Bruno ANGELES
Slide presentation IV
12 November 2009

Polyphonic Transcription

Introduction
The polyphonic transcription of music is the conversion of audio signals with musical content into a symbolic format. It has applications in automatic DJing, genre classification (possibly for library indexing), and database querying, among other fields.

Monophonic vs. Polyphonic
Monophonic transcription consists of identifying the notes of a single instrument. Polyphonic transcription can be interpreted in different ways: as a single instrument playing concurrent notes, as an instrument playing (individual or concurrent) notes within a polyphonic context, or as several instruments playing concurrent notes.

While monophonic transcription has already been implemented with respectable success, the same cannot be said about polyphonic transcription. It remains an open problem, and has seen an increased interest from Music Information Retrieval experts in recent years.

The Human Method
Hainsworth and Macleod (2004) describe an informal investigation of the steps that expert musicians follow in order to transcribe polyphonic music. Their findings reveal that musicians perform an initial sketch of the musical piece in order to identify sections and key phrases, before the chord scheme or bass line are transcribed. A final step consists in finding the melody and counter-melodies.

This study of nineteen musicians also reveals the fact that musicians support their transcription with instrumental or software playback, musical knowledge, beat tracking, style detection, and instrument identification.

Issues
Existing systems often place strict restrictions on the music that can be transcribed, in the form of the type of instrument supported, musical genre, or maximum polyphony.

It should be noted that any form of transcription is not only interested in the pitch of notes, but also on their timing, attack, and release. Not all transcriptions focus on all of these parameters. Frequency-based analysis is also complicated by overlapping harmonics.

Klapuri (2004) argues in his Ph.D. dissertation that the problem of polyphonic transcription will not be solved until the complex way in which humans perform this task has been better understood and implemented.

Methods
The methods used are often divided into three categories. The bottom-up methods use only low-level analysis, and were the first ones to be developed (by Moorer in the 1970s). Blackboard systems develop hypotheses at various levels before making a decision, and model-based algorithms combine low-level and high-level analysis as well as parameter extraction to fit a particular model.

A common method is to perform pre- and post-processing (such as lowpass filtering and the removal of outliers), with event extraction and classification using Support Vector Machines, Hidden Markov Models, Neural Networks, Gaussian Mixture Models, and other classifiers. This classification tends to rely on previous knowledge of the instruments, thus placing restrictions on the system's input.

Niedermayer (2008) uses the matrix factorization of power spectrum components to perform the transcription and incorporates background knowledge of the musical piece into the matrix of basis vectors. Nichols and Raphael (2007) use a probabilistic approach in order to predict notes, and combine this with amplitude tracking.

Evaluation of Transcription Methods

Precision, recall, and f-measure are used in order to evaluate the quality of transcription systems. If N_C is the number of correct events detected, N_D the total number of events detected, and N the actual number of events, then precision is $P = \frac{N_C}{N_D}$, recall is $R = \frac{N_C}{N}$, and the f-measure is $F = \frac{2RP}{R+P} = \frac{2N_C}{N+N_D}$. The most common instruments in available papers are drums and the piano, although research exists with bass guitar, the violin, the voice, and the classical guitar.

Ryynänen and Klapuri (2005) obtain precision and recall values of about 40%. Although these are low, they are considered as encouraging because the described system tracks multiple notes by multiple instruments.

Bruno et al. (2003) describe a recall of about 50% by using 30 neural networks, and explain the low success rate by the low quality of the training material.

Hainsworth and Macleod (2005) obtain a precision of 78.7% when transcribing a bass guitar's monophonic sound in a polyphonic context.

Better results are obtained with percussions. Gillet and Richard (2005) obtain a precision and recall of about 84%, but only follow the bass and snare drums. The higher results can be explained by the fact that the almost constant pitch of percussions makes the transcription easier, compared to other types of instruments.

Conclusion

The task of transcribing polyphonic music is an ongoing problem, and is likely to result in several commercial applications when it can be performed with better precision and fewer restrictions. Despite being a difficult task, it can be done by expert musicians, and the success of a software solution is likely to depend on an improved understanding of the human methods to perform such transcription. Combining polyphonic transcription with source separation could improve the usefulness of the end system.

Reference

Bruno, I., S. L. Monni, and P. Nesi. 2003. Automatic music transcription supporting different instruments. In *Proceedings of the Third International Conference on Web Delivering of Music*. Leeds, UK. 37–44.

Cemgil, A. T., B. Kappen, and D. Barber. 2003. Generative model based polyphonic music transcription. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, NY, USA. 181–4.

Gillet, O. K., and G. Richard. 2005. Drum track transcription of polyphonic music using noise subspace projection. In *Proceedings of the International Conference on Music Information Retrieval*. London, UK. 92–9.

Hainsworth, S. W. 2001. Analysis of musical audio for polyphonic transcription. In *First Year Ph.D. Report*.

Hainsworth, S. W., and M. D. Macleod. 2001. Automatic bass line transcription from polyphonic music. In *Proceedings of the International Computer Music Conference*. Havana, Cuba.

Hainsworth, S. W., and M. D. Macleod. 2004. The automated music transcription problem. In *Cambridge University Engineering Department*.

Klapuri, A. 2004. Signal processing methods for the automatic transcription of music. In *Ph.D. Dissertation*.

Lidy, T., A. Rauber, A. Pertusa, and J. M. Iñesta. 2007. Improving genre classification by combination of audio and symbolic descriptors using a transcription system. In *Proceedings of the International Conference on Music Information Retrieval*. Vienna, Austria. 61–6.

Marolt, M. 2001. SONIC: Transcription of polyphonic piano music with neural networks. In *Proceedings of the Workshop on Current Research Directions in Computer Music*. Barcelona, Spain. 217–24.

Nichols, E., and C. Raphael. 2007. Automatic transcription of music audio through continuous parameter tracking. In *Proceedings of the International Conference on Music Information Retrieval*. Vienna, Austria. 387–92.

Niedermayer, B. 2008. Non-negative matrix division for the automatic transcription of polyphonic music. In *Proceedings of the International Conference on Music Information Retrieval*. Philadelphia, PA, USA. 544–9.

Paulus, J. 2006. Acoustic modelling of drum sounds with Hidden Markov Models for music transcription. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Toulouse, France.

Pertusa, A., and J. M. Iñesta. 2003. Polyphonic music transcription through dynamic networks and spectral pattern identification. In *IAPR International Workshop on Artificial Neural Networks in Pattern Recognition*. 19–25.

Ryynänen, M., and A. Klapuri. 2006. Transcription of the singing melody in polyphonic music. In *Proceedings of the International Conference on Music Information Retrieval*. Victoria, BC, Canada.