the more difficult dynamic conditions. When the baseline already maintains a relatively intact threshold-satisfying region, the room for improvement is naturally smaller. In contrast, under conditions with stronger visual difficulty or more severe dynamic mismatch, NV is more likely to suffer from shrinkage and fragmentation of the usable band, and the benefit of internal model updating becomes more apparent. This trend supports the intended engineering role of the proposed method: it acts as a lightweight compensation mechanism that is particularly valuable when time-varying geometry would otherwise degrade control continuity.

Taken together, the results in Figs. 2 and 3 show that the proposed vision-guided update improves two complementary aspects of dynamic active headrest performance. First, it enlarges the stability margin by allowing larger admissible step sizes under time-varying secondary paths. Second, it improves attenuation continuity by preserving a more continuous and practically usable mid-to-high-frequency control region. These two effects are mutually consistent. A better-matched internal secondary-path model not only reduces the risk of step-size-induced instability, but also helps maintain the integrity of the threshold-satisfying time-frequency region during dynamic operation. Therefore, the proposed method provides an effective first-order compensation for geometry-induced secondary-path drift while preserving the simplicity of the underlying playback-referenced 2×2 MIMO FxLMS controller.

6 CONCLUSION

This article presented a vision-guided secondary-path model updating scheme for a playback-referenced feedforward 2×2 MIMO FxLMS active headrest controller. By mapping latency-compensated ear–loudspeaker distance changes to an equivalent delay shift and amplitude scaling of nominal impulse responses, the method updates only the internal models used for filtered-x generation and normalised adaptation, without modifying the controller structure. Using a synchronized offline replay protocol for controlled comparisons, experiments on a semi-anechoic platform show a 67% –300% increase in the maximum stable step size and improved attenuation continuity, yielding a median threshold-bandwidth gain of approximately 10–90 Hz across six conditions. Limitations include the use of offline replay and a first-order parametric path model that does not explicitly capture scattering- or pose-induced spectral-shape changes; these effects will be addressed in future real-time, in-the-loop validation and richer low-dimensional modelling [10, 11].

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Elliott S J, Nelson P A. Active noise control. IEEE Signal Processing Magazine, 1993, 10(4): 12-35.

[2] Elliott S J. Signal Processing for Active Control. London, U.K. Academic Press, 2001.

[3] Lu L, Yin K L, de Lamare R C, et al. A survey on active noise control in the past decade-Part I: Linear systems. Signal Processing, 2021, 183: 108039. DOI: 10.1016/j.sigpro.2021.108039.

[4] Hu M, Xue J, Lu J. Online multi-channel secondary path modeling in active noise control without auxiliary noise. Journal of the Acoustical Society of America, 2019, 146(4): 2590-2595. DOI: 10.1121/1.5129380.

[5] Im S, Kim S, Woo S, et al. Deep learning-assisted active noise control in a time-varying environment. Journal of Mechanical Science and Technology, 2023, 37(3): 1189-1196. DOI: 10.1007/s12206-023-0206-2.

[6] Ma Y, Xiao Y, Ma L, et al. A robust feedforward hybrid active noise control system with online secondary-path modelling. IET Signal Processing, 2023, 17(1): e12183. DOI: 10.1049/sil2.12183.

[7] Oh J Y, Jung H W, Lee M H, et al. Enhancing active noise control of road noise using deep neural network to update secondary path estimate in real time. Mechanical Systems and Signal Processing, 2024, 206: 110940. DOI: 10.1016/j.ymssp.2023.110940.

[8] Elliott S J, Jung W, Cheer J. Head tracking extends local active control of broadband sound to higher frequencies. Scientific Reports, 2018, 8: 5403. DOI: 10.1038/s41598-018-23531-y.

[9] Jung W, Elliott S J, Cheer J. Combining the remote microphone technique with head-tracking for local active sound control. Journal of the Acoustical Society of America, 2017, 142(1): 298-307. DOI: 10.1121/1.4994292.

[10] Lai C K, Cheer J, Shi C. The effect of head-tracking resolution on the stability and performance of a local active noise control headrest system. Journal of the Acoustical Society of America, 2025, 157(2): 766-777. DOI: 10.1121/10.0035576.

[11] Veronesi F, Lai C K, Cheer J. Interpolation between plant responses in a head-tracked local active noise control headrest system. Mechanical Systems and Signal Processing, 2025, 240: 113401. DOI: 10.1016/j.ymssp.2025.113401.

[12] Jung W, Elliott S J, Cheer J. Local active control of road noise inside a vehicle. Mechanical Systems and Signal Processing, 2019, 121: 144-157. DOI: 10.1016/j.ymssp.2018.11.003.

[13] Karthik M L N S, Pradhan S, George N V. Performance evaluation of an active headrest system using a filtered-x least mean square/fourth algorithm with virtual sensing. Journal of the Acoustical Society of America, 2023, 154(5): 2878-2891. DOI: 10.1121/10.0022329.

[14] Liu Y, Li H, Zou H, et al. Active headrest combined with a depth camera-based ear-positioning system. Journal of the Acoustical Society of America, 2025, 157(1): 519-526. DOI: 10.1121/10.0034860.

[15] Zhang Z, Wu M, Yin L, et al. Robust parallel virtual sensing method for feedback active noise control in a headrest. Mechanical Systems and Signal Processing, 2022, 178: 109293. DOI: 10.1016/j.ymssp.2022.109293.

[16] Samarasinghe P N, Zhang W, Abhayapala T D. Recent advances in active noise control inside automobile cabins: Toward quieter cars. IEEE Signal Processing Magazine, 2016, 33(6): 61-73. DOI: 10.1109/MSP.2016.2601942.

[17] Jiang H, Chen H, Tao J, et al. Accuracy requirements of ear-positioning for active control of road noise in a car. Applied Acoustics, 2024, 225: 110164. DOI: 10.1016/j.apacoust.2024.110164.