Agència
de Gestió d'Ajuts
Universitaris
i de Recerca

# Memòria justificativa de recerca de les convocatòries BE, PIV, BCC, NANOS i BP

**La memòria justificativa consta de les dues parts que venen a continuació:**
**1.- Dades bàsiques i resums**
**2.- Memòria del treball (informe científic)**

Tots els camps són obligatoris

## 1.- Dades bàsiques i resums

**Nom de la convocatòria**

**BE**

**Llegenda per a les convocatòries:**

| | |
|---|---|
| BCC | Convocatòria de beques per a joves membres de comunitats catalanes a l'exterior (BCC) |
| BE | Beques per a estades per a la recerca fora de Catalunya (BE) |
| BP | Convocatòria d'ajuts postdoctorals dins del programa Beatriu de Pinós (BP) |
| CTP-AIRE | Ajuts per accions de cooperació en el marc de la comunitat de treball dels Pirineus (CTP). Ajuts de mobilitat de personal investigador. |
| NANOS | Beques de recerca per a la formació en el camp de les nanotecnologies (NANOS) |
| PIV | Beques de recerca per a professors i investigadors visitants a Catalunya (PIV) |

**Títol del projecte:** ha de sintetitzar la temàtica científica del vostre document.
Processament de música basat en el seu contingut

**Dades de l'investigador**


| Nom | Cognoms |
|---|---|
| Rafael | Ramirez Melendez |

Correu electrònic
rramirez@iua.upf.edu

**Dades del centre d'origen**

Universitat Pompeu Fabra


**Número d'expedient**
2007 BE-100252

**Paraules clau:** cal que esmenteu cinc conceptes que defineixin el contingut de la vostra memòria.
Processament de música, aprenentatge automàtic

**Data de presentació de la justificació**
01/06/2008

**Resum del projecte:** cal adjuntar dos resums del document, l'un en anglès i l'altre en la llengua del document, on s'esmenti la durada de l'acció

**Resum en la llengua del projecte** (màxim 300 paraules)
Donada l'explosió de la música en l'internet i la ràpida expansió de les col•leccions de música digital, un repte clau en l'àrea de la informació musical és el desenvolupament de sistemes de processament musical eficients i confiables.L'objectiu de la investigació proposada ha estat treballar en diferents aspectes de l'extracció, modelatge i processat del contingut musical. En particular, hem treballat en l'extracció, l'anàlisi i la manipulació de descriptors d'àudio de baix nivell, el modelatge de processos musicals, l'estudi i desenvolupament de tècniques d'aprenentatge automàtic per a processar àudio, i la identificació i extracció d'atributs musicals d'alt nivell. Hem revisat i millorat els nostres components d'anàlisis d'àudio i revisat components per a l'extracció de descriptors inter-nota i intra-nota en enregistraments monofónics d'àudio. Hem aplicat el nostre treball previ en Tempo a la formalització de diferents tasques musicals. Finalment, hem investigat el processat d'alt nivell de música basandonos en el seu contingut. Com exemple d'això, hem investigat com músics professionals expressen i comuniquen la seva interpretació del contingut musical i emocional de peces musicals, i hem usat aquesta informació per a identificar automàticament intèrprets. Estudiem desviacions en paràmetres com to, temps, amplitud i timbre a nivell inter-nota i intra-nota.

**Resum en anglès** (màxim 300 paraules)
A key challenge in the area of music information, given the explosion of online music and the rapidly expanding digital music collections, is the development of efficient and reliable music processing systems. The objective of the proposed research has been to work on different aspects of the extraction, modelling and processing of musical content. In particular, we worked on the extraction, analysis and manipulation of low-level audio descriptors, the modeling of musical processes, the study and development of machine learning techniques for processing audio, and the identification and computation of high-level semantic musical attributes. We revised and improved our current audio analysis components and refined components for the extraction of both inter-note and intra-note descriptors in monophonic music audio recordings. We applied our previous work on Tempo to the formalization of different musical tasks, e.g. musical processes. Finally, we investigated high-level content-based music processing in the context of monophonic audio recordings. As an instance of this last point, we investigated how skilled musicians express and communicate their view of the musical and emotional content of musical pieces and how to use this information in order to automatically identify performers. We study deviations of parameters such as pitch, timing, amplitude and timbre both at an inter-note level and at an intra-note level.

**Resum en anglès** (màxim 300 paraules) – continuació -.

**2.- Memòria del treball** (informe científic sense limitació de paraules). Pot incloure altres fitxers de qualsevol mena, no més grans de 10 MB cadascun d'ells.

## 1. Introduction

A key challenge in the area of music information, given the explosion of online music and the rapidly expanding digital music collections, is the development of efficient and reliable music search and retrieval systems. One of the main deficiencies of current music search and retrieval systems is the gap between the simplicity of the content descriptors that can be currently extracted automatically and the semantic richness in music information. Conventional information retrieval has been mainly based on text, and the approaches to textual information retrieval have been transferred into music information retrieval. However, music contents and text contents are of a very different nature which very often makes textual information retrieval unsatisfactory in a musical context. It has been widely recognized that music retrieval techniques should incorporate high-level music information. Our research in this project focused on high-level content-based music processing in the context of monophonic audio recordings. In particular, we investigated how famous musicians express and communicate their view of the musical and emotional content of musical pieces and how to use this information in order to automatically identify performers. We study deviations of parameters such as pitch, timing, amplitude and timbre both at an inter-note level and at an intra-note level. The identification of performers by using the expressive content in their performances raises particularly interesting questions but has nevertheless received relatively little attention in the past.

The rest of the report is organized as follows: Section 2 sets the background for our research. Section 3 describes components for music content description. Section 4 describes our research on programming language design applied to the especification of musical processes and the manipulation of audio analysis components. Section 5 describes our research on performance-driven performer identification, and finally, Section 6 presents some conclusions.

## 2. Background

Music performance plays a central role in our musical culture today. Concert attendance and recording sales often reflect people's preferences for particular performers. The manipulation of sound properties such as pitch, timing, amplitude and timbre by different performers is clearly distinguishable by the listeners. Expressive music performance studies the manipulation of these sound properties in an attempt to understand expression in performances. Understanding and formalizing expressive music performance is an extremely challenging problem which in the past has been studied from different perspectives, e.g. (Seashore, 1936), (Gabrielsson, 1999), (Bresin, 2002).

This report describes a machine learning approach to investigate how skilled musicians (Jazz saxophone players in particular) express and communicate their view of the musical and emotional content of musical pieces and how to use this information in order to automatically distinguish among performers. We study deviations of parameters such as pitch, timing, amplitude and timbre both at an inter-note-level and at an intra-note-level. This is, we analyze the pitch, timing (onset and duration), amplitude (energy mean) and timbre of individual notes, as well as the timing and amplitude of individual intra-note events. We focus on saxophone performance where timing and pitch measurements present a greater challenge compared to the measurements in piano performances (this is due to the fact that in piano performances certain expressive resources, e.g. vibrato and glissando, are absent).

Our approach to performer identification is motivated by our pervious work (Ramirez, 2005b) on expressive music performance synthesis. In (Ramirez, 2005b) we consider a set of inflections (characterized by intra-note features) and use the note musical context (characterized by inter-note features) in order to predict the type of inflection to be used in that context. We use particular instances, i.e. audio samples, of the type of inflection predicted to synthesize expressive performances from inexpressive score descriptions. It is clear that by using a particular performer's samples the synthesized pieces 'sound' like played by that performer. Thus, it seems reasonable to apply the inverse process for performer identification.

Previous research addressing expressive music performance using machine learning techniques has included a number of approaches. Lopez de Mantaras et al (Lopez de Mantaras, 2002) report on SaxEx, a performance system capable of generating expressive solo saxophone performances in Jazz. Their system is based on case-based reasoning, a type of analogical reasoning where problems are solved by reusing the solutions of similar, previously solved problems. In order to generate expressive solo performances, the case-based reasoning system retrieves from a memory containing expressive interpretations, those notes that are *similar* to the input inexpressive notes. The case memory contains information about metrical strength, note duration, and so on, and uses this information to retrieve the appropriate notes. One limitation of their system is that it is incapable of explaining the predictions it makes and it is unable to handle melody alterations, e.g. ornamentations.

Ramirez et al (Ramirez, 2006) have explored and compared diverse machine learning methods for obtaining expressive music performance models for Jazz saxophone that are capable of both generating expressive performances and explaining the expressive transformations they produce. They propose an expressive performance system based on inductive logic programming which induces a set of first order logic rules that capture expressive transformation both at an inter-note level (e.g. note duration, loudness) and at an intra-note level (e.g. note attack, sustain). Based on the theory generated by the set of rules, they implemented a melody synthesis component which generates expressive monophonic output (MIDI or audio) from inexpressive melody MIDI descriptions.

With the exception of the work by Lopez de Mantaras et al and Ramirez et al, most of the research in expressive performance using machine learning techniques has focused on classical piano music where often the tempo of the performed pieces is not constant. The works focused on classical piano have focused on *global* tempo and loudness

transformations while we are interested in both *intra-note* and *inter-note* level tempo and loudness transformations.

Widmer (Widmer, 2001) (Widmer, 2002) reported on the task of discovering general rules of expressive classical piano performance from real performance data via inductive machine learning. The performance data used for the study are MIDI recordings of 13 piano sonatas by W.A. Mozart performed by a skilled pianist. In addition to these data, the music score was also coded. The resulting substantial data consists of information about the nominal note onsets, duration, metrical information and annotations. When trained on the data an inductive rule learning algorithm discovered a small set of quite simple classification rules that predict a large number of the note-level choices of the pianist.

Nevertheless, the use of expressive performance models, either automatically induced or manually generated, for identifying musicians has received little attention in the past. This is mainly due to two factors: (a) the high complexity of the feature extraction process that is required to characterize expressive performance, and (b) the question of how to use the information provided by an expressive performance model for the task of performance-based performer identification. To the best of our knowledge, the only group working on performance-based automatic performer identification is the group led by Gerhard Widmer. Saunders et al (Saunders, 2004) apply string kernels to the problem of recognizing famous pianists from their playing style. The characteristics of performers playing the same piece are obtained from changes in beat-level tempo and beat-level loudness. From such characteristics, general performance alphabets can be derived, and pianists' performances can then be represented as strings. They apply both kernel partial least squares and Support Vector Machines to this data.

Stamatatos and Widmer (Stamatatos, 2005) address the problem of identifying the most likely music performer, given a set of performances of the same piece by a number of skilled candidate pianists. They propose a set of very simple features for representing stylistic characteristics of a music performer that relate to a kind of 'average' performance. A database of piano performances of 22 pianists playing two pieces by Frédéric Chopin is used. They propose an ensemble of simple classifiers derived by both subsampling the training set and subsampling the input features. Experiments show that the proposed features are able to quantify the differences between music performers.

## 3. Components for music content description

The objective here is to revise and improve our current audio analisis components. In this respect, we developed components for the extraction of both inter-note and intra-note descriptors in monophonic music audio recordings. The set of intra-note features includes descriptors such as the note's attack level, sustain duration, sustain slope, amount of legato with the previous note, amount of legato with the following note, mean energy, spectral centroid and spectral tilt. The set of inter-note features includes the relative pitch and duration of the neighboring notes (i.e. previous and following notes) as well as the musical structures (obtain by a musical analysis) to which the note belongs. Our interest in obtaining descriptors both at the intra and inter-note level is that as an ensemble they can provide a high level semantic description of music content. The intra-note features represent the internal structure of performed notes in a musical piece, while the inter-note features representing information about the music context in which musical events occur.

### 3.1 Extraction of Inter-note Features

We perform a spectral analysis of a portion of sound whose size is a parameter of the algorithm. This spectral analysis consists of multiplying the audio frame with an appropriate analysis window and performing a Discrete Fourier Transform (DFT) to obtain its spectrum. In this case, we use a frame width of 46 ms, an overlap factor of 50%, and a Keiser-Bessel

25dB window. Then, we compute a set of low-level descriptors for each spectrum: energy and an estimation of the fundamental frequency. From these low-level descriptors we perform a note segmentation procedure. Once the note boundaries are known, the note descriptors are computed from the low-level values.

*Energy computation.* The energy descriptor is computed on the spectral domain, using the values of the amplitude spectrum at each analysis frame. In addition, energy is computed in different frequency bands as defined in (Klapuri, 1999), and these values are used by the algorithm for note segmentation.

*Fundamental frequency estimation.* For the estimation of the instantaneous fundamental frequency we use a harmonic matching model derived from the Two-Way Mismatch procedure (TWM) (Maher, 1994). For each fundamental frequency candidate, mismatches between the harmonics generated and the measured partials frequencies are averaged over a fixed subset of the available partials. A weighting scheme is used to make the procedure robust to the presence of noise or absence of certain partials in the spectral data. The solution presented in (Maher, 1994) employs two mismatch error calculations. The first one is based on the frequency difference between each partial in the measured sequence and its nearest neighbor in the predicted sequence. The second is based on the mismatch between each harmonic in the predicted sequence and its nearest partial neighbor in the measured sequence. This two-way mismatch helps to avoid octave errors by applying a penalty for partials that are present in the measured data but are not predicted, and also for partials whose presence is predicted but which do not actually appear in the measured sequence. The TWM mismatch procedure has also the benefit that the effect of any spurious components or partial missing from the measurement can be counteracted by the presence of uncorrupted partials in the same frame. We perform a spectral analysis of all the windowed frames. Once this is done, the prominent spectral peaks of the spectrum are detected from the spectrum magnitude. These spectral peaks of the spectrum are defined as the local maxima of the spectrum which magnitude is greater than a threshold. The spectral peaks are compared to a harmonic series and a two-way mismatch (TWM) error is computed for each fundamental frequency candidates. The candidate with the minimum error is chosen to be the fundamental frequency estimate.

*Note descriptors.* We compute note descriptors using the note boundaries and the low-level descriptors values. The low-level descriptors associated to a note segment are computed by averaging the frame values within this note segment. Pitch histograms have been used to compute the pitch note and the fundamental frequency that represents each note segment, as found in (McNab, 1996).

*Musical Analysis.* After having computed the note descriptors as above, and in order to provide an abstract musical structure for monophonic recordings, we decided to use Narmour's theory of perception and cognition of melodies (Narmour 1990), (Narmour, 1991) to analyse the audio recordings. The Implication/Realization model proposed by Narmour is a theory of perception and cognition of melodies. The theory states that a melodic musical line continuously causes listeners to generate expectations of how the melody should continue. The nature of these expectations in an individual is based on intervallic and direction characteristics of musical fragments.

### 3.2 Extraction of Intra-note Features

The intra-note segmentation method is based on the study of the energy envelope contour of the note. Once onsets and offsets are located, we study the instantaneous energy values of the analysis frames corresponding to each note. This study is carried out by analyzing the envelope curvature and characterizing its shape, in order to estimate the limits of the intra-note segments. For each note in a monophonic audio recording, we identify the segments attack, sustain and release (Bernstein, 1976). In order to extract these three characteristic segments, we study the smoothed derivatives in a similar way that presented in (Jenssen, 1999), where partial amplitude envelopes are modeled for isolated sounds. The main difference is that we analyze the notes in their musical context, rather than isolated. In

addition, only three linear segments are considered. Moreover, instead of studying the contribution of all the partials, we obtain general intensity information from the total energy envelope characteristic.

*Intra-note segment characterization.* Once we have found the intra-note segment limits, we describe each one by its duration (absolute and relative to note duration), start and end times, initial and final energy values (absolute and relative to note maximum) and slope. For the stable part of each note (sustain segment), we extract an averaged spectral centroid and spectral tilt in order to have timbral descriptors related to the brightness of a particular execution.

**4. Programming/specification language for musical processes**

We applied our previous work on Tempo (Ramirez, 2000, 2005), a logic-based concurrent language, to the formalization of different musical tasks, e.g. musical processes. Many researchers, e.g. (Kowalski 1986, Pratt 1986), have proposed methods for reasoning about temporal phenomena in other contexts using partially ordered sets of events. Our approach to modeling musical processes is based on the same general idea. The basic idea here is to use a constraint logic program to represent the (usually infinite) set of constraints of interest. The constraints themselves are of the form X<Y, read as ``X precedes Y'', where X and Y are either events or positive integers, and < is a partial order. We consider a musical system as an instance of a reactive concurrent system which, in general, is not required to terminate. The constraint logic program as defined above has a procedural interpretation that allows a correct specification to be executed in the sense that processes run only as permitted by the constraints represented by the program. This procedural interpretation is based on an incremental execution of the program and a lazy generation of the corresponding partial orders.

Processes interact with each other by performing simple operations on a shared constraint store. Conceptually, processes can be distributed in several machines or executed concurrently in one machine.

*Program verification.* Event annotations naturally divide musical processes in a (often very small) number of states. Transitions among these states are explicitly determined by the system constraints. Taking into account these constraints it is quite simple to translate a musical system into a transition system. It is then straightforward to further translate the transition system of the musical system into a description in a model checker description language and directly apply the model checker to verify program properties. We have applied the SMV model checking system (McMillan, 1993) to verify properties of our programs. We have implemented a system which automatically translates an extended Java program (i.e. a Java program synchronizing using Tempo constraints) into a model M in SMV's description language. Thus, it suffices to code the property we want to verify using the specification language of SMV resulting in a CTL (computation tree logic) formula p, and run SMV with inputs M and p.

*Musical applications.* We applied our proposed language to the specification and execution of musical process networks. With this aim, we extended the language to allow language events trigger musical events and processes. We considered systems in which a number of musical processes interact with each other respecting a number of constraints in order to produce a musical output. Also, we considered networks of audio processing modules which are linked to each other and must respect a set of temporal constraints (e.g. some processing tasks have to be done before others) by assigning language events to processing modules and specify the execution order of the system as a set of Tempo constraints over these events.

**5. High-level content-based music processing: Performer identification**

As an example of high-level content-based music processing we investigated the identification of famous musicians by their playing styles. In particular, we investigated the feasibility of recognizing famous jazz saxophonists by their expressive playing style using the descriptors extracted in Section 3.

Before attempting to build a system to automatically identify a musician by his or her playing style we asked ourselves how is this task performed by a music expert? In the case of Jazz saxophonists our hypothesis is that most of the cues for performer identification come from the timbre or 'quality' of the notes performed by the saxophonist. That is to say, while timing information is certainly important and is useful to identify a particular musician most of the information relevant for identifying a performer is the timbre characteristics of the performed notes. In this respect, the saxophone is similar to the singing voice in which most of the information relevant for identifying a singer is simply his or her voice's timbre. Thus, the algorithm we designed aimed at detecting patterns of notes based on their timbre content. Roughly, the algorithm consists of generating a performance alphabet by clustering similar (in terms of timbre) individual notes, inducing for each performer a classifier which maps a note and its musical context to a symbol in the performance alphabet (i.e. a cluster), and given an audio fragment identify the performer as the one whose classifier predicts best the performed fragment.

The data used were monophonic audio commercial recordings of improvisations performed by four famous Jazz saxophonists: Billie Pierce (*Aria's Prance, Chelsea Bridge, In Your Own Sweet Way*), Joe Henderson (*Lush Life, Modinha*), Branford Marsalis (*St Thomas*) and Kenny Garrett (*Last Sax*). In order to have a similar number of instances (notes and musical phrases) we selected a subset of the recordings.

We segmented all the recorded pieces into audio segments representing musical phrases. Given an audio fragment and a set of possible performers, we use a classifier to identify the performer: for each note in the melody fragment the classifier computes the set of its intra-note features, the set of its inter-note features and, based on the note's intra-note features, the cluster membership of the note for each of the clusters. Once this is done, for each performer a classifier predicts a cluster representing the expected type of note the performer would have played in that musical context. This prediction is based on the note's inter-note features. The score for each performer is updated by taking into account the cluster membership of the predicted cluster (i.e. the greater the cluster membership of the predicted cluster, the more the score of the performer is increased). Finally, the performer with the higher score is returned.

**Results**

In average, there were a total of 820 notes available for each performer. We segmented each of the performed pieces in phases and obtain an average of 130 short phrases and 17 long phrases for each performer. The length of the short phrases and long phrases ranged from 4 to 10 notes and 30 to 60 notes, respectively. The expected classification accuracy of the default classifier (one which chooses randomly one of the four performers) is 25% (measured in correctly classified instances percentage). In the short phrase case, the average accuracy and the accuracy obtained for the most successful trained classifier was 71.95% and 74.76%, respectively. In the long phrase case, the average accuracy and the accuracy obtained for the most successful trained classifier was 71.19% and 74.92%, respectively. The correctly classified instances percentage for each learning method is presented in Table 1. The results for short and long phrases are statistically significant which indicates that the considered intra-note and inter-note features are indeed useful for training successful classifiers to identify performers from their playing style.

| | 1-note | Short-phrase | Long-phrase |
|---|---|---|---|
| Decision Trees | 23.65 | 68.34 | 67.44 |
| Support Vector Machines | 27.64 | 74.76 | 74.92 |
| Artificial Neural Networks | 25.48 | 72.65 | 71.90 |
| k-Nearest Neighbor | 24.32 | 71.06 | 72.23 |
| Bagging (decision trees) | 26.46 | 70.44 | 66.37 |
| Boosting (decision trees) | 24.87 | 71.21 | 67.84 |
| Voting (decision trees, SVM, ANN, 1-NN) | 28.31 | 70.84 | 72.47 |
| Stacking (decision trees, SVM, ANN, 1-NN) | 29.03 | 76.36 | 76.39 |

Table 1: Classification accuracy for the 1-note, short-phrase and long-phrase cases (in correctly classified instances percentage)

**Discussion**

The difference between the results obtained in the case studies and the accuracy of a baseline classifiers, i.e. the classifier guessing at random, indicates that the intra-note and inter-note features presented contain sufficient information to identify the studied set of performers, and that the machine learning methods explored are capable of learning performance patterns that distinguish these performers. It is worth noting that every learning algorithm investigated (decision trees, SVM, ANN, k-NN and the ensemble methods reported in Table 1) produced significantly better than random classification accuracies. This supports our statement about the feasibility of training successful classifiers for the case studies reported. However, note that this does not necessary imply that it is feasible to train classifiers for arbitrary performers.

We have selected three types of musical segment lengths: 1-note segments, short-phrase segments (4-12 notes), and long-phrase segment (30-62 notes). As expected, evaluation using 1-note segments results in poor classification accuracies, while short-phrase segments and long-phrase segment evaluation results in accuracies well above the accuracy of a baseline classifier. Interestingly, there is no substantial difference in the accuracies for short-phrase sand long-phrase segment evaluation which seems to indicate that in order to identify a particular performer it is sufficient to consider a short phrase segment of the piece, i.e. the identification accuracy does not increase substantially by considering a longer segment.

**6. Conclusions**

We worked on different aspects of the extraction, modelling and processing of musical content. We focused on the task of identifying famous performers from their playing style using note descriptors extracted from audio recordings. In particular, we concentrated in identifying Jazz saxophonists and explored and compared different machine learning techniques for this task. We characterized performances by representing each note in the performance by a set of intra-note features corresponding to the internal structure of the note, and a set of inter-note features representing the context in which the note appears. The results obtained indicate that the intra-note and inter-note features investigated contain sufficient information to identify the studied set of performers, and that the machine learning methods explored are capable of learning performance patterns that distinguish these performers.

**References**

(Berstein, 1975) Bernstein, A. D., Cooper E. D., "The piecewise-linear technique of electronic music synthesis". *J. Audio Eng. Soc.* Vol. 24, No. 6, July/August 1976.

(Bresin, 2002) Bresin, R. 2002. Articulation Rules for Automatic Music Performance. In Proceedings of the 2001 International Computer Music Conference. San Francisco, International Computer Music Association.

(Gabrielsson, 1999) Gabrielsson, A. (1999). The performance of Music. In D.Deutsch (Ed.), The Psychology of Music (2nd ed.) Academic Press.

(Jenssen, 1999) Jenssen, K., "Envelope model of isolated musical sounds". *Proceedings of COST G-6 Workshop on Digital Audio Effects (DAFx)*, Trondheim, 1999.

(Klapuri, 1999) Klapuri, A. (1999). Sound Onset Detection by Applying Psychoacoustic Knowledge, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP

(Kowalski, 1986) Kowalski, R.A. and Sergot, M.J. (1986). A logic-based calculus of events. New Generation Computing 4, pp. 67--95.

(Lopez de Mantaras, 2002) Lopez de Mantaras, R. and Arcos, J.L. (2002). AI and music, from composition to expressive performance, AI Magazine, 23-3.

(Maher, 1994) Maher, R.C. and Beauchamp, J.W. (1994). Fundamental frequency estimation of musical signals using a two-way mismatch procedure, Journal of the Acoustic Society of America, vol. 95 pp. 2254-2263.

(McMillan, 1993) McMillan, K.L. (1993). Symbolic Model Checking. Kluwer Academic Publishers.

(McNab, 1996) McNab, R.J., Smith Ll. A. and Witten I.H., (1996). Signal Processing for Melody Transcription,SIG working paper, vol. 95-22.

(Narmour, 1990) Narmour, E. (1990). The Analysis and Cognition of Basic Melodic Structures: The Implication Realization Model. University of Chicago Press.

(Narmour, 1991) Narmour, E. (1991). The Analysis and Cognition of Melodic Complexity: The Implication Realization Model. University of Chicago Press.

(Pratt, 1986) Pratt, V. (1986). {\it Modeling concurrency with partial orders}. International Journal of Parallel Programming 15, 1, pp. 33--71.

(Ramirez, 2000) Ramirez, R., Santosa, A.E., Yap, R. (2000). Concurrent programming made easy, IEEE International Conference on Engineering of Complex Computer Systems, IEEE Press.

(Ramirez, 2005) Ramirez, R. and Santosa, A.E. (2005). Formal Verification of Concurrent and Distributed Constraint-based Java Programs, IEEE International Conference on Engineering of Complex Computer Systems, IEEE Press.

(Ramirez, 2006) Ramirez, R., Hazan, A., Maestre, E., Serra, X. **"**A Data Mining Approach to Expressive Music Performance Modeling**"**, book chapter in *Multimedia Data mining and Knowledge Discovery*, Springer-Verlag

(Ramirez, 2005) Ramirez, R. and Hazan, A. (2005). Modeling Expressive Music Performance in Jazz, Proceedings of the 18th Florida Artificial Intelligence Research Society Conference (FLAIRS 2005), Clearwater Beach, Florida.

(Ramirez, 2005b) Ramirez, R., Hazan, A. (2005). A Learning Scheme for Generating Expressive Music Performances of Jazz Standards, Proceedings International Joint Conference on Artificial Intelligence, Edinburgh.

(Seashore, 1936) Seashore, C.E. (ed.) (1936). Objective Analysis of Music Performance. University of Iowa Press.

(Saunders, 2004) Saunders C., Hardoon D., Shawe-Taylor J., and Widmer G. (2004). Using String Kernels to Identify Famous Performers from their Playing Style, Proceedings of the 15th European Conference on Machine Learning (ECML'2004), Pisa, Italy, 2004.

(Stamatatos, 2005) Stamatatos, E. and Widmer, G. (2005). Automatic Identification of Music Performers with Learning Ensembles. *Artificial Intelligence* 165(1), 37-56.

(Widmer, 2002) Widmer, G. (2002). Machine Discoveries: A Few Simple, Robust Local Expression Principles. Journal of New Music Research 31(1), 37-50.

(Widmer, 2001) Widmer, G. (2001). Discovering Strong Principles of Expressive Music Performance with the PLCG Rule Learning Strategy. Proceedings of the 12th European Conference on Machine Learning (ECML'01), Freiburg, Germany. Berlin: Springer Verlag.