Music Information Acquisition, Preservation, and Retrieval (MUMT-611)

Voice Separation Summary

Denis Lebel – 110125645 – denis.lebel@mail.mcgill.ca

Introduction

Voice separation consists of adequately separating notes of a musical piece into different voices. Applications of this technique include theme finding, music analysis, and transcription of low-level musical data into score notation. One of the main challenges of voice separation is to differentiate between chords and voices when overlapping notes have the same duration and onset time. In the literature, various approaches have been proposed to correctly separate voices given a stream of notes in a symbolic representation. In this document, the split-point, rule-based, local optimization, contig mapping, and predicate approaches are briefly described.

Split Point Approach

A trivial solution to voice separation consists of splitting the range of all possible pitches into a set of disjoint intervals corresponding to voices. Each note of the input stream is assigned to the corresponding pitch interval. Although very simple, this technique does not always produce correct results since it assumes that voices are not overlapping in the pitch range. According to (Kilian 2002), the split point separation method is mostly used in commercial sequencer software packages.

Rule-based Approach

Another solution is to take advantage of the voice-leading rules used by the composer. Examples of such rules include limiting the number of voices used, preferring small pitch intervals between successive notes, and avoiding overlapping voices pitch range. One problem with this technique lies in the large number of possible rules, which can increase the complexity of the system quite a bit. Moreover, since these rules might be specific to a composer, for example, a given piece needs to be assigned a subset of all possible rules that apply to it in order for the system to produce good results. Rules may additionally apply only to a certain section of a piece (e.g., number of voices varying over time), in which case the separation may be erroneous. Therefore, the input stream may also need to be fragmented to reflect these rules changes. A rule-based approach was noticeably used with dynamic programming in the Melisma Music Analyzer system as described in (Temperley 2001).

Local Optimization Approach

The local optimization is an approach proposed in (Kilian 2002) that uses a heuristic algorithm (i.e., an iterative process that finds the best solution from a given set at each iteration). The idea consists of fragmenting the input stream into small slices containing overlapping notes and assigning these notes to voices using a randomized local search algorithm that minimizes a parametric cost function. This approach does not find the correct voice separation solution for a given piece, but rather provides a reasonable solution in different contexts (mainly for transcription). The solution can also be tuned in realtime through a set of user-controllable parameters. Although presenting better results than the two previous approaches, the local optimization method still encounters problems with musical pieces with overlapping pitch ranges. On the other hand, chords are correctly identified, which isn't the case in all other approaches. The algorithm presented in (Kilian 2002) is implemented in the midi2gmn (Midi to Guido Music Notation) program.

Contig Mapping Approach

In contrast with the method proposed by Kilian, the contig mapping approach aims at providing the correct analysis rather than an appropriate result for transcription. Perceptual principles are exploited to reduce the computational complexity of the algorithm, which makes it more appealing than the previous solutions. As described in (Chew 2004), the algorithm consists of segmenting the piece into collections of overlapping fragments (contigs) and then reconnecting fragments of adjacent contigs using a shortest distance method. This technique is implemented in a Java-based voice separation analyzer system named VoSA. Results seem to be promising. In addition to the new algorithm, three metrics have been introduced in (Chew 2004) to measure voice separation performance, namely: the average fragment consistency, the correct fragment connection, and the average voice consistency.

Predicate Approach

Similar to the contig mapping approach, the same-voice predicate technique focuses on finding the correct voice separation solution for music analysis or theme finding purposes. This approach is implemented in the VoiSe system and uses a same-voice predicate (implemented as a learned decision tree) to determine whether or not two notes are in the same voice, as well as a separate algorithm to perform the actual voice separation. The predicate examines various features related to distance and rhythm. The voice-numbering algorithm then uses the predicate values to assign each note to a voice. A nice feature of this algorithm is that it is not restricted to a pre-determined number of voices. Furthermore, as explained in (Kirlin 2005) the algorithm works in both explicit and implicit polyphony. The experiments results show that their method does produce decent results. However, as any other learned algorithm, it rarely produces error-free results.

Conclusion

Finally, although different techniques have been presented in the past few years and progress has definitely been made, there still doesn't exist the perfect solution that produces correct results no matter which musical piece is being processed. The question one could ask then is whether or not a fully automated voice separation algorithm is really that useful.

References

Kirlin, P., and P. Utgoff. 2005. VoiSe: Learning to segregate voices in explicit and implicit polyphony. *Proceedings of the International Conference on Music Information Retrieval*. 552–7.

Chew, E., and X. Wu. 2004. Separating voices in polyphonic music: A contig mapping approach. *Proceedings of the International Symposium on Computer Music Modeling and Retrieval*. 1–20.

Kilian, J., and H. Hoos. 2002. Voice separation: A local optimisation approach. *Proceedings of the International Conference on Music Information Retrieval*. 39–46.

Temperley, D. 2001. *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

Cambouropoulos, E. 2000. From MIDI to traditional musical notation. *Proceedings of the AAAI Workshop on Artificial Intelligence and Music*.