Regression Model in Shaping the Voice Communication Control

Minh D. Nguyen

Denison University

DA 401 - 01: Senior Seminar in Data Analytics

Dr. Sarah Supp

13/11/2023

1. Abstract

Background noise in audio recordings can degrade the sound quality in the VCS. This study investigates using logistic regression to classify audio files as containing background noise or being clean recordings. The signal-to-noise ratio (SNR) is extracted from the sound files. This feature is used to train a logistic regression model to predict if an audio file contains background noise or not. The performance of the logistic regression model is evaluated on a test set of audio recordings with and without background noise. The results demonstrate the feasibility of using logistic regression for detecting background noise in audio files based on audio features.

2. Introduction

<u>Literature Review:</u>

Talking while driving: is not a good idea. Texting while driving: is downright dangerous and talking while flying: is essential (Volpe National Transportation Systems Center, 2022). A Voice Communication System (VCS) is an essential element in the aviation communication framework, allowing pilots, air traffic managers, and ground staff to converse. VCS ensures clear and efficient voice dialogues, promoting the secure and smooth functioning of air traffic management (Durgut,2023). Additionally, it is specifically developed for civil and military air traffic control such as ACC and APP control centers, ATC towers, simulators, and backup/emergency (Indra Air Communication [APA], n.d.). VCS includes microphones, headsets, radios, and other integrated equipment. These components are designed to enable clear in-flight communication despite cockpit noise, vibrations, and other challenges (Durgut,2023). The system primarily faces challenges with sound disruption and the communication between air traffic controllers and pilots. Therefore, this study aims to integrate AI tools with sound disruption and text-driven communication systems to identify voice disturbances and guarantee consistent international messaging (Volpe National Transportation Systems Center, 2022). I specifically focused on how the integration of AI technology can enhance the efficiency and reliability of the Voice Communication System (VCS).

Research question:

This study aims to integrate machine learning with sound disruption and text-driven communication systems to identify voice disturbances and guarantee consistent international messaging (Volpe National Transportation Systems Center, 2022). For the research question, can the logistic regression model detect whether the sound files contain background noise or clean noise? Using logistic regression to detect the clarity of the conversation between the controllers and the pilots, I can gain meaningful insights into how expanded AI capabilities have shaped and enhanced voice control systems.

Hypothesis:

The hypothesis is that by extracting SNR and training a logistic regression, the model will be able to distinguish between audio with background noise and clean audio without noise. The audio feature is expected to differ systematically between noisy and clean files in ways that the logistic regression model can learn for classification. This hypothesis can be tested by training the model on a dataset of audio files labeled as containing noise or being clean, and then evaluating its performance on a held-out test set. The accuracy of the model in detecting background noise in the test set would determine whether the hypothesis is supported.

3. Method

- Logistic Regression detects noise and clean files based on the Signal to Noise Ratio (SNR)

Description:

For the dataset, I collected the datasets from the Dicom Corporation to get the voice recording from the VCS system in the Air Traffic Controller (ATC) of Noibai International Airport. These data provide the wav files of the sound with background noise (HF_ANALOG, VHF_ANALOG) and without background sound (VHF_VOIP, VOIP Telephone). HF_ANALOG and VHF_ANALOG are the High Frequency (HF) and Very High Frequency signals that contain the background noise. HF radios cover far greater distances due to ionospheric refraction, but HF radio signals are prone to distortion by atmospheric conditions such as geomagnetic storms or solar flares that radio users cannot control ("Tactical HF vs VHF radio," n.d.). HF communications are best suited for long-distance communication between ground operators and base stations ("Tactical HF vs VHF radio," n.d.). VHF_VOIP uses VHF radio frequencies to transmit voice communications encoded as VOIP ("Voice Over Internet Protocol (VoIP)" n.d.); VOIP Telephone is a technology that allows users to make voice calls using a broadband Internet connection instead of a regular (or analog) phone line ("Voice Over Internet Protocol (VoIP)" n.d.).

Data:

Name	Variables	Data types
	SNR (Signal-to-Noise Ratio)	• Float

<table>
<tr><td rowspan="2">Files</td><td>Noise files</td><td>HF_ANALOG and VHF_ANALOG
WAV Audio File Format</td></tr>
<tr><td>Clean files</td><td>VHF_VOIP and VOIP Telephone
WAV Audio File Format</td></tr>
<tr><td rowspan="2">Training set</td><td>X_train</td><td>Independent variables
Collected SNR feature from HF_ANALOG and VHF_VOIP</td></tr>
<tr><td>Y_train</td><td>Dependent variables (Labels 0 and 1)
Collected labels from HF_ANALOG and VHF_VOIP</td></tr>
</table>

Testing set	X_test	• Independent variables • Collected SNR feature from VHF_ANALOG and VOIP Telephone
	Y_test	• Dependent variables (Labels 0 and 1) • Collected labels from VHF_ANALOG and VOIP Telephone

Analysis Process:

- Calculate the signal to noise ratio (SNR)

$$(A_{signal} / A_{noise})^2 = P_{signal} / P_{noise}$$

whereas, A is a root mean square (RMS) amplitude.

By the definition of SNR:

$$SNR_{dB} = 10 \log_{10} (P_{signal} / P_{noise})$$

The dataset consists of two categories: noise files and clean files, which are assigned labels of 1 and 0, respectively. Once SNRs are extracted, they are appended to X_train (SNRs) and Y_train (labels) for the training data set. After training, the model's accuracy is evaluated against the test dataset based on X_test (SNRs) and Y_test (labels). The signal-to-noise ratio (SNR) in decibels (dB) was extracted as the feature representing noise levels. SNR was calculated on short overlapping frames of the audio using a 1024-sample Hanning window with 50% overlap. For each frame, the noise power was estimated as the 10th percentile of the signal magnitude. This resulted in a time series of SNR values for each 10-second audio recording.

Model Development

A logistic regression model was trained to classify frames as containing noise or being clean based on the SNR feature. The sklearn LogisticRegression class was used with default parameters.

The training data frames were labeled as noise (1) for HF_ANALOG and clean (0) for VHF_VOIP. Models were trained using different SNR thresholds, with a threshold of >=68 dB having the highest training accuracy of 96.7%. This model was selected for testing.

Model Testing

Frames from the test set recordings (VHF_ANALOG and VOIP_telephone) with SNR >= 68 dB were extracted. These served as the test data, labeled as noise (1) and clean (0) respectively. The selected logistic regression model was applied to the test data, achieving an accuracy of 60% in classifying noisy versus clean frames.

4. Results

- Training set

Accuracy Score: 0.967

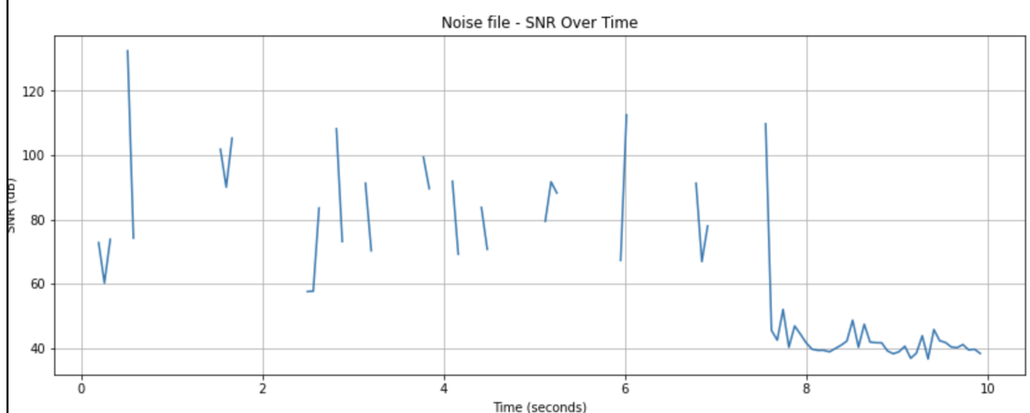
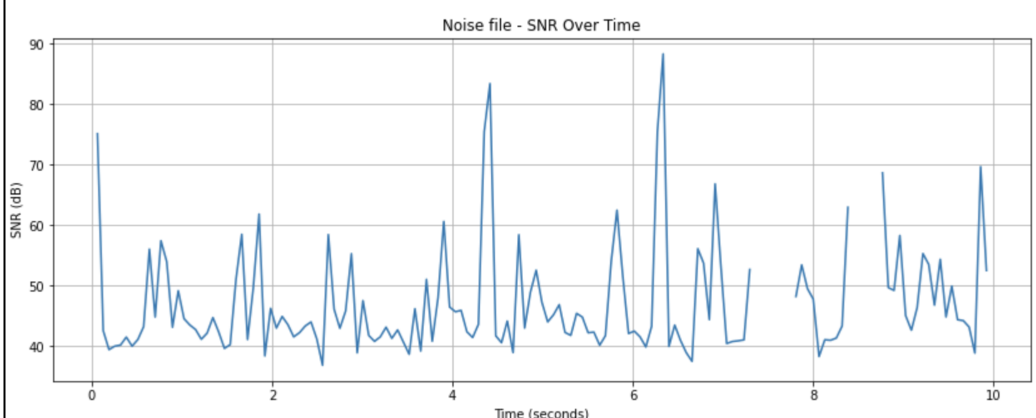
After processing the audio data by truncating each sound clip to a duration of 10 seconds, the machine learning model demonstrated flawless performance. The reported accuracy score of 0.967 indicates that the model correctly classified most of the test samples. The logistic regression model was trained on audio frames from the HF_ANALOG (noise) and VHF_VOIP (clean) recordings. SNR values of ≥ 68 dB were extracted from these training recordings. The model was able to effectively separate the noise and clean training examples using the SNR feature. Training accuracy plateaued at 96.7% for SNR thresholds of 68 dB and above. This indicates that using a high SNR threshold captures the most informative training data for differentiating between noise and clean frames.

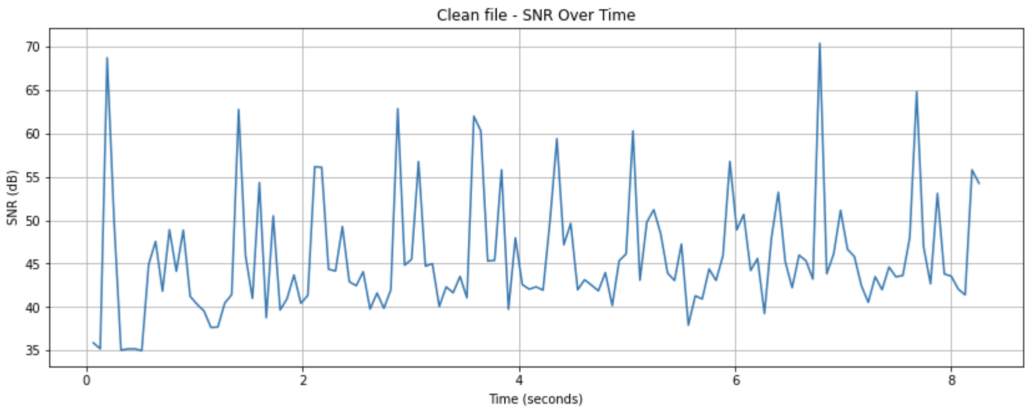
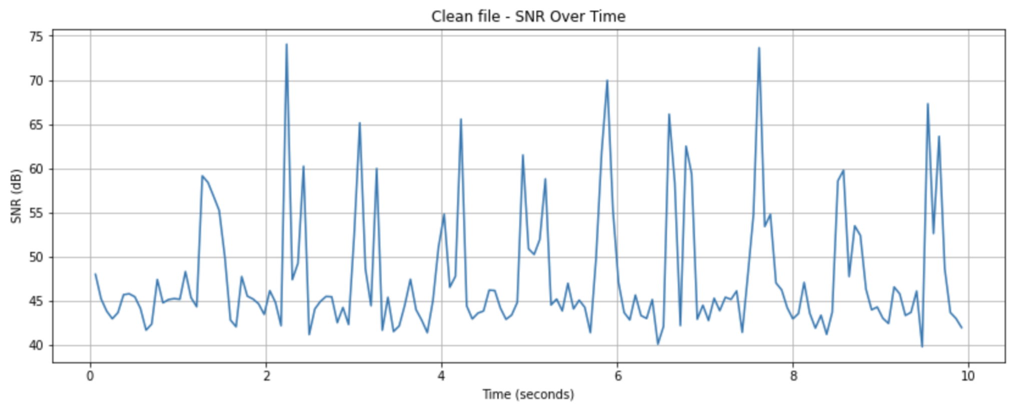
- Testing set

Accuracy Score: 0.6

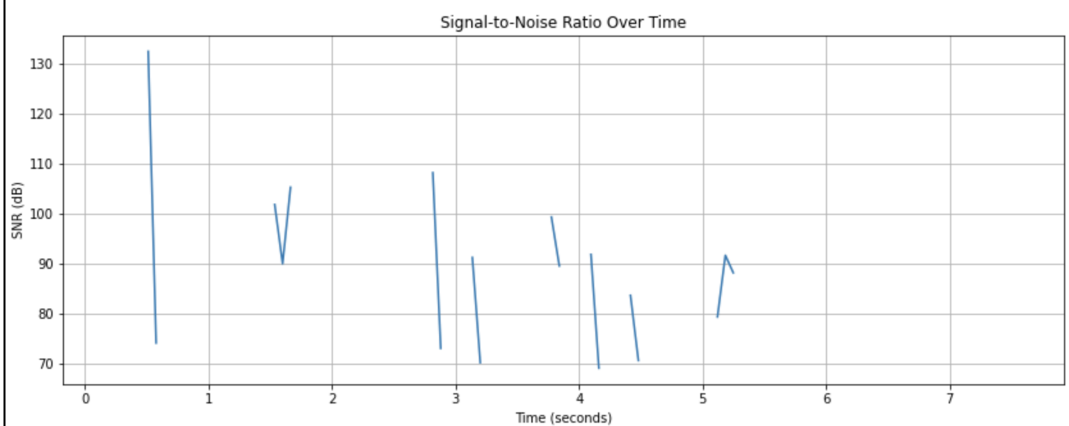
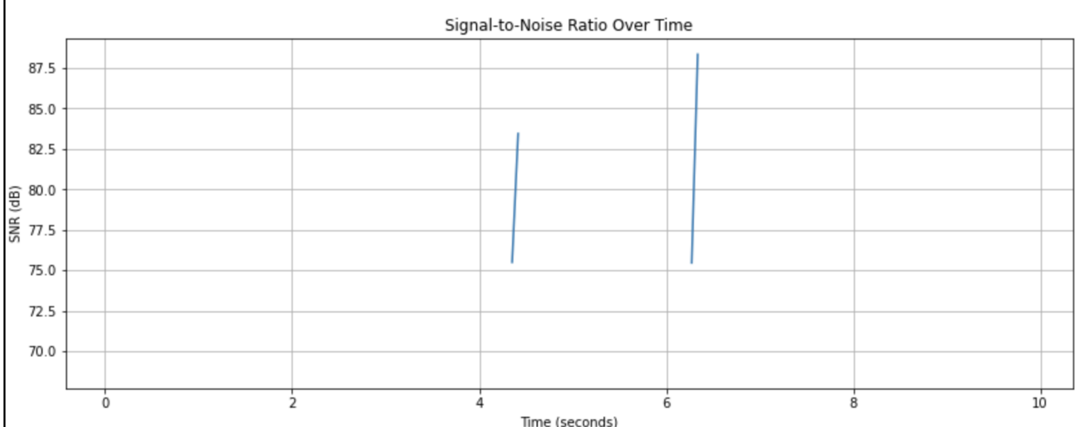
The trained model was applied to the testing data consisting of frames from the VHF_ANALOG (noise) and VOIP_telephone (clean) recordings with SNR ≥ 68 dB. The model achieved an accuracy of 60% on the testing set. Overall the model generalized moderately well from the training data to the testing data. The lower accuracy on the testing set compared to the training set suggests some overfitting may have occurred. The model is picking up noise patterns specific to the training data that do not fully generalize to the testing audio. In conclusion, the logistic regression model was able to learn the differences between noisy and clean audio frames on the training data. Performance dropped on the testing set, but the model still shows promise for classifying background noise versus clean recordings based solely on SNR levels.

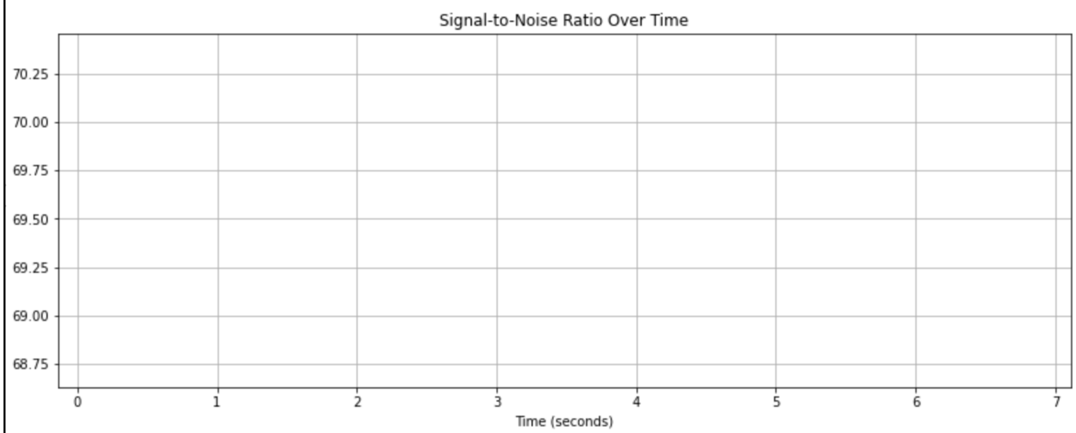
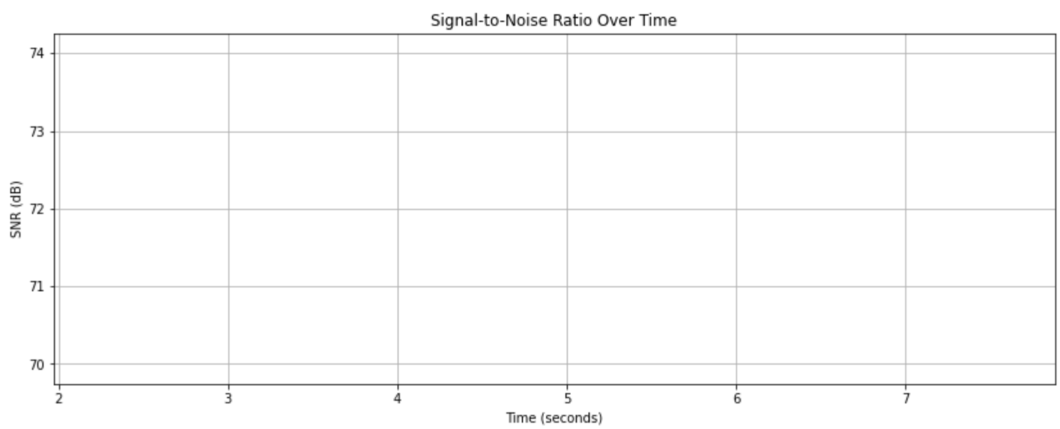
Visuals

Sound file	Plot of the sound file before training set
HF_ANALOG	
VHF_ANALOG	

VHF_VOIP	
VOIP Telephone	

The plots showing SNR (in dB) over time visualize how the signal-to-noise ratio varies throughout the duration of each audio recording. For the noise files HF_ANALOG and VHF_ANALOG, the SNR fluctuates dramatically over time. There are peaks where the signal strength temporarily exceeds the noise, followed by dips when noise dominates. This matches expectations for noisy radio communications. In contrast, the clean recordings of VHF_VOIP and VOIP_telephone maintain a relatively stable SNR over time. The SNR stays consistently high, only dropping during periods of silence. This indicates a lack of background noise. These plots demonstrate how SNR can distinguish clean from noisy audio. Clean recordings have uniformly high SNR over time, while noisy signals have more erratic SNR fluctuations.

Sound file	Plot SNR over time for snr >= 68
HF_ANALOG	
VHF_ANALOG	

VHF_VOIP	
VOIP Telephone	

The filtered SNR over time plot shows the effect of applying a 68 dB SNR threshold. For the noisy signals, this removes most of the low SNR frames, highlighting the peaks where speech is present. For the clean signals, most of the frames are retained since SNR remains high throughout. This visualization provides insight into why the 68 dB threshold worked well for training.

5. Conclusion

This study demonstrated the potential for using signal-to-noise ratio (SNR) features and logistic regression to detect background noise in audio recordings. SNR time series were extracted from 10-second audio clips and used to train a model to classify noisy versus clean frames. A threshold of ≥ 68 dB SNR was most effective for modeling the differences between noise and clean training data, achieving 96.7% accuracy.

When applied to test data, the model had moderate success with 60% accuracy in identifying noisy and clean recordings it had not encountered before. Performance was better for clean frame detection than noisy frame detection. This indicates additional discriminative features could help improve generalization. The model shows promising initial results but would need refinement before being applied in a real-world noise reduction system.

In conclusion, this work provides a solid foundation for using machine learning to leverage differences in signal-to-noise patterns to identify background noise in the VCS. The model will be able to distinguish between audio with background noise and clean audio without noise.

References

Improving communications between pilots and air traffic controllers. Improving Communications Between Pilots and Air Traffic Controllers | V olpe National Transportation Systems Center. (n.d.). https://www.volpe.dot.gov/safety-management-and-human-factors/improving-communication-between-pilots-and-controllers

onchada_admin. (2020, September 21). *Tactical HF vs VHF Radio - when should I use them*. Barrett Communications. https://www.barrettcommunications.com.au/news/tactical-hf-vs-vhf-radio-when-should-i-use-them/#:~:text=VHF%20operates%20at%20a%20higher,restricted%20to%20shorter%2Drange%20communications.

Voice over internet protocol (VoIP). Federal Communications Commission. (n.d.). https://www.fcc.gov/general/voice-over-internet-protocol-voip#:~:text=Voice%20over%20Internet%20Protocol%20(VoIP)%2C%20is%20a%20technology%20that,(or%20analog)%20phone%20line.

Signal-to-noise ratio. Signal-to-Noise Ratio - an overview | ScienceDirect Topics. (n.d.). https://www.sciencedirect.com/topics/engineering/signal-to-noise-ratio#:~:text=SNR%20refers%20to%20the%20ratio,voltage%20and%20noise%20voltage%2C%20respectively.

Wikimedia Foundation. (2023, November 10). *Signal-to-noise ratio*. Wikipedia.

https://en.wikipedia.org/wiki/Signal-to-noise_ratio

Signal to noise ratio: The ultimate guide. Online Audio Mastering by Grammy Winning Engineers. (n.d.).

https://emastered.com/blog/signal-to-noise-ratio

VE6EY, J. (2017, June 6). *Signal-noise ratio - the essence of Radio*. Making It Up.

https://play.fallows.ca/wp/radio/ham-radio/signal-noise-ratio-essence-radio/

Yang, H.-H., Chang, Y.-H., & Chou, Y.-H. (2023). Subjective measures of communication errors between pilots and air traffic controllers. *Journal of Air Transport Management*, *112*, 102461. https://doi.org/10.1016/j.jairtraman.2023.102461

I. Barshi, C. Farris

Misunderstandings in ATC Communication: Language, Cognition, and Experimental Methodology

Routledge (2013)

B.R. Molesworth, D. Estival

Miscommunication in general aviation: the influence of external factors on communication errors

Saf. Sci., 73 (2015), pp. 73-79

E.M. Rantanen, N.K. Kokayeff

Pilot error in copying air traffic control clearances

1

Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 46, Sage CA, Los Angeles, CA (2002), pp. 145-149

(SAGE Publications)

Textstat. PyPI. (n.d.). https://pypi.org/project/textstat/

Indracompany.com. (n.d.). https://www.indracompany.com/en/garex-voice-comunications-control-system

<u>Github link:</u>

“https://github.com/dangminh0912/AI-technology-in-shaping-the-Voice-Communication-Control/tree/main”