From Raw Polyphonic Audio to Locating Recurring Themes

We survey several approaches to the task of transcribing polyphonic music on the diatonic scale and introduce some new ones. We do not address the issue of instrument identification at all; instead we limit our analysis to a single keyboard string instrument with discrete pitch such as piano or harpsichord. The second part of our work is concerned with comparison of musical sequences of polyphonic music using a higher-level encoding.

Experiments on polyphonic music transcription

We consider transcription algorithms which convert raw audio into a list of fundamental frequencies over, and possibly varying in, time. We believe that for the purposes of retrieval this is sufficient to capture some essential (if crude) details of a performance, while avoiding the more involved interpretation problems usually associated with transcription.

Conceptually we divide the task into two subtasks: *time-frequency spectral analysis* and *fundamental line extraction*. Of the many algorithms that have been used or proposed for time-frequency analysis of musical and speech signals, we implemented and tested *least mean squared filtering*; a decimated version of the *constant-Q spectrogram* due to Brown (1991); and a decimated version of the *Phase Vocoder*. However, none of them gave satisfactory results for polyphonic signals. Thus we eventually decided to abandon Fourier methods altogether in favour of *auto-regressive* (AR) estimators. We believe that we are the first to have applied AR methods to musical signals. Of the four AR methods we tested, Marple's *MODCOVAR* algorithm (Marple 1987) turned out the most accurate (von Schroeter 2000). We therefore used Marple's algorithm as the basis for all further experiments.

Based on what little prior expertise was available on fundamental line extraction, and guided by experimental results, we developed a suite of transcription algorithms, starting from a generalisation of the frame-based *correlation approach*, but finding ourselves naturally led to applying concepts with increasingly topological content (Fig 1).

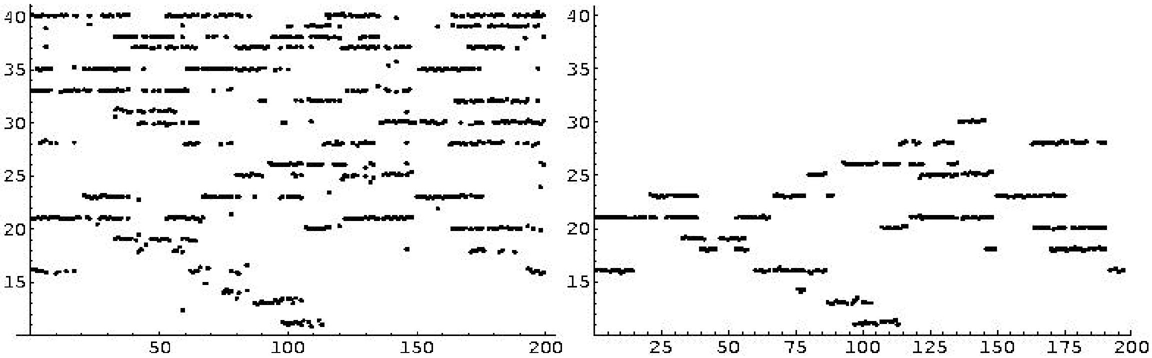

Figure 1: Time-pitch spectra of bars 2-3 of Bach's Fugue I from Part I of the WTC. (a) Left: AR spectrum using Marple's algorithm. (b) Right: fundamentals of connected components.

Best results were achieved with an algorithm which models tones as *two-dimensional subsets of the time-pitch spectrum* whose points are connected by two kinds of relations, namely connectivity in time as continuation of a spectral line across neighbouring frames, and connectivity in pitch as a simultaneous pattern relation between points in the same frame. There is a two-level hierarchy of such combined time-pitch relations: (P) two points can belong to a single tone pattern; (T) they can belong to a single tone pattern with specified fundamental. (T) implies (P). Thus (P) can be used to break down the input spectrum into connected components such that each tone pattern belongs to exactly one of these components; no prior knowledge of the fundamentals is necessary for this step. Within the (P) components we then list the maximal (T) components based at each of the lines, where delays of the fundamentals within the (T) components are tolerated up to a tunable threshold. This list is what we call a *covering table*; the partial ordering of its entries by inclusion reflects a partial ordering of possible tone hypotheses by their "explanatory power," although not necessarily in any probabilistic sense. The result of the corresponding analysis is a point spectrum in which each point indicates an instantaneous fundamental pitch of a note hypothesis; thus the output format is richer in pitch details than the ordinary MIDI format and would in principle also accommodate pitch-variable instruments.

Locating recurrent themes

We studied the algorithm of Mongeau and Sankoff (1990) for computing the similarity between melodies (Doraisamy 1995). Our experiments with Mongeau and Sankoff's algorithm used the first few notes of Bach's Fugue I of The Well-Tempered Clavier, Book I, with a number of variations such as key change, skipping notes, augmentation and diminution. We found this to be a good overall dissimilarity measure except for the following variations: 1) changing the rhythm of a melody line and 2) transposing a melody line into a different key. Both, unfortunately, are quite common variations employed by composers. We therefore suggest using an invariant coding using the *melodic contour* (with pitch offsets instead of absolute values) and the *rhythmic contour* (using time ratios as opposed to absolute times).

Although the output of our polyphonic analysis is somewhat richer in format than ordinary MIDI, it does not contain the details and the quality of a high-level encoding such as Humdrum. We are currently investigating how this gap can be closed with an automatic procedure. Once this challenges have been overcome, a traditional approach based on monophonic melody comparisons as outlined in this section could be used to locate recurring themes or, more generally, to compare musical sequences.

Acknowledgements: This work is partially supported by the EPSRC, UK.

Author Information

Thomas von Schroeter[1] , Shyamala Doraisamy[2] and Stefan M Rüger[3]

[1]T H Huxley School, [3]Department of Computing
Imperial College of Science Technology and Medicine
London SW7 2BZ, England
ts9@ic.ac.uk, s.rueger@ic.ac.uk
http://km.doc.ic.ac.uk/

[2]Department of Multimedia
Faculty of Computer Science and IT
University Putra Malaysia
43400 UPM Serdang, Selangor D.E., Malaysia
mailto:shyamala@fsktm.upm.edu.my

Suggested Readings

Brown, J C (1991). Calculation of a constant Q spectral transform. *J Acoust Soc Am* 89, 425-434

Doraisamy, S (1995). *Locating recurrent themes in musical sequences.* MSc Thesis, U Malaysia Sarawak

Marple, S L (1987). *Digital spectral analysis with applications*. Prentice-Hall

Mongeau, M and D Sankoff (1990). Comparison of musical sequences. *Computers and the Humanities* 24

von Schroeter, T (2000). Auto-regressive spectral line analysis of piano tones. TR DoC-7 Imperial College