

# Interactive Music Transcription based on Rhythm Tree Languages

Florent Jacquemard  
INRIA – Sorbonne Universités  
STMS (IRCAM-CNRS-UPMC), Paris, France  
`florent.jacquemard@inria.fr`

Adrien Ycart  
School of Electronic Engineering and Computer Science  
Queen Mary University of London, UK  
`adrien.ycart@gmail.com`

January 27, 2017

We consider symbolic rhythm transcription: the problem of the conversion of a sequence of timestamped notes, *e.g.* a MIDI file, into conventional Western music notation [5]. This scenario can follow a transcription from audio to MIDI, or occur in a context of score edition (input from a MIDI keyboard), or algorithmic composition (when the sequence is the result of a computation).

The fitness of a notation to a given a MIDI input is an intrinsically ambiguous notion, as it depends strongly on the the musical context. We propose a system available as a graphical component of the assisted composition environment OpenMusic [2], which privileges user interactions in order to search for an appropriate balance between different criteria, in particular the precision of the transcription and the readability of the musical scores produced.

We followed for the development of this system [7] a uniform approach where a set of preferred rhythm notations is specified a priori as a formal tree language. We did this for two reasons. First, rhythm trees (RT) [1] are a natural representation of the proportional rhythmic notation, encoding in their hierarchical structure note durations defined by successive divisions of the beat. Second, this approach benefits from well established formalisms and algorithms for representing and processing tree languages (often applied in the context of natural language processing).

More precisely, a language of preferred rhythm notations is defined as a weighted tree automaton [3], associating to every RT a weight value representing its complexity. Such automata are an internal representation and

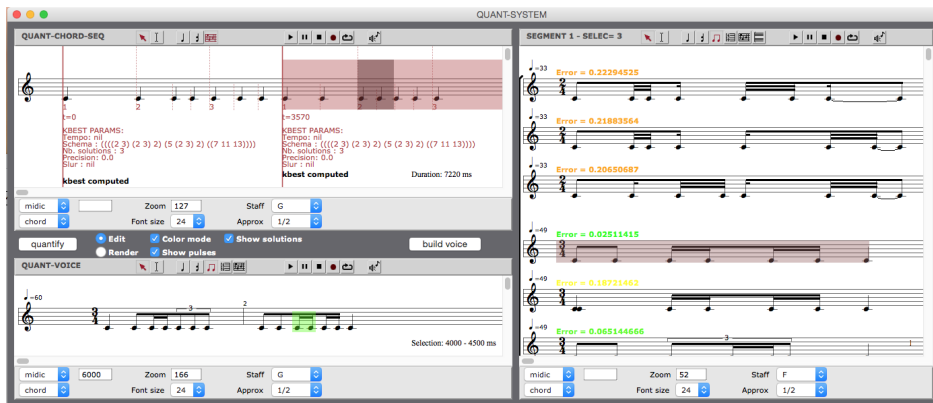


Figure 1: <http://repmus.ircam.fr/cao/rq>

can be specified by various parameters and combined by operations such as union or intersection, providing a modular approach for the specification of RT languages. Given such an automaton and an input sequence of timestamped notes, we can construct a second automaton associating to each RT a rank defined as a combination of its complexity and its fitness to input.

Then, we use efficient dynamic programming algorithms for the lazy enumeration of RT, in ascending rank (*aka k-best parsing* [4]). This provides a procedure for the on-the-fly enumeration of a (possibly very large) set of transcription solutions, from the best one to the worst one.

Its implementation is carried out via a dedicated interface in OpenMusic, offering the user the possibility to choose her favorite transcriptions in an interactive way, by exploring the solutions space, with visualizing and local editing. A particular attention has been paid to the processing of grace notes and rests in this setting. Moreover, the choices of the user permit to refine the initial model of RT language.

The above procedure of rhythm quantization has been extended to deal with tempo detection. Both problems have been tightly coupled thanks to our modular approach based on RT language enumeration. Indeed, assuming that the input is divided into segments with constant tempi, we can encode in the weighted tree automata an initial step of division of each segment in beats (as an higher tree stage), defining a tempo. Initial segmentation can be user defined or rely on a procedure such as [6].

## References

- [1] Carlos Agon, Karim Haddad, and Gérard Assayag. Representation and rendering of rhythm structures. In *Proc. 2nd Int. Conf. on Web Delivering of Music*, pages 109–113, Darmstadt, Germany, 2002. IEEE Computer Society.
- [2] Jean Bresson, Carlos Agon, and Gérard Assayag. Openmusic: visual programming environment for music composition, analysis and research. In *Proc. of the 19th ACM Int. Conf. on Multimedia*, pages 743–746. ACM, 2011.
- [3] Zoltán Fülöp and Heiko Vogler. Weighted tree automata and tree transducers. In Manfred Droste, Werner Kuich, and Heiko Volger, editors, *Handbook of Weighted Automata*, pages 313–403. Springer, 2009.
- [4] Liang Huang and David Chiang. Better k-best parsing. In *Proc. of the 9th Int. Workshop on Parsing Technology*, pages 53–64. Association for Computational Linguistics, 2005.
- [5] Martin Piszczalski and Bernard A Galler. Automatic music transcription. *Computer Music Journal*, 1(4):24–31, 1977.
- [6] Aaron C Yang, Elaine Chew, and Anja Volk. A dynamic programming approach to adaptive tatum assignment for rhythm transcription. In *Multimedia, Seventh IEEE International Symposium on*, pages 8–pp. IEEE, 2005.
- [7] Adrien Ycart, Florent Jacquemard, Jean Bresson, and Slawomir Stawko. A Supervised Approach for Rhythm Transcription Based on Tree Series Enumeration. In *International Computer Music Conference (ICMC)*, Proceedings of the 42nd International Computer Music Conference (ICMC), Utrecht, Netherlands, September 2016.