Requirements for Music Notation regarding Music-to-Score Alignment and Score Following

Diemo Schwarz
schwarz@ircam.fr

December 3, 2003

Contents

Preface

This document describes music alignment and real-time score following, and the resulting requirements for a score representation or music notation format. It is a chapter extracted from my current PhD thesis *Data-Driven Concatenative Sound Synthesis*, to be finished middle of December of this year. I performed only minimal editing, so it is quite long. If you want to get right to the point, I suggest reading section 2 for a list of applications of music alignment, and thus applications for a music notation standard, and certainly sections 5.1 and 5.6 with the requirements for the score format for alignment and a conclusion. If you need more motivation for why a standardised score format is so important, please go on reading section 6 relating standardisation efforts for evaluation of alignment and section 3 giving an overview of the large amount of work done on this topic.

However, if you have the time to read it all, I'd be glad to hear any suggestions or remarks. The list of score formats in section 5 is rudimentary and not very well researched. It is obsoleted anyway by the enormous amount of references listed on the very helpful web pages of Musicnetwork. The complete draft of the report is at `http://www.ircam.fr/anasyn/schwarz/thesis`.

1 Introduction

Music alignment is the association of events in a musical score (in our case, notes) with points in the time axis of an audio signal. The signal is a digital recording of the score being played by musicians and is referred to as the *performance*. An alignment implies a segmentation of the performance according to the events in the score. Additionally, we can associate the symbolic information contained in the score to the segments obtained. Music alignment is sometimes also called *score–performance matching* and has many other applications evoked in section 2.

The two principal methods used are Dynamic Time Warping and Hidden Markov Models. Other methods perform a beat alignment, see for instance (Scheirer 1998; Dixon 2001), but do not treat music alignment in general.

Both methods share a score modeling described in section 4, and can cope with polyphonic and multi-instrument performances as well as with performances where fast sequences or trills are present. Normally, blind segmentation methods, which use only the information from the audio signal, are not very accurate with these kinds of performances.

The alignment is done using a model of the musical score, which is built from an external score coding. To code the score, we have several formats to choose from, which are compared in section 5, after defining the *requirements for a good score format* in section 5.1. Evaluation of alignment lays the ground for obtaining objective results and their comparison. Evaluation methods are discussed in section 6.

2 Applications of Alignment

A great part of the research in computer science is devoted to the automation of processes carried out by humans. For instance, the segmentation of a large collection of recordings, which may last several hours, can not feasibly be done manually because of the large amount of data. The same situation applies for difficult signals (i.e., fast sequences of notes with legato) where manual segmentation may be tedious or imprecise.

Automatic music alignment has many applications in various fields, the most important of these are briefly described here, followed by a comparison of the required accuracy for each of them.

2.1 Music Information Retrieval (MIR)

- Alignment allows the *indexing* of continuous media through the implied segmentation, and labeling of the segments with the score information for content-based retrieval (CBR) (Foote 1999; Dannenberg, Foote, Tzanetakis, and Weare 2001).

- The total alignment cost between pairs of sequences can be considered as a *distance measure* (as in early works on speech recognition), allowing to find the best matching scores or performances from a database.

2.2 Musical Analysis

- For a researcher in musicology working on a symbolic score, alignment allows them to listen to a performance of the score at a specific position of interest. One such application is the *CUIDADO Music Browser* (see Vinet, Herrera, and Pachet 2002).

  Moreover, linking the symbolic score the audio signal allows to test hypotheses about the acoustic effects of compositional and orchestration techniques.

- For the field of *performance study*, the comparison of different performances regarding the expressive parameters related to timing and interpretation characteristics of the musician, can be of great help.

2.3 Sound Analysis

- Alignment is related to the problem of real-time synchronisation between performers and computers, usually called *score following*, when additional constraints of low-latency and only local knowledge of the performance are introduced. Off-line alignment can be used as a bootstrap procedure for the training of real-time statistical models, to get a good segmentation to start from.

Score following (Orio and Déchelle 2001; Schwarz and Orio 2002; Orio, Lemouton, Schwarz, and Schnell 2003) is the synchronisation of a computer with a performer playing a known musical score. It now has a history of about twenty years as a research and musical topic, and is an ongoing project at Ircam (ATR 2003).

In order to transform the interaction between a computer and a musician into a more interesting experience, research for a virtual musician has as goal to simulate the behaviour of a musician playing with another, to create a virtual accompanist or "synthetic performer", that will follow the score of the human musician. Score following is often addressed as "real-time automatic accompaniment". This problematic is well defined in (Dannenberg 1984; Vercoe 1984; Vercoe and Puckette 1985), where we can find the first use of the term "score following". Since the first formulation of the problem, several solutions have been proposed, some academic, others in commercial applications. See for instance (Baird, Blevins, and Zahler 1990; Baird, Blevins, and Zahler 1993; Bryson 1995; Dannenberg and Mont-Reynaud 1987; Dannenberg and Mukaino 1988; Grubb and Dannenberg 1994; Grubb and Dannenberg 1997; Grubb and Dannenberg 1998; Puckette 1990; Puckette 1995; Puckette and Lippe 1992; Raphael 1999b; Raphael 2001a; Raphael 2001b; Orio and Déchelle 2001; Orio, Lemouton, Schwarz, and Schnell 2003; Izmirli, Seward, and Zahler 2003).

- Alignment can help in determining very precise single or multiple *fundamental frequency* values, by reducing the number of candidates according to the expected values from the score. This comes especially handy for descriptor calculation for unit database building (see the last application below): Many descriptors depend on the fundamental frequency, so a very precise fundamental frequency estimation is essential. This is still not always assured by the current methods, especially in the attack phases and slightly polyphonic parts.

- *Source separation* (Vincent 2003) is closely intertwined with segmentation: knowing the temporal location of notes in a multi-instrument piece, for instance, helps singling them out in the frequency domain. On the other hand, to find the notes' time positions, we need to distinguish them from other instruments' notes.

2.4 Musical Synthesis

- The creation of an *augmented score* (or retranscription) means rewriting the score to be closer to the expressiveness of the performance, e.g. using the aligned note times, the dynamics, and frequency deviations (vibrato) from the performance to write a new Midi file based on the score with e.g. the rhythmic groove and the dynamics of the performance.

- Our main application, described in (Schwarz 2000; Schwarz 2003a; Schwarz 2003b), is the *segmentation* of a performance into notes and labeling (tagging) of the notes with the information from the score for building unit databases. Along with the note pitch and length, there can be additional symbolic information attached to the score, such as dynamics, articulation, or lyrics.

  For selection based synthesis to succeed, we need large databases of musical material, that are well segmented into our synthesis units, i.e. notes in the case of the application to instrument synthesis. Only automatic methods can provide that much material, but to obtain the required preciseness, blind segmentation methods, such as those described in (Rossignol, Rodet, Soumagne, Colette, and Depalle 1998; Rossignol 2000), are too error-prone. Moreover, with blind segmentation, we'd know the place of the segments, but would know nothing more about their content. That's why much work was devoted to segmentation by alignment, described in (Orio and Schwarz 2001; Soulez, Rodet, and Schwarz 2003), for music where the score was available.

2.5 Required Accuracy of Alignment

The required accuracy of the alignment varies greatly with the application. Table 1 gives a rough estimation of the requirements for the above applications.

Field	**Application**	**Accuracy**
MIR	Indexing	+
	Distance measure	−
Musical Analysis	Music Browser	o
	Performance study	+
Sound Analysis	Score following	+
	Fundamental frequency estimation	+
	Source separation	+
Musical Synthesis	Augmented score	++
	Segmentation	++

Table 1: Comparison of required accuracy of alignment for different applications

3 Previous Work

In general, automatic alignment of sequences has been a popular research topic in many fields, such as string analysis, molecular biology, and notably speech recognition. The literature is considerably vast, and we only mention two comprehensive overviews on the different approaches in speech recognition (Rabiner and Juang 1993) and in biological sequence analysis (Durbin, Eddy, Krogh, and Mitchison 1998).

Alignment has also been used in speech synthesis research, as a useful tool for preparing unit databases for concatenative speech synthesis. The results of the MBROLIGN technique from the MBROLA project (Malfrere and Dutoit 1997), has been the motivation for our method.

Concerning automatic music alignment specifically, the main works are score following techniques tuned for off line use, mostly based on stochastic models (Grubb and Dannenberg 1997; Grubb and Dannenberg 1998; Raphael 1999b; Raphael 1999a; Raphael 2001a; Raphael 2001b; Loscos, Cano, and Bonada 1999b; Loscos, Cano, and Bonada 1999a; Orio and Déchelle 2001; Shalev-Shwartz, Dubnov, Friedman, and Singer 2002; Dannenberg and Hu 2003; Turetsky 2003; Turetsky and Ellis 2003). All of these techniques consider mainly monophonic recordings.

For music containing a drum or percussion part, aligning the notes is more difficult for the proposed methods. Beat alignment algorithms (Hainsworth and Macleod 2003; **?**) specialise on aligning the drum hits in a performance with a percussion score representation, trying to ignore the interleaved melodic notes.

For note recognition, there are many pitch detection techniques using signal (Rodet and Doval 1992; Doval 1994; Prudham 2002) or auto-correlation (de Cheveigné and Kawahara 2002; de Cheveigné and Henrich 2002; de Cheveigné 2002), for instance. These techniques are often efficient in monophonic cases but none of these use score information and are therefore sub-optimal in our situation.

4 Score Parsing

In order to build the internal score representation (the model), the external score file has to be parsed. As implicitly introduced in (Orio and Schwarz 2001), the result of the score parsing is a time-ordered sequence of *score events* at every change of polyphony, i.e. at each note start and end, as exemplified in figure 1.

Many score files, especially those generated by recording a MIDI-performance, contain score events with slight desynchronisations: For instance, the notes of a chord played on a keyboard are never triggered perfectly synchronous, but are slightly arpeggiated (figure 2). Equally, passages presumed legato can show a short overlap or gaps between notes (figure 3).

To avoid to create too many very short states at these score events, a *quantisation* is performed that fuses score events within a window of usually 30 ms into one single event, and eliminates pauses shorter than 100 ms altogether. The result is that all circled events in figures 2 and 3 are moved to the time of the earliest event for each circle.

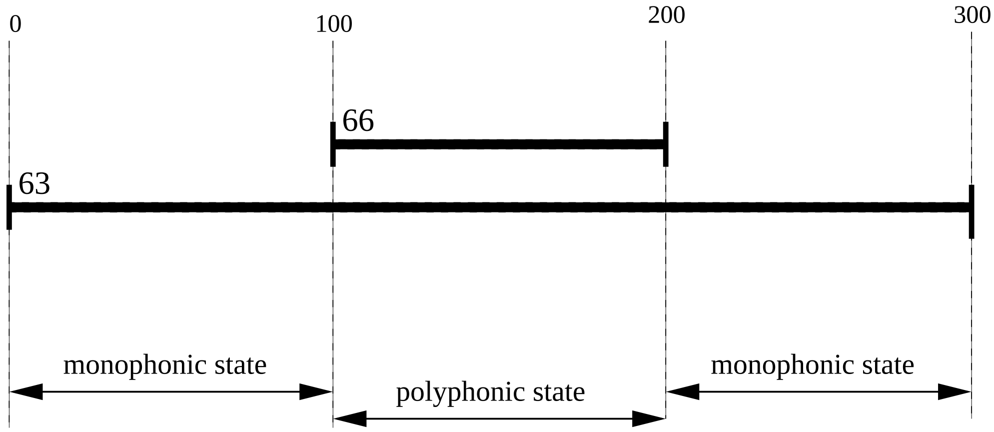

Figure 1: Score parsing into score events and the states between them.

5 Score Formats

The scores to be aligned are coded in a digital score file format. The internal score representation (the model) is built from this external coding.

5.1 Requirements for a Score Format

The definition of the imported score format is essential for the ease of use of automatic alignment. The requirements and constraints are multiple:

1. It has to be powerful, flexible, and extensible enough to represent all the things we want to align. These are essentially the notes, but some exceptions apply: trills, for instance, should not be represented as the notes actually played, but as a single object. For singing voice performances, there are musical events that are insufficiently represented as notes, e.g. fricatives, because they have no pitch.

   Another important information are the *cues*. A cue is a numeric or symbolic label attached to a position or a note in the score, that is output by a score following system to trigger an event in the electronic part of the piece whenever the corresponding score position is reached in the performance.

2. The score format should correspond as closely as possible to the printed score the musician had at her disposal, when she performed the score. It should be, above all, on the same level of abstraction, i.e. it should bear easily accessible high-level musical information intended for the musician.

3. It should be easy to generate. Export from popular score editors, or recording it from live input should be easily possible.

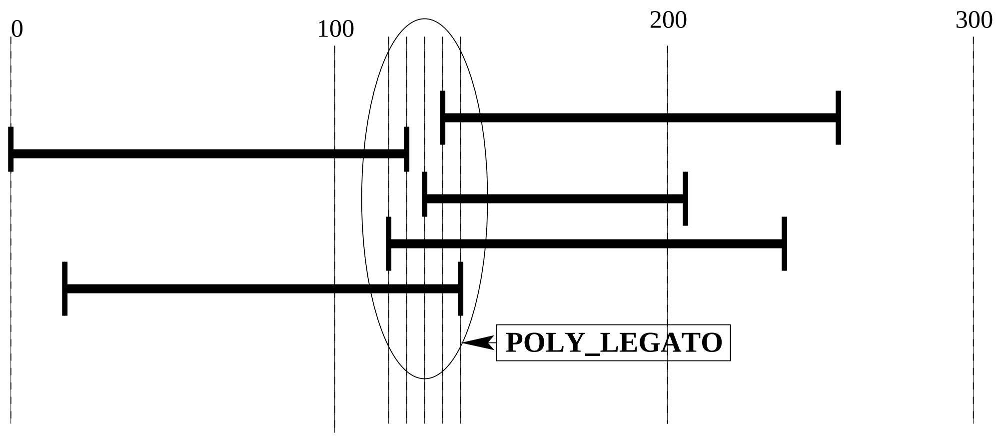

Figure 2: Desynchronised chord.

4. It should be easy to read: There should be an existing parser for importing it, preferably as an open source library under a license qualifying for free (libre) software[1].

5. It is convenient to be able to fine-tune imported scores within the alignment system, without re-importing.

6. A large corpus of pieces coded in this format should be available, covering at least classical music where recordings are readily available.

In the following, we give an overview of several current score-representation formats, with remarks about their suitability for music alignment, and a conclusion in section 5.6. For an overview of all sorts of musical data formats, see (Selfridge-Field 1997).

5.2 Score Editor Formats

Finale, Sibelius

Finale by *Coda Music* and *Sibelius* by *Sibelius*[2] are the most widely used commercial score editors. Their file formats are proprietary and even if they could be decoded, they would most probably not be suitable for high-level score coding, since they focus on graphical printed score layout.

NIFF

Notation Interchange File Format (NIFF Consortium 1995) is a graphical interchange format for music notation, and therefore not suitable either for abstract score representation.

[1] The free software community (see `www.fsf.org`) is about to adopt the term *libre software* to distinguish the sense of "free" as in "free speech" from the "free beer" sense.

[2] `www.sibelius.com`

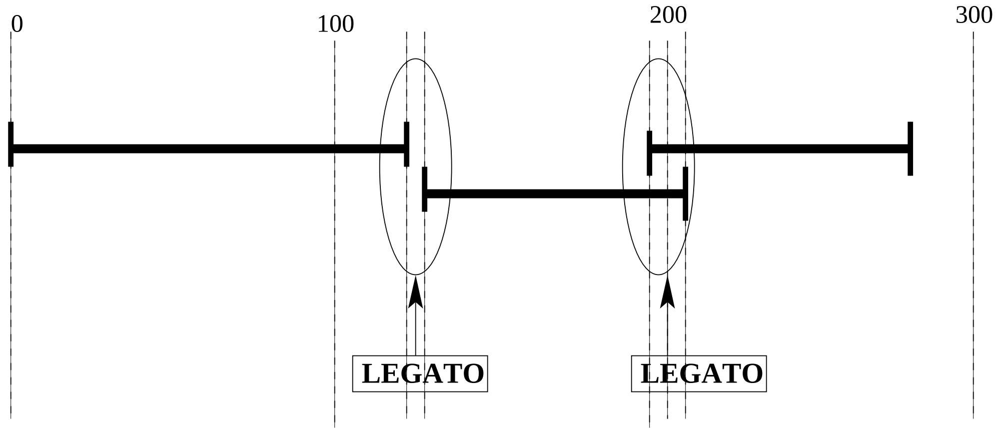

Figure 3: Desynchronised legato notes.

Guido

Guido (Hoos, Hamel, Renz, and Kilian 1998) (Guido 2000) is an open non-commercial format for high-quality score layout and representation. Many free software tools and libraries for reading, displaying, and editing exist.

5.3 Mark-Up Languages

Standard Music Description Language (SMDL)

The Standard Music Description Language (Newcomb 1990; Newcomb 1991; ISO 1995) is a ISO standard developed by the Music in Information Processing Standards (MIPS) committee to enable interchange of musical documents, and attach external information to cues for musical comedy and opera. It is an SGML (Standard Generalized Markup Language) document type definition (DTD), i.e. a format defined in the meta-language SGML and not very widely used in the community.

MusicML

MusicML

MuTaTeD

Music Tagging Type Definition (MuTaTeD!) is the integration of SMDL (Standard Music Description Language) and NIFF (Notation Interchange File Format) to support the representation of music as a time-structured entity and at the same time to provide a standard for high-quality display via the NIFF format. MuTaTeD'II (Böhm, MacLellan, and Hall 2000), started in November 1999, adds music information retrieval capabilities with delivery/access services for encoded music. The wider aim is to establish a standard Meta-DTD for music tagging languages.

Wedelmusic XML Format

The Wedelmusic XML Format (Bellini and Nesi 2001) has been developed within the *WEDELMUSIC—Web Delivering of Music* project (Wedelmusic 2003). It offers a standalone editor, and plugins to save Wedelmusic XML format files for *Finale* and *Sibelius*.

MusicXML

MusicXML

Music Notation in MPEG

In August 2003, the Moving Picture Experts Group (MPEG) has started an Ad Hoc Group (AHG) on music notation requirements[3] to study the requirements of integrating music notation into MPEG.

This activitiy is supported by the MUSICNETWORK project in workshops of music notation[4].

Most promising, integrated in standard MPEG-7

5.4 Frameworks

A framework is a library of classes or functions destined to facilitate writing music applications. Some existing frameworks also provide data-structures for musical score representation, access and manipulation functions for these, and an external file representation, which makes them eligible for the score format we are looking for. However, although they are quite high-level, their principal aim is not interchange, and due to their strong link with an existing application there are doubts about their generality and extensibility.

Common Practice Music Notation

Common Practice Music Notation (CPNview) (cpnview DAFx 2001)

Allegro

The Allegro music and sound processing framework (Dannenberg and van de Lageweg 2001) defines a high-level internal data structure for musical scores that can be saved to disk and reloaded.

5.5 MIDI

The *Standard MIDI File* (SMF) format (Midifile) is an extension to the MIDI (*Musical Instrument Digital Interface*) control protocol (midi, mma web) to represent MIDI performances in a file. It adds to the pure MIDI control messages the possibility to group them into several tracks, to represent tempo and musical meter, and to add metadata (comments, copyright, author information) and lyrics.

[3] `http://www.dsi.unifi.it/\home{nesi}/mpeg/ahg-mn-65-66.html`

[4] `http://www.interactivemusicnetwork.org/events/notation_workshop_2003.html`

5.6 Conclusion

Unfortunately, MIDI is currently still the most practical format, even with its shortcomings and restrictions. It does indeed fulfill all the requirements mentioned above: Defining some conventions, it can code everything we want to align, e.g. using special Midi channels, controllers, or text events[5]. It can be exported from every score editor, and can be read and fine-tuned in many simple and embedded midi editors.

The result is that we stay with Midi for the time being, but research has to be done for a higher-level representation that inserts itself well into the composer's and musical assistant's workflow.

There is an overwhelming number of MIDI files available on the net (however, sometimes in bad quality[6]). For all the other formats, although much more advanced, the lack of sufficiently large corpora is the most severe disadvantage. The latter condition, however, could change rapidly as soon as the musical information retrieval (MIR) and musicology communities decide to use one of the formats, and are able to gain institutional support from several sufficiently large research facilities, and, more important, commercial support from one the leading score editor vendors.

6 ICMC 2003 Panel Session on Evaluation

Evaluation gives an indication of the quality of the alignment algorithm and allows to compare different methods, parameters, etc., and to quantify the improvements gained by training.

The International Computer Music Conference in September 2003 hosted a 90 minute panel session with the topic of evaluation of score following and audio alignment systems in the aim of creating a standard for evaluation and a standard set of test cases. These test cases would immensely profit from a standardized high level score format other than midi. It was organised by Noel Zahler, the participants were Roger Dannenberg, Cort Lippe, Ozgur Izmirli, Xavier Rodet, and Diemo Schwarz. Nicola Orio, Miller Puckette, Christopher Raphael, and Barry Vercoe were also invited but could not attend.

Over the following year, the participants will exchange the necessary data to agree on such a standard and constitute a database of test sounds and their reference alignments.

References

AS (2000). Analysis–Synthesis Team, Ircam—Centre Pompidou. WWW page. `http://www.ircam.fr/anasyn`.

[5] We could—if we really wanted to—use text events to carry XML tags, such as to code structured information in a canonical extensible way, linked to precise points in time in the midi score.

[6] This constitutes good test examples for score–performance mismatches

ATR (2003). Real-Time Applications Team/Équipe Applications Temps-Réel, Ircam—Centre Pompidou. WWW page. `http://www.ircam.fr/equipes/temps-reel`

Baird, B., D. Blevins, and N. Zahler (1990). The Artificially Intelligent Computer Performer: The Second Generation. In *Interface – Journal of New Music Research*, Number 19, pp. 197–204.

Baird, B., D. Blevins, and N. Zahler (1993). Artificial Intelligence and Music: Implementing an Interactive Computer Performer. *Computer Music Journal 17*(2), 73–79.

Bellini, P. and Paolo Nesi (2001, November). WEDELMUSIC Format: An XML Music Notation Format for Emerging Applications. In *First International Conference on WEB Delivering of Music (WEDELMUSIC'01)*, Florence, Italy, pp. 79. `http://www.wedelmusic.org/`

Böhm, Carola, Donald MacLellan, and Cordy Hall (2000). MuTaTeD'II: A System for Music Information Retrieval of Encoded Music. In *Proceedings of the International Computer Music Conference (ICMC)*, Berlin. ICMA.

Bryson, Joanna (1995). The Reactive Accompanist: Adaptation and Behavior Decomposition in a Music System. In L. Steels (Ed.), *The Biology and Technology of Intelligent Autonomous Agents*. Springer-Verlag: Heidelberg, Germany.

Dannenberg, Roger, Jonathan Foote, George Tzanetakis, and Christopher Weare (2001, September). Panel: New directions in music information retrieval. In *Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba.

Dannenberg, R. and N. Hu (2003). Polyphonic audio matching for score following and intelligent audio editors. In *Proceedings of the International Computer Music Conference (ICMC)*.

Dannenberg, Roger and Patrick van de Lageweg (2001, September). A system supporting flexible distributed real-time music processing. In *Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba.

Dannenberg, Roger B. (1984). An On-Line Algorithm for Real-Time Accompaniment. In *Proceedings of the ICMC*, pp. 193–198.

Dannenberg, Roger B. and B. Mont-Reynaud (1987). Following an Improvisation in Real Time. In *Proceedings of the ICMC*, pp. 241–248.

Dannenberg, Roger B. and Mukaino (1988). New Techniques for Enhanced Quality of Computer Accompaniment. In *Proceedings of the ICMC*, pp. 243–249.

de Cheveigné, Alain (2002). Two voice fundamental frequency estimation. *Journal of the Acoustical Society of America (JASA) 111*.

de Cheveigné, Alain and Nathalie Henrich (2002). Fundamental frequency estimation of musical sounds. *Journal of the Acoustical Society of America (JASA)*.

de Cheveigné, Alain and Hideki Kawahara (2002). Yin, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America (JASA) 111*, 1917–1930.

Dixon, Simon (2001). An empirical comparison of tempo trackers. In *8th Brazilian Symposium on Computer Music*, Fortaleza, Brazil.

Doval, Boris (1994, March). *Estimation de la fréquence fondamentale des signaux sonores*. Ph. D. thesis, LAFORIA, Université Paris VI.

Durbin, R., S. Eddy, A. Krogh, and G. Mitchison (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press.

Foote, Jonathan (1999, January). An overview of audio information retrieval. *Multimedia Systems 7*(1), 2–10.

Grubb, Lorin and Roger B. Dannenberg (1994). Automating Ensemble Performance. In *Proceedings of the ICMC*, pp. 63–69.

Grubb, Lorin and Roger B. Dannenberg (1997). A Stochastic Method of Tracking a Vocal Performer. In *Proceedings of the ICMC*, pp. 301–308.

Grubb, Lorin and Roger B. Dannenberg (1998). Enhanced Vocal Performance Tracking Using Multiple Information Sources. In *Proceedings of the ICMC*, pp. 37–44.

Hainsworth, S. and M. Macleod (2003). Onset detection in musical audio signals. In *Proceedings of the International Computer Music Conference (ICMC)*.

Hoos, H. H., K. A. Hamel, K. Renz, and J. Kilian (1998). The GUIDO Music Notation Format — A Novel Approach for Adequately Representing Score-level Music. In *Proceedings of the International Computer Music Conference (ICMC)*, Berlin, pp. 451–454. ICMA.

ISO (1995, July). Music Description Language (SMDL). International Organization for Standardization (ISO), `http://www.oasis-open.org/cover/smdlover.html`.

Izmirli, O., R. Seward, and N. Zahler (2003). Melodic pattern anchoring for score following using score analysis. In *Proceedings of the International Computer Music Conference (ICMC)*.

Loscos, A., P. Cano, and J. Bonada (1999a). Low-Delay Singing Voice Alignment to Text. In *Proceedings of the International Computer Music Conference (ICMC)*.

Loscos, A., P. Cano, and J. Bonada (1999b). Score-Performance Matching using HMMs. In *Proceedings of the International Computer Music Conference (ICMC)*, pp. 441–444.

Malfrere, F. and T. Dutoit (1997). Speech synthesis for text-to-speech alignment and prosodic feature extraction. In *Proc. ISCAS 97*, Hong-Kong, pp. 2637–2640. www (?? 2000).

Newcomb, S. R. (1990, ????). Explanatory cover material for section 7.2 of X3V1.8M/SD-7, fifth draft (music description language). In J. Moline, D. Benigni, and J. Baronas (Eds.), *Proceedings of the Hypertext Standardization Workshop (NIST SP 500-178)*, Gaithersburg, MD, USA, pp. 179–188. National Institute for Standards and Technology.

Newcomb, Steven R. (1991, July). Standards: Standard Music Description Language complies with hypermedia standard. *Computer 24*(7), 76–79.

NIFF Consortium (1995, July). NIFF 6a: Notation Interchange File Format. Technical report, NIFF Consortium.

Orio, Nicola and François Déchelle (2001). Score Following Using Spectral Analysis and Hidden Markov Models. In *Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba. International Computer Music Association.

Orio, Nicola, Serge Lemouton, Diemo Schwarz, and Norbert Schnell (2003, May). Score Following: State of the Art and New Developments. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, Montreal, Canada.

Orio, Nicola and Diemo Schwarz (2001). Alignment of Monophonic and Polyphonic Music to a Score. In *Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba.

Prudham, Boris (2002). Estimation de la fréquence fondamentale d'un signal.

Puckette, Miller (1990). EXPLODE: A User Interface for Sequencing and Score Following. In *Proceedings of the ICMC*, pp. 259–261.

Puckette, Miller (1995). Score Following Using the Sung Voice. In *Proceedings of the ICMC*, pp. 199–200.

Puckette, Miller and Cort Lippe (1992). Score Following in Practice. In *Proceedings of the ICMC*, pp. 182–185.

Rabiner, Lawrence R. and Biing-Hwang Juang (1993). *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall.

Raphael, Christopher (1999a). A Probabilistic Expert System for Automatic Musical Accompaniment. *Jour. of Comp. and Graph. Stats 10*(3), 487–512.

Raphael, Christopher (1999b). Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 21*(4), 360–370.

Raphael, Christopher (2001a). A Bayesian Network for Real Time Music Accompaniment. *Neural Information Processing Systems (NIPS)* (14).

Raphael, Christopher (2001b). Music Plus One: A System for Expressive and Flexible Musical Accompaniment. In *Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba.

Rodet, Xavier and Boris Doval (1992, October). Maximum likelihood harmonic matching for fundamental frequency estimation. In *124th meeting of ASA, workshop on "Automatic Tracking of Musical Frequency"*, Volume 92 of *2*, New Orleans, Louisiana, USA. Journal of the Acoustical Society of America (JASA).

Rossignol, Stéphane (2000, July). *Segmentation et indexation des signaux sonores musicaux*. Ph. D. thesis, Université Paris VI. www (AS 2000).

Rossignol, S., X. Rodet, J. Soumagne, J-L. Colette, and P. Depalle (1998, October). Feature Extraction and Temporal Segmentation of Acoustic Signals. In *Proceedings of the International Computer Music Conference (ICMC)*, Ann Arbor, MI, USA.

Scheirer, Eric (1998). Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America (JASA)* (50), 588–601.

Schwarz, Diemo (2000, December). A System for Data-Driven Concatenative Sound Synthesis. In *Digital Audio Effects (DAFx)*, Verona, Italy, pp. 97–102.

Schwarz, Diemo (2003a, October). New Developments in Data-Driven Concatenative Sound Synthesis. In *Proceedings of the International Computer Music Conference (ICMC)*, Singapore.

Schwarz, Diemo (2003b, September). The Caterpillar System for Data-Driven Concatenative Sound Synthesis. In *Digital Audio Effects (DAFx)*, London, UK.

Schwarz, Diemo and Nicola Orio (2002, December). Project report score following. Technical report, Ircam, Paris, France.

Selfridge-Field, Eleanor (Ed.) (1997). *Beyond Midi: The Handbook of Musical Codes*. Cambridge, Massachusetts, USA: MIT Press.

Shalev-Shwartz, Shai, Shlomo Dubnov, Nir Friedman, and Yoram Singer (2002). Robust temporal and spectral modeling for query by melody. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 331–338. ACM Press.

Soulez, Ferréol, Xavier Rodet, and Diemo Schwarz (2003, October). Improving Polyphonic and Poly-Instrumental Music to Score Alignment. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, Baltimore, Maryland, USA.

TCTS (2000, September). TCTS (Circuit Theory and Signal Processing) Lab, Faculté Polytechnique de Mons. WWW page. www (?? 2000).

Turetsky, Robert (2003). MIDIAlign: You did *what* with MIDI? Web page `http://www.ee.columbia.edu/~rob/midialign`

Turetsky, Robert J. and Daniel P.W. Ellis (2003, October). Ground-Truth Transcriptions of Real Music from Force-Aligned MIDI Syntheses. In *Pro-*

ceedings of the International Symposium on Music Information Retrieval (ISMIR), Baltimore, Maryland, USA.

Vercoe, Barry (1984). The Synthetic Performer in the Context of Live Performance. In *Proceedings of the ICMC*, pp. 199–200.

Vercoe, Barry and Miller Puckette (1985). Synthetic Rehearsal: Training the Synthetic Performer. In *Proceedings of the ICMC*, pp. 275–278.

Vinet, Hugues, Perfecto Herrera, and François Pachet (2002, September). The Cuidado Project: New Applications Based on Audio and Music Content Description. In *Proceedings of the International Computer Music Conference (ICMC)*, Gothenburg, Sweden, pp. 450–453.