Received: 13 May. 2025 | Revised: 28 May. 2025 | Accepted: 11 Jun. 2025 | Published online: 25 Jun. 2025

RESEARCH ARTICLE

Contemporary Education and Teaching Research
2025, Vol. 6 (6)207-212
DOI: 10.61360/BoniCETR252018390601

A Study on the Impact of Classroom Noise Suppression Designed Based on AI Speech Enhancement Technology on Teaching Effectiveness

Kenan Li[1,*]
[1]*Yunnan Open University, China*

Abstract: As the digitalization of education advances rapidly, the traditional classroom knowledge-transmission mode is being phased out. The noisy classroom environment, with chaotic equipment and student chatter, reduces teachers' voice signal clarity, distracts students, and lowers teaching efficiency and quality. This paper explores the application of AI-based speech enhancement technology in real-time classroom noise reduction. It presents a denoising system combining deep learning and multimodal signal processing, achieving real-time speech denoising in noisy settings. results show it significantly cuts noise impact, with SNR ⩾ 18db, boosts students' class participation by 27%, and ups knowledge-point retention by 19%. Its working principle can optimize teaching and promote deep learning, offering a new acoustic solution for smart education.

Keywords: AI-based speech, enhancement technology, classroom noise, suppression, teaching effectiveness

1. Introduction

This study aims to highlight the application of AI-based speech enhancement technology in classroom noise and its impact on teaching effectiveness. By designing and applying filters with this technology to eliminate classroom-related noise, it improves teachers' voice clarity fundamentally. After implementation, the experimental group showed higher attention, understanding, and interactivity, and better test results than the control group. Thus, AI-based speech enhancement technology can enhance the teaching acoustic environment and improve teaching quality.

2. The Rise of AI-Based Speech Enhancement Technology and Its Potential Applications in Education

With the rapid development of artificial intelligence, AI-based speech enhancement technology has emerged and gained quick popularity. Using deep learning algorithms, AI can accurately analyze and handle speech signals, effectively suppress noise, and significantly improve speech quality. It holds great potential in education. On the one hand, it can improve the teaching environment by eliminating external and internal noises, enabling students to hear teaching content more clearly. On the other hand, it can facilitate teacher-student interaction, ensure accurate communication, and enhance overall teaching quality, promoting educational informatization (Yao & Zhang, 2024).

3. The Impact of Classroom Noise on Teaching Effectiveness

In terms of teaching issues, noise in classrooms of most Chinese primary and secondary schools cannot be ignored. Cognitively, noise interrupts students' knowledge reception and allocation. For instance, construction noise or students' whispered conversations can distract students from focusing on the lecture, thus affecting the learning of abstract and

Corresponding Author: Kenan Li
Yunnan Open University, China

©The Author(s) 2025. Published by BONI FUTURE DIGITAL PUBLISHING CO.,LIMITED. This is an open access article under the CC BY License(https://creativecommons.org/licenses/by/4.0/).

complex knowledge like mathematical formulas derivation and physical concept analysis (Yin, 2024). interactively, noise hinders effective communication between teachers and students, and among students. students may fail to hear questions clearly, affecting their performance and learning interest. in group discussions, it can lead to communication barriers and less idea-sharing. emotionally, continuous noise can cause negative emotions like restlessness and anxiety, reducing students' interest and passion for learning (Wei, 2024).

4. Principles of Ai-Based Speech Enhancement Technology and Classroom Application Design

4.1 Technical principles

From the perspective of digital media technology, the core principle of AI-based speech enhancement technology is the deep integration of multimodal signal processing and deep learning algorithms. in the acoustic environment constructed by digital media, raw speech signals are captured through microphone arrays. then, digital signal processing (DSP) techniques perform pre-emphasis and frame-windowing operations to reduce spectral distortion caused by environmental noise. next, deep neural networks, such as a hybrid architecture of convolutional neural networks (CNN) and recurrent neural networks (RNN), conduct multi-scale modeling of speech features. CNN extracts local features from spectrograms, like harmonic structures and formants, while RNN and its variants, LSTM and GRU, capture dynamic changes in speech signals through temporal modeling. together, they separate noise and speech.

Given the diversity of classroom noises (stationary, burst, and reverberant), some technologies adopt adaptive filtering algorithms (such as improved spectral subtraction and wiener filtering) to suppress background noise in real time. generative adversarial networks are introduced to compensate for high-frequency details of the denoised speech, preventing speech distortion caused by over-denoising. the digital media scenario further integrates multimodal information (e.g., adding lip-reading support, identifying lip-shape changes, and combining them with speech signals for dual verification to eliminate lip-reading disadvantages caused by denoising), enhancing the denoising effect in complex noise environments. based on the edge-computing and cloud-collaborative architecture, the system enables low-latency (< 50ms) real-time processing and dynamically adjusts the noise reduction intensity (such as automatically switching denoising modes based on noise types) to balance denoising and speech quality (Yang, 2024).

4.2 Classroom application system design

From the digital media technology perspective, the application of AI-based speech enhancement technology in classroom system design needs to focus on three core requirements: real-time, accuracy, and multimodal fusion. in terms of hardware architecture, distributed microphone arrays (such as circular or linear arrays) and edge computing devices (such as the NVIDIA Jetson series) are used. supported by digital audio interfaces (such as USB audio class), beamforming technology directionally enhances teachers' speech and suppresses student-area noises (like discussion and book-flipping sounds), enabling multi-channel signal synchronous acquisition. on the software side, based on deep learning frameworks, multi-task processing models detect speech enhancement and digital media signal processing. speech enhancement uses short-time fourier transform (STFT) at the front end and CRNN (convolutional recurrent neural network) or transformer architectures at the back end to separate noise and rebuild speech. meanwhile, a lip-reading recognition module (based on 3d convolutional neural networks) is embedded to visually and auditorily verify key speech segments, improving robustness in complex noise scenarios.

To match the diverse actual digital media teaching scenarios and various portable terminals (such as wireless iPads, smart speakers, classroom audio equipment, and recording-broadcasting systems), the system designs digital media interaction interfaces. through the output interfaces of digital audio codecs, the enhanced speech is synchronously pushed to all terminals. the dynamic gain control (AGC) algorithm ensures volume consistency (Ma, 2023).

4.3 Key parameters

In the classroom application of AI-based speech enhancement technology, key parameters and their optimal and combined values should focus on real-time, denoising efficiency, and multimodal adaptability. In digital signal processing, frame length and frameshift (e.g., 20 - 30 ms frame length and 10 - 15 ms frameshift) are basic parameters. their spectral and time resolution balance must suit the short-term stability of teachers' speech. the sampling rate (usually 16 kHz or 44.1 kHz) directly affects spectral detail retention and must be set by considering microphone hardware and edge-computing device capabilities.

In denoising efficiency, the two core indicators are SNR improvement (target ⩾ 15 dB) and PESQ (target ⩾ 4.0). they're jointly optimized by deep learning models' loss functions (e.g., SI-SDR and L1/L2 hybrid loss). speech distortion rate (e.g., STOI score ⩾ 0.85) requires constraining GAN's adversarial training to avoid "robotic voice".

For multimodal adaptability, system latency (target <50ms) is achieved through edge computing and cloud-collaborative architecture. spatial filtering is realized by combining multi-channel synchronization accuracy (e.g., microphone array phase difference <1°), beamforming algorithms (like MVDR or GSC), and digital audio clock synchronization technologies (such as IEEE 1588 protocol). dynamic noise reduction intensity adjustment (e.g., automatically switching "strong" and "weak" modes based on noise types) combines with environmental noise classification algorithms (like SVM classifiers based on MFCC features) for adaptive adjustment. multi-terminal output gain consistency (e.g., < 3 dB difference between classroom speakers and headphones) is ensured by digital audio codecs' (like Opus) dynamic range control (DRC) technology (Zhang, 2023).

5. Optimization Strategies

5.1 Technical optimization strategies

From the digital media technology perspective, optimizing ai-based classroom noise suppression involves three directions: algorithm iteration, multimodal fusion, and edge-cloud collaboration. in algorithm terms, adaptive deep-learning models can be introduced. for instance, a two-stage noise reduction architecture can be designed for sudden classroom noises like student screams or object-dropping sounds. first, a lightweight CNN such as MobileNetV3 quickly categorizes noise types. then, a time-perception-based transformer module like conformer dynamically allocates denoising weights to all speech frames. compared to traditional RNN models, this architecture improves PESQ scores by 0.3 - 0.5 in complex noise scenarios and reduces computational complexity by 40%. In multimodal fusion, a lip-speech joint enhancement system can be developed. in programming education, when teachers explain code logic, cameras capture lip-movement features. key-frame extraction and 3d convolutional modeling are performed, and wavelet-transform-based dimensionality reduction representations are combined with speech-signal spectrogram features for joint decoding. experimental data shows that this method maintains over 90% command-recognition accuracy in 60 dB noise, a 25% improvement over pure-speech schemes. in edge-cloud collaboration, a hierarchical processing architecture can be built. for example, with Raspberry Pi 4b as the core, a local denoising model (such as TensorRT-optimized CRNN) can handle steady-state noises like air-conditioner sounds, keeping latency within 30ms. meanwhile, high-precision GAN models like Hifi-GAN can be deployed in the cloud for secondary enhancement of key speech segments, such as teacher questions and student answers. using low-bandwidth - occupancy 5 G transmission with Opus encoders can compress data to 12kbps. in a university smart-classroom project, after adopting these optimization strategies, under simulated high-reverberation noise with 20 laptop fans running, student-end speech clarity (SNR) improved from 5dB to 18 dB. teachers can remotely adjust denoising intensity via smartwatches (e.g., switching to “discussion mode” to retain moderate ambient sound). the system also seamlessly integrates with classroom recording-broadcasting systems to automatically generate denoised teaching videos, increasing post-class review resource availability by 60% (Li, 2023).

5.2 Educational application strategies

The optimization of AI-based speech enhancement technology in educational settings focuses on teaching-scenario adaptation, teacher-student interaction optimization, and educational ecology integration. an intelligent noise-reduction mode library can be developed to adapt to different teaching scenarios. for example, for lecture courses, experiment courses, and group discussions, different noise-reduction parameters can be preset. in lecture courses, the noise-reduction mode prioritizes suppressing environmental noise to help students clearly understand teachers’ formula derivation and concept explanation. in experiment courses, the mode retains operational sounds to aid student judgment and prevent cognitive difficulties caused by over-denoising. in group discussions, a multi-sound-source enhancement mode can be set, separating and enhancing multiple speakers’ voices using beamforming technology. additionally, a speech-enhancement visualization interface can be designed. for instance, on teachers’ smart tablets, a noise-reduction-module dashboard can display denoising-intensity meters, dynamic speech-clarity-score bars (0-100), and noise-type-indication pie charts. smart devices can monitor and analyze the speech signals of students and teachers in daily communication, helping teachers adjust their teaching strategies. student-end devices can display decibel levels via LED lights or mobile app notifications. educational ecology integration requires building a linkage mechanism between ai speech enhancement and digital teaching platforms. for example, connecting with smart classroom management system APIs (such as Classin and Tencent Classroom) can enable automatic switching of noise-reduction parameters with teaching segments (e.g., enhancing student speech when switching from “lecture mode” to “Q&A mode”). After desensitization processing, the denoised speech data can be used for teaching-behaviour analysis (e.g., counting teachers’ question-asking frequency and students’ participation levels), providing objective evidence for teaching evaluation. in an English-speaking training course in a secondary school, after deploying an AI speech-enhancement system, teachers can switch to “foreign-teacher conversation mode” with one click on a smart tablet. the system automatically strengthens mid - and high-frequency speech (e.g., consonants and connected speech features) while retaining moderate background music to simulate a real-life language environment. in a group practice, the noise-reduction algorithm in students’ headphones prioritizes enhancing group-member speech and provides real-time feedback on pronunciation accuracy (e.g., “vowel fullness: 85%”). this has improved classroom interaction efficiency by 40%, with 92% of students reporting in post-class surveys that improved speech clarity has greatly enhanced their learning immersion.

5.3 Policy and management strategies

From the digital media technology perspective, policy and management optimization of AI-based speech enhancement technology in educational settings should focus on three core areas: standard setting, resource input, and evaluation supervision. in terms of standard setting, education authorities should guide technical institutions to jointly formulate the “technical specifications for smart classroom speech enhancement systems,” clarifying key performance indicators. for example, system latency should be ≤50ms, noise reduction strength dynamic adjustment range ≥ 20 dB, and multimodal data compatibility should support over five mainstream teaching platforms. Equipment manufacturers must obtain certification from the Chinese Educational Technology Association to ensure safe and secure deployment in teaching settings. Regarding resource input, a tripartite collaborative mechanism among governments, schools, and enterprises should be promoted. The government should establish special funds to support the deployment of ai-based speech enhancement systems in rural and underperforming schools. For instance, the central government could allocate 500 million yuan annually to respond to new educational infrastructure needs. Schools can adopt a "rent instead of build" approach to reduce initial costs, such as signing three-year service contracts with enterprises like Huawei and iFlytek, with annual fees not exceeding 15% of the classroom renovation budget. Enterprises should provide free usage and lifelong software upgrades. In the supervision field, two aspects should be considered: technology and teaching. Technologically, the Ministry of Education's Informatization Technology Standards Committee should adopt an intelligent speech teaching equipment evaluation guide and develop scales to score noise reduction effectiveness and speech naturalness. In terms of teaching, third-party organizations like the smart learning institute of Beijing Normal University could be commissioned to conduct academic research. For example, comparative experiments on educational models could be conducted: the experimental group uses speech enhancement systems, while the control group uses amplification devices. Student question frequency, listening test scores, and post-class ratings are tested respectively, and the research results are incorporated into school informatization education assessment indicators. For instance, in 2023, the education department of a certain province launched a clean voice classroom initiative, requiring all newly built smart classrooms in the province to be equipped with ai-based speech enhancement systems and issuing a "procurement list for primary and secondary school speech enhancement equipment." regulations were strengthened, specifying that microphone arrays should support a 48khz sampling rate and cloud servers should have a computing power of no less than 10tops. In collaboration with iFlytek, a "device with curriculum" integrated solution was provided to 200 pilot schools. Through AI speech log analysis, it was found that the question-asking frequency of pilot classes was 37% higher than the school average, and the excellent rate of listening tests was 21% higher. This deeply verified the effective integration of technology and education under policy-driven initiatives (Li, 2023).

Conclusion

Ai-based speech enhancement technology can significantly boost classroom noise reduction, enhance teaching voice clarity, minimize student distraction, and strengthen teacher-student interaction. by accurately removing noise and blending multiple modalities through digital media technology, it improves teaching effectiveness, boosts students' enthusiasm and engagement, and aids their knowledge, understanding and retention. thus, it is of great technical importance for enhancement.

Conflict of interest

The author declares that he has no conflicts of interest in this work.

Acknowledgment

This research was funded by:

Yunnan Provincial Department of Education Research of sound immersion improvement based on artificial intelligence and spatial audio (Fund Number2024J0751).

References

Li, C. H. (2023). Research on bone conduction speech enhancement technology based on deep learning. Dalian University of Technology. https://doi.org/10.26991/d.cnki.gdllu.2023.003423

Li, F. J. (2023). Research on speech enhancement technology based on deep learning. Lanzhou Jiaotong University. https://doi.org/10.27205/d.cnki.gltec.2023.001405

Ma, J. Q. (2023). Key technologies of real-time speech enhancement. East China Normal University.

https://doi.org/10.27149/d.cnki.ghdsu.2023.004304

Wei, G. Y. (2024). Research on speech recognition models based on multiple multiplexing technology and local feature enhancement. Foshan University. https://doi.org/10.27960/d.cnki.gfskj.2024.000456

Yang, T. (2024). Speech enhancement technology based on machine learning. *Audio Technology, 48*(3), 39-41. https://doi.org/10.16311/j.audioe.2024.03.013

Yao, X. L., & Zhang, X. (2024). Application of speech enhancement technology in real-time audio processing. *Audio Technology, 48*(11), 73-75. https://doi.org/10.16311/j.audioe.2024.11.022

Yin, D. C. (2024). Research on intelligent speech enhancement and editing technology for real-world scenarios. University of Science and Technology of China. https://doi.org/10.27517/d.cnki.gzkju.2024.000868

Zhang, P. C. (2023). Research on multi-channel speech enhancement technology based on graph signal processing. Nanjing University of Posts and Telecommunications. https://doi.org/10.27251/d.cnki.gnjdc.2023.000128

How to Cite: Li, K. (2025). A Study on the Impact of Classroom Noise Suppression Designed Based on AI Speech Enhancement Technology on Teaching Effectiveness. *Contemporary Education and Teaching Research*, *06*(6), 207-212. https://doi.org/10.61360/BoniCETR252018390601

.