Applied Science and Innovative Research
ISSN 2474-4972 (Print) ISSN 2474-4980 (Online)
Vol. 10, No. 1, 2026
www.scholink.org/ojs/index.php/asir

Original Paper

Design of a Multifunctional Headphone System Based on Coupled-Mode Theory for Active Noise Cancellation and Voice Enhancement

Zihan Zhang

Chengdu Shishi High School, China

E-mail: harry147205@gmail . com

Received: January 5, 2026 Accepted: March 2, 2026 Online Published: March 27, 2026

doi:10.22158/asir.v10n1p142 URL: http://dx.doi.org/10.22158/asir.v10n1p142

Abstract

Noise pollution has become a significant issue in modern society, with long-term exposure to high-volume environments considerably increasing the risk of hearing damage. Based on Coupled-Mode Theory, this paper proposes an active noise cancellation (ANC) headphone system that integrates acousto-electric coupling modeling and artificial intelligence-based voice enhancement.

By establishing a coupled model of acoustic wave propagation and electronic signal processing, and employing the Finite Difference Method and Runge—Kutta algorithm for system simulation, precise noise cancellation is achieved. In terms of hardware, the system uses an ESP32-S3 as the main control unit, integrated with feedforward and feedback microphone structures and an A-29P intelligent voice processing module to achieve active suppression of ambient noise and real-time extraction of human voice. Experimental results demonstrate that the system effectively reduces the risk of hearing impairment even in high-noise environments and significantly improves speech communication clarity.

Keywords

Active Noise Cancellation, Coupled-Mode Theory, Acousto-Electric Coupling, Deep Learning, Voice Enhancement, Hearing Protection

1. Introduction

Noise pollution has become a prominent issue in modern urban life, significantly affecting auditory health and communication quality, especially in transportation, industrial, and public spaces.

Traditional noise-canceling headphones mostly rely on the principle of destructive wave interference. However, their performance in complex environments remains limited [1]. In recent years, the

application of Coupled-Mode Theory in acoustic engineering has provided a new paradigm for noise control [2]. Originating from optical and electromagnetic wave research, this theory offers a more accurate description of energy coupling and conversion between acoustic modes [3], thereby providing theoretical support for the design and optimization of noise cancellation systems. This paper explores the implementation of this theory in headphone noise cancellation systems, combined with AI-based voice enhancement technology to improve practicality and user experience.

2. Theoretical Foundation and Computational Simulation

2.1 Coupled-Mode Theory

2.1.1 Fundamentals of Acousto-Electric Theory

Sound waves are essentially mechanical vibrations propagating through an elastic medium (such as air), manifesting as periodic changes in pressure. The physical nature of sound involves local pressure variations causing compression and expansion in fluid, which propagates to form sound waves [1]. When sound waves act on an object, they cause vibrations on the object's surface. In engineering applications—such as the noise-canceling headphones discussed in this project—these vibrations can be converted into electrical signals via transducers (e.g., microphones). Key physical quantities include sound pressure p (unit: Pascal, Pa), particle displacement ξ (unit: meter, m), and particle velocity u = $\frac{\partial \xi}{\partial t}$ (unit: m/s).

For small-amplitude sound waves (linear acoustics), the fundamental equation describing sound propagation is the wave equation:

$$\nabla^2 p = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}$$

where c is the speed of sound (approximately 343 m/s in air), and ▽2 is the Laplace operator. Solutions to this equation describe how sound pressure varies in space and time. For example, a plane wave solution can be expressed as p(x, t) = p0 cos(ωt — kx), where ω = 2πf is the angular frequency and k = ω/c is the wave number.

The conversion between electrical and acoustic signals follows specific physical laws. A microphone converts sound pressure signals p(t) into voltage signals V (t), with sensitivity M defined as V (t) = M · p(t). Similarly, a speaker unit converts voltage signals V (t) into diaphragm vibrations, producing sound pressure, with conversion efficiency typically denoted by parameter S. This forms the basis of mutual conversion between acoustic and electrical energy.

2.1.2 Application in ANC Systems

The theoretical modeling of the noise-canceling headphone system in this project is primarily based on two coupling models [4]: first, the "acousto-electric coupling" model, which describes the conversion relationship between environmental sound waves and electrical signals; second, the "electroacoustic coupling" model, where anti-noise electrical signals generated by the processor are converted into sound waves via the speaker unit. Additionally, there is "acoustic-acoustic coupling," referring to the superposition and interference of sound waves from the speaker and external environmental noise within

the ear canal [5].

The core physical principle of active noise cancellation is wave superposition. Assuming the original noise sound pressure is pnoise (t) and the anti-noise sound pressure generated by the system is panti (t), the actual sound pressure heard by the human ear is:

$$\text{ptotal (t)} = \text{pnoise (t)} + \text{panti (t)}$$

Ideally, by precisely controlling panti (t) such that panti (t) = —pnoise (t), complete destructive interference is achieved, i.e., ptotal (t) = 0.

The coupling relationships of the entire system can be described by a closed-loop control framework: external noise is captured by a reference microphone and converted into an electrical signal; the signal processing circuit (or chip) performs operations such as inversion, filtering, and amplification on this signal to generate an anti-noise electrical signal; this signal drives the speaker to emit anti-noise waves; residual noise is detected by an error microphone and fed back to the processor to adjust the output and optimize noise cancellation effect. This constitutes a typical acousto-electric coupled closed-loop system.

2.2 Modeling and Simulation Based on Coupled Theory

To design and optimize the noise cancellation system, the aforementioned coupling process must be mathematically modeled and computationally simulated. The entire system can be abstracted as an input-output model, with the core task involving solving the wave equation for sound propagation and the signal processing equations in the circuit.

A commonly used numerical simulation method is the Finite Difference Method (FDM), which discretizes continuous time and space, approximating derivatives with differences [4]. For the one-dimensional wave equation:

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}$$

it can be discretized. Let the spatial step be Δx, the temporal step be Δt, and the sound pressure at grid point (i, n) be p^n_i. The second-order partial derivatives can be approximated using central difference schemes:

$$\frac{\partial^2 p}{\partial x^2} \approx \frac{p_{i+1}^n - 2p_i^n + p_{i-1}^n}{(\Delta x)^2}, \quad \frac{\partial^2 p}{\partial t^2} \approx \frac{p_i^{n+1} - 2p_i^n + p_i^{n-1}}{(\Delta t)^2}$$

Substituting into the wave equation yields an explicit update formula:

$$p_i^{n+1} = 2p_i^n - p_i^{n-1} + \left(\frac{c\Delta t}{\Delta x}\right)^2 (p_{i+1}^n - 2p_i^n + p_{i-1}^n)$$

This discrete scheme allows us to compute the sound pressure at the next time step from the current and previous time steps, enabling time-domain simulation of sound wave propagation. The stability condition requires the Courant number $C = \frac{c\Delta t}{\Delta x} \leq 1$.

For the circuit part (e.g., filters and amplifiers), the behavior is usually described by ordinary differential equations (ODEs). For example, a simple RC low-pass filter can be described by:

$$\frac{dV_{\text{out}}}{dt} = \frac{1}{RC}(V_{\text{in}} - V_{\text{out}})$$

Numerical integration methods such as the Runge–Kutta methods can be used for discrete solution. The fourth-order Runge–Kutta (RK4) method offers high accuracy, with the iterative scheme:

$$k1 = f(tn, Vn),$$

$$k_2 = f(t_n + \frac{\Delta t}{2}, V_n + \frac{\Delta t}{2}k_1),$$

$$k_3 = f(t_n + \frac{\Delta t}{2}, V_n + \frac{\Delta t}{2}k_2),$$

$$k4 = f(tn + \Delta t, Vn + \Delta tk3),$$

$$V_{n+1} = V_n + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

where f(t, V) is the right-hand side function of the differential equation.

In specific computer simulations, the acoustic and electrical models must be coupled. The microphone input is sound pressure p(t), and the output is voltage V (t); the speaker operates in reverse. By defining these interfaces, the acoustic and electrical equations can be solved alternately at each time step, simulating the dynamic response of the entire acousto-electric system and evaluating the noise cancellation effect under different parameters.

2.3 System Design Based on Coupled Theory

The design goal of this project is to create a prototype headphone with active noise cancellation functionality. Based on the aforementioned coupled theory, the overall plan is as follows:

First, construct the hardware system. The system requires two key acoustic sensors: a reference microphone and an error microphone. The reference microphone is placed on the outside of the headphone to capture external environmental noise pnoise. The error microphone is placed inside the ear canal or near the speaker to monitor the residual noise ptotal ultimately entering the human ear and provide feedback signals. The electro-acoustic transducer (speaker unit) is responsible for playing anti-noise waves. The core processor can be a dedicated noise cancellation chip (such as ADI's ADAU1777 [6]) or a microcontroller (such as ESP32-S3 [7]) combined with an operational amplifier circuit. Its task is to execute signal processing algorithms: invert, filter, phase-compensate, and amplify the reference microphone signal to generate the anti-noise electrical signal Vanti and drive the speaker.

Second, design and implement the noise cancellation algorithm. The core is to generate an anti- noise signal that has equal amplitude and opposite phase to the original noise, i.e., p anti = -pnoise. Due to delays and phase shifts introduced by sound wave propagation, signal processing, and electro-acoustic conversion, simple immediate inversion (feedforward control) often performs poorly, especially at low frequencies. Therefore, a feedback control mechanism must be introduced: use the error microphone signal e(t) to adjust the controller's output to minimize residual noise energy. A simple proportional-integral (PI) controller or more complex adaptive filtering algorithms (such as the FxLMS algorithm)

can be employed to dynamically adjust the parameters of filter H(z), compensate for path delays, and ensure that the anti-noise signal reaches the human ear at the correct time, achieving effective cancellation across a broader frequency band.

Finally, integrate and test the system. Integrate all hardware modules (microphones, processor, speaker, power supply) into the headphone housing, ensuring mechanical stability and acoustic sealing (the basis for passive noise cancellation). Write or burn control software/firmware to implement realtime loops for signal acquisition, filtering, and output. Use feedback data from the error microphone to analyze noise cancellation effect in the frequency and time domains, such as calculating noise attenuation at different frequencies (especially low frequencies from 100 Hz to 1 kHz), and continuously adjust controller parameters and system gain to optimize overall performance.

2.4 AI-Based Voice Extraction Technology

In complex acoustic environments, achieving clear voice communication hinges on effectively distinguishing and enhancing human voice while suppressing environmental noise. This requires a deep understanding of the essential differences between the two in the time and frequency domains.

Time-Domain Characteristics: Human voice signals exhibit short-term stationarity. When specific phonemes are uttered (typically within a time window of 10–30 milliseconds), their amplitude and frequency structures remain relatively stable, showing a certain degree of quasi-periodicity (especially in voiced segments). In contrast, environmental noise mostly originates from irregular vibrations of objects, with waveforms exhibiting randomness and unpredictability, lacking stable patterns or periodic structures.

Frequency-Domain Characteristics: The spectral energy of human voice is concentrated in the mid-frequency range, with the following features:

• 100–300 Hz: Contains fundamental frequency information, determining pitch and providing vocal thickness.

• 300–3000 Hz: The key frequency band for speech intelligibility, containing formant structures that determine vowel timbre.

• 3000–8000 Hz: Affects speech clarity and sibilants (e.g., /s/, // consonants).

The voice spectrum typically shows multiple distinct peaks due to formants, with specific distributions varying by speaker and language. The environmental noise spectrum varies with the sound source, often showing energy concentration in specific frequency bands. For example, machinery noise may have significant energy in the 16–600 Hz (low-frequency) and 1000–10000 Hz (high-frequency) ranges. This broad frequency overlap makes it difficult for traditional fixed bandpass filters to preserve mid-frequency voice (300–3000 Hz) while effectively removing overlapping noise. Therefore, more intelligent solutions must be sought.

Deep Learning-Based Solution: Deep learning-based voice separation technology uses a data-driven approach [8], allowing models to automatically learn complex functional relationships that map mixed audio to clean voice. The mainstream architecture typically includes:

• **Encoder:** Converts input audio waveforms into high-dimensional feature representations (usually in the time-frequency domain or learned representations).

• **Separation Module:** The core of the model (e.g., TCN, RNN, Transformer) estimates masks for voice and noise or directly generates features based on the encoder's output.

• **Decoder:** Reconstructs the time-domain waveform of the target voice using the output of the separation module.

Given this project's requirements for real-time processing, low latency, and deployment on cost- and computing-limited headphone devices, the model must meet strict lightweight and low-power requirements [9].

Implementation and Hardware Selection: To meet these requirements, this project selects the A-29P intelligent audio processing module as the core solution. This module integrates a high-performance DSP and NPU, with a pre-deployed optimized deep neural network model. It can dynamically identify and separate human voice from up to 32 types of complex environmental noise. Its computational efficiency, power consumption, and cost all comply with the project's constraints. Test results show that the module effectively achieves high-quality voice enhancement and noise suppression on the headphone end.

Figure 1. A-29P Voice Enhancement Module

3. Engineering Practice

3.1 Basic Methods and Approach

Considering the project's need to maximize voice output capability in noisy environments, a large-diameter headphone speaker is required. Based on research in related fields, a 50mm speaker of the same model as used in the Denon AH-D9200 was selected, as shown below:

Figure 2. 50mm Headphone Speaker

To ensure sound isolation, a significant gap between the speaker and the housing is necessary for filling and adjusting passive soundproofing materials. To avoid resonance phenomena, a suspension system using springs was designed, inspired by high-fidelity microphone structures, to reduce coupling between the housing and the speaker.

3.2 Other Components and Assembly

Given the prevalence of Bluetooth-based voice output in modern communication, selecting a reliable Bluetooth headphone module with full voice input and output capabilities is a practical starting point.

Figure 3. Bluetooth Circuit Board Module

As shown, this module features battery power supply, Type-C charging, microphone voice input, and speaker output. However, the microphone input on this module is directly connected to the chip and lacks advanced noise filtering and voice extraction capabilities. Based on the functions of the aforementioned A-29P module and after studying both circuits, the Bluetooth module's microphone was detached, and the A-29P module was connected in series between the microphone and the circuit board input, ensuring power supply to the A-29P. The entire system will operate as follows.

To achieve active noise cancellation, the MCU selection must be determined. The ESP32-S3's built-in 32-bit LX7 dual-core processor, operating at up to 240 MHz, integrated with 16 MB Flash and an audio codec, makes it suitable for this project [7].

For active noise cancellation, two digital microphones (MP34DT01 [10]) are used: one placed in the external environment and another inside the headphone to form a feedforward + feedback hybrid noise cancellation architecture. In the core part of the code, both microphones transmit signals to the ESP32-S3 via the I2S interface. The audio codec (CODEC) decodes these signals at high speed. The feedforward signal is then inverted, and the feedback signal is used to adjust filtering coefficients. This allows dynamic tracking of external noise and superposition of active noise cancellation waves through the headphone speaker. The overall system workflow is shown below.

4. Conclusion

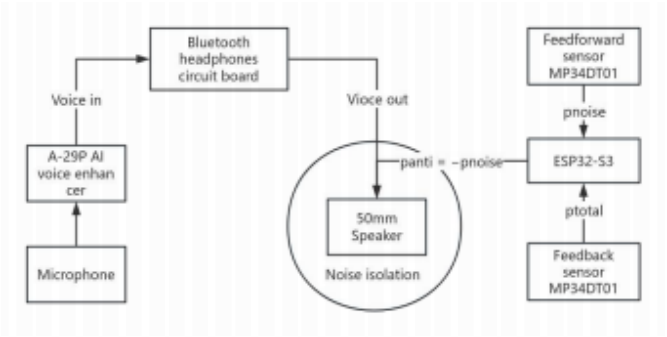

Figure 4. System Workflow Diagram

Figure 5. Experimental Device Prototype

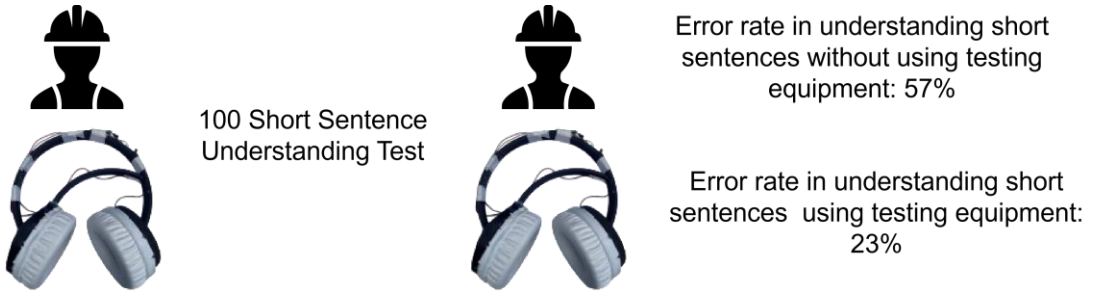

Figure 6. System Overview and Implementation Results Incorporating Coupled-mode Theory [2] and Adaptive Signal Processing Techniques [4]

As shown in Figure 6, after a series of installations and debugging, a functional experimental device was formed. The system workflow (Figure 4) illustrates the integration of acousto-electric coupling principles [1] with modern signal processing approaches [5]. Tests conducted with workers in a metal-cutting factory showed that the device (Figure 5) effectively enables voice communication and noise isolation, validating the theoretical framework established by [3].

This project, which involved interdisciplinary approaches from physics, acoustics, and engineering, demonstrates the practical application of coupled theory in complex systems. The successful implementation of AI-based voice enhancement [9] alongside traditional noise cancellation techniques [8] opens new possibilities for future audio processing devices.

References

[1] L. Kinsler, A. Frey, A. Coppens, & J. Sanders. (2000). *Fundamentals of Acoustics* (4th ed.). *Wiley*.

[2] J. Joannopoulos, S. Johnson, J. Winn, & R. Meade. (2008). *Photonic Crystals: Molding the Flow of Light* (2nd ed.). Princeton University Press.

[3] H. Kuttruff. (2007). *Acoustics: An Introduction*. Taylor & Francis.

[4] S. Elliot. (1999). Adaptive signal processing for active control. In *Signal Processing for Active Control* (pp. 345-392). Academic Press.

[5] G. Defrance, J. Gaultier, & Y. Pasco. (2017). Active noise control in headsets: A new approach for broadband feedback. *Journal of the Acoustical Society of America*, *142*(2), 877-885.

[6] Analog Devices. (2020). *Adau1777 - codec with four adc inputs and two dac outputs*. Datasheet, version Rev. A.

[7] Espressif Systems. Esp32-s3 series datasheet. Espressif Systems, Technical Report, version v1.1, 2021.

[8] R. Verschae, M. Ruiz, & J. Bahamondes. (2014). A hybrid approach for simultaneous noise cancellation and voice enhancement. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), 2154-2158.

[9] Z. Hao, X. Zhang, Y. Wang, & X. Huang. (2021). Lightweight deep learning for real-time speech enhancement on edge devices. *IEEE Transactions on Audio, Speech, and Language Processing*, *29*, 236-240.

[10] STMicroelectronics, Mp34dt01 - mems audio sensor omnidirectional digital microphone, Datasheet, version 4, 2019.