CATCHING Z'S PROPOSAL

By

Srikar Palani

Prineet Parhar

Suprathik Vinayakula

Final Report for ECE 445, Senior Design, Spring 2026

TA: Zhuchen Shao

6 May 2026

Project No. 88

Abstract

Catching Z's is a bedside embedded system designed to stop sleep fragmentation caused by sudden environmental noises. Unlike conventional white noise machines that operate continuously, this system monitors ambient acoustics and activates masking noise only when disruptive sound events occur. The device utilizes a high sensitivity electret microphone to establish a baseline noise profile and an ESP32-S3 microcontroller to detect amplitude and frequency spikes. When a noise disturbance is detected that exceeds the baseline by 10 decibels, the system generates a masking audio, either white, pink, or brown noise within 100 milliseconds. This report goes over the design and verification of the acoustic sensing, signal processing, and power regulation subsystems.

Contents

1. Introduction

Sleep fragmentation caused by sudden environmental noises such as police sirens or barking dogs negatively impacts cognitive performance and long-term health. Conventional white noise machines attempt to solve this by operating continuously at a pre-set volume. However, this approach can be ineffective against short, high-volume disturbances and unnecessary during quiet periods. The purpose of this project is to engineer a smart system that continuously monitors noise and counteracts it only in response to disruptive sound events.

The following chapters detail the engineering solution to this problem. Chapter 2 describes the design of the sensing, processing, and power subsystems. Chapter 3 presents the verification procedures used to ensure the system meets latency and output pressure requirements. Chapter 4 details the cost analysis of the prototype, and Chapter 5 summarizes any uncertainties and future work.

1.1 Solution

As a solution we propose Catching Z's, a bedside system that monitors ambient audio in and adaptively generates masking noise in response to disruptive sound events. The device uses a high sensitivity electret microphone to constantly sample room audio. An onboard microcontroller (ESP32-S3) processes this data to establish a baseline for the ambient noise. When the system detects a sudden spike in amplitude that exceeds the established baseline by a set threshold it activates the audio output subsystem.

The audio subsystem generates masking audio that smoothly fades in to neutralize the disturbance. Once the environmental noise returns to the baseline level, the masking noise gradually fades out. This adaptive response minimizes unnecessary noise exposure while preventing the masking system itself from waking the user. The user can adjust the sensitivity of the detection and the maximum volume of the masking noise using a rotary encoder interface.

1.2 Visual Aid

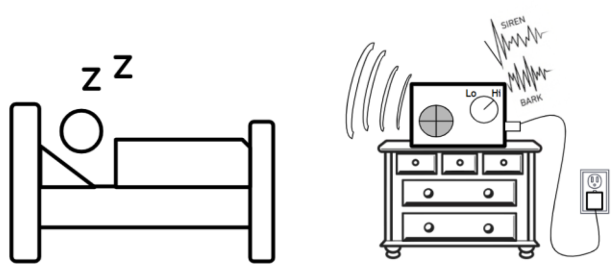

Figure 1: Representation of the Catching Z's system in a bedroom environment. The device (on the nightstand) detects incoming noise and emits a masking sound around the user, powered by a standard USB-C wall connection.

1.3 High-level Requirements List

- **Detection Latency:** The system must trigger the masking noise playback within 100 milliseconds of a sound event that exceeds a baseline by ≥ 10 dB to effectively mask the disturbance.
- **Output Sound Pressure Level:** The audio subsystem must be capable of producing a masking noise over an adjustable range of 40 dB to 75 dB measured at 0.5 meters (standard bedside distance).
- **Continuous Operation Stability:** The system must operate continuously for a minimum duration of 8 hours (a standard sleep cycle) without any software or hardware crashes.

2 Design

2.1 Block Diagram

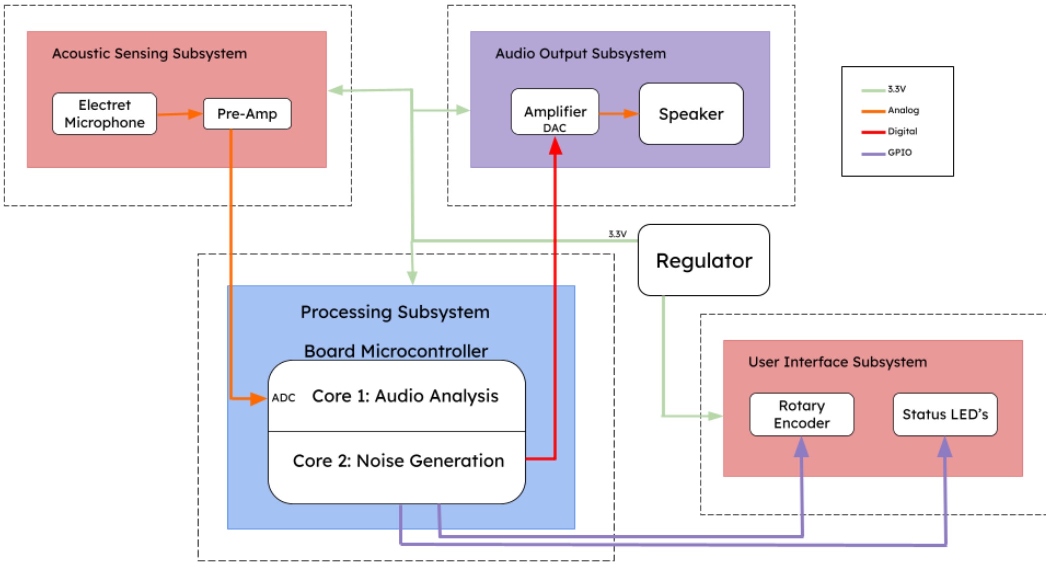

Figure 2: Block Diagram of the Catching Z's Subsystems interfacing with each other

The design of Catching Z's is divided into three primary subsystems: Sensing, Processing/Audio Output, and User Interface/Power. The sensing subsystem produces a digitized stream that the processing subsystem analyzes to maintain a moving baseline of ambient noise. When an event exceeds the baseline by **≥ 10 dB**, the processing subsystem works with the audio output chain to generate masking noise (white/pink/brown) that is faded in and then faded out. The user interface configures sensitivity and maximum masking volume, while the power subsystem ensures low noise sensing and reliable audio output for an entire sleep cycle.

2.1.1 Audio Sensing Subsystem

The acoustic sensing subsystem uses an electret microphone with a MAX4466 preamplifier. The microphone output is centered around mid-supply and fed into the ESP32-S3 ADC. The ADC samples at approximately 16 kHz. A moving average baseline is maintained. A disturbance is detected when the RMS value exceeds the baseline by 10 dB. This subsystem directly supports the 100-millisecond latency requirement by using short analysis windows and fast sampling. If the gain is too low, disturbances will not be detected. If it clips or is too noisy, false triggers will occur. A design risk is incorrect gain selection. If the microphone gain is too high, loud sounds will saturate the ADC. If too low, the 10 dB increase may not be measurable. This can be evaluated by measuring peak voltages at expected sound levels and ensuring they remain within the ADC range.

2.1.2 Processing Subsystem

The processing subsystem uses the ESP32-S3-WROOM-1 microcontroller. One core handles sampling and detection, while the second core generates masking noise. A simple state machine is used with listening, masking, and release states. When a disturbance is detected, the processor sends audio data to the amplifier through I2S. The total delay includes sampling time, RMS calculation, decision time, and audio startup time, and is kept below 100 ms by limiting frame size and minimizing buffering.

The processor must continuously compute RMS values, update the baseline, and respond quickly to changes. If processing falls behind or buffers overflow, the latency requirement will not be met. The subsystem must also apply a smooth fade-in and fade-out to avoid sudden sounds. The envelope can be implemented as a simple ramp where gain increases gradually over a short period. Without stable processing timing and correct task handling, the system would miss disturbances or respond too slowly.

2.1.3 Audio Output Subsystem

The audio output subsystem consists of a MAX98357A I2S Class-D amplifier and a 4 Ω, 3 W speaker. The microcontroller generates white, pink, or brown noise digitally and streams it to the amplifier. The expected sound pressure level can be estimated using $SPL = S + 10\log(P) - 20\log(r)$, where S is speaker sensitivity in dB at 1 W and 1 m, P is power in watts, and r is distance in meters. At 0.5 meters, the distance adds about 6 dB compared to 1 meter. With typical small speaker sensitivity, the required 75 dB SPL is achievable with well under 1 W of power. This subsystem must produce adjustable output from about 40 dB to 75 dB SPL at 0.5 meters. It must respond immediately to processor commands and avoid clicks or distortion. A design risk is speaker sensitivity variation. If the speaker is less efficient than expected, the maximum SPL may not reach 75 dB. This risk will be tested using a sound level meter at 0.5 meters and adjusting digital gain as needed.

2.1.4 User Interface & Power Subsystem

The user interface consists of a rotary encoder connected to GPIO pins. The encoder adjusts detection sensitivity and maximum masking volume. Status LEDs indicate power and masking activity. The power subsystem uses a 5 V USB-C input regulated to 3.3 V with an AMS1117-3.3 regulator. The system must supply at least 500 mA without voltage drops. Proper decoupling capacitors are required to prevent amplifier switching noise from affecting the microphone signal.

This subsystem supports the continuous operation requirement by maintaining stable voltage and preventing resets during audio output. If the regulator overheats or voltage drops occur, the microcontroller could reset and fail to meet the 8-hour stability requirement. Thermal and voltage measurements will verify that the regulator operates within safe limits during extended operation.

3. Design Verification

The verification process focuses on ensuring the system meets the criteria for success.

3.1 Latency Verification

To verify the requirement that masking noise triggers within 100 milliseconds of a detected event, we utilize an oscilloscope to monitor both the analog microphone input and the I2S output line. A controlled noise impulse is generated, and the time difference between the input spike and the activation of the audio output is measured.

3.2 Audio Output Verification

The audio subsystem is tested using a calibrated decibel meter placed at a standard bedside distance (approx. 0.5 meters). The system will output noise at minimum and maximum settings to verify the controllable range of 40 dB to 75 dB SPL.

4. Costs

4.1 Parts

Parts Costs

Part	Manufacturer	Retail Cost ($)	Bulk Purchase Cost ($)	Actual Cost ($)
ESP32-S3-WROOM-1	Espressif Systems	6.56	4.71	6.56
MAX4466 Mic Amp	Maxim Integrated	6.95	5.56	6.95
MAX98357A Amp	Maxim Integrated	5.95	4.76	5.95
4Ω 3W Speaker	Adafruit	1.95	1.56	1.95
PEC11R Encoder	Bourns	2.20	1.14	2.20
AMS1117-3.3	UMW	0.29	0.13	0.29
Total		**23.9**	**17.86**	**23.9**

5. Conclusion

5.1 Uncertainties

A primary risk is the potential for the system to overreact to brief, harmless sounds (false positives). While we hope to mitigate this, environmental variability remains a challenge. We may require more complex machine learning models to distinguish between disruptive noises (sirens) and non-disruptive spikes (coughing).

5.2 Ethical considerations

The design of Catching Z's follows the ethical principles outlined in the IEEE Code of Ethics and the ACM Code of Ethics. Because the system uses a microphone in a private bedroom environment, privacy is a central concern. All audio processing occurs locally on the ESP32-S3, and the device does not record, store, or transmit audio data. No wireless communication is enabled in the baseline design, reducing the risk of misuse or data breaches.

The system operates at low voltage, minimizing electrical hazards. Proper grounding and insulation will reduce shock and overheating risks. The maximum sound output is limited in firmware to 75 dB at 0.5 meters to prevent hearing damage, and masking noise fades in and out gradually to avoid startling the user.

Engineers are responsible for accurate performance reporting and avoiding exaggerated health claims. While the device may help mitigate sleep disruption from environmental noise, it is not a substitute for broader noise control solutions. Safety and privacy are at the core of this project, and we as engineers aim to meet professional engineering and societal standards.

References

[1] B. Berglund, T. Lindvall, and D. H. Schwela, "Guidelines for Community Noise," World Health Organization, Geneva, Switzerland, 1999. [Online]. Available: https://apps.who.int/iris/handle/10665/66217

[2] *ESP32-S3-WROOM-1 / ESP32-S3-WROOM-1U Datasheet*, v1.1, Espressif Systems, Shanghai, China, 2023. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32-s3-wroom-1_wroom-1u_datasheet_en.pdf

[3] *MAX4465–MAX4469: Low-Cost, Micropower, SC70/SOT23-8, Microphone Preamplifiers with Complete Shutdown*, Maxim Integrated, San Jose, CA, USA, 2001. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/MAX4465-MAX4469.pdf

[4] *MAX98357A/MAX98357B: Tiny, Low-Cost, PCM Class D Amplifier with Class AB Performance*, Maxim Integrated, San Jose, CA, USA, 2015. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/MAX98357A-MAX98357B.pdf

[5] *PEC11R Series - Contacting Encoder*, Bourns, Inc., Riverside, CA, USA, 2015. [Online]. Available: https://www.bourns.com/docs/Product-Datasheets/PEC11R.pdf

[6] *AMS1117: 1A Low Dropout Voltage Regulator*, Advanced Monolithic Systems, Livermore, CA, USA, 2016.

Appendix A

Requirement and Verification Table

System Requirements and Verifications

Requirement	Verification	Verification status (Y or N)
1. The system shall trigger masking noise playback within 100 ms of detecting a sound event exceeding the ambient baseline by ≥ 10 dB.	1. Connect oscilloscope to Mic Output and I2S Data Line. 2. Generate a 1kHz tone burst >10dB above baseline. 3. Measure time delta between input spike and I2S activity. 4. Verify delta < 100ms.	
2. The audio subsystem shall produce masking noise over a controllable range of 40 dB to 75 dB SPL at the bedside.	5. Place decibel meter 0.5m from speaker. 6. Set system to Min volume; verify reading approx 40 dB. 7. Set system to Max volume; verify reading approx 75 dB.	
3. The system shall operate continuously for overnight use (8 hours) without performance degradation.	8. Power system via USB-C. 9. Run continuously for 8 hours. 10. Check for overheating or audio artifacts every 2 hours.	