William Victor Simmons

St. Catharine's College

wvs22

Part II Individual Project Proposal

Techniques for Separation of Harmonic Sound Sources

11 October 2016

Project Originator: William Victor Simmons

Project Supervisor: Dr D.J. Greaves

Director of Studies: Dr S.N. Taraskin

Overseers: Dr T.G. Griffin and Dr P. Lio

Introduction and Description of the Work

The ability to isolate and focus on an individual sound source for identification and localisation is an ability humans possess which has proved non-trivial to simulate computationally. In particular, the "cocktail party" problem [1] is the instance of this sound separation problem for a collection of human voices in conversation. Good solutions to such problems are desirable in the fields of music and audio analysis and editing, amongst others.

The key focus of this project is to investigate and compare a selection of techniques for sound separation. The specific case of this problem being considered is where the system is provided with knowledge of the number of sound sources present and that they are purely harmonic sounds but the input contains sources with varying waveforms, time offset, base frequency and stereo positions.

Resources Required

No special resources will be required. I intend on using my own PC (Windows 10 Home, Intel Core i7-3537U, 8.0GB RAM). I accept full responsibility for this machine and I have made contingency plans to protect myself against hardware and/or software failure. In particular, all documents, sound samples and source code will be stored on a Microsoft OneDrive cloud space and version control handled through git with a repository on GitHub and further backups made to a dedicated flash drive after each work slot (see timetable). In the event of hardware damage to my PC, the MCS PCs shall suffice.

For the sound samples in use, I shall be obtaining a variety of synthetic sounds from tools within the free Audacity software and more natural sounds from the website of the Philharmonia Orchestra (provided under a Creative Commons licence), composing these using Audacity.

Starting Point

Prior to starting this project, I have read the contents of the Part II Digital Signal Processing course in addition to reading the referenced papers on Sinusoidal Modelling. The intention of the initial research period is to familiarise myself with Non-negative Matrix Factorisation methods and an another alternative method for the extension.

Substance and Structure of the Project

The main technique at the centre of this project is using Sinusoidal Modelling [2] to separate sinusoidal trajectories (sinusoids with slow amplitude and frequency modulation over time) into collections representing the original sources based on some relative "distance" between the trajectories.

Non-negative Matrix Factorisation [3] has been shown to be useful for sound separation methods [4] and so a solution using this will also be designed and implemented. This will be compared to the first solution under statistical and perceptive tests of quality.

Since the Sinusoidal Modelling method does not preserve phase information, the similarity between the separated output and the original sounds will be determined based on the average cross-correlation between their power spectra. We can also consider the cross-correlation of their spectrograms to capture the similarity of the sounds as they evolve through time. These will be averaged over the training set created from synthetic mixes of single note recordings of musical instruments and considering how this performance is affected by the mixing parameters.

As, in many circumstances, the output of a solution to sound separation would be used for human consumption, it would be beneficial to know how perceptually different the output and the original sounds are. For this purpose, some human tests will be carried out to judge the quality of the reconstruction.

A third implementation of another technique may be completed as an extension if given sufficient time.

The solutions will be developed as standalone C++ applications for easy extensibility if they are to be used in further study. However, prototype solutions will be created using Matlab to ensure the principles of the solutions are suitable and aid design of the final versions. Whilst the intention of the C++ solutions is to be more extensible, quality of

the separation should not be lost in doing so. The cross-correlation tests will be carried out on both the prototypes and the solutions to gauge their respective accuracies.

In most multimedia applications, it is often viewed as appropriate to discard details as long as the result is not perceptually invasive, hence the prevalence of lossy compression for visual and audio data. Human tests will be conducted to consider how perceptually different the original and reconstructed sounds are by comparing how identifiable they are. Hearing ability typically peaks before the age of 25 [5], so the testing population will be of this age group as they are more likely to be able to detect any distortion in the reconstructions.

Reference

[1] *Some Experiments on the Recognition of Speech, with One and with Two Ears*, C. Cherry, Imperial College, University of London
(`http://www.ee.columbia.edu/ dpwe/papers/Cherry53-cpe.pdf`)

[2] *Separation of Harmonic Sound Sources Using Sinusoidal Modeling*, T. Virtanen and A. Klapuri, Tampere University of Technology
(`http://www.cs.tut.fi/sgn/arg/music/tuomasv/sssep.pdf`)

[3] *Algorithms for Non-negative Matrix Factorisation*, D.D. Lee and H.S. Seung
(`https://papers.nips.cc/paper/1861-algorithms-for-non-negative-matrix-factorization.pdf`)

[4] *Sound Source Separation Using Sparse Coding with Temporal Continuity Objective*, T. Virtanen, Tampere University of Technology
(`http://www.cs.tut.fi/sgn/arg/music/tuomasv/icmc2003.pdf`)

[5] *Presbycusis Values in Relation to Noise Induced Hearing Loss*, A. Spoor, University of Leiden

Success Criteria

The project's success will be judged by having completed the following:

- A solution to the discussed sound separation problem using the Sinusoidal Modelling technique should be designed and implemented.

- A similar solution to the discussed sound separation problem using NMF should be designed and implemented.

- A collection of test sound files should be designed and assembled to cover a range of values for the test parameters (additive noise levels and relative pitches, time offsets and stereo positions between the sources).

- Measurements of performance of the solutions should be obtained using the test set.

- Human tests should be performed to measure for the audible similarity between the original sounds and their reconstructions after separation and perceptive quality of the separation.

- The dissertation describing this project must be written.

Timetable and Milestones

The work units for this project will be split into segments of typically a fortnight in length. The planned starting date is Thursday 20th October 2016.

Segment 1: 20th October - 2nd November

Read around the topics of Sinusoidal Modelling, Non-Negative Matrix Factorisation and other techniques for sound separation. Prepare sound sample set for tests.

Deliverable: A selection of individual harmonic sounds, at a range of pitch and time offsets, and sound files containing synthetic mixes of pairs of these.

Segment 2: 3rd November - 16th November

Prototype the Sinusoidal Modelling solution in Matlab.

Deliverable: A Matlab source code file which performs sound separation using Sinusoidal Modelling and the output files from successful separation of a subset of the test files.

Segment 3: 17th November - 30th November

Prototype the NMF solution in Matlab.

Deliverable: A Matlab source code file which performs sound separation using NMF and the output files from successful separation of a subset of the test files.

Segment 4: 1st December - 28th December

Rebuild the Sinusoidal Modelling solution in C++.

Deliverable: The C++ source code files which perform sound separation using Sinusoidal Modelling and the output files from successful separation of a subset of the test files.

Segment 5: 29th December - 18th January

Rebuild the NMF solution in C++.

Deliverable: The C++ source code files which perform sound separation using NMF and the output files from successful separation of a subset of the test files.

Segment 6: 19th January - 1st February

Slack time for covering any setbacks when building the solutions. Investigate and optimise the distance weightings in the Sinusoidal Modelling solution based on the majority of the test set. Write the progress report.

Deliverable: A set of graphs describing the effects of changing the weightings on the quality of separation and reconstruction. A completed progress report ready for submission.

Segment 7: 2nd February - 15th February

Use the remainder of the test data set to investigate how both solutions handle different pitch intervals between the sounds, additive noise levels and quantity of sounds.

Deliverable: A collection of statistics describing the performance of both solutions under a range of values for each of these properties of the test data.

Segment 8: 16th February - 1st March

Slack time for producing evaluation statistics. If this is not needed, then this period will be for the extension task of implementing another alternative algorithm.

Segment 9: 2nd March - 15th March

Begin the first draft of the dissertation, starting with the Introduction, Preparation and Conclusion sections.

Deliverable: The first draft of the Introduction, Preparation and Conclusion sections of the dissertation.

Segment 10: 16th March - 5th April

Complete the first draft of the dissertation.

Deliverable: The first draft of the Implementation and Evaluation sections of the dissertation.

Segment 11: 6th April - 26th April

Acquire feedback from first draft and apply this into second draft of the dissertation.

Segment 12: 27th April - 10th May

Make final adjustments and polish the dissertation.

Segment 13: 11th May - 19th May

Slack time for dissertation writing in case of delays due to Tripos preparation.

Deliverable: The completed dissertation.