Using Blackboard Systems for Polyphonic Transcription: A Literature Review

Cory McKay
Faculty of Music, McGill University
555 Sherbrooke Street West
Montreal, Quebec, Canada H3A 1E3
cory.mckay@mail.mcgill.ca

1. Introduction

Automated general-purpose polyphonic music transcription systems have a great deal of potential utility, both to music technologists and to traditional music researchers who would find it convenient to avoid having to manually transcribe performances. Such systems have yet to be successfully implemented, however, although some limited success has been achieved with blackboard systems. There is ample opportunity for researchers to build on the work that has already been done in this field. This paper serves as both an introduction to the use of blackboard systems for polyphonic transcription and as a literature review.

2. Polyphonic transcription

Ideally, a polyphonic transcription system would take in an arbitrary musical audio signal and produce a notated score, complete with pitch, rhythm, dynamics and tempo information for each voice in the signal. Such a system could perhaps even glean information from audio signals that is not traditionally notated, but is perceived by humans.

Unfortunately, current systems are very far from being able to achieve these goals. Problems such as spectral variations within a given voice, voice crossing and difficulties in identifying the correct octave of a note have prevented the successful implementation of a general-purpose transcription system.

Some success, however, has been achieved using simplified models. As will be discussed later in this paper, systems have been implemented that transcribe audio signals derived from only one or a few instruments, that extract only a given instrument from a complex signal or that deal with music that obeys very restrictive rules. In addition to these limitations, most systems that have been implemented produce very simplified scores, often with information limited to pitch and time of note onsets for each voice. Such systems have reportedly achieved success rates of between 70% and 90%, although these numbers may be inflated due to the limited testing suites that have been applied.

3. Blackboard systems

Blackboard systems have been used for some time by researchers in artificial intelligence, but have only been applied to music transcription since the early 1990's. The term "blackboard" comes from the notion of a group of experts standing around a blackboard working together to solve a problem. Each expert writes contributions on the blackboard based on his/er expertise. The experts watch the problem evolve on the blackboard until a solution is achieved.

In terms of computing, the "blackboard" is a central dataspace that is usually arranged in a hierarchy so that input is at the lowest level and output is at highest. The "experts" are called "knowledge sources," and they generally consist of a set of heuristics and preconditions whose satisfaction result in a hypothesis that is written to the blackboard. Each knowledge source forms hypotheses based on information from the front end of the system and hypotheses presented by other knowledge sources. The problem is considered solved when all knowledge sources are satisfied with all hypotheses on the blackboard to within a given margin of error.

Blackboard systems eliminate the need for a global control module and allow problems to be solved through a combination of top-down and bottom-up processing. These systems are also highly adaptable, as new knowledge sources can be added or existing ones updated with minimal changes needed to the rest of the system.

Most music has a naturally hierarchal structure that lends itself well to blackboard systems. Blackboard systems allow the easy integration of signal processing knowledge at the low level, information about human perception at the middle level and music theory at the high level.

It should be noted that some caution should be exercised in regards to the influence of knowledge sources with expertise in music theory, as too much knowledge in a given style of music may limit the general applicability of a system. The recent trend, however, has been to increase knowledge about music theory in order to improve success rates with limited test suites.

4. The work of Keith Martin

Keith Martin was one of the first researchers to apply blackboard systems to music transcriptions (Martin, 1996 a). Although the early work of Kashino (Kashino, Nakadia, Kinoshita and Tanaka, 1995) slightly preceded his, Martin's system is simpler and provides a better introduction to blackboard systems. His system was limited to analyzing performances of piano music and was tested on four-voice Bach chorales.

The front-end of Martin's system applied short-time Fourier transforms to the input signal to generate associated sets of onset times, frequencies and amplitudes that were fed to the blackboard system. The blackboard system consisted of thirteen knowledge sources, each falling into one of three types: garbage collection, physics and musical practice. The hypotheses made by the knowledge sources fell into five hierachally-organized classes, namely tracks, partials, notes, intervals and chords.

Knowledge sources with access to upper-level hypotheses were able to put pressure on knowledge sources with lower-level access to make certain hypotheses, and vice versa. For example, if the hypotheses had been made that the notes C and G are present in a given beat, a knowledge source with information about chords might put forward the hypothesis that there is a C chord present, thus putting pressure on other knowledge sources to find an E or Eb.

A sequential scheduler was used to coordinate the knowledge sources. It allowed each knowledge source to make its contribution in turn until all were satisfied with the hypotheses on the blackboard.

One of the greatest weaknesses of this system was that it tended to misidentify octaves. In order to resolve this problem, Martin proposed changing the front-end of the system (Martin, 1996 b). He suggested first using a bank of filters to produce log-lag correlograms in the front-end, and then determining pitch by measuring the periodic energy in each filter channel as a function of lag. The correlograms could then be fed as the basic unit to the blackboard system. Martin did not achieve any definitive experimental results indicating whether this approach is better than his original approach.

5. The work of Kunio Kashino

Rather than using a sequential scheduler to coordinate the blackboard system, Kashino used a Bayesian probability network (Kashino, Nakadia, Kinoshita and Tanaka, 1995). Bayesian networks are well known for producing good results, despite noisy input or missing data. They are often used in implementing learning methods that trade off prior belief in a hypothesis against its agreement with current data. They therefore seem to be well suited to coordinating blackboard systems. There has not yet been any experimental research directly comparing the success of the sequential approach used by Martin to this Bayesian network approach, however.

Aside from this important difference between Kashino's work and that of Martin, Kashino also used knowledge sources with information about stream segregation taken from research in human auditory scene analysis. Kashino also gave his knowledge sources more high-level musical knowledge than Martin.

Kashino also set up his system so that it could analyze signals containing more than one instrument. In order to accomplish this, he used knowledge sources programmed with the frequency components of different instruments played with different parameters.

In a later publication (Kashino and Hagita, 1996), Kashino suggested replacing the Bayesian network with a Markov Random Field hypothesis network. This allowed information to be integrated on a multiply connected hypothesis network, unlike the Bayesian network that only allowed singly connected networks. This made it possible to deal with two kinds of transition information within a single hypothesis network, namely chord transitions and note transitions.

This approach was successful in correcting problems relating to misidentification of octaves and of instruments that plagued the previous system, although it did introduce some new errors. The new system performed achieved a recognition rate of 71.7% on a three-part arrangement of Auld Lang Syne, an overall improvement of roughly 10% over the old system.

Kashino later suggested a shift away from strict blackboard systems by performing more work in the front-end of the system and mathematically formalizing the work previously done by the knowledge sources (Kashino and Murase, 1998). Adaptive template matching was used in this new system. This system found the correlation between the output of a bank of filters arranged in parallel and a set of templates corresponding to particular notes played by particular instruments.

Although this approach did achieve an average recognition rate of 88.5% on recordings of piano, violin and flute, abandoning the blackboard approach seriously compromised the scalability and adaptability of the system. It is unlikely that the system would continue to perform well if more templates were added, particularly ones with similar frequency spectra or a great deal of spectral variation from note to note.

6. Recent research

As mentioned in section 2, an alternative approach is to take an input signal containing arbitrary instruments and extract information relating to only one of them. Some success has been achieved in extracting basslines (Hainsworth and Macleod, 2001). High frequencies were filtered out of the signal and simple mathematical relations were used to trim hypotheses.

Bello and Sandler (2000) have designed a system based on Martin's design, using a sequential scheduler. Aside from refining the knowledge sources and adding high-level musical knowledge, they implemented a chord recognizer knowledge source as a feed-forward neural network. The network was trained using spectrographs of different chords of a piano and it produced candidate chords. The network could output more than one hypothesis at each iteration, allowing the system to perform a parallel exploration of the solution space. Preliminary testing showed that the system had a tendency to misidentify octaves and make incorrect identification of note onsets, but these problems could potentially be solved by modifications to the signal processing system that feeds the blackboard system data and by refining the knowledge sources.

7. Conclusions

Research on polyphonic transcription systems appears to be proceeding along two paths. The first path involves producing specialized systems that extract only a certain type of instrument or are only able to deal with signals containing limited instruments. The second path involves producing more general systems. Specialized systems of the first type have shown some promise, and can provide useful tools until better general-purpose systems are developed. General-purpose systems have much more potential however, which justifies further research despite the many difficulties that have been encountered.

The blackboard paradigm seems to be the best model that has been proposed so far for general-purpose transcription systems. Further refinements of the knowledge sources could improve results. The incorporation of further information relating to human perception of sound, specifically auditory streaming, could prove valuable. More information about music theory could also be used but, as discussed earlier, caution should be taken not to compromise the generality of the systems.

The neural network approach suggested by Bello and Sandler could prove to be particularly fruitful. Experiments involving different types of networks applied to different types of knowledge sources could be informative.

It might be particularly interesting to do research into producing semi-automatic systems with interchangeable knowledge source modules that could be switched for different types of music. Since it is relatively easy for humans to identify the instruments present in a signal and the style of the music, humans could specify this information to the system, which would then use the appropriate knowledge source modules for the given styles and instrumentations. Neural-network-based knowledge sources would be particularly well suited to this type of system, as the networks could be automatically trained to deal with new instruments or styles. This would eliminate manual programming of knowledge

sources for the many possible styles and instrumentations.

8. Bibliography

Bello, J. P. and M. B. Sandler. 2000. Blackboard Systems and Top-Down Processing for the Transcription of Simple Polyphonic Music. *Proceedings of the COST G-6 Conference on Digital Audio Effects*.

Hainsworth, S. W. and M. D. Macleod. 2001. Automatic Bass Line Transcription from Polyphonic Music. *Proceedings of the Interntational Computer Music Conference*.

Kashino, K., K. Nakadia, T Kinoshita and H. Tanaka. 1995. Application of Bayesian probability network to music scene analysis. *Proceedings of the International Joint Conference on AI, CASA workshop,* 52-59.

Kashino, K. and Norihiro Hagita. 1996. A Music Scene Analysis System with the MRF-Based Information Integration Scheme. *Proceedings of the International Conference on Pattern Recognition*, vol. 2. 725-729.

Kashino, K. and Hiroshi Murase. 1998. Music Recognition Using Note Transition Context. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, vol. 6. 3593-3596.

Martin, K. D. 1996 a. A Blackboard System for Automatic Transcription of Simple Polyphonic Music. *M.I.T. Media Lab Perceptual Computing Technical Report #385*, July 1996.

Martin, K. D. 1996 b. Automatic Transcription of Simple Polyphonic Music: Robust Front End Processing. *M.I.T. Media Lab Perceptual Computing Technical Report #399*, November 1996.