



Audio Engineering Society

Convention Express Paper 349

Presented at the AES 158th Convention
2025 May 22–24, Warsaw, Poland

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Supervised Machine Learning for Quality Assurance in Loudspeakers: Time Distortion Analysis

Gregg O'Donnell¹

¹Undergraduate Student, University of New Brunswick, 15 Dineen Dr, Fredericton, NB, Canada, E3B 5A3

Correspondence should be addressed to Gregg O'Donnell (gregg@soundbud.ca)

ABSTRACT

Measuring a speaker's ability to respond to an instantaneous pulse of energy will result in distortion at its output. Factors such as speaker geometry, material properties, equipment error, and the conditions of the environment will create artifacts within the captured data. This paper explores the extraction of time-domain features from these responses, and the training of a predictive model to allow for classification and rapid quality analysis of non-sonic parameters.

1 Introduction

When a speaker produces distortion, the recorded waveform can correlate to non-sonic design parameters of the system. Consider the step response of a speaker with high inertia. The geometry of the distorted waveform in the output could be modulated by the amount of windings in the coil, the mass of the driver, or the rigidity of the cone material [1].

This paper demonstrates that supervised machine learning can be applied in the analysis of speaker distortion to deduce a physical design parameter for the purposes of quality assurance. A classifier using Linear Discriminant Analysis [2] will demonstrate that non-sonic parameters can be correlated to features extracted from recorded audio. The non-audio parameter selected for



Fig. 1: Square wave half-cycle: 8 Ω added series resistance (Top), speaker only (Bottom)

classification is additional series resistance, as it can be quickly changed. As shown in Figure 1, the change in damping associated with increasing resistance [3] is also easy to visualize.

The paper then discusses the methodology behind the audio capture, windowing, and pre-processing [4]. Fea-

ture extraction, reduction, and selection are discussed [5]. The results will be presented and a confusion matrix elaborating on the error distribution will be provided and explained [2].

The paper concludes with an explanation of the use case of supervised machine learning in quality assurance.

2 Methods

The speaker selected was a 6.5" Tannoy 608 bookshelf speaker. To eliminate additional energy storage, the tweeter and crossover circuit were removed. Figure 2 shows the microphone and speaker prior to placement.



Fig. 2: Microphone and speaker in anechoic chamber

Each recording involved the speaker playing a 90 Hz square wave measured at 100 dBA. During playback, the microphone slowly but constantly moved radially across the speaker to avoid over-training on specific positions. A Shure SM-57 was selected due to the need for low handling noise. A Dynamount X2-R hosting a secondary mount provided additional shock protection [6]. The captures were recorded at a sampling rate of

96 kHz and a bit depth of 24 bits. The travel time across the speaker is approximately 2 full minutes from edge to edge. Captures were performed at five fixed axial distances ranging from 2" to 14", and under three conditions: 0, 1.5, and 3 Ω of added series resistance. The training dataset was recorded in anechoic conditions, and the testing dataset in non-anechoic conditions [4].

For pre-processing, the signals were separated into individual square wave cycles. To maintain proper alignment, the second derivative of each cycle was calculated and 1067-sample windows were created beginning at the first inflection point of each cycle [7]. Once the samples were separated into windows, each window was normalized to -13 dBFS. No equalization or other effects were used. Each distance setting is equally represented within the dataset.

Audio was converted to CSV format and processed using LibEMG [5]. While LibEMG was initially developed for myoelectric control systems, the feature extractor was effective at extracting more than 50 features. Minimum Redundancy Maximum Relevance (mRMR) [8] reduced the feature set to a ranked list of the best 30 features. Sequential feature selection [9] further reduced this, only including features that improved cross-validation accuracy by over 1% [2].

Table 1: Selected features.

<i>Feature</i>	<i>Name</i>
PAP	Peak Absolute Power [10]
LS	L-Score [5]
TM	Temporal Moment, 3rd Order [5]

A biplot of the selected features in principal component space [2] was created to ensure that the vectors were not redundant and pointed in the directions best describing the variances within the principal components. Figure 3 shows the first two principal components, Figure 4 shows the second and 3rd. It was observed that the PAP feature [10] explained the most variance out of all of the features.

A basic LDA classifier was applied using the selected features, as shown in Figure 5.

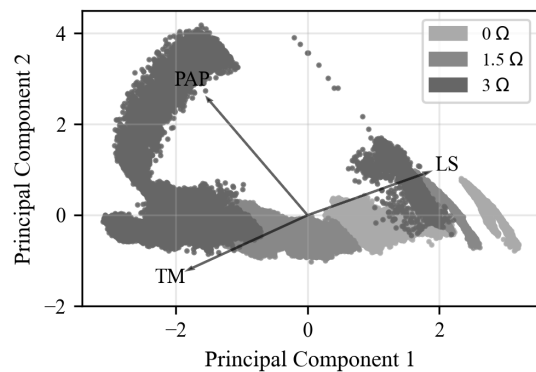


Fig. 3: Biplot over first two principal components

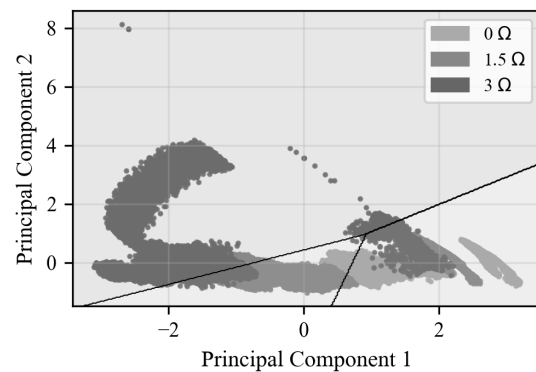


Fig. 5: LDA decision boundaries over first two principal components

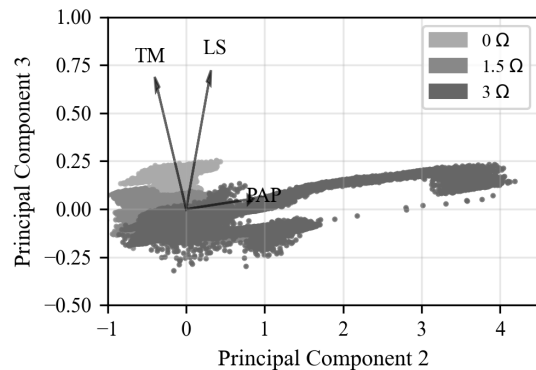


Fig. 4: Biplot over second and third principal components

3 Results

Figure 6 shows that prediction accuracy on the non-anechoic test dataset reaches 79% when using the three selected features. The confusion matrix is shown in Figure 7. Error is lowest when no additional resistance is present.

4 Discussion

It was observed that most of the error is associated with added resistance. It is suspected that the classifier is occasionally having difficulty with separability between the samples containing proximity effect and the samples containing added resistance. This effect can also be observed by the chaotic distribution of the 3Ω values in Figure 7. Nonetheless, 79% accuracy on a

classifier would be considered a successful model. It is expected that if all of the windows captured at 2" were removed from the dataset, this error would be significantly reduced.

5 Conclusion

This study shows that machine learning can provide a practical, data-driven alternative to traditional QA testing. Pushing a speaker into distortion and analyzing its time-domain response may allow for component-level diagnostics without relying on inference from standard performance metrics.

6 End Notes

The author would like to thank Dr. Chris Rouse, Murray Ferris, Bruce LeGrow, Dr. Erik Scheme, the libEMG team, UNB ECESES, The Engineering Endowment Fund, Dr. Khatami, and the faculty at the University of New Brunswick for their support throughout this project.

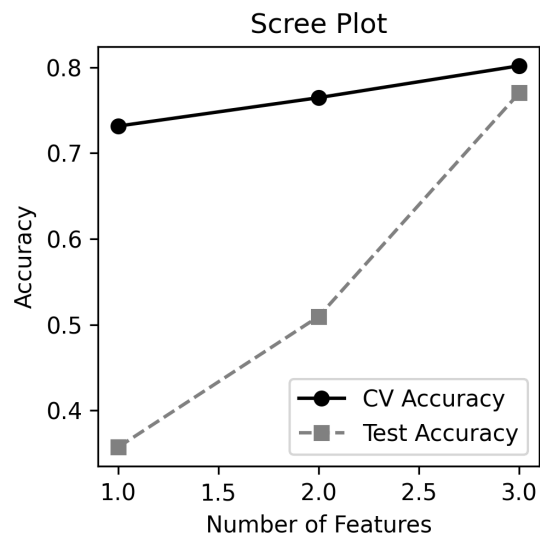


Fig. 6: Scree plot after feature reduction

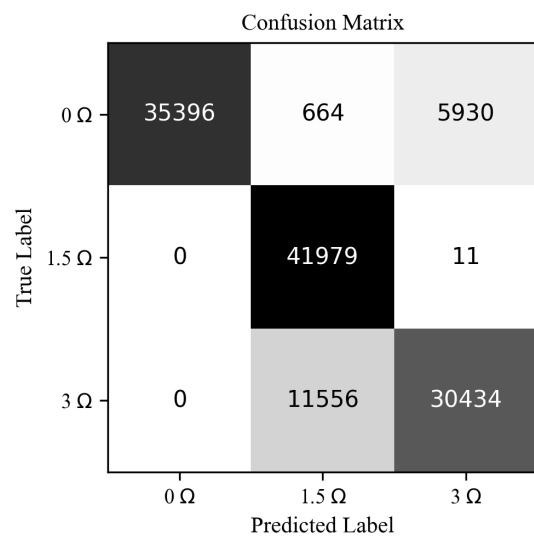


Fig. 7: Confusion matrix for LDA classification

References

- [1] Iain ragnar, “time distortion in loudspeakers,” *journal of the audio engineering society*, (1207), 1977.
- [2] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, 12, pp. 2825–2830, 2011.
- [3] OpenStax, *University Physics Volume 1*, OpenStax, Rice University, 2023, section 15.5: Damped Oscillations.
- [4] O'Donnell, G., “Near-Field Speaker Time Distortion - Anechoic and non-Anechoic,” 2025, doi: 10.21227/bgrk-r252.
- [5] Eddy, E., Campbell, E., Phinyomark, A., Bateman, S., and Scheme, E., “LibEMG: An Open Source Library to Facilitate the Exploration of Myoelectric Control,” *IEEE Access*, 11, pp. 87380–87397, 2023, doi:10.1109/ACCESS.2023.3304544.
- [6] Shure, “Microphone Techniques for Live Sound Reinforcement,” 2025, accessed: 2025, see p. 21.
- [7] Oppenheim, A. V. and Schaffer, R. W., *Discrete-Time Signal Processing*, Prentice Hall, 3rd edition, 2010, see discussion on second-derivative peak detection and signal alignment.
- [8] Ding, C. and Peng, H., “Minimum Redundancy Feature Selection From Microarray Gene Expression Data,” volume 3, pp. 523– 528, 2003, ISBN 0-7695-2000-6, doi:10.1109/CSB.2003.1227396.
- [9] Rückstieß, T., Osendorfer, C., and van der Smagt, P., “Sequential Feature Selection for Classification,” 2011, ISBN 978-3-642-25831-2, doi: 10.1007/978-3-642-25832-9_14.
- [10] Pancholi, S., Jain, P., Varghese, A., and Joshi, A., “A Novel Time-Domain based Feature for EMG-PR Prosthetic and Rehabilitation Application,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6046–6049, IEEE, 2019, doi: 10.1109/EMBC.2019.8857399.