New applications of content processing of music

Philippe Aigrain

European Commission[1] DGIS/E2

Rue de la Loi, 200

B-1049 Brussels, Belgium

Tel : +32.2.296.0365

Fax : +32.2.296.7018

Email : Philippe.Aigrain@cec.eu.int

Running heads: New applications of content processing

[1] Views presented in this paper are only the author's and do not necessarily represent the official position of the European Commission.

SUMMARY

Though digital processing of music has been a reality for many years, from studio production to home reproduction, from musical instruments to editing, we are still very far from having a clear picture of what role content processing of music is likely to play in tomorrow's music world. *Content processing* is meant as a general term covering feature extraction and modelling techniques[2] for enabling basic retrieval, interaction and creation functionality. The aim of using this term is to establish a direct analogy with similar concepts in image and video processing. This paper proposes to study how a variety of music-related activities could make use of some existing or quickly maturing content processing technologies. In particular, it focuses on new aspects of listening, interacting with music, finding and comparing music, performing it, editing it, exchanging it with others or selling it, teaching, analysing and criticising it.

INTRODUCTION

20 years of digital audio have made little difference to everyone's music world until quite recently. The benefits of better sound quality and denser distribution carriers were not different in nature to those brought by older revolutions in analogue recording. Digital music meant only digital bytes representing musical signals, which were then played back just like analogue signals. The main changes a sociologist of listening would have detected are the growing segmentation of listening enabled by easier direct access to tracks and segments, and the inability of half of the population to read red-on-black/small-font CD leaflets. Of course, in the musical studio, composers and engineers dealt with the synthesis of new sound forms, the computer representation of musical languages, new ways of interacting with music and controlling performance, and new transcription

systems for non-written music. But this was an isolated practice, developed in technology environments that could not be disseminated to general use. Similarly, ethnomusicologists, who were always pioneers in the use of technology to understand and transcribe music (Ellington, 1992), were using digital technologies to build or validate models (Arom, 1985), in an even more confidential environment. Two changes have profoundly modified this landscape, and now put content processing of music in a position to deliver new functionality. The first one is the general development of media technology. Now that the information infrastructure has matured, that processing power, low-level representation formats, large memory and writable storage devices are readily available, the technology developments can finally focus on what to do in the digital world. Many researchers have identified the medium level, dealing with features and models of time segments and sound objects that can be recognised, manipulated, created, presented and exchanged as the right layer to develop new approaches: see for instance (Hawley, 1993). The second change is the large-scale use of MIDI instruments, editors and sequencers has accustomed users (a "limited" group, but still in the range of millions) to the explicit manipulation of digital representations of musical objects. This paper claims that we are now witnessing the progressive birth of new musical media. These media can take very diverse forms. They are today no more than exploratory drafts that will be profoundly reshaped through their maturation, their cultural dissemination and their inscription in the economy. The approach proposed in this paper to analyse this emerging domain is to focus on musical activities, and to hint at how new functionality can be brought in these activities by use of content processing of music, in a way that could give users new capabilities. But before such scenarios can be described, it is necessary to identify what are the key enabling technologies that are available or at hand.

[2] Signal modelling, pattern recognition, sound production models, and perceptive or cognitive modelling.

There would be little possibility of meaningful breakthroughs in applications if investigations had not been conducted for the past 30 years in a number of scientific fields to develop the key concepts and technologies of content processing of music. As will appear in the short catalogue that follows, we are particularly indebted to ethnomusicologists, specialists of psycho-acoustics, electro-acoustical music composers and sound analysis/synthesis engineers. Many concepts and techniques have also been borrowed from speech processing and recognition research. In the following, we do not document the signal processing techniques linked to time-frequency analysis or auditory front-ends used to better approximate human perceptive input. One should also note that some techniques have been designed for real-time applications and other apply only to off-line processing, either because they are too computationally intensive, or most often because they have inherent delays. In the future, it is possible that sound recording and publishing formats will include in-stream coding of meaningful features (segments, associated scores, sound source descriptions, etc.) that will make them readily available for applications.

Pitch recognition

Unsurprisingly, considering the predominance of pitch in western music and its notation, pitch recognition is a particularly active field. Pitch recognition was already successfully tackled in the monophonic case with analogue devices such as Charles Seeger's *Melograph* (Ellington, 1992) and has been brought to new degrees of precision in the digital domain by work such as Judith Brown's (Brown, 1992). The extension of monophonic pitch detection to the polyphonic case has proven to be difficult. It is only recently that breakthroughs have occurred, using completely different techniques such as projection of spectra on virtual chords (Carreras et al. in this issue) or on virtual spectral models

corresponding to unambiguous perception of pitch (Lepain in this issue). We can now consider that multiple pitch recognition in polyphonic contexts (including orchestral) delivers results that approximate a piano reduction with errors limited to cases in which a naïve listener would also be faced with a difficult task. Some issues such as temporal segmentation associated with pitches have also seen interesting recent developments: see for instance (Rossignol et al.) in this issue.

Beats, rhythm and dynamics

After pitch, the next obvious choice is to look for the automatic identification of time/energy events such as attacks, beats, rhythmical structure and prosody or phrasing. Curiously, though it can be tackled to a certain extent with relatively simple techniques, it has received less attention. Beat tracking in real-time – more precisely tracking *tactus* – has received much attention and is today successfully achieved in reasonably complex rhythmical environments: see f. i. (Large, 1995) and work by Keith Martin and Eric Sheirer at the MIT Machine Listing Group. The recognition of full rhythmical or metrical structures is still a challenge. Approaches such as presented in (Cambouropoulos, 1997) deliver convincing results on scores, but it is still unclear whether similar results can be obtained from musical signals. Achieving automatic metrical indexing of a musical recording (indexing in bars for instance), even with some assistance of a human user, would be a major enabler for some applications (see below automatic score following). An interesting technique for constructing visual representations of dynamics and prosody (including accent and *rubato*) has been proposed in (Todd, 1994), and could be applied to rhythm analysis and segmentation.

Automatic score following

Originally addressed in the context of interactive performance, (Vercoe, 1984) (Dannenberg, 1984), automatic score following has been tackled mostly from the angle of matching pitches in a sound input to pitches indicated in a score. (Denain et al., 1997), have proposed techniques for matching music against a score even in

presence of performance errors or intentional deviations. They also uses pitches as the primary input, but by representing scores as tree structures (bringing together chords in one node), it differentiates better between essential structure and secondary events.

Timbre classification

Timbre received early attention, because timbre quality was the main flaw of musical sound synthesis. (Risset, 1991) and many experimental studies in psycho-acoustics summarised in (Handel, 1989) have set a good scientific basis on which one can try to develop automatic timbre classifiers. Eric Sheirer and Keith Martin have tackled this difficult problem

Temporal segmentation

Temporal segmentation, that is the breakdown of musical content into discrete (and possibly multiscale or overlapping) time segments, has always been a subject of interest for researchers in musical perception (see f. i. (Thoresen, 1985). (Lerdahl-Jackendoff, 1983) defined rules to model how listeners segment music, some of which are specific of tonal music, and others are general. These rules have been further elaborated for instance by (Deliège, 92), but they remain quite far from possible application to real musical signals. On the contrary, (Aigrain, 1995), (Aigrain et al., 1995), (Lepain, 1996) and (Rossignol et al. in this issue) have proposed techniques for achieving automatic temporal segmentation based on dynamics and prosody, or on recognised pitches. One should note that temporal segmentation is a key enabler of discrete manipulation of segments in user interfaces (see below).

Melody extraction and melody matching

Querying for a giving melodic *motif* has been seen as the Holy Grail of content-based access to music. (Kageyama et al., 1993) and (Ghias et al., 1995) have proposed techniques that work on simple cases. Assuming that polyphonic pitch detection is available, querying for melodies (for instance from an example or humming) calls for extracting some salient melody either by voice segregation, such as proposed by (Uitdenbogerd-Zobel, 1998), or by harmonic analysis, and for conducting approximated matching. The dominant matching model is contour-based rather than interval-based. Matching in the presence of ornamentation and other modifications to the searched motif, or matching using rules that fit some non-western music perception models are difficult challenges. Recent approaches try to tackle direct matching in polyphonic space.

Chords, harmony, and chord prediction

Assisted analysis of harmony for tonal music can be achieved when one starts with a decent pitch detection (or of course with a score), as illustrated by (Mouton-Pachet, 1995). The most difficult problem is the correct choice of time scale at which to conduct the analysis, which can easily be provided by the human user. In the context of jazz, and in particular of "trading fours", researchers have developed chord prediction systems. See (Pachet, 1998) and (Thom, 1995). Though these harmony analysis or prediction techniques deliver useful assistance or modelling, much work is still needed before they can consistently perform on wider sets of real music.

Gestural input and multi-parameter control

Musical performance on instruments calls for the continuous fine-tuned control of input parameters. Controlling multiple parameters in real-time using simple controls is another side of the complexity of musical interfaces. Though it is not content processing of music per se, gestural input is a key component of

interactive performance and interactive music installations. An extended account of the history of gesture interfaces and other sensors for music interfaces can be found in (Paradiso, 1998). (Camurri-Ferrentino, 1999) presents an example of integration of such interfaces in interactive multimedia installations. Seen from the angle of content processing of music, the key question related to gestural input is the choice of features and models to be mapped with parameters controlled by gestures.

Streams and voices

Stream segregation, that is the ability to separate in perception and track in time particular streams or voices in a musical input, is one of most impressive achievements of human listening. Understanding or modelling it has been a key research agenda for the auditory scene analysis community (McAdams-Bregman, 1979) (Bregman, 1990). Researchers in automatic analysis of musical signals have tried to work along these lines to achieve stream-based analysis in the time-frequency domain, using for instance correlation in the time evolution of features of sound objects associated in the same stream (Tanguiane,). This approach has proved to be difficult, because of computational complexity, but even more because the input analysed was too low-level to allow for identification of perception-level objects. When starting with analysed pitches or an existing score, simple strategies based on rules such as tracking highest pitches, or streaming based on pitch contiguity give interesting results on some types of musical contents. The debates on whether and at which step of a processing chain it makes sense to recognise discrete musical objects will not reach a conclusion soon: see for instance (Scheirer, 1998).

Spatialisation

Sound spatialisation techniques open new dimensions to listening. It is only when constraining the possibilities of sound spatialisation or placing them in the frame

of a given reference model than meaningful applications are found. (Pachet, 1998) presents a very interesting interactive listening system based on constrained control of a sound spatialisation system. The constraints express for instance the need to maintain some balance between sound sources. Sound spatialisation is also an enabling technology for sonic browsing (see below).

Multi-feature classification and retrieval

In sound databases, for instance databases of Foley sounds used in movie production, but also when navigating in a sound recording, for instance to find the next repetition of some motif, multi-feature classification and retrieval has a much greater potential than search on an isolated feature. For instance, searching simultaneously on rhythmical pattern and melody provides much better results that searching on each in isolation. (Blum et al., 1996) is a key reference on sound retrieval using multiple low-level features. When one moves to more complex music segments, it is obvious that much progress is still needed to achieve reliable retrieval, considering the variety of time scales and instantiations of features.

Visual representation

Visualising musical sound was explored even before recording techniques where available. A large part of music signal processing and some feature analysis were achieved with music visualisation in mind. Spectrography and other sonographical techniques were first developed for off-line production of a sound image illustrating low-level, small-time scope features of musical sound. (Cogan, 1984) showed that it could be extended to provide wider time-scope representations of musical contents. More recently, spectrography was made real-time, used for direct control of listening, and developed towards highlighting salient or meaningful components (Deutsch, 1999), or for direct control of sound processing (Audiosculpt, 1995). Work conducted at the GRM (Besson, 1991) uses a sonographical image as a background for user produced graphical transcription.

(Lepain, 1996) and (Aigrain et al. 1995) describe an interactive listening interface based on visualisation of higher level features and time segmentation with gravity of selection in each feature space.

Time stretching, pitch shifting and other processing effects

When content processing is used not only to analyse sound, but also to generate it (in applications such as editing, interactive music, performance or installations), sound features and models are used in synthesis or analysis/re-synthesis processes. Time stretching and pitch shifting are commonly used processing techniques, for which exist both time domain techniques (cheaper but with lower quality of results) and frequency domain (or combined domain) techniques. See for instance software developed by the company Muscle Fish[3]. (Serra-Smith, 1990) has proposed a method for the decomposition of musical signals into a stochastic component and a deterministic component, that enables a wide variety of effects. More generally, many analysis techniques try to produce sound representations that can be used in re-synthesis for effects (Rodet et al., 1995) or for perceptive validation of the analysis.

Sonic spaces

Browsing directly musical spaces is of course a very appealing idea. (Fernstrom-Bannon, 1997) has proposed an approach to sonic browsing with a clear generic potential, though its range of applicability remains to be explored. Sound documents (for instance musical recordings) are analysed and the features extracted are projected in 3-dimensional space. A document becomes then a point in this space, which is assumed to be absorbent for sound. The user can then browse through this space, using a sound spatialiser, hearing each document with intensity according to its distance. The quality of such browsing interfaces rests

[3] http://www.musclefish.com/

on the relevance of the extracted features, and the ability of the user to navigate meaningfully in the projected feature space.

THE RESHAPING OF SOME MUSIC ACTIVITIES

The surveys that follows does not include the most obvious applications of content processing of music for synthesis and new musical instruments, and for composition, that have been explored at depth in the last 30 years of musical research. It rather focuses on emerging applications of potentially wide usage, building on trends in other areas of musical activities (listening, interactive music, performance, teaching, etc.)

Listening to and interacting with music

One may think that listening is a very stable activity. Of course, "simple" listening to a musical flow that imposes its organisation of time on the listener and whose contents are independent of any action from the listener will remain a dominant and essential musical activity. It is nonetheless well known that beyond its passive appearance, listening is an active exploration of the sound surface. Furthermore, as soon as there is a technical mediation, listening is always associated with some form of control on what is listened to, if only through choosing and sequencing. A first dimension of the applications of content processing of music is to support and empower active listening.

Even without modifying the auditory contents, associating it to synchronised visual representations enables new dimensions of active listening. Lyrics, simplified scores adapted to real-time reading, graphical scores supporting listening to some particular aspect in a pedagogical perspective are bringing sound

visualisation from the spectrography lab or interactive listening research prototypes to wider usage. The dissemination potential of these visual interfaces is very wide, from music-on-demand sites to musical multimedia publishing, but also to the audio-visual hi-fi home environment. The main challenge is in the production of new graphical representations that effectively support listening.

Of course, it is when user actions actually control how or which music is being played back that we enter the realm of interactive listening. We have already mentioned how sound spatialisation, such as implemented by (Pachet, 1998) represents a first step in this direction. The ability to explore spatially a full chamber music or orchestral piece by moving closer to some instrumental parts can be seen as building upon the human perception ability to focus on particular streams in a sound input, illustrated in the *cocktail party effect*, and bringing it to new dimensions. A difficult challenge is how to design the user interface used to enable virtual displacement, which must not distract attention. It is unclear whether visual immersion in the virtual space is the right choice to enable the auditory immersion that is looked for.

Interactive listening per se occurs when the listener builds the auditory contents being listened to. As such it has no clear delineation from performance. But mixing provides a model of an intermediate situation between listening and performance, whose interest has been illustrated by some recent realisations. One example is the *Shift-Control* (Audiorom, 1998) interactive listening CD-ROM produced by the London-based Audiorom® company. In this product, which is derived from interactive music installations developed by the same group, 16 different visual interfaces enable music creation through mixing of pre-created elementary sounds. The quality of this realisation lies in the underlying models of musical pieces that enable naïve interactive listeners to produce what they will consider as meaningful musical contents. Some of the interfaces are very close to a mixing console metaphor, while others use physical sound production

metaphors (controllable moving objects bouncing against obstacles and each other) or language-based metaphors (playing a text on a musical alphabet). The key question here is whether this or other similar achievements are one-time feats, or whether they can be brought to the status of constructive art forms by developing explicit models of musical components and their meaningful associations.

Music performance

The scope of this paper does not allow for an account of how gestural input and force feedback techniques can be used, together with physical models of sound production, to build new musical instruments. But interactive performance can also be based on interaction at other levels: either because gestures or motion are used to control higher level parameters of a musical piece, or because interaction occurs between a human performer and a pre-written computer piece, or between several human performers.

The first model has often been used in interactive installations, for instance by capturing motion of a dancer, or of visitors, and using them to control sound synthesis or composition level parameters. Camurri (Camurri-Ferrentino, 1999) has developed systems that analyse motion/gesture input to extract intention-level features or features correlated with emotions and map in music or sound production at this level. The second model (interaction between a performer and a pre-written piece) was explored by Boulez, but could clearly be brought to new levels by the availability of computing power and new recognition techniques. The third model (mediated interaction between performers) has been explored by Eliëns (Eliens et al., 1997) who developed his *Jamming on the Web* system for distributed jam sessions using an networked MIDI server. While it is unclear that distributed jamming can really develop to be a common practice, it is a very interest paradigm for exploring the requirements of shared music spaces. Due to

network delays, a distributed jamming system must not only be reactive, but also predictive. Using representations of music features that are richer than possible in a MIDI environment would enable new levels of interaction.

Manipulating musical contents

The back office functions of music publishing are primary candidates for taking great benefits of content processing music technologies. Curiously, most sound editing software are still using pure waveform or spectrographical representations of musical sound, incorporating some real-time signal processing abilities such as edit during playback[4]. They do not exploit fully the possibility of recognising higher level musical features, and use them to enable or assist the manipulation. Segmentation and the related direct manipulation functionality (Lepain, 1996) bring immediate benefits. It is in the domain of sound restoration[5] and sound effects that content processing technologies have been put to wider use. The sound effects research community (DAF'X, 1998) has been particularly active in developing open libraries of software that represent a key step towards generic usage.

Finding, exchanging and selling music

Content-based indexing and retrieval is the application of content processing of music that has received most attention. It has been envisaged within the frame of large musical repositories such as those of libraries or music-on-demand services as a classical information retrieval problem. This type of approach can be found for instance in the Music Library of the Future project (Pennycook, 1997) or at IRCAM Médiathèque (IRCAM, 1999). But also, in smaller repositories such as those of a Karaoke system (Kageyama, 1993), in which retrieval must occur almost real-time to enable the start and synchronisation of the musical

[4] See for instance Digidesign® products: http://www.digidesign.com/
[5] See for instance CEDAR® products: http://www.cedar-audio.com/ and their integration as plug-ins to sound editing software.

accompaniment to the singer. Content-based navigation within different performances of the same musical piece or even within a single musical recording (for navigation from one extract to a similar one) is a variation on the content-based retrieval paradigm.

As with content-based retrieval of images, a technology push approach has been dominant which has underestimated the difficulty of large-scale retrieval, and has neglected the importance of non-query oriented features (content-based visualisation, summarisation and auditory browsing) and of user interfaces for assistance to query formulation. Furthermore, researchers have sometimes aimed at the automatic identification of features (such as genres) that can be so easily produced as descriptive metadata, that it is not really worth building a whole automatic system to rediscover them. But despite these initial signs of immaturity, content-based retrieval and navigation have real usefulness. The potential for libraries, is obvious. In the case of music-on-demand services, for usage paradigms such as "find what you have just listened to but missed the title", or "find similar music that you might like", the ability of content-based retrieval to deliver better than other schemes still has to be proved. But other aspects of content-based navigation, such as for comparison between extracts of different performances of the same musical piece[6], may prove to deliver an essential added value to a service.

One key step towards making content-based retrieval of music deliver real benefits, is the incorporation of temporal indexing, and the development of content-description and metadata standards, that can make it possible to use metadata information (for instance a score, or meter and bars) and map it to sound contents. Once again, the experience with content-based access to video has illustrated the importance of structural information to enable feature-based access.

[6] Not necessarily restricted to classical music, also applies to pop songs, traditional music, etc.

Music teaching and critic

The human and economical importance of instrument tuition, music teaching and music critic (by professionals as well as by simple listeners) is often overlooked. In this domain, content processing of music and the assistance it can give to human perception can support new dimensions and new activities.

Computer-assisted instrument tuition based on content processing has attracted special attention, in particular for support to practice by the pupil between human-delivered courses. It has been explored in work by Stephen Smoliar for keyboards, and in the ESPRIT European project MUSTUTOR for non-keyboards instruments at beginner level (Mustutor, 1997). When trying to develop instrument tuition systems that analyse the pupil production and recommend adequate modification or practice, several difficult issues must be addressed. One has to achieve feature extraction from poor music production in noisy sound environments, to be able to match it to exercises and to correctly identify causes of difficulties for deriving sound pedagogical advice. Furthermore, the user interaction design is no less challenging, since a classical PC interaction environment is obviously inadequate.

Regarding analysis and comparison of performances, one can find in (Aigrain-Lepain, 1996) an account of experimental usage of content processing-based interactive listening by music teachers, musicologists, performers and critics, in the frame of a working group created by the *Bibliothèque Nationale de France*. It is clear that content processing can extend the range of human perception and the scientific basis of affirmations regarding music. This calls for technical functionality for producing illustrative examples, including musical extracts, and for adequate provision in intellectual property regulation, management systems and practice. There cannot be proof without the possibility to present evidence.

Music analysis

As already mentioned, ethnomusicologists have been pioneers of content processing techniques. Regarding music transcription, they are faced with a need common to all analysts of non-written music, or music written in non-musical notation such as programming languages. Using feature analysis and associated content processing (navigation from time segment to time segment, time stretching, pitch shifting, re-synthesis from models), one can today give a new impulse to assisted transcription. Assisted transcription and analysis call also for annotation functionality which has only be very partially explored, one interesting exception being the already mentioned *Acousmographe* developed at GRM (Besson, 1991). The ability to import and export feature representations from one representation space to another, for instance from a feature-based signal representation to a score-like representation, while keeping the essential underlying information, are key to efficient annotation. A difficulty when developing annotation systems is that meaningful annotations are attached to contents of very different time-scale and have a non-linear structure: see for instance the example of structural representation metadata proposed by Lindsay and Kriechbaum in this issue.

CONCLUSION

This paper has tried to give a short overview of the technology pool and of some promising applications of content processing of music outside of music synthesis and composition. It is probable that the real applications will turn out to be different of what we can now imagine. But the search will not be in vain.

ACKNOWLEDGEMENTS

The author is very grateful to the anonymous referees for suggesting important references, and for their useful comments on drafting.

REFERENCES

Aigrain, Ph. (1995). Segmentation de documents musicaux en vue de leur écoute et de leur analyse. In *5 Œuvres face à leur public*. Marseille: MIM.

Aigrain, Ph., Joly, Ph., Lepain, Ph. & Longueville V. (1995). Representation-based user interfaces for the audio-visual library of year 2000. *In Proceedings of the Multimedia Computing and Networking Conference* (pp. 35-45). San Jose: SPIE proceedings 2417.

Aigrain, Ph., & Lepain, Ph. (1996). Le groupe Ecoute Interactive de la Musique de la Bibliothèque Nationale de France. *In Actes des Journées d'Informatique Musicale* (pp. 128-138). Caen: GREYC, Mai 1996.

Arom, S. (1985). De l'écoute à l'analyse des musiques centro-africaines. *Analyse Musicale*, 1, 35-39.

Audiorom (1998). http://www.audiorom.com

Audiosculpt (1995). Software developed at IRCAM. See http://www.ircam.fr/

Besson, D. (1991). La transcription des musiques électro-acoustiques: que noter, comment et pourquoi. *Analyse Musicale*, 24, 37-41.

Blum, T, Keisler, D., Wheaton J., & Wold, E. (1996). Audio databases with content-based retrieval. IEEE Multimedia 3(3), 27-36.

Bregman, A. (1990). *Auditory Scene Analysis*. Cambridge, Massachussets: MIT Press.

Brown, J. (1992). Musical fundamental frequency analysis tracking using a pattern recognition method. *Journal of the Acoustical Society of America*, 89(1), 425-434.

Cambouropoulos, E. (1997). Musical rhythm: A formal model for determining local boudaries, accents and metre in a melodic surface. In M. Leman (ed.), *Music, Gestalt and Computing*. Berlin: Springer Lecture Notes in Computer Science vol. 1317.

Camurri, A. & Ferrentino, P. (1999). Interactive environments for music and multimedia.

Cogan, R. (1984). New images of musical sounds. Cambridge: Harvard University Press.

DAF'X 98: First COST-G6 Workshop on Digital Audio Effects (DAFX98), November 19-21, 1998, Barcelona, Spain http://www.iua.upf.es/dafx98/

Deliège, I. (1992). Perception et analyse de l'œuvre musicale: point de rencontre. *Analyse Musicale*, 26, 7-13.

Dannenberg, R. (1984). An On-Line Algorithm for Real-Time Accompaniment. In Proceedings of the 1984 International Computer Music Conference. Snas Francisco, ICMA, pp. 193-198.

Desain, P.,Honing, H., & H. Heijink (1997). Robust Score-Performance Matching: Taking Advantage of Structural Information. In Proceedings of the 1997 International Computer Music Conference. San Francisco: ICMA.

Deutsch, W. (1998). Visualisaiton of Musical Signals. Presented at the ACM MM'98 workshop on Content Processing of Music for Multimedia Applications. http://www.kfs.oeaw.ac.at/PSA/bristol0998/index.htm

Eliëns, A., van Welie, M., van Ossenbruggen, J. & Schönhage B., (1997). Jamming on the Web. In *Proceedings of the 6th International World-Wide Web Conference*, http://decweb.ethz.ch/WWW6/Technical/Paper038/Paper38.html

Ellington, T. (1992). Transcription. In H. Myers, (ed.), *Ethno-Musicology: An introduction* (pp. 110-152). London: W.W. Norton.

Fernstrom, M., & Bannon, L. (1997). Explorations in sonic browsing. Proceedings of HCI'97. Bristol.
http://www.ul.ie/~idc/library/papersreports/MikaelFernstrom/hciuk97/sonicbrowse.html.

Ghias, A., Logan J., Chamberlain D., & Smith, B.C. (1995). Query by humming – music information retrieval in an audio database. In *Proceedings of the 2nd ACM International Multimedia Conference*. San Francisco: ACM.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge: MIT Press.

Hawley, M. (1993). *Structure out of sound*. Ph. D. dissertation. Cambridge: MIT Media Lab.

IRCAM (1999). http://mediatheque.ircam.fr/

Kageyama, T. Mochisuki, & K. Takashima, Y. (1993). Melody retrieval by humming. In *Proceedings of the International Computer Music Conference 1993* (pp. 349-351). San Francisco: International Computer Music Association.

Large, E. W. (1995). Beat tracking with a non-linear oscillator. In Working Notes of the IJCAI'95 Workshop on AI and Music, 24-31.

Lepain, Ph.(1996). SATIE: An interactive software for listening to musical recordings. In *Proceedings of the 4th ACM International Multimedia Conference* (pp. 413-414). Boston: ACM.

Lepain, Ph. 1998). Ecouté interactive de documents musicaux numériques. In M. Chemillier & F. Pachet (eds.), Recherches et applications en informatique musicale. Paris: Hermès, 1998.

Lerdahl, F. & Jackendoff, R., (1983). A generative theory of tonal music, Cambridge: MIT Press.

McAdams, S., Bregman, A. (1979). Hearing Musical Streams. *Computer Music Journal* 3(4):26-44.

Mouton, R. Pachet, F. (1995). *Numeric vs. Synbolic controversy in automatic analysis of tonal music*. Working Notes of the IJCAI'95 Workshop on AI and Music. Montréal. http://www.poleia.lip6.fr/~fdp/papers.html

Mustutor, (1997). ESPRIT project 24909 MUSTUTOR. http://www.ilsp.gr/mustutor/MusTutor.htm

Pachet. F. (1998), Computer analysis of jazz chord sequences. Is Solar a Blues? In *Readings in Music and Artifical Intelligence*. Harwood Academic Publishers, 1998.

Pachet, F. & Delerue, O. (1998). MidiSpace: A constraint-based music spatializer. In *Proceedings of the 6th ACM International Multimedia Conference* (pp. 351-360). Bristol: ACM.

Paradiso, J.A. (1998). *Electronic Music Interfaces*. http://www.media.mit.edu/~joep/SpectrumWeb/SpectrumX.html/. Extended version of a paper from IEEE Spectrum.

Pennycook, B. (1997). The Music Library of the Future: A Pilot Project. http://www.music.mcgill.ca/newHome/mlfProject/html/mlfProposal.html

Risset, J-C. (1991). Timbre analysis by synthesis: representations, imitations and variants for musical composition". In G. Poli (ed.), *Representations of musical signals* (pp. 7-43). Cambridge, Massachussets: The MIT Press.

Rodet, X., Depalle, Ph., & Garcia, G. (1995). New possibilities in sound analysis/synthesis, Proceedings of the ISMA'9' Conference. Dourdan. On-line version accessible at the IRCAM Médiathèque server. http://www.ircam.fr/

Scheirer, E.D. (1998). *Music Perception Systems*. Ph. D. Dissertation. Cambridge, Massachussets: MIT Media Lab.

Serra, X., & Smith, J. (1990). Spectral modelling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal* 14(4), 12-24.

Thom, B. (1995). Predicting chords in jazz: The good, the bad and the ugly. . *Working Notes of the IJCAI'95 Workshop on AI and Music*. Montréal. http://www.cs.cmu.edu/afs/cs.cmu.edu/misc/mosaic/common/omega/Web/people/bthom/

Thoresen, L. (1985). Un modèle d'analyse auditive, *Analyse musicale* 1, 44-60.

Todd, N.P.M. (1994). The auditory "primal sketch": a multiscale model of rhythmic grouping". *Journal of New Music Research*, 23, 25-70.

Uitdenbogerd, A. & Zobel J. (1998). Manipulation of music for melody matching. In *Proceedings of the 6th ACM International Multimedia Conference* (pp. 235-240). Bristol: ACM.

Vercoe, B. (1984). The Synthetic Performer in the Context of Live Performance. In Proceedings of the 1984 International Computer Music Conference. Snas Francisco, ICMA, pp. 199-200.