Classification of Sound Levels Based on Evoked Responses
ELG 7172

Farnaz Jazaeri,
Niloofar Jazaeri

2023 October

Summary of the Proposal

In this project proposal, we outline our research endeavor titled "Classification of Sound Levels from Evoked Responses." Our objective is to develop a machine learning model capable of categorizing sound levels based on evoked responses, particularly the envelope following response (EFR). The dataset at our disposal contains EFR data recorded at various sound levels, ranging from 55 to 85. Our project will focus on the selection and extraction of pertinent features, including spectral content and time-domain waveforms, to enable precise classification. We will also explore the possibility of employing end-to-end learning, where the raw time-domain signal or spectrum is directly provided to our machine learning models. The collaboration of our two team members, with one specializing in feature engineering and the other in model development and evaluation, will be instrumental in achieving the project's objectives. Regular meetings will facilitate a dynamic exchange of ideas and findings. Additionally, we have confirmed access to the provided dataset and will execute the project in the Python programming environment, with the essential libraries readily available. Finally, we have referenced relevant literature to guide our methodology and ensure we remain aligned with the current state of the field.

Background

In the background section, write briefly using as many paragraphs, lists, tables, figures as you can about the main problem you select to study for your research. Remember that this will need to be a quantitative research study. In the background section, you will generally organise your writing along the following:

- Start by writing what do we know about the topic
- Then write about what do we NOT know about the topic
- Then write about what about the unknowns that you are going to address

The field of auditory neuroscience and the classification of sound levels based on evoked responses, particularly the envelope following response (EFR), have garnered increasing interest due to their wide-ranging applications, from hearing research to speech processing and even emotional bio-signal classification. In this two-page background and literature review section, we will explore the foundational concepts of EFR, signal processing, and machine learning as they pertain to the project's goals. Additionally, we will provide a brief overview of relevant research papers, starting with the foundational work of Purcell and Munhall (2006), which examined adaptive control of vowel formants in response to real-time formant perturbations, and Blankertz et al.'s (2011) tutorial on single-trial analysis and classification of ERP components. We will then delve into the study by Längkvist et al. (2014) concerning unsupervised feature learning and deep learning for time-series modeling, followed by Jafari and Dutoit's (2010) investigation into the relationship between fundamental frequency and mel frequency cepstral coefficients in speaker recognition. Finally, we will explore Zhang and Luo's (2015) work on the classification of emotional biosignals evoked by watching music videos, which is relevant to the broader context of evoked response classification.

Purcell, D. W., Munhall, K. G. (2006)

The study by Purcell and Munhall (2006) delves into the adaptive control of vowel formants in response to real-time formant perturbations, offering valuable insights into the adaptability of the human auditory system. The research focused on how individuals adjust their vocal tract configurations to accommodate formant changes, revealing the intricate nature of auditory processing. While this work primarily pertains to speech and vocal tract adaptations, it underscores the broader concept of the brain's dynamic response to acoustic stimuli. These findings provide a foundational understanding of how the human auditory system can adapt to variations in acoustic signals, which is highly relevant to the broader context of auditory neuroscience, including the EFR responses used in our classification task.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., Müller, K. R. (2011)

Blankertz et al. (2011) present a tutorial on single-trial analysis and classification of event-related potentials (ERPs), offering valuable guidance on how to extract and classify neural responses. ERPs are highly relevant to our project as they share similarities with EFR, both representing the brain's response to specific stimuli. This tutorial underscores the importance of feature extraction, model development, and classification techniques, which align with our project's objectives. The study provides a robust foundation for understanding the intricacies of single-trial analysis and the challenges associated with classifying neural responses, which will be invaluable for our classification of sound levels based on EFR.

Längkvist, M., Karlsson, L., Loutfi, A. (2014)

Längkvist et al. (2014) offer a comprehensive review of unsupervised feature learning and deep learning for time-series modeling. This review is significant to our project as it highlights the potential of deep learning techniques in modeling time-series data, which is pertinent to the EFR signals we are working with. The study introduces various methodologies for extracting informative features from time-series data and showcases the growing role of deep learning in pattern recognition tasks. Deep learning approaches, including convolutional neural networks and recurrent neural networks, are increasingly prominent in signal processing and classification tasks, providing the framework for end-to-end learning, which is one of our project's key explorations.

Jafari, M. G., Dutoit, T. (2010)

Jafari and Dutoit's (2010) work explores the relationship between fundamental frequency and mel frequency cepstral coefficients (MFCCs) in the context of speaker recognition. While their focus is on speaker recognition, the study highlights the significance of acoustic features in classification tasks. Fundamental frequency and MFCCs are both acoustic attributes that contribute to speaker recognition, and this research suggests their potential relevance to other auditory classification tasks, including sound level classification based on EFR. Understanding the link between acoustic features and classification is fundamental to our project's success.

Zhang, Z., Luo, Z. (2015)

Zhang and Luo (2015) focus on the classification of emotional biosignals evoked by watching music videos, illustrating the broader applications of classifying biosignals based on auditory stimuli. This study is relevant to our project's context, as it emphasizes the feasibility of classifying biosignals from evoked responses. In our case, we aim to classify sound levels based on EFR, but the methodologies and insights from this research are applicable. Understanding how biosignals can be effectively classified based on auditory inputs provides a broader perspective on the possibilities and challenges in our auditory neuroscience and machine learning task.

Recent Researches

In addition to these foundational papers, recent research in the field has been instrumental in advancing our understanding of evoked responses and signal classification. Recent studies have explored the complexities of evoked responses in various domains, such as auditory neuroscience, speech processing, and emotional signal classification. These studies have employed diverse techniques, from traditional machine learning to state-of-the-art deep learning models. They have also emphasized the significance of feature selection, model development, and feature engineering, which align with our project's objectives.

Recent researches have contributed to the development of more robust and accurate classification models, particularly for tasks involving neural and auditory responses. By building upon these recent advancements and considering the specific characteristics of our EFR dataset, we aim to contribute to this body of research and further enhance our understanding of evoked response classification. These recent studies underscore the dynamic nature of the field, where innovation and adaptation are key drivers of progress.

Conclusion

In conclusion, the background and literature review provide essential context for our project, bridging the gap between auditory neuroscience and machine learning. By drawing from foundational research, such as the work of Purcell and Munhall (2006) and Blankertz et al. (2011), as well as recent studies in the field, we aim to navigate the complexities of evoked response classification effectively. These studies collectively reinforce the significance of feature extraction, model development, and classification techniques and highlight the potential of deep learning methods in addressing the challenges of sound level classification based on evoked responses. The synthesis of past and current research positions us to contribute to this dynamic and interdisciplinary field, fostering innovation in auditory neuroscience, signal processing, and machine learning. [4, 1, 3, 2, 5]

Goal and Objectives

The primary goal of our project is to develop a robust machine learning system for classifying sound levels based on evoked responses, with a specific focus on the envelope following response (EFR). The central objective is to explore and assess the effectiveness of different feature extraction techniques, including spectral content and time-domain waveforms, and to evaluate the potential of end-to-end learning for this classification task. Collaborating as a team, our two members will concentrate on complementary aspects, with one dedicated to feature engineering and the other to the development and evaluation of machine learning models. The aim is to identify which features or approaches yield the most accurate and reliable results in categorizing sound levels. Additionally, the project is committed to maintaining ethical data handling practices and ensuring data privacy and confidentiality, in accordance with standard guidelines. By leveraging the confirmed dataset, Python programming environment, and relevant references, our research endeavors to contribute to the field of auditory neuroscience and signal processing by providing insights into the classification of sound levels based on evoked responses.

Available Datasets in Auditory Neuroscience and Signal Processing

[label=–]**PhysioNet Databases**

- - Website: https://physionet.org/
- **Common Voice Dataset**
  - Website: https://commonvoice.mozilla.org/en
- **UrbanSound Dataset**
  - Website: https://urbansounddataset.weebly.com/urbansound8k.html
- **FreeSound**
  - Website: https://freesound.org/
- **ESC-50 Dataset**
  - Website: https://github.com/karoldvl/ESC-50 </enditemize
  - **DESED Dataset**
    - Website: http://dcase.community/challenge2019/task-sound-event-detection-in-domestic-e
  - **Urban Sound Tagging Dataset (UST)**
    - Website: https://urbansoundtagging.github.io/
  - **DCASE - Detection and Classification of Acoustic Scenes and Events**
    - Website: http://dcase.community/challenge2021/
  - **Audioset**
    - Website: https://research.google.com/audioset/
  - **TESS - Toronto Emotional Speech Set**
    - Website: https://tspace.library.utoronto.ca/handle/1807/24487

Task Distribution for "Classification of Sound Levels Based on Evoked Responses" Project

Niloofar (Classification)

- **Model Selection**: Niloofar will be responsible for selecting and implementing classification models, such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), or other relevant algorithms. The selection will depend on the specific dataset and research goals.
- **Data Preprocessing**: Niloofar will manage data preprocessing tasks, including data cleaning, normalization, and handling missing values. This ensures that the data is prepared for feature extraction and model training.
- **Model Training**: She will train the selected classification models using the preprocessed data. This includes fine-tuning hyperparameters and evaluating the model's performance using appropriate metrics.
- **Hyperparameter Tuning**: Niloofar will explore hyperparameter tuning techniques to optimize the model's performance. This may involve techniques like grid search or random search.
- **Model Evaluation**: She will be responsible for thoroughly evaluating the models' performance, including metrics like accuracy, precision, recall, and F1-score. Cross-validation and hold-out validation methods can be employed.
- **Ensemble Methods**: Niloofar can explore ensemble methods like Random Forest or Gradient Boosting to enhance the classification accuracy.
- **Documentation**: Documenting the entire classification pipeline, including the models selected, parameters, and results, is crucial for the project's reproducibility.

Farnaz (Feature Extraction)

Feature Selection: Farnaz will focus on selecting relevant features from the evoked responses data. She will investigate different features that may be indicative of sound levels and choose the most informative ones.

Signal Processing: Implementing signal processing techniques to extract features from the evoked responses, such as spectral content and time-domain characteristics. This may involve techniques like Fast Fourier Transform (FFT), wavelet transforms, and time-domain analysis.

Dimensionality Reduction: Farnaz can explore dimensionality reduction techniques like Principal Component Analysis (PCA) to reduce the number of features while preserving relevant information.

Feature Engineering: Creating new features or transforming existing ones to enhance their discriminative power. This may include feature scaling and normalization.

Feature Visualization: Visualizing the extracted features to gain insights into their distribution and potential separability.

Data Integration: Collaborating with Niloofar to ensure that the extracted features are appropriately integrated into the classification pipeline.

Documentation: Documenting the feature extraction methods and choices, as well as the rationale behind selecting specific features.

References

[1] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Müller. Single-trial analysis and classification of erp components—a tutorial. *NeuroImage*, 56(2):814–825, 2011.

[2] Mohammad GH Jafari and Thierry Dutoit. Exploring the relationship between fundamental frequency and the mel frequency cepstral coefficients in the context of speaker recognition. In *Interspeech*, pages 1010–1013, 2010.

[3] Martin Längkvist, Lars Karlsson, and Amy Loutfi. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11–24, 2014.

[4] David W Purcell and Kevin G Munhall. Adaptive control of vowel formants in response to real-time formant perturbations. *The Journal of the Acoustical Society of America*, 120(6):966–977, 2006.

[5] Z Zhang and Z Luo. On the classification of emotional biosignals evoked by watching music videos. *IEEE Transactions on Affective Computing*, 7(2):145–158, 2015.