Research Design 2 – Task 1 – Section 1

The development of machine learning (ML) has greatly influenced many different areas, including the areas of sound recognition and music analysis. Musical instruments can be quite hard to identify, such as the clarinet, the instruments possess complex tonal qualities and has a wide variety of playing styles. Paired with acoustic resemblance to other woodwind instruments, the clarinet provides unique challenges when it comes to identification. Solutions to sound identification often requires feature extraction and sound analysis, which may not be robust enough for a variety of recording situations such as polyphonic audio or a performers style of playing. Deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have greatly improved musical sound identification, including the clarinet’s. Recent advancements have shown an improvement in accuracy and precision in audio classification tasks as opposed to traditional methods such as manual feature extraction [1]. Models for sound identification are frequently sensitive to the size of the dataset provided. The use of large-scale labelled audio datasets combined with sophisticated machine learning techniques has further improved the ability to identify musical instruments with more reliability and precision [2].

The purpose of this experimental study is to test the theory of machine learning-based sound classification, which predicts that the choice of neural network architecture relates to the accuracy and effectiveness of clarinet sound identification. This study aims to utilize clarinet and similar-pitched instrument audio datasets from publicly available music datasets or curated sources. The independent variables (CNN, RNN, and CNN-RNN models) will be defined as different deep learning architectures used for audio classification, where CNNs extract spatial patterns from spectrograms, RNNs capture temporal dependencies, and CNN-RNN hybrids combine both approaches. The dependent variables (classification accuracy, precision, recall, and F1-score) will be defined as performance metrics used to evaluate the effectiveness of each model in identifying clarinet sounds from a given dataset.

The aim of this research is to explore and provide a comparison between already existing machine learning techniques and algorithms and differentiate between the respective results obtained by providing reliable and consistent audio data which can bring insightful analysis of future audio classification projects for any musical instruments as this field is still relatively underrepresented. Specifically, the study hypothesis what machine learning algorithm is best at identifying sound from the clarinet by using sound recognition techniques on several waveform audio files from the clarinet and several other instruments with similar pitch and frequency.

Considering the previously mentioned factors, as shown in Fig 1, during this research, the positivism philosophy will be followed due to an objective data analysis and a deductive research approach will be taken in regards to testing various machine learning models and comparing their effectiveness. Additionally, an experimental strategy will be utilised, in company with a mono-method choice and a cross-sectional time horizon.

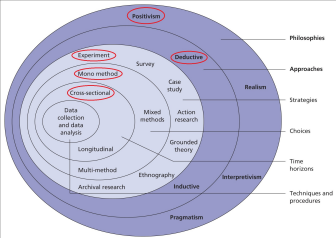


Fig 1. The Research Onion

References

[1] J. Salamon and J. P. Bello, "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279-283, Mar. 2017.

[2] A. Mesaros, T. Heittola, and T. Virtanen, "Metrics for Polyphonic Sound Event Detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 11, pp. 2066-2079, Nov. 2017.