M.Eng Project Proposal: Machine learning applied to timbral acoustic analysis

*Topic: applying signal processing and machine learning methods to analysing the acoustical timbre of voices or musical instruments in audio recordings.*

Acoustical timbre can be described as the qualities characteristic to a sound, allowing it to be recognised and “distinguished from other sounds at the same pitch and loudness” [1]. Timbral analysis consists of automatically extracting this information from a digital audio signal, such that the sound’s unique qualities can be expressed.

**Aims and objectives of the project**

The aim of the project is to develop a method to automatically recognise the timbral qualities intrinsic to a signal, for the purposes of categorising a sound in absolute terms (labelling) or in relative terms to cluster signals and differentiate sounds from one another.

Timbral analysis has applications in speech processing, as well as in the classification and synthesis of musical sounds. Precisely describing the timbre specific to sounds such as an individual human voice or instrument is a challenging problem not usually tackled by a single traditional method such as spectral or probabilistic signal analysis. The combination of these methods with machine learning constructs such as neural networks could allow for a system to infer the embeddings encapsulating timbral information. Developing such a system would prove directly helpful in real-world downstream tasks such as Automatic Speech Recognition (ASR) and the generation of audio metadata.

**Outline of the proposed method**

I expect this project to have minimal requirements: a laptop (python, MATLAB), using publicly available data, etc. - no special apparatus or funding is required. Any data or existing work used or repurposed will be sourced with permission from its creator or using open-source licensing.

Steps and intermediate milestones could include:

* *Literature Review*  
  - Developing my understanding of the theory and practical considerations involved in the project by researching existing approaches and methods for timbral analysis in the literature.   
  - Categorising (by application, or by methods applied) the existing work on the topic.  
  - Testing any publicly available implementations to evaluate their relevance to the project.  
  - Researching public datasets that could be used for reference, training and evaluation.
* *Specialisation*- Selecting a specific focus and methods to apply. This will be informed by what I deem most interesting and feasible in the scope of the project.   
  - Narrowing the types of sounds considered for timbral analysis, for instance to only include speech or a single musical instrument. This will be contingent on available datasets.
* *Development*, including several iterations of increasing complexity, beginning with a simplistic baseline system to benchmark further progress against.
* *Evaluation* of methods developed, potentially includes evaluating the gains brought to downstream tasks such as speaker diarization for ASR.

**Academic and industrial background**

Dr Patrick Naylor, of the Speech & Audio Processing Lab, confirms he will supervise this project. I have also completed relevant work in my Industrial Placement project with Nuance Communications Inc. on audio-visual speaker diarization (labelling “who is speaking when” in dialogue), which was supervised by Dr Naylor. The experience I gained in machine learning applied to signal processing will inform how I approach this project, especially since timbral speech analysis can be applied to diarization.

**References**

|  |  |
| --- | --- |
| [1] | N. M. McLachlan, "Timbre, Pitch, and Music," *Oxford Handbooks online,* 2016. From the American National Standards Institute (ANSI) definition of timbre. |