Expressive Performance in the Human Tenor Voice

Maria Cristina Marinescu*, Rafael Ramirez†
*IBM T.J. Watson Research Center/USA
†Universitat Pompeu Fabra/Barcelona, Spain

***Abstract*—This paper presents preliminary results on expressive performance in the human tenor voice. This work investigates how professional opera singers manipulate sound properties such as timing, amplitude, and pitch in order to produce expressive performances. We also consider the contribution of features of prosody in the artistic delivery of an operatic aria. Our approach is based on applying machine learning to extract patterns of expressive singing from performances by Josep Carreras. This is a step towards recognizing performers by their singing style, capturing some of the aspects which make two performances of the same piece sound different, and understanding whether there exists a correlation between the occurrences correctly covered by a pattern and specific emotional attributes.**

I. Introduction

One of the most interesting and elusive questions in music is what makes two expressive interpretations of the same musical piece sound like two different songs even when performed by the same singer. Given a set of expressive performances of the same piece which have different interpretation styles — and possibly different emotional attributes[1] — are the patterns learned from each performance similar or very different? How distinguishable is a singer based on the patterns extracted from his interpretations? What do the patterns that are similar for multiple singers capture? Which patterns are a matter of timber, which are based in specific expressive techniques that a singer employs, and which are a combination of the two — by choice or because the pattern is more readily realizable given the characteristics of a specific voice? Is there a correlation between the occurrences correctly covered by a pattern and specific emotional attributes associated with those music pieces? How do singers resolve possible conflicts between the music and the prosody of the lyrics?

This work investigates how professional opera singers manipulate sound properties such as timing, amplitude, and pitch in order to produce expressive performances of music fragments. In the initial phase we are interested in note-level manipulations; we therefore define a set of note-level descriptors of interest and we focus on the differences between their measured values in the actual performance and the written score, given the context of the surrounding notes. Previous approaches exists that are looking at expressive instrumental performances. There are a couple of important differences between instrumental music and voice in operatic music. First, one guitar may subtly differ from another one, but the timbre of the instrument is relatively fixed. Human voice displays a great variety in timbre; it is partly because of timbre that a voice is well-suited to a type of song but not to another, and it may be because of our preference in timbre that we prefer a singer over another. Distinguishing which features of expressiveness are the singer's interpretation choice and which ones are typical of his timbre is an issue that doesn't come up in instrumental music.

Secondly, instrumental music does not have lyrics. Lyrics convey a more specific meaning to a song than it would otherwise have. Therefore they can both add to and detract from the expressivity of a performance. Several aspects are at work here: how appropriate in meaning is the performance given the lyrics, and how to reconciliate possibly contradicting prosodic, metric, and score cues. For instance, adopting the wrong intonation or grouping the lyrics into the wrong prosodic units can ruin an otherwise good interpretation. In this work we are looking at a couple of preliminary descriptors for the lyrics which are syllable-specific: stress and syllable type.

Our approach is based on applying various machine learning (ML) techniques to extract patterns of expressive singing from different performances of the same, or different arias, sung by several world-class tenors. As a first step, we start with a test suite consisting of twelve interpretations of six different aria fragments performed by Josep Carreras. Using sound analysis techniques based on spectral models we extract high-level descriptors representing properties of each note, as well as of its context. A note is characterized by its pitch and duration. The context information for a given note consists of the relative pitch and duration of the neighbouring notes, as well as the Narmour structures to which the note belongs. In this work, our goal is to learn under which conditions a performer shortens or lengthens a note relative to what the score indicates, and when he sings a note louder or softer than what would be expected given the average energy level of the music fragment. Some of the most interesting rules that the ML algorithm learns are presented in the result section.

The rest of the paper is organized as follows. Section II describes related work in expressive performance. Section III describes our test suite, introduces the note-level descriptors, and explains how we extract the data that

[1]Emotional attributes are similar to what other researchers refer to as *moods* or *affective labels* and can simultaneously take one value for each aspect that they reflect.

is used as the input to the ML algorithms. Section IV presents the learning algorithms; Section V discusses some of the most interesting results. We conclude in Section VI.

II. Related Work

Understanding and formalizing expressive music performance is an extremely challenging problem which in the past has been studied from different perspectives, e.g. [16], [6], [3]. The main approaches to empirically studying expressive performance have been based on statistical analysis (e.g. [15]), mathematical modeling (e.g. [19]), and analysis-by-synthesis (e.g. [5]). In all these approaches, it is a person who is responsible for devising a theory or mathematical model which captures different aspects of musical expressive performance. The theory or model is later tested on real performance data in order to determine its accuracy. This paper describes a machine learning approach to investigate how opera singers express and communicate their view of the musical and emotional content of musical pieces.

Previous research addressing expressive music performance using machine learning techniques has included a number of approaches. Widmer [20] 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.

Tobudic et al. [18] describe a relational instance-based approach to the problem of learning to apply expressive tempo and dynamics variations to a piece of classical music, at different levels of the phrase hierarchy. The different phrases of a piece and the relations among them are represented in first-order logic. The description of the musical scores through predicates (e.g. *contains(ph1,ph2)*) provides the background knowledge. The training examples are encoded by another predicate whose arguments encode information about the way the phrase was played by the musician. Their learning algorithm recognizes similar phrases from the training set and applies their expressive patterns to a new piece.

Ramirez et al. [13], [14] explore and compare different machine learning techniques for inducing both, an *interpretable* expressive performance model (characterized by a set of rules) and a *generative* expressive performance model. Based on this, they describe a performance system capable of generating expressive monophonic Jazz performances and providing 'explanations' of the expressive transformations it performs. The work described in this chapter has similar objectives but by using a genetic algorithm it incorporates some desirable properties: (1) the induced model may be explored and analyzed while it is 'evolving', (2) it is possible to guide the evolution of the model in a natural way, and (3) by repeatedly executing the algorithm different models are obtained. In the context of expressive music performance modeling, these properties are very relevant.

Lopez de Mantaras et al. [8] report on SaxEx, a performance system capable of generating expressive solo 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. However, their system does not allow one to examine or understand the way it makes predictions.

Other inductive machine learning approaches to rule learning in music and musical analysis include [4], [1], [9] and [7]. In [4], Dovey analyzes piano performances of Rachmaninoff pieces using inductive logic programming and extracts rules underlying them. In [1], Van Baelen extended Dovey's work and attempted to discover regularities that could be used to generate MIDI information derived from the musical analysis of the piece. In [9], Morales reports research on learning counterpoint rules. The goal of the reported system is to obtain standard counterpoint rules from examples of counterpoint music pieces and basic musical knowledge from traditional music. In [7], Igarashi et al. describe the analysis of respiration during musical performance by inductive logic programming. Using a respiration sensor, respiration during cello performance was measured and rules were extracted from the data together with musical/performance knowledge such as harmonic progression and bowing direction.

III. Expressive singing in the tenor voice

Our choice of studying the human singing voice in the operatic context is not arbitrary; in fact, we believe that operatic music is an ideal environment to start getting some answers to our questions. First, there is a constrained environment in which the music is performed and which is given by the written score and the meaning of the lyrics. Keeping such variables fixed makes the results and comparisons between different singers more meaningful. It also makes it easier for a listener to characterize different performances from the point of view of their emotional attributes. Secondly, good operatic singers tend to have both better voice and better technique than singers in most other genre, and can employ them more efficiently for expressive interpretations. In this context, we choose to focus on the most sought-after role in operas, the human tenor voice, arguably the role for which the most famous arias have even been written.

A. Training data

We have chosen six fragments of arias from Rigoletto, Un Ballo in Maschera, and La Traviata. For four of the fragments we have selected two different interpretations; one of the remaining two fragments has three different interpretations, while the remaining one has a single interpretation. In total the twelve fragments consist of 415 notes in which the tenor and the orchestra do not overlap. The choice of interpretations is not random; we have tried to incorporate very different, yet expressive, performances of the same piece. One of the questions we are interested in answering is whether the expressivity patterns we learn from interpretations of the same aria by the same singer are similar despite the different feel of each performance we choose.

One of the reasons we chose to focus on Josep Carreras as a test case is our subjective observation that his interpretations are highly expressive, yet at the same time they can exhibit a wide variation in emotional attributes even over different performances of the same aria. Another reason why he is the ideal candidate for us is that both the timbre of his voice and his delivery have changed considerably over time. In general, we make the assumption that timber does not vary significantly over short periods of time, but it may change dramatically over long periods. By studying recordings that are close in time we can compare expressivity patterns while controlling over the timbre. Studying recordings that are chronologically far but exhibit the same emotional attributes can on the other hand help understanding which of the patterns we learn are greatly affected by changes in timbre and which are not. We therefore keep track of the recording date of the interpretations that we are processing.

A secondary reason to record this information has to do with what we call *appropriateness* of an interpretation — the capacity of a singer to inhabit a musical piece. Defining this measure is an interesting topic in itself, and touches on many aspects including the question of meaning in music. Our assumption is that recordings closer in time of arias sung in a language familiar to the tenor will minimize appropriateness variations. Future experiments aim to selectively control over the effect of such factors.

B. Musical analysis

We use sound analysis techniques based on spectral models [17] for extracting high-level symbolic features from the recordings. We characterize each performed note by a set of features representing both properties of the note itself and aspects of the musical context in which the note appears. Information about the note includes note pitch and note duration, while information about its melodic context includes the relative pitch and duration of the neighboring notes (i.e. previous and following notes) as well as the Narmour structures to which the note belongs.

In order to provide an abstract structure to our performance data, we decided to use Narmours theory [10]

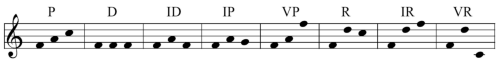

Fig. 1. Prototypical Narmour structures

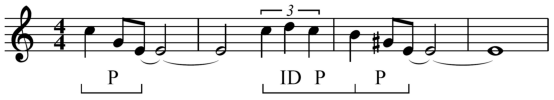

Fig. 2. Narmour analysis of a musical fragment

to analyze the performances. 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. According to Narmour, any two consecutively perceived notes constitute a melodic interval, and if this interval is not conceived as complete, it is an implicative interval, i.e. an interval that implies a subsequent interval with certain characteristics. That is to say, some notes are more likely than others to follow the implicative interval. Based on this, melodic patterns or groups can be identified that either satisfy or violate the implication as predicted by the intervals. Figure 1 shows prototypical Narmour structures. We parse each melody in the training data in order to automatically generate an implication/realization analysis of the pieces. Figure 2 shows the analysis for a fragment of *All of me*.

We additionally annotate the lyrics with syllable-specific information. In our fragments it is overwhelmingly the case that a syllable corresponds to a note in the score. The exceptions are few; in one instance two syllables of a word correspond to a single note. The rest of the ten cases are instances in which the last syllable of a word ends in a vowel and the first syllable of the following one starts with a vowel and they together correspond to a single note in the score. For the beginning we simply specify which syllables are stressed or unstressed, and whether they are open or closed. The librettos for all the fragments in the test suite are written in Italian. If any of the syllables which correspond to a note is stressed then the note will be stressed. In Italian a syllable is open if it ends in a vowel and closed otherwise.

Lastly, we want to see how prosody interacts with the score and the meter of the lyrics. We consider that prosody can give important clues about the emotional content that the singer wants to communicate as it reflects aspects that are not inherent in the lyrics: intonation, rhythm, and 'prosodic' stress. For instance, many have observed that stress may be a matter of the prosodic unit rather than the actual stress of the words. A prosodic unit is a unit of meaning which can be as short as a word and as long as a statement; it is a chunk of speech that may in fact

reflect how the brain processes speech. Acoustically, a prosodic unit is characterized by a few phonetic cues: (1) a typical pitch contour which gradually declines towards the end of the unit and resets itself at the beginning of the next unit, (2) perceptual discontinuities between units, (3) long final unit words. We are interested in where the actual stress falls in a performance, which syllables are over-articulated, what the pitch contour can tell us about the emotional state that the singer transmits, and how are potential conflicts solved between the stress in a prosodic unit and the meter of the lyrics. To make such observations we need to (1) establish the meter of the lyrics and (2) split the lyrics into prosodic units.

C. Learning task

For each expressive transformation, we approach the problem both as a regression and a classification problem. As a regression problem we learn a model for predicting the lengthening ratio of the performed note wrt the score note. This is, a predicted ratio greater than 1 corresponds to a performed note longer than as specified in the score, while a predicted ration smaller than 1 coresponds to a shortened performed note (e.g. a 1.15 prediction corresponds to a 15% performed note lengthening wrt the score). As a classification problem, the performance classes of interest are *lengthen, shorten* and *same* for duration transformation, and *soft, loud* and *same* for energy variation. A note is considered to belong to class *lengthen*, if its performed duration is 20% longer (or more) that its nominal duration, e.g. its duration according to the score. Class *shorten* is defined analogously. A note is considered to be in class *loud* if it is played louder than its predecessor and louder than the average level of the piece. Class *soft* is defined analogously. We decided to set these boundaries after experimenting with different ratios. The main idea was to guarantee that a note classified, for instance, as lengthen was purposely lengthened by the performer and not the result of a performance inexactitude.

IV. Learning Algorithm

We used Tilde's top-down decision tree induction algorithm [2]. Tilde can be considered as a first order logic extension of the C4.5 decision tree algorithm: instead of testing attribute values at the nodes of the tree, Tilde tests logical predicates. This provides the advantages of both propositional decision trees (i.e. efficiency and pruning techniques) and the use of first order logic (i.e. increased expressiveness). The increased expressiveness of first order logic not only provides a more elegant and efficient specification of the musical context of a note, but it provides a more accurate predictive model [12].

We apply the learning algorithm with target predicates: `duration/3` and `energy/3`. (where `/n` at the end of the predicate name refers to the predicate arity, i.e. the number of arguments the predicate takes). Each target predicate corresponds to a particular type of transformation: `duration/3` refers to duration transformation and `energy/3` to energy transformation.

For each target predicate we use as example set the complete training data specialized for the particular type of transformation, e.g. for `duration/3` we used the complete data set information on duration transformation (i.e. the performed duration transformation for each note in the data set). The arguments are the musical piece, the note in the piece and performed transformation.

We use (background) predicates to specify both note musical context and background information. The predicates we consider include `context/8`, `narmour/2`, `succ/2` and `member/3`. Predicate `context/8` specifies the local context of a note. i.e. its arguments are *(Note,Pitch, Dur, MetrStr, PrevPitch, PrevDur, NextPitch, NextDur)*. Predicate `narmour/2` specifies the Narmour groups to which the note belongs. Its arguments are the note identifier and a list of Narmour groups. Predicate `succ(X,Y)` means `Y` is the successor of `X`, and Predicate `member(X,L)` means `X` is a member of list `L`. Note that `succ(X,Y)` also means that `X` is the predecessor of `Y`. The `succ(X,Y)` predicate allows the specification of arbitrary-size note-context by chaining a number of successive notes:

$$succ(X_1, X_2),\ succ(X_2, X_3), \ldots,\ succ(X_{n-1}, X_n)$$

where X_i $(1 \leq i \leq n)$ is the note of interest.

V. Results

The induced classification rules are of different types. Both, rules referring to the local context of a note, i.e. rules classifying a note solely in terms of the timing, pitch and metrical strength of the note and its neighbors, as well as compound rules that refer to both the local context and the Narmour structure were discovered. We discovered a few interesting duration rules:

IF Metrical_Strength = veryweak AND
Note_Duration ∈ (-inf, 0.425] AND
Next_Interval ∈ (-1.5, 0.6] AND
Syllable_Stress = stressed
THEN Stretch_Factor = 2.515625

The note duration is measured as the fraction of a beat, where a beat is a quarter note. The interval is measured in number of semitones. The metrical strength is *verystrong* for the first beat, *strong* for the third beat, *medium* for the second and fourth beats, *weak* for the offbeat, and veryweak for any other position of the note. The rule above says that the notes that are in a very weak metrical position, are shorter or equal then 0.425 of a beat (roughly an eight of a note or less), are followed by a note that is lower by at most 1.5 semitones or higher by at most 0.6 semitones, and correspond to a syllable which is stressed, are performed as a 2.5 times longer note than the duration of the note in the score. What is interesting is that a rule with precisely the same *Metrical_Strength*, *Note_Duration*, and *Next_Interval* is performed only 1.3 longer if the corresponding syllable is not stressed.

The next interesting rule has the following form:

IF Metrical_Strength = medium AND

Note_Duration ∈ (0.425, 0.6] AND
Next_Interval ∈ (2.7, 4.8] AND
Syllable_Stress = unstressed AND
narmour(VR, gr_2)
THEN Stretch_Factor = 2.5

narmour(VR, gr_2) says that the note is in the last (third) position of the registral reversal Narmour structure (VR). Informally this rule says that a note that signals a change of register direction between two intervals of moderate to large size is performed 2.5 longer than the duration of the note in the score if it corresponds to a syllable that is not stressed and it is in the second or fourth beat position. The algorithm also learns two interesting rules about note duration shortening:

IF Metrical_Strength = weak AND
Note_Duration ∈ (0.425, 0.6] AND
Next_Interval ∈ (-1.5, 0.6] AND
Syllable_Stress = stressed AND
narmour(R, gr_2)
THEN Stretch_Factor = 0.328125

IF Metrical_Strength = weak AND
Note_Duration ∈ (0.425, 0.6] AND
Next_Interval ∈ (-1.5, 0.6] AND
Syllable_Stress = unstressed AND
narmour(P, gr_2)
THEN Stretch_Factor = 0.40625

These rule indicate that a note corresponding to a stressed syllable immediately following a higher note, and which will be followed by a note close in frequency will be reduced in length to 0.3 of its duration in the score. This technique would accentuate the final note of the largest local ascending interval. Similarly, a small ascending interval that comes after another small interval in the same direction and which corresponds to an unstressed syllable will be shortened to 0.4 of its duration in the score. According to the Narmour principles, a small interval will be followed by another small interval in the same direction; therefore if the note corresponds to a syllable which is not stressed then its importance will be diminished by shortening its duration. On the other hand, if the unstressed note is at the end of a short descending interval followed by a larger descending interval then the note's duration will be lengthened to 1.9 of its duration in the score, in preparation for the downward 'plunge':

IF Metrical_Strength = weak AND
Note_Duration ∈ (0.425, 0.6] AND
Next_Interval ∈ (-3.6, -1.5] AND
Syllable_Stress = unstressed AND
narmour(IP, gr_2)
THEN Stretch_Factor = 1.90625

An example of energy classification rule is:

IF succ(C, D) AND
narmour(A, D, [nargroup(d, 1)| E]) AND
narmour(A, C, [nargroup(d, 1)| E]) .
THEN energy(A, C, loud) :-

This is, "perform a note loudly if it belongs to an D Narmour group in first position and if its successor belongs to a D Narmour group in first position".

while examples of energy regresion rules are:

IF Note_Duration ∈ (0.425, 0.6] AND
Prev_Interval ∈ (-4.8, -2.7] AND
narmour(IP, gr_1)
THEN Energy = 109.3799415

That is, "perform a note loudly if it belongs to an IP Narmour group in the second position and if its predecessor interval is a large ascending interval". A similar interpretation has the following rule for a R Narmour group:

IF Note_Duration ∈ (0.425, 0.6] AND
Prev_Interval ∈ (-inf, -6.9] AND
narmour(R, gr_1)
THEN Energy = 103.715628

Intuitively these two rules say that there is usually a low note that prepares a high, loud note.

A. Prosody vs. Meter

Let us consider one of the three interpretations of the aria **Forse la soglia attinse** from **Un Ballo in Maschera** by Giuseppe Verdi, specifically the recording from 1975 at La Scala. Let us analyze the fragment consisting of *Ah l'ho segnato SILENCE Ah l'ho segnato SILENCE il sacrifizio mio.* There are three prosodic units (PU) here, separated by the silences. The rhythm is iambic. The stress will therefore fall on *l'ho*, *gna*, *sa*, *fi*, and *mi*; these positions are said to be strong and the rest are weak. In the actual interpretation the second "Ah" is stressed, and according to the iambic meter it raises a conflict between the stress of the meter and the prosody. Accentuating a syllable which is in a weak position creates forward motion towards the next stressed syllable in a strong position, namely *gna* (in what is called a *stress valley* [21]). The strong stress on *gna* gives a sense of positive closure. On the other hand the frequency at which the second prosodic unit ends is high (above 300Hz). This is not a typical terminal shape for a prosodic unit as the high pitch suggests something more to come, an arousing rather than settling interest. This is the qualification of the action in PU2 and arrives in form of PU3 — *il sacrifizio mio*.

The pitch shape of PU2 is different from the shape of PU1 in several respects. PU1 has a terminal shape and the notes are sung relatively flat (i.e. with not much vibrato). The syllable *Ah* is not greatly accentuated nor particularly loud, and it is short. In fact, it is five times shorter than the *Ah* note in PU2, even though in the score the ratio is a quarter note to a half note. The emotional state that it transmits points towards decisiveness. On the other hand, the pitch contour of PU2 goes up, involves a lot of vibrato, over-articulates *Ah* and ends at a very high frequency. In fact PU2 ends at considerably higher pitch then it begins at; something not apparent from the score. These features

all imply some form of forward movement, continuation, and doubt.

VI. Conclusions

This paper presents an approach for detecting expressive patterns of the human tenor voice. We employ machine learning methods to investigate how professional opera singers manipulate sound properties such as timing, amplitude, and pitch in order to produce expressive performances of particular music fragments. We present preliminary results for performances of twelve arias by Josep Carreras. Our approach also takes into consideration features of the lyrics associated with the arias in our test suite. Currently we are considering syllable stress and type, and we are starting to look at the interplay between prosody, meter, and score, in creating lyric-dependent expressive patterns.

References

[1] Van Baelen, E. and De Raedt, L. (1996). Analysis and Prediction of Piano Performances Using Inductive Logic Programming. International Conference in Inductive Logic Programming, 55-71.

[2] H. Blockeel, L. De Raedt, and J. Ramon. Top-down induction of clustering trees. In ed. J. Shavlik, editor, Proceedings of the 15th International Conference on Machine Learning, pages 53-63, Madison, Wisconsin, USA, 1998. Morgan Kaufmann.

[3] Bresin, R. (2000). Virtual Visrtuosity: Studies in Automatic Music Performance. PhD Thesis, KTH, Sweden.

[4] Dovey, M.J. (1995). Analysis of Rachmaninoff's Piano Performances Using Inductive Logic Programming. European Conference on Machine Learning, Springer-Verlag.

[5] Friberg, A.; Bresin, R.; Fryden, L.; 2000. Music from Motion: Sound Level Envelopes of Tones Expressing Human Locomotion. Journal of New Music Research 29(3): 199-210.

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

[7] Igarashi, S., Ozaki, T. and Furukawa, K. (2002). Respiration Reflecting Musical Expression: Analysis of Respiration during Musical Performance by Inductive Logic Programming. Proceedings of Second International Conference on Music and Artificial Intelligence, Springer-Verlag.

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

[9] Morales, E. (1997). PAL: A Pattern-Based First-Order Inductive System. Machine Learning, 26, 227-252.

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

[11] Quinlan, J.R. (1993). C4.5: Programs for Machine Learning, San Francisco, Morgan Kaufmann.

[12] Ramirez, R. et al. (2006). A Tool for Generating and Explaining Expressive Music Performances of Monophonic Jazz Melodies, International Journal on Artificial Intelligence Tools, 15(4), pp. 673-691

[13] Ramirez, R. Hazan, A. Gómez, E. Maestre, E. (2005). Understanding Expressive Transformations in Saxophone Jazz Performances, Journal of New Music Research, Vol. 34, No. 4, pp. 319-330.

[14] Rafael Ramirez, Amaury Hazan, Esteban Maestre, Xavier Serra, A Data Mining Approach to Expressive Music Performance Modeling, in Multimedia Data mining and Knowledge Discovery, Springer.

[15] Repp, B.H. (1992). Diversity and Commonality in Music Performance: an Analysis of Timing Microstructure in Schumann's 'Traumerei'. Journal of the Acoustical Society of America 104.

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

[17] Serra, X. and Smith, S. (1990). "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition", Computer Music Journal, Vol. 14, No. 4.

[18] Tobudic A., Widmer G. (2003). Relational IBL in Music with a New Structural Similarity Measure, Proceedings of the International Conference on Inductive Logic Programming, Springer Verlag.

[19] Todd, N. (1992). The Dynamics of Dynamics: a Model of Musical Expression. Journal of the Acoustical Society of America 91.

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

[21] Tsur, R. (1997). Poetic Rhythm: Performance Patterns and Their Acoustic Correlates. Versification. An Electronic Journal of Literary Prosody.