INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2026, VOL. 72, NO. 1, PP. 1–9
Manuscript received January 19, 2026; revised February, 2026. doi: 10.24425/ijet.2026.157864

Using Vowel Characteristics for Multi-channel Signal Decorrelation and Reverberation

Michele Pizzi, and Bartłomiej Mróz

***Abstract*—This paper presents a novel audio decorrelation method that integrates velvet noise with parametric modeling of vowel filters derived from recorded speech. By capturing vocal timbre, the technique extends decorrelation and artificial reverberation tools, offering new creative possibilities. Velvet noise is filtered to model vowel resonances, enabling users to imprint speech or singing characteristics onto multichannel effects. Inspired by choir acoustics, the system synthesizes distinct vowels per channel, producing speech-like textures that are convolved with input audio. The approach emphasizes immersive, voice inspired sound design, with discussion of implementation challenges, creative applications, listening tests, and directions for future research.**

***Keywords*—audio decorrelation; multi-channel reverberation; linear predictive coding; timbre control; spatial audio; vowel synthesis**

I. Introduction

THE creation of immersive audio environments requires techniques that can convincingly spread sound across multiple channels, making listeners feel surrounded by a realistic acoustic space.

Audio decorrelation is a crucial processing technique where a signal is processed to produce multiple outputs that possess different waveforms than the original sound source but sound the same [1]. One of the applications of audio decorrelation is to upmix mono signals to stereo or multi-channel sound sources.

Multi-channel reverbs and decorrelation are core tools for creating spacious, diffuse, and convincing acoustic scenes in surround and immersive audio. Decorrelation transforms correlated channel content into perceptually distinct signals; multi-channel reverbs generate directional and enveloping late reflections that define room size, envelopment, and source diffuseness [2]. Velvet noise is quasi-random type of sparse noise consisting of values of +1, 0 and -1 positioned at irregular distance in time. Most of the samples of the signal are comprised of zeroes [3]. Velvet noise techniques have emerged as efficient tools for decorrelation and reverberation design. The velvet noise decorrelator demonstrated that sparse noise sequences can achieve effective decorrelation with significantly reduced computational cost compared to traditional methods [4]. This idea was extended through velvet noise feedback delay networks, which increased echo density and realism in artificial reverberation while maintaining efficiency [5]. Further developments introduced dark velvet noise, enabling the modeling of non-exponential reverberation decays with improved spectral control [6]. Most recently, multichannel interleaved velvet noise has been proposed to distribute sparse sequences across channels, producing low-correlation, spatially diffuse reverberation suitable for immersive audio applications [7].

While multi-channel reverberation and decorrelation are widely used to define room size and source diffuseness in immersive audio, most existing approaches focus primarily on technical aspects such as statistical independence and transparency, often overlooking the creative potential of timbre control [1], [8].

To address this gap, parametric modeling techniques, especially Linear Predictive Coding (LPC), provide powerful tools for timbre manipulation.

LPC models the vocal tract as an all-pole filter whose coefficients capture the spectral envelope of the speech signal, thus approximating the resonant characteristics (formants) that distinguish different vowels [9], [10]. By representing the speech waveform through a compact set of predictor parameters, LPC provides both an efficient encoding method and a powerful analytical tool to study vowel production [11].

This parametric framework not only reduces the dimensionality of the data but also aligns closely with the source–filter theory of speech [12], making it particularly suitable for vowel modeling in both synthesis and recognition tasks [13]. Vowels can be modeled as a quasi-periodic glottal excitation passed through a time-varying vocal tract filter [14]; the LPC model represents the speech signal as an excitation driving the filter, yielding a compact representation of vowel timbre. [15].

This alignment with the source–filter theory underpins LPC's effectiveness for vowel analysis and synthesis, where the poles of the LPC filter correspond to vocal tract resonances [9], [10], and the resulting envelope cleanly separates timbral structure from pitch and fine harmonic detail [11]–[13], [15].

Beyond speech synthesis, LPC has been successfully applied in music and sound design to transfer the timbral qualities of speech to other sound sources. The ability to control timbre through LPC filters separately from the pitch is an attractive approach to manipulate timbre over time in

M. Pizzi is an Independent Researcher (e-mail: info@pizzimusic.com).
B. Mróz is with the Department of Multimedia Systems, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland (e-mail: bartlomiej.mroz@pg.edu.pl).

audio production. Furthermore, it allows us to use different sound sources as input of a formant filter, therefore transferring the timbre of speech to other signals. LPC has been used successfully in music to capture the timbre of speech sound sources to create novel sound transformations and new timbres. This capability has been harnessed by composers and sound designers within various sound design methodologies. As a result, LPC serves as a powerful tool for exploring and shaping new timbres, expanding the palette of sounds available for musical composition and audio production [16]–[18]. Parametric modeling has also been explored as a method to synthesize reverb-like textures by convolving stereo white noise with synthesized vowel-like sounds with random character produced by randomized distance between synthetic glottal pulses [19].

Building on these foundations, this work introduces a novel approach to multi-channel reverberation that integrates the spectral shaping capabilities of LPC with velvet noise-based decorrelation methods.

By leveraging audio analysis of user-provided recordings of vowel sounds, the system enables the creation of fully customized reverb effects tailored to the unique characteristics of each input signal.

By fusing vowel-inspired parametric modeling with velvet noise decorrelation, this research introduces a novel step forward in audio signal processing. Decorrelation and reverberation techniques are typically optimized for statistical independence and perceptual transparency, but they rarely explore the expressive potential of timbre. By embedding the resonant structures of human vowels into the decorrelation process, the proposed method not only achieves effective spatial diffusion but also introduces a new dimension of timbral control. A distinctive competitive feature of this approach is that users can employ their own recordings to design customized 5.1 reverb effects, ensuring that each project is fully unique. This capability transforms decorrelation from a purely technical process into a generative medium, bridging rigorous algorithmic design with experimental sound design. The result is an innovative contribution that expands the functional scope of reverberation, offering both engineers and artists a technically robust yet creatively personal tool for immersive audio production.

While previous work has explored velvet noise for decorrelation and LPC for timbral modeling, no existing approach integrates these methods to leverage the qualities of vowels for customizable multi-channel reverb.

Spatial awareness has played a role across many musical epochs. As early as in the Baroque period, composers incorporated space related concepts, such as the placement of choirs in Venetian polychoral traditions [20]. This effect is inspired by Venetian polychoral music where performers were positioned around the listener to create striking spatial contrasts. Translating this concept into the algorithm, each channel carries two distinct vowels, which are deliberately blended as part of the surround reverb process.

The motivation behind this blending is to mirror the way overlapping vowel sounds in choral settings fuse into the reverb tail of a church into evolving textures, producing unique flavors of decorrelation. In this way, the effect leverages the spectral richness of vowel resonances while transforming them into a versatile tool for spatial and timbral shaping.

This paper aims to develop and evaluate a system that enables users to design personalized multi-channel reverberation effects by combining LPC-based vowel modeling with velvet noise decorrelation. This approach has broad potential across music production, film, gaming, and virtual reality. By enabling user-driven, timbre-rich reverb effects, it allows sound designers and engineers to create more immersive and personalized audio experiences. Such flexibility allows for both exacting control and imaginative sound design, enhancing its relevance for various media applications. The goal of this method is to explore speech synthesis within the context of audio decorrelation and creative design of audio effects.

II. Objectives

The primary objective of this study is to advance the field of immersive audio production by developing a novel system that integrates LPC-based vowel modeling with velvet noise decorrelation. This approach aims to enable the creation of customizable multi-channel reverberation effects that leverage both technical rigor and creative timbral control. Specifically, the research seeks to address the following questions:

1) How can LPC-based vowel modeling be integrated with velvet noise decorrelation to create customizable multi-channel reverberation effects?
2) How does user-driven customization of reverb effects, using personal vowel recordings, impact the perceived envelopment of the multi-channel reverb?
3) Does embedding vowel-inspired spectral shaping into the decorrelation process provide a viable method to enhance timbral control in immersive audio environments?

By investigating these questions, the study aims to expand both the technical and creative possibilities of multi-channel reverberation, offering engineers and artists a robust yet highly personal tool for immersive audio design.

III. Method

A. Effect Implementation and Design

The algorithm integrates three complementary techniques to transform a mono input signal into a 5.1-channel reverberation effect. First, a Feedback Delay Network (FDN) is employed to produce a dense, natural-sounding mono reverberation. Second, LPC is used to design filters that capture the spectral characteristics of vowel recordings, thereby capturing the timbral qualities to transfer to the reverb tail.

Finally, convolution with the LPC-filtered velvet noise sequences achieves multi-channel decorrelation, distributing the reverberation spatially across the 5.1 output channels. This process not only imparts the spectral qualities of two distinct vowels to the reverberated signal but also transforms the mono source into a decorrelated 5.1 surround effect. The audio effect prototype discussed in this paper has been developed and tested in GNU Octave [21]. Figure 1 outlines the overall architecture of the proposed implementation for the audio effect presented here.

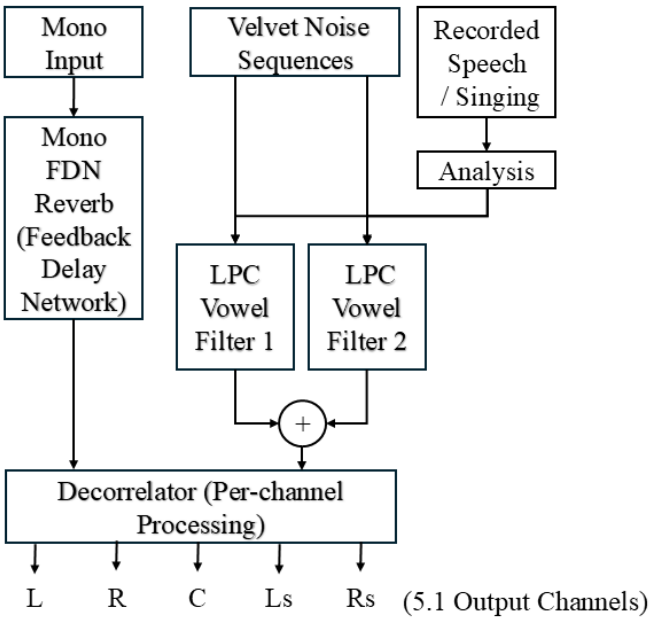

Fig. 1. Block diagram of the proposed multi-channel reverberation system, integrating mono input FDN reverb, LPC-based vowel filtering from speech, and velvet noise decorrelation across 5.1 channels.

The first stage applies reverberation to a mono input source using an FDN reverb. An FDN is a signal processing architecture composed of multiple interconnected delay lines whose outputs are recursively mixed through a feedback matrix. This structure is widely employed in artificial reverberation due to its ability to generate dense, natural-sounding echoes with high computational efficiency. In the present implementation, four delay lines are combined using a well-established methodology for feedback gain matrix design, as implemented in [22], [23]. The output of this stage remains monophonic.

Building on this, the second stage extracts the target vowel resonances from an audio recording. The user can select two 30-millisecond segments of vowel sounds, where a Bartlett window is applied to the amplitude envelope of each segment.

The default duration for the speech segments has been set to 30 milliseconds; users can customize the length of the analysis frame. This is achieved using the *lpc* function and *bartlett* from the signal package in GNU Octave [21]. Two parallel all-pole LPC models of order 10 are employed, each representing five resonances (formants) corresponding to two distinct vowel sounds.

Subsequently, the third stage synthesizes velvet noise sequences in parallel. To produce a 5.1 surround output, the system generates ten distinct sequences, two for each channel of the target configuration.

The velvet noise generator follows the implementation described in [4]. Each sequence is then used as excitation for a vowel filter, with two vowel filters assigned to each channel. One unique sequence excites each filter, producing two vowel-like signals that resemble vocal fry voice production. The two filtered signals are summed to produce a single output signal, after which a Bartlett window is applied to the signal envelope to reduce unwanted resonances in the reverb timbre. Unlike the implementation in [4], a Bartlett window is used as the temporal envelope of the velvet noise sequence instead of an exponentially decaying envelope. This approach has been empirically evaluated, drawing on sound design techniques described in [19]. This process is repeated in parallel across all channels of the 5.1 system. Figure 2 illustrates the per-channel processing signal flow.

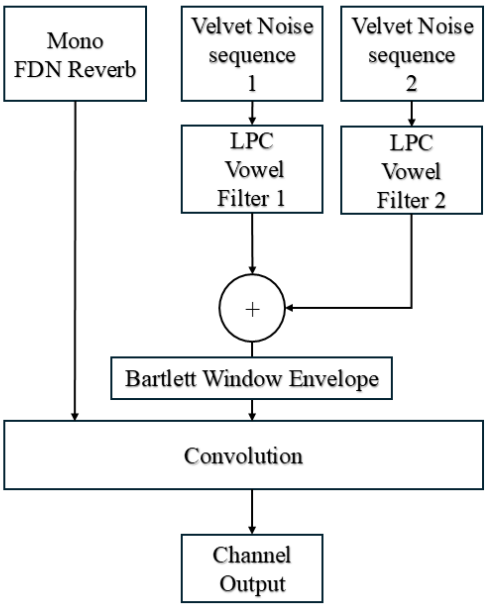

Fig. 2. Signal flow for per-channel decorrelation in the 5.1 reverb effect. Each channel receives two velvet noise sequences, each filtered through distinct LPC vowel models, summed and windowed before convolution with the mono FDN reverb output. In total, the system uses ten distinct filtered velvet noise sequences to achieve the decorrelation effect of the reverberated signal for a 5.1 configuration.

Finally, the mono output of the FDN reverb is convolved with the decorrelated, vowel-filtered sequences for each channel, thus achieving 5.1 spatial decorrelation.

This design broadens creative possibilities by allowing users to adjust both the length and density of velvet noise sequences, with particular emphasis on longer durations of approximately 100 milliseconds and over. Such durations have proven especially effective for sound design experiments involving customized reverb tails. In addition, employing a lower density of non-zero values appears to reduce unpleasant resonances. Taken together, these features enable the design to exploit the spectral characteristics of vowel resonances while diminishing their intelligibility, yielding a broader and more versatile timbre that adapts effectively to diverse input sources.

To the best of our knowledge, this is the first approach to integrate these methods into a single architecture that employs two distinct vowel sounds to enable customized vowel-controlled timbral shaping in a surround-reverb context.

B. Listening Tests Overview

This study employed a MUSHRA-like procedure to carry out listening tests investigating two perceptual attributes, Reverb Envelopment and Timbral Balance, in the context of

the novel audio effect. Because the effect under investigation is entirely new, the objective was not to test predefined hypotheses but to explore how listeners perceive and describe the effect along these perceptual dimensions. By focusing on mapping the perceptual space rather than confirming specific predictions, the procedure provides an initial framework for understanding the auditory consequences of the effect and establishes a foundation for subsequent confirmatory studies.

C. Participants

The listening panel consisted of twelve (N = 12) advanced listeners with task-specific attribute training. To ensure a consistent understanding of the perceptual dimensions under investigation, the participants underwent a short training session prior to the test. In this session, exaggerated examples of reverberant envelopment and timbral imbalance were presented and discussed. This familiarization phase clarified the constructs of Envelopment and Timbral Balance and established a shared perceptual frame of reference across the panel. The Participants were volunteers from the Department of Multimedia Systems, Gdańsk University of Technology, and volunteers with some previous sound design or musical experience. Inclusion criteria required self-reported normal hearing and no history of hearing impairment.

D. Stimuli Preparation

The audio material for this study was drawn from the EBU Sound Quality Assessment Material (SQAM) database [24], which provides standardized reference recordings for evaluating sound quality. From this collection, two instrument tracks were selected: piano (track number 39), representing a percussive and polyphonic source, and violin (track number 08), representing a monophonic sustained source. To prepare the stimuli, mono signals were derived from the left channel of these recordings, and short extracts were created for processing with the new 5.1 reverberation effect.

The same extracts have been processed using the same FDN reverb length and the same velvet noise sequences for all different audio examples. The only variation was the choice of the vowel combination for the filter. The vowel sounds used for designing the experiment stimuli were isolated from the singing recording from SQAM tracks 44, 45, 46, and 47 to mimic a music production scenario where the sound engineer was capturing reverb timbres from the vocal tracks [24]. The ALOFON corpus [25] served as the source material for this study, offering a wide range of speaker voices suitable for investigating vocal characteristics. From this resource, vowel recordings were selected to provide diverse phonetic material. These recordings were then combined to design exaggerated timbral balance examples, enabling controlled exploration of timbre manipulation.

In the experimental design, the first condition was the original sound material drawn from a mono source, either piano or violin. The second condition was processed with velvet noise without the application of vowel filters. For the remaining conditions, three distinct mixing variations were explored. The first mix (*'Mix1'*) maintained vowel consistency across different voices, combining a tenor and a bass both articulating the vowel */a/*. The second mix (*'Mix2'*) introduced variation by assigning different vowels to different voices, pairing an alto on */e/* with a tenor on */a/*. The third mix (*'Mix3'*), by contrast, focused on variation within a single voice, with the bass pairing the vowels */e/* and */a/*. Figure 3 illustrates the relationship between the two spectral envelopes captured to design the vowel filters for *'Mix3'*.

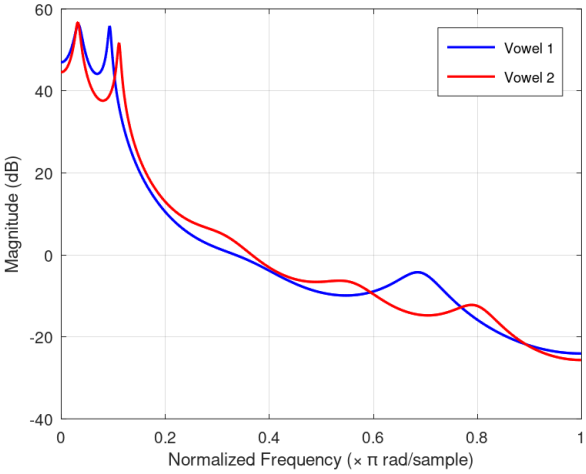

Fig. 3. Spectral envelopes of two LPC-modeled vowels used as filters in the multi-channel reverb effect for *'Mix 3'*: variation within a single voice (*/e/* and */a/*). The plots illustrate how the spectral characteristics of each vowel blend, highlighting differences in formant structure that influence the timbral quality and envelopment of the resulting reverberation. These spectral profiles underpin the perceptual differences explored in the listening tests.

Finally, the mono dry (unprocessed) signal was presented from the Center channel as indicated by ITU-R BS.775-4 [26], while the individual decorrelated channels have been mapped across the 5.1 system. A balance between the dry mono source and the 5.1 reverb effect has been applied with normalization and gain reduction to create an approximation of 50-50 dry-wet mix. For the first condition where only the original mono source is presented, the same mono signal has been multiplied across all loudspeakers.

E. Procedure

In the experiment, participants evaluated two perceptual attributes of reproduced sound: the sensation of reverberant envelopment and the timbral balance between bass and treble. Both attributes were assessed using structured rating scales specifically designed to capture subtle variations in auditory perception. The use of MUSHRA-like procedure and selection of parameters, along with their definitions and corresponding scales, was based on established methodologies for describing sound quality [27]–[29].

Reverb envelopment refers to the sensation of being spatially surrounded by reverberation. As the envelopment becomes more pronounced, it becomes increasingly difficult for listeners to assign a specific position, limited extension, or preferred direction to the reverberant field. The resulting impression is one of diffuse spatial immersion. Such perceptions can arise under both diotic and dichotic (uncorrelated)

presentations of reverberant audio material. To quantify this experience, participants rated the degree of envelopment on a continuum ranging from “almost not pronounced” through “balanced or just right” to “too pronounced”. Psychoacoustic spaciousness is a critical perceptual dimension in both performance and reproduction contexts; this motivates the evaluation metrics chosen here, [20].

Timbral balance denotes the perceived equilibrium between bass and treble. When the balance shifts toward the low end, the sound may be described as dark, marked by excessive bass or insufficient treble. A milder version of this is somewhat dark, where bass remains prevalent. At the centre of the scale is the neutral condition, where bass and treble are perceived as equally loud, regardless of whether they are strong or weak. If this balance alters the prominence of the midrange, the effect is instead assessed under the parameter of midrange strength. Toward the high end of the spectrum, the sound may be judged as somewhat bright, with treble more prominent, and at the extreme, as bright, characterized by excessive treble or weak bass.

The listening experiment was conducted using SAPETool, which presented the stimuli in randomized order [30]. The tool enabled participants to focus on looped segments of audio according to their preference, facilitating a detailed exploration of the perceptual qualities of the samples. Figure 4 presents the SAPETool interface employed in the MUSHRA-like evaluation procedure.

SAPETool is intended for MUSHRA testing. For this study, the software was adapted to follow a modified MUSHRA-like procedure. Participants were presented with multiple stimuli of the same excerpt (*Velvet noise*, *Mix1*, *Mix2*, *Mix3*, and the original unprocessed file) for direct comparison on a continuous rating scale. Unlike a standard MUSHRA test, no hidden reference or explicit low-quality anchor was included; instead, the original file was presented as a perceptual baseline. The scoring scale was adapted to capture timbral balance, with 50 as the neutral case, 0 as “dark,” and 100 as “bright.” These adaptations make the procedure MUSHRA-inspired while tailored to the perceptual dimension under investigation. A similar scale was used in the Envelopment test.

For each parameter, five conditions were tested across two trials: one with piano sounds and the other with violin sounds, resulting in a total of ten audio samples per parameter.[1] Additionally, the definition of the parameter being evaluated and its scale was displayed on the testing interface of SAPETool throughout the duration of the test. Upon completion of the test, SAPETool stored the participants’ responses, making the data readily available for subsequent analysis.

Listening tests were performed in the recording studio of the Department of Multimedia Systems (Gdańsk University of Technology) as a controlled 5.1 listening environment.

[1]The stimuli used for the MUSHRA-like listening tests are publicly available for download and listening at https://www.pizzimusic.com/en/news-37/vowel-reverb.html

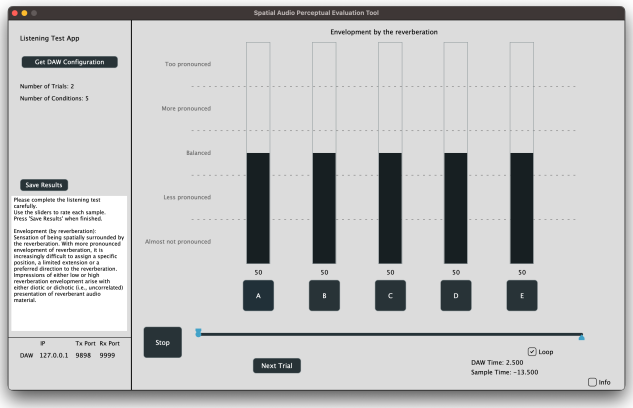

Fig. 4. Example of the SAPETool interface for test envelopment. Participants could listen to and rate each case, or repeatedly examine specific details for comparison. The definition of the parameter under evaluation was continuously displayed in the left panel throughout the test.

IV. Results and Discussion

A. Introduction

This section reports the results of listening experiments on auditory perception of reverberation, with a focus on the perceptual attributes of Envelopment and Timbral Balance. Linear mixed-effects models (LMMs) are employed to assess how variations in stimuli, instrument type, and listener characteristics influence subjective ratings, while simultaneously accounting for individual differences in perception.

B. Model Specification

For each perceptual attribute, responses were analyzed using LMMs in R. The general form of the model is:

$$Rating_{ijk} = \beta_0 + \sum_{p=1}^{P} \beta_p X_{p,ijk} + b_{0i} + \sum_{q=1}^{Q} b_{qi} Z_{q,ijk} + \epsilon_{ijk} \quad (1)$$

where β_p represents the fixed effects, b_{0i} is the random intercept for participant i, b_{qi} are random slopes, and ϵ_{ijk} is the residual error for participant i, stimulus j, and trial k.

Models were fitted using the lme4 package and simplified using backward selection based on likelihood ratio tests ($p > 0.05$). *Post-hoc* analyses were performed with the emmeans package, and diagnostics were assessed using the DHARMa package [31]–[33].

Following a top-down model selection procedure as described by [34], a maximal model was used as the starting point and was systematically simplified based on likelihood ratio tests. The final models for Envelopment (Equation 2) and Timbral Balance (Equation 3) were selected as the most parsimonious structures that adequately fit the data. Equation 2 includes the Stimulus-by-Instrument interaction, main effects of Sex and Age, and participant-level random intercepts and slopes for Stimulus. Equation 3 includes only the main effects of Stimulus and Test Duration, with participant-level random intercepts, reflecting the less complex structure for this perceptual attribute.

$$Rating \sim Stimulus + Instrument + Sex + Age + (1 + Stimulus \mid Participant) \quad (2)$$

$$Rating \sim Stimulus + Test\ Duration + (1 \mid Participant) \quad (3)$$

C. Envelopment Results

Model diagnostics confirmed that residual assumptions were met, Fig. 5. The final model 2 revealed significant effects, detailed in Table I.

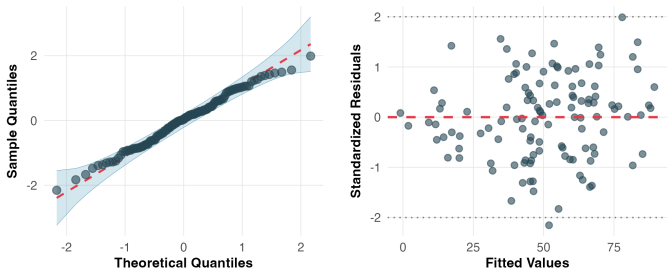

Fig. 5. Model diagnostics for the Envelopment linear mixed-effects model using DHARMa. The left panels show the Q–Q plot of residuals and residuals versus predicted values, indicating no meaningful deviations from normality or from the assumed linear relationship. The right panel shows a scatterplot of residuals versus fitted values; the random, noise-like structure of the points is consistent with homogeneity of variance and supports the adequacy of the model fit.

TABLE I
FIXED EFFECTS ESTIMATES FOR THE ENVELOPMENT LMM

Term	Estimate	Std. Error	t-value
Intercept	18.29	8.96	2.04
Stimulus:Mix1	35.92	10.26	3.50
Stimulus:Mix2	19.92	7.88	2.53
Stimulus:Mix3	42.17	11.60	3.63
Stimulus:Velvet Noise	25.83	9.53	2.71
Instrument:Violin	2.83	4.53	0.63
Sex:Male	11.21	5.54	2.02
Approximate age.L	−14.52	5.77	−2.51
Stimulus:Mix1 × Instrument:Violin (interaction)	−14.25	6.41	−2.22
Stimulus:VelvetNoise × Instrument:Violin (interaction)	12.17	6.41	1.90

Post-hoc comparisons showed significant differences between several stimuli, notably *'Original'* vs. *'Mix3'*, $t_{11} = -3.58$, $p = 0.029$, and *'Mix2'* vs. *'Mix3'*, $t_{11} = -3.83$, $p = 0.019$. Figures 6 and 7 visualize the main and interaction effects.

D. Envelopment Discussion

The analysis demonstrates that the perception of Envelopment is a multifaceted auditory phenomenon shaped by several interrelated factors. Key factors include the characteristics of the audio processing itself (*Stimulus*), the demographic attributes of the listeners (*Sex*, *Age*), and the interaction between the stimulus and the *Instrument*. The use of a mixed-effects model with random slopes for *Stimulus* proved particularly

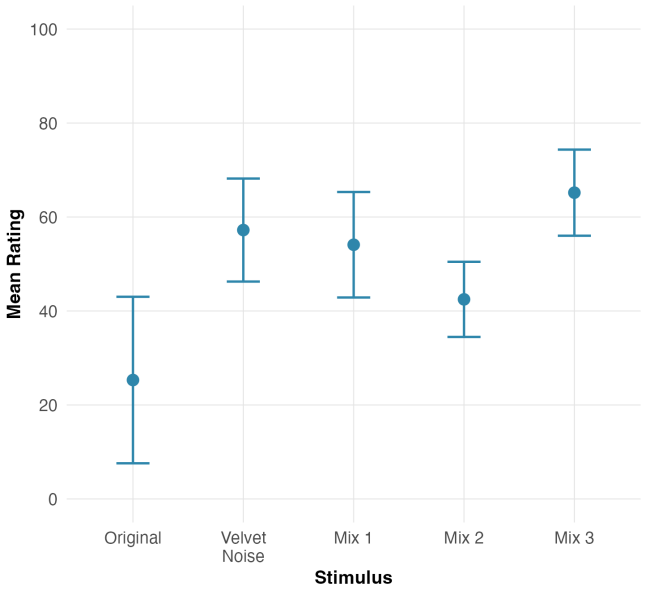

Fig. 6. Confidence Intervals for Envelopment ratings of perceived envelopment across five auditory stimuli. *'Velvet Noise'* and mix conditions (*'Mix 1'*, *'Mix 2'*, *'Mix 3'*) yielded higher envelopment ratings compared to the Original stimulus, with Mix 3 achieving the highest mean score.

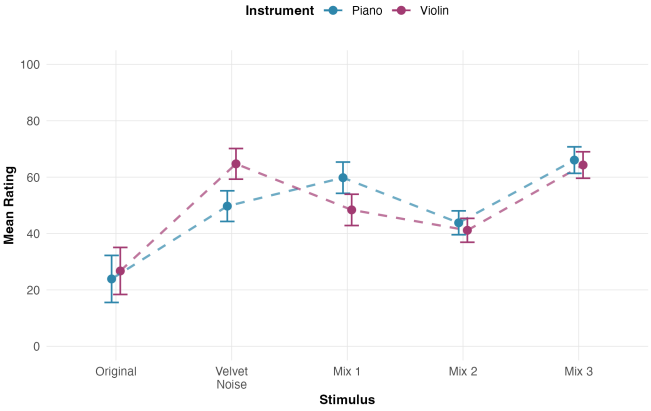

Fig. 7. Interaction plot of stimulus type and instrument (piano, violin) on Envelopment ratings for Piano (blue circles) and Violin (pink circles) across five stimulus conditions (*'Original'*, *'Velvet Noise'*, *'Mix 1'*, *'Mix 2'*, *'Mix 3'*). Both instruments show increased envelopment with added processing. Piano consistently scored slightly higher for *'Mix 1'*, *'Mix 2'*, and *'Mix 3'*, while Violin scored higher for *'Velvet Noise'*.

effective, highlighting the considerable inter-individual variability in how listeners experience this perceptual dimension. Notably, participants consistently rated *Mix1*, *Mix3*, and *Velvet Noise* as significantly more enveloping than the unprocessed Original. This finding suggests that the processing strategies employed in these stimuli systematically alter spatial and reverberant cues in ways that enhance the sensation of immersion. In other words, the manipulations appear to strengthen auditory attributes—such as diffuseness, spaciousness, or the impression of being surrounded by sound—that listeners reliably interpret as heightened envelopment. Taken together, these results underscore both the sensitivity of envelopment perception to subtle acoustic manipulations and the importance

of accounting for listener-specific differences when modeling auditory experience. They also point toward practical implications for audio engineering and sound design, where targeted processing techniques can be leveraged to create more immersive listening environments across diverse audiences.

E. Timbral Balance Results

Diagnostics showed no model violations, as shown in Fig. 8. The model 3 indicated a significant positive effect of *'Test Duration'*, $t_{104} = 2.62$, $p = 0.01$. Stimulus differences were not significant (all $p > 0.05$), as shown in Fig. 9.

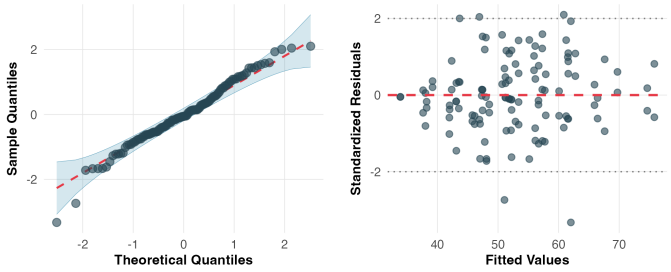

Fig. 8. Model diagnostics for the Timbral Balance linear mixed-effects model using DHARMa. The left panels show the Q–Q plot of residuals and residuals versus predicted values, indicating no meaningful deviations from normality or from the assumed linear relationship. The right panel shows a scatterplot of residuals versus fitted values; the random, noise-like structure of the points is consistent with homogeneity of variance and supports the adequacy of the model fit.

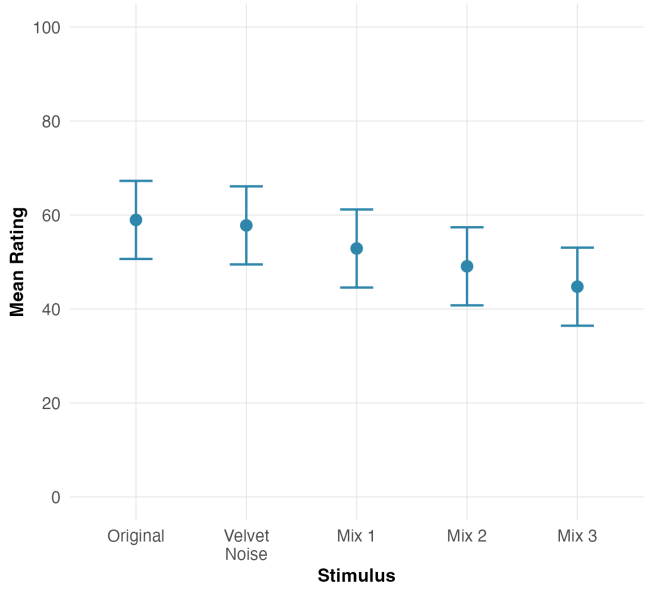

Fig. 9. Confidence Intervals for Timbral Balance ratings for timbral balance (0 = Dark, 50 = Balanced, 100 = Bright) across five stimulus conditions (*'Original', 'Velvet Noise', 'Mix 1', 'Mix 2', 'Mix 3'*). Results indicate a progressive shift from brighter perception in the Original stimulus toward darker perception in *'Mix 3'*, with *'Mix 2'* closest to the balanced reference point.

F. Timbral Balance Discussion

In contrast, ratings of Timbral Balance were not strongly differentiated by the stimuli tested. Instead, the *'Test Duration'* emerged as a significant predictor, suggesting that factors related to the listening session itself, such as fatigue or adaptation, may play a more critical role than the specific audio processing for this attribute. The Timbral Balance assessment revealed no reliable differences between processing variants, though it did show a modest session-related upward drift in ratings. This suggests that combined-vowel processing imparts audible character without disturbing the overall spectral equilibrium, supporting its recognition as a subtle coloration control that enhances presence and vowel-like colour without materially altering perceived brightness or darkness. Considering the system's capacity to yield more extreme outcomes, as demonstrated in the training cases, there remains a broader range of timbral possibilities to explore. Constructing deliberately degraded reverberation is nontrivial, since the perceptual cues that define 'poor quality' are highly context-dependent and may overlap with creative or stylistic choices, making it difficult to design anchors that are both controlled and universally perceived as low-fidelity.

V. Conclusions

The analysis reveals that the perception of Envelopment is a complex phenomenon, significantly influenced by the specific audio processing (*'Stimulus'*), participant demographics (*'Sex'*, *'Age'*), and the interaction between the stimulus and the *'Instrument'*. The successful application of a mixed-effects model with random slopes for *'Stimulus'* underscores the high degree of individual listener variability in this perceptual dimension. Participants perceived *'Mix1'*, *'Mix3'*, and *'Velvet Noise'* as substantially more enveloping than the Original. This implies that the processing used in those stimuli changes spatial or reverberant cues in a way listeners consistently interpret as greater envelopment.

Timbral Balance ratings showed little differentiation across stimuli, with session-related factors such as fatigue or adaptation exerting greater influence than processing variants. The results suggest that combined-vowel processing adds subtle coloration without disturbing spectral equilibrium, while highlighting both the potential for broader timbral exploration and the challenges of defining anchors that are both controlled and consistently perceived as low-fidelity.

These divergent findings highlight the multi-faceted nature of reverberation perception and demonstrate the utility of LMMs in dissecting complex experimental data in psychoacoustics.

In interpreting the results, the small panel of 12 participants constitutes a clear limitation, constraining statistical power and limiting generalizability beyond the tested cohort. Given the novelty of the audio effect and the goal of mapping its perceptual implications, the study is best positioned as exploratory, providing initial evidence of perceptual trends rather than population-level conclusions. These outcomes provide a structured map of perceptual attributes, candidate metrics, and task parameters that can anchor hypotheses and power analyses for subsequent large-scale experiments, including targeted investigations with musically trained listeners and deployments in immersive formats with height channels. While the present

results are exploratory, they suggest promising directions for future work. One such avenue is the development of audio effects that leverage user imported vowel recordings to derive formant filters for customized surround reverberation. This approach would extend speech inspired timbre control [19], [17] into spatial audio contexts, aligning with psychoacoustic insights into spaciousness [2], [20] and the perceptual salience of timbre [35]. Such a system could enable personalized timbral coloration of reverberant fields, offering composers and sound designers new creative tools.

VI. Summary

This work has introduced a novel audio effect that integrates LPC synthesis with velvet noise decorrelation in a new configuration, enabling refined timbre control within immersive reverberation contexts. Through a MUSHRA-like listening procedure, the perceptual dimensions of the effect were systematically explored, with particular attention to the sensations of Envelopment and Timbral Balance.

The findings indicate that the method can impart subtle yet perceptible coloration while maintaining overall spectral equilibrium, thereby offering a flexible tool for shaping the immersive qualities of reverberant sound fields. Beyond its immediate application, the approach highlights a broader potential for combining signal-modeling and decorrelation techniques to expand the creative and perceptual palette of spatial sound design.

Several directions for future research emerge from the present study. One important avenue is the extension of the proposed audio effect to immersive audio formats that incorporate height information, such as Dolby Atmos and Ambisonics. This will enable evaluation of the perceptual impact of the effect in three-dimensional sound fields and allow exploration of its applicability in contemporary spatial audio production environments.

Another direction involves refining the design of the listening experiment by focusing on participants with formal musical training. Restricting the participant pool in this way is expected to yield more nuanced insights into how individuals with heightened auditory sensitivity and domain-specific expertise perceive the effect. Such a refinement would also make it possible to investigate whether musical background systematically influences perceptual judgments. The integration of the exaggerated training cases as anchors within the trials may also be considered for improving future iterations of the study.

At present, only the overall duration of the experiment is recorded, which limits the ability to analyze response behavior in detail. Capturing the time taken for each individual trial will make it possible to identify which stimuli or instrument categories (e.g., violin versus piano) require longer decision-making processes. This additional layer of data will provide a more comprehensive understanding of listener engagement and cognitive load during evaluation.

In sum, the study lays the groundwork for a versatile effect that not only enriches timbral control in immersive reverberation but also opens fertile directions for future research in spatial sound perception and production.

Data Availability

All stimuli employed in the MUSHRA-like listening tests are publicly available for download and playback at https://www.pizzimusic.com/en/news-37/vowel-reverb.html.

Acknowledgements

The authors would like to thank Tomasz Rudzki for valuable advice and guidance in working with SAPETool, which greatly facilitated the identification of a strategy to implement experimental design within the software.

References

[1] G. Kendall, “The decorrelation of audio signals and its impact on spatial imagery,” *Computer Music Journal*, vol. 19, Dec. 1996. [Online]. Available: https://doi.org/10.2307/3680992

[2] S. Disch, “Decorrelation for immersive audio applications and sound effects,” in *Proceedings of the 26th International Conference on Digital Audio Effects (DAFx-23)*, Copenhagen, Denmark, Sep. 2023.

[3] J. Fagerström, “Velvet noise in audio processing,” Ph.D. dissertation, Aalto University, 2025. [Online]. Available: https://aaltodoc.aalto.fi/handle/123456789/134157

[4] B. Alary, A. Politis, and V. Välimäki, “Velvet-noise decorrelator,” in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, Edinburgh, UK, Sep. 2017, pp. 405–411. [Online]. Available: http://www.dafx17.eca.ed.ac.uk/papers/DAFx17_paper_96.pdf

[5] J. Fagerström, B. Alary, S. J. Schlecht, and V. Välimäki, “Velvet-noise feedback delay network,” in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Vienna, Austria, Sep. 2020, pp. 219–226. [Online]. Available: https://www.dafx.de/paper-archive/2020/proceedings/papers/DAFx2020_paper_23.pdf

[6] J. Fagerström, N. Meyer-Kahlen, S. J. Schlecht, and V. Välimäki, “Dark velvet noise,” in *Proceedings of the 25th International Conference on Digital Audio Effects (DAFx20in22)*, G. Evangelista and N. Holighaus, Eds., Vienna, Austria, Sep. 2022, pp. 192–199. [Online]. Available: https://dafx2020.mdw.ac.at/proceedings/papers/DAFx20in22_paper_31.pdf

[7] K. Prawda, S. J. Schlecht, and V. Välimäki, “Multichannel interleaved velvet noise,” in *Proceedings of the 25th International Conference on Digital Audio Effects (DAFx20in22)*, G. Evangelista and N. Holighaus, Eds., Vienna, Austria, 2022, pp. 208–215. [Online]. Available: https://dafx2020.mdw.ac.at/proceedings/DAFx20in22Proceedings.pdf

[8] C. Anemüller, O. Thiergart, and E. A. P. Habets, “Multi-channel neural audio decorrelation using generative adversarial networks,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 58, Nov. 2024. [Online]. Available: https://asmp-eurasipjournals.springeropen.com/articles/10.1186/s13636-024-00378-y

[9] G. Fant, *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations*. Walter de Gruyter, 1971.

[10] J. Makhoul, “Linear prediction: A tutorial review,” *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975. [Online]. Available: https://doi.org/10.1109/PROC.1975.9792

[11] J. D. Markel and A. H. J. Gray, *Linear Prediction of Speech*. Springer Science & Business Media, 2013.

[12] K. N. Stevens, *Acoustic Phonetics*. MIT Press, 2000.

[13] B. Gold, N. Morgan, and D. Ellis, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. Wiley, 2011.

[14] P. Ladefoged and S. F. Disner, *Vowels and Consonants*. John Wiley & Sons, 2012.

[15] B. S. Atal and J. Remde, “A new model of lpc excitation for producing natural-sounding speech at low bit rates,” in *ICASSP ’82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1982, pp. 614–617. [Online]. Available: https://doi.org/10.1109/ICASSP.1982.1171649

[16] P. Lansky and K. Steiglitz, “Synthesis of timbral families by warped linear prediction,” in *ICASSP ’81. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1981, pp. 576–578. [Online]. Available: https://doi.org/10.1109/ICASSP.1981.1171240

[17] P. Lansky, “Compositional applications of linear predictive coding,” in *Current Directions in Computer Music Research*. Cambridge, MA, USA: MIT Press, 1989, pp. 5–8.

[18] J. A. Moorer, "The use of linear prediction of speech in computer music applications," *Journal of the Audio Engineering Society*, vol. 27, pp. 134–140, 1979.
[19] M. Pizzi, "New speech-inspired tools for exploring timbre in computer-based composition and music production," Ph.D. dissertation, University of York, 2018. [Online]. Available: https://etheses.whiterose.ac.uk/id/eprint/29104/
[20] T. Ziemer, *Psychoacoustic Music Sound Field Synthesis: Creating Spaciousness for Composition, Performance, Acoustics and Perception*. Springer, 2019.
[21] J. W. Eaton, D. Bateman, S. Hauberg, and R. Wehbring, *GNU Octave version 10.2.0 manual: a high-level interactive language for numerical computations*, 2025. [Online]. Available: https://www.gnu.org/software/octave/doc/v10.2.0/
[22] J. Stautner and M. Puckette, "Designing multi-channel reverberators," *Computer Music Journal*, vol. 6, no. 1, p. 52, 1982. [Online]. Available: https://doi.org/10.2307/3680358
[23] E. Tarr, *Hack Audio: An Introduction to Computer Programming and Digital Signal Processing in MATLAB*. Routledge, 2018.
[24] European Broadcasting Union, "Sound quality assessment material (SQAM)," EBU, 2008. [Online]. Available: https://tech.ebu.ch/publications/sqamcd
[25] B. Kostek, T. Ciszewski, P. Spaleniak, A. Czyżewski, and S. Zaporowski, "Alofon corpus," Gdańsk University of Technology, 2020. [Online]. Available: https://doi.org/10.34808/7v2c-2y58
[26] International Telecommunication Union, "Recommendation itu-r bs.775-4: Multichannel stereophonic sound system with and without accompanying picture," ITU, Dec. 2022.
[27] ITU-R, "Methods for selecting and describing attributes and terms in the preparation of subjective tests," International Telecommunication Union, Geneva, Switzerland, Tech. Rep. Report ITU-R BS.2399-0, 2017.
[28] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, and S. Weinzierl, "A spatial audio quality inventory (SAQI)," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 984–994, Sep. 2014. [Online]. Available: https://doi.org/10.3813/aaa.918778
[29] N. Zacharov, *Sensory Evaluation of Sound*, 1st ed. United States: CRC Press, 2018. [Online]. Available: https://doi.org/10.1201/9780429429422
[30] T. Rudzki, D. Murphy, and G. Kearney, "A daw-based interactive tool for perceptual spatial audio evaluation," in *Audio Engineering Society Convention 145*, New York, USA, 2018.
[31] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
[32] R. Lenth, "emmeans: Estimated marginal means," 2021, r package version 1.7.0.
[33] F. Hartig, "Dharma: Residual diagnostics for hierarchical models," 2022, r package version 0.4.5.
[34] B. T. West, K. B. Welch, and A. T. Gałecki, *Linear Mixed Models: A Practical Guide Using Statistical Software*, 3rd ed. Chapman and Hall/CRC, 2022, pp. 15–58. [Online]. Available: https://doi.org/10.1201/9781003181064
[35] P. Tużnik, P. Augustynowicz, and P. Francuz, "Electrophysiological correlates of timbre imagery and perception," *International Journal of Psychophysiology*, vol. 129, pp. 9–17, 2018. [Online]. Available: https://doi.org/10.1016/j.ijpsycho.2018.05.005