Methodologies for Expressiveness Modelling of and for Music Performance

Article *in* Journal of New Music Research · September 2004

DOI: 10.1080/09298210420003177796

1 author:

Giovanni De Poli
University of Padua
154 PUBLICATIONS **2,728** CITATIONS

SEE PROFILE

Methodolgies for Expressiveness Modeling of and for Music Performance

Giovanni De Poli
Centro di Sonologia Computazionale
Department. of Information Engineering Padova University
University of Padova

ABSTRACT

Expression is an important aspect of music performance. It is the added value of a performance, and is part of the reason that music is interesting to listen to and sounds alive. Moreover, understanding and modeling expressive content communication is important in many engineering applications. In human musical performance, acoustical or perceptual changes in sound are organized in a complex way by the performer in order to communicate musical content to the listener. The same piece of music can be performed trying to convey a specific interpretation of the score by adding mutable expressive intentions. The analysis of these systematic deviations has led to the formulation of several models that try to describe their structures, with the aim of explaining where, how and why a performer modifies, sometime in an unconscious way, what is indicated by the notation of the score. Modeling paradigms and problems are reviewed and issues for future research efforts are discussed.

1. INTRODUCTION

Music is an important means of communication where three actors participate: the composer, the performer and the listener. The composer instills into the works his/her own emotions, feelings and sensations, and the performer communicates them to the listeners. The composer describes his musical ideas by a score or a process. The information contained in the score (or produced by the process) has a double function: a descriptive one, as a symbolic representation of the cognitive elements constituting the composition, and a functional one, as a way to convey instructions to the performer. Other information is implicit in the score and regards performance style and interpretative conventions. The performer interprets these symbols, taking into account the implicit information and the personal artistic feeling and aim, and produces the sounds by using a musical instrument. Music performance includes all the human activity that lies between the symbolic score and the music instrument.

Music performance is an interesting topic to study for its multi-disciplinary significance. In this paper paradigms and issues emerged in research on modeling expressiveness in music performance will be reviewed and future research perspectives will be discussed. The literature in music performance analysis and modeling is large. After a first pioneering phase in the forties when Seashore and his group (Seasore, 1936) collected and analyzed many objective measurements from performances, not much activity was done until the seventies. Then a new interest on this field arose, with a wealth of empirical work. Good overviews are presented by Gabrielsson (1999); **?** and Palmer (1997). In the following we will discuss performance modeling approaches mainly from a information processing point of view. In Section 2 we will present the basic issue on what models and computational models are used for, and we will discuss expression communication in music performance. In Section 3 we will introduce the aspect of how musical information is represented for modeling purposes. Finally in Section 4 the main strategies used in model development will be presented in detail. Models for understanding, performance synthesis, and artistic creation will be discussed.

This is an Author's Accepted Manuscript of an article published in the Journal of New Music Research, vol. 33, issue 3, pp. 189-202, 2004 , available online at: http://www.tandfonline.com/ [DOI:10.1080/09298210420003177 96]

2. BASIC ISSUES

2.1 Models

Frequently in science, *models* are employed to evidence and abstract some relations that can be hypothesized, discarding details that are felt to be not relevant for what is being observed and described. Models can be used to predict the behavior in certain condition and compare these results with observations. In this sense, they serve to generalize the findings and have both a descriptive and predictive value.

In the study of music performance, models for describing music performance start being developed soon. The possibility, offered by technology, of implementing the models and to experiment with their behavior by simulation gave rise to an increased use of technology in music research. Moreover, computer science and music technology developed many conceptual frameworks and practical tools in the last decades, that are very useful for music performance investigation. For example artificial intelligence, knowledge engineering, soft computing methodologies, physics based models, MIDI instruments, signal processing analysis methods, computer controlled performance, motion capture devices, constitute paradigms and tools that are basic of many performance models.

The idea of developing *computational models* of music performance dates back to the first music application of computers. It can be mentioned the Groove system by

Matthews and Moore (1970) that allowed real time control and editing of performer actions described (graphically or symbolically) by time functions. The first models were mainly dedicated to music production and experimentation, and were embedded in computer programs for music synthesis or representation and for interactive performance. Their theoretical assumptions and conceptual foundation were often not explicit. Later models for performance understanding started to be developed (e.g. KTH performance rule system (Sundberg et al., 1983)) and now we can expect a convergence of efforts toward models that are oriented toward both performance understanding and production.

We can distinguish two kinds of models. The *complete model* tries to explain all of the observed performance deviations on the basis of the given data. This approach tends to give very complex models and thus poor insight into the relevant relations. In fact, note level analysis cannot explain all the observed deviations. The other kind is the *partial model*, which aims only to explain what can be explained at note level, giving a small and robust set of rules. Moreover, when rules for categorical decisions (e.g., play faster of slower) rather than for computing an exact value are used, more understandable results can be obtained.

2.2 From mathematical models to information processing models

The classic way to describe relations in models is by using mathematical expressions composed of observable (and often measurable) facts called variables or parameters. Developing and then validating *mathematical models* is the typical way to proceed in science and engineering. Often the variables are divided into input variables, supposed known, and output variables, which are deduced by the model. In this case, inputs can be considered as the causes and output the effect of the phenomenon. A mathematical model can be implemented on a computer by numerical analysis techniques. In this way, we can compute the values of output variables corresponding to the provided values of inputs. This process is called *simulation* and it is widely used to predict the behavior of the phenomenon in different circumstances.

However, a computer does not only deal with numerical values. More generally, it can be considered as information processing engine. From this perspective, models describe relations between different kinds of information about the phenomenon. Thus, a fundamental problem in developing *information processing* models is to define which kind of information we want to deal with and how we may represent it on a computer.

The case of music performance is quite interesting; in fact, the information that can be considered regards many aspects. We can distinguish three layers. The first is the *physical information* that can be measured, as timing or performers movements. This information can be represented as numbers and is typically used and processed by mathematical tools. The second layer is the *symbolic information* as the score, where the notes are represented by symbols in the common music notation. These symbols refer more to a cognitive organization of the music than to an exact physical value. For example, the duration symbol indicates a division of the meter, while the actual duration of a performed note can vary. Processing at this level uses typical symbolic and logic representations of computer science. At a higher level, we have the *expressive information* more related to the affective and emotional content of the music. Recently computer science and engineering started paying attention to this level of information and developing suitable theories and processing tools. Music and music performance in particular, attracted the interest of researchers for developing and testing such tools. Moreover in performance modeling, all the information levels should be taken into account in a coordinated way. As a consequence, information representation and model structure are crucial topics in model design and will be discussed in section 3.

2.3 Expressiveness in music performance

The communication of expressive content by music can be studied at three different levels: considering composer message, performer expressive intentions and listener perceptual experience. Studies of the first kind are historically more developed. Generally, they analyze the elements of the musical structure and the musical phrasing that are critical for a correct interpretation of composers message.

The contribution of the performer to expression communication has two facets: to clarify the composers message by enlightening the musical structure and to add his personal interpretation of the piece. A mechanical performance of a score is perceived as lacking of musical meaning and is considered dull and inexpressive as a text read without any prosodic inflexion. Indeed, human performers never respect tempo, timing and loudness notations in a mechanical way when they play a score: some deviations are always introduced, even if the performer explicitly wants to play mechanically (Palmer, 1989).

Thus in general *expressiveness* refers both to the means used by the performer to convey the composer message and to his own contribution to enrich the musical message. However many music performance studies concentrate on the first aspect trying to understand the performer actions to better convey the musical structure. Simulation models are often evaluated by the musical acceptability of their results, or in other words how well a supposed ideal interpretation of that particular piece is approached. Expressiveness related to the musical structure may depend on the dramatic narrative developed by the performer, on physical and motor constraints or problems (e.g. fingering), on stylistic expectation based on cultural norm (e.g. jazz vs. classic music) and on the actual performance situation (e.g. audience engagement) (Clarke, 1995). Figure 1 shows the relation between dynamics profiles and the main elements of music structure of the first measures of a piano performance of Mozart sonata K 545 (figure 2). It is particularly evident that the musician emphasized with a decrescendo the end of the first inciso (bar 2), the first semi-phrase (bar 4), the first phrase (bar 8) and the period (bar 16).

Recently interest is growing also in taking into account

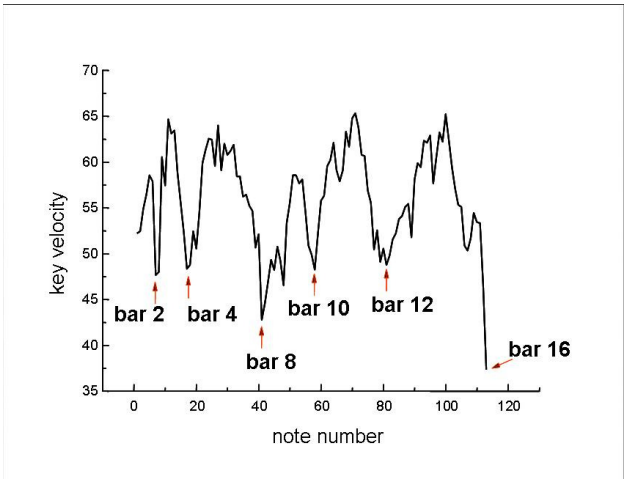

Figure 1. Dynamics profiles and the main elements of music structure of the first measures of a piano performance of Mozart sonata K 545 (figure 2). It is particularly evident that the musician emphasized with a decrescendo the end of the first inciso (bar 2), the first semi-phrase (bar 4), the first phrase (bar 8) and the period (bar 16).

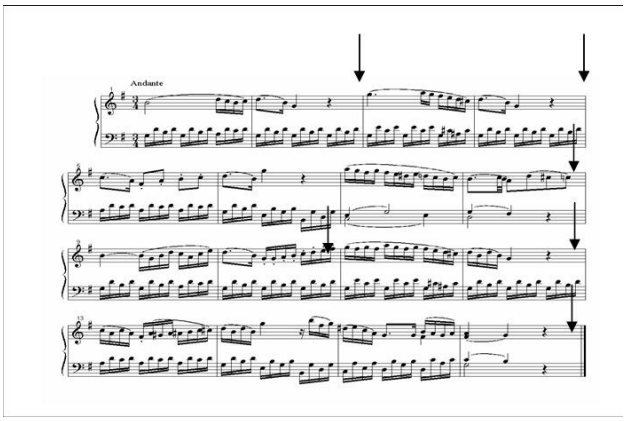

Figure 2. Score of the first 16 measures Mozart sonata K 545. The arrows indicate the end of the first inciso (bar 2), the first semi-phrase (bar 4), the first phrase (bar 8) and the period (bar 16).

the expression component added by the performer. Some aspects are still strongly related to the musical piece, as performer specific style, and influences of stylistic expectation based on cultural norm (e.g. jazz vs. classic music) or actual performance situation (e.g. audience engagement). Nevertheless, other communicative aspects can be taken into account. Experiments are carried out by asking performers to play the same piece according diverse specific adjectives or nuances or trying to convey different content. The researcher then seeks to understand and model the strategies used in these performances. Often basic emotions are chosen as possible expressions (Gabrielsson and Juslin, 1996). and in this case the term expressive performance refers to *emotional performance*. Notice that sometimes the emotions the performer tries to convey can be in contrast with the character of the musical piece. A slightly broader interpretation of expression as *KANSEI* (Japanese term indicating sensibility, feeling, sensitivity) (Suzuki and Hashimoto, 2004) or affective communication (Picard, 1997) is proposed in some Japanese or American studies. We prefer the broader term *expressive intentions* that include emotion, affections as well as other sensorial and descriptive adjectives or actions. Furthermore, this term evidences the explicit intent of the performer in communicating expression.

Understanding of specific *artistic intentions* of top-level performers is more challenging. While artists aim to express aesthetic value, We feel that these qualities are probably impossible to model, without loosing their real essence.

3. INFORMATION AND MUSIC PERFORMANCE

3.1 Expressive performance parameters

When we want to develop an information processing model, it is important to define which is the relevant information we will use. This choice depends on the phenomenon we are observing and on the available detection techniques. In our case, we want to describe music performance and we can observe the variations a music performer is doing when he plays. This kind of information is often called expressive parameters. The most relevant information used in performance models are discussed in this section.

At a physical information level, the main expressive parameters considered in the models are related to timing of musical events and tempo, dynamics (loudness variation), and articulation (the way successive notes are connected). These parameters are particularly relevant for keyboard instruments. Moreover, they are the basic parameters of the MIDI protocol and thus are easily measurable on electronic music instruments or employable for obtaining a music performance. For some instruments and for the singing voice other acoustic parameters are taken into account such as vibrato and micro-intonation, or pedalling on the piano. In contemporary music, timbre is often an essential expressive parameter; sometimes also virtual space location or movement of the sound source is used as expression features..

These parameters can be measured directly by a MIDI musical instrument or (with more effort) by detecting the performer movements. However, it should be noted that these measurements depend on an accurate instrument calibration. In fact, the relation between MIDI commands and their sonic realizations depends greatly on the instrument. For example, the Note-off command indicates the beginning of the sound decay and not the ending of the note, as might often be desired.

Physical information can also be gathered from audio recordings. Additional expressive parameters can be taken into account, such as timbre. However, the performance parameters are more difficult to collect automatically, especially for multi-voice music, and depend on the recording conditions. Different methods are often used and thus the measurements reported in the literature may be not directly comparable. This fact contributes to make the accumulation of knowledge difficult. For instance, it

is not always clear exactly when a tone starts, nor when the attack phase can be considered completed. The amplitude envelope inspection is not sufficient. Therefore, the attack duration of a note can be measured in different ways, leading to dissimilar values. On the other hand, in real-time applications we need effective but not too complex feature analysis algorithms. It is advisable that the progress of computational analysis techniques should provide useful and standardized tools for performance parameter detection.

The interrelation of these physical parameters is not well understood. Therefore, models often try to separate the parameters and to model their effect separately, or to deal with a combination of very few of them. The problem is particularly evident when we want to model some effects that can be rendered in different ways. For example, the performer can emphasize a note by increasing its loudness, or by lengthening its duration or by a slight time shift, or by a particular articulation or timbre modification. The use of more abstract representations could probably help in separating the lowlevel features from higher-level ones. This approach would call for multi-level models or a combination of models acting at different abstraction levels. For instance, in the previous example a model can decide that a note should be emphasized because of its structural importance, and a second model will decide how to realize the emphasis taking into account the context, the expressive resources of the instrument, stylistic expectations etc. While the performer probably uses such multi-level strategies intuitively in his/her musical practice, a precise definition of intermediate parameters, effective for modelling purposes, is still partial. More research is needed for the selection of these intermediate parameters, for finding a possible quantification, and for assessing their effectiveness.

The interrelation of these physical parameters is not well understood. Thus, often models try to separate the parameters and to model their effect separately or to deal with a combination of very few of them. The problem is particularly evident when we want to model some effects that can be rendered in different ways. For example, the performer can emphasize a note by increasing its loudness, or by lengthening its duration or by a slight time shift, or by a particular articulation or timbre modification. Probably the use of more abstract representations could help in separating the low-level features from higher-level ones. This approach would call for multilevel models or a combination of models acting at different abstraction level. For instance in the previous example, a models can decide that a note should be emphasized because of its structural importance and a second model will decide how to realize the emphasis taking into account the context, the expressive resources of the instrument, stylistic expectations etc.. While probably, the performer uses intuitively such multilevel strategies in his/her musical practice, a precise definition of intermediate parameters, effective for modeling purpose, is still partial. More research is needed for the selection of these intermediate parameters, for finding a possible quantification, for assessing their effectiveness.

As regards symbolic information, the score is a typical reference and it is usually represented as a list of time events. More difficult is the representation of the musical structure. The knowledge is only partially formalized, especially toward classical music. Very few computational models have been proposed for automatic (or semiautomatic) structure extraction from the score (Lerdahl and Jackendorff, 1983; Narmour, 1900; Cambouropoulos, 1997) and their results are not very reliable. Thus the segmentation and the structure is often introduced by hand. The classic paradigm derives from early language modelling and consists in musical grammars represented as a hierarchical tree structure (e.g., phrase, sub-phrase, melodic gesture, note). This paradigm is much less applicable for contemporary music where other musical parameters and constructs are more pertinent. Music performance research will greatly benefit from theoretic advancements in contemporary music analysis.

The understanding of the expressive information is still vague. While its importance is generally acknowledged, the basic constituents are less clear. Often the simple range expressiveinexpressive is used. The most frequently used paradigms for representing emotions in music performance modelling are the basic emotions and the dimensional approach, e.g., the valence-arousal space (Juslin, 2001). The dimensional approach was also used with success for other kinds of expressive intentions (Canazza et al., 2003). In this field too, more research and experimental insight will be very fruitful (see Scherer (2004) for a discussion). On the other hand continuous measurements of subject reactions during a performance, recently used in psychological research (see, e.g., Schubert (2001)), may provide useful data and parameters for performance research.

3.2 Information representation

A key issue is how the model represents the information. The most important aspect is the representation of time. Time can be considered from both a physical and a symbolic point of view. The first one, *performance-time*, refers to the actual time that can be measured during a performance. The second refers to the position in the score (e.g., phrase or measure) and is often called *score-time* or score position; it is often measured in units (or subunits) of measure. Models of *timing* normally aim to describe the relation between performance and score time (Honing, 2001). Performers adapt performance time of musical events in subtle way. Understanding models try to explain these variations, while synthesis models compute these variations.

Another important aspect of time representation is tempo that is the reciprocal of durations as a function of score position. Traditionally it is measured by a metronome (M.M.) number indicating the number of beats per minute (bpm) of performance time. A distinction may be made between the *mean* tempo (i.e., the average tempo across the whole piece disregarding possible variations); the *main* tempo (i.e., the prevailing tempo when passages with momentary variations such as slow start, final ritard, fermatas, and amorphous caesuras are excluded); the local tempo, which is maintained only for a short time and is measured as the

inverse of the inter-onset interval relative to its nominal length in the score (Repp, 1992; Gabrielsson, 1999). Although it is still unclear what exactly constitutes the perception of tempo, it seems to be related at least in metrical music to the notion of beat or tactus; the speed at which the pulse of the music passes at a moderate rate (i.e., the metrical level at which one counts the beat (Honing, 2002)).

Models for understanding usually describe tempo as function of score position and measure it in seconds per metrical or score unit. In this case, global and local tempos are considered, depending on the time scale. Typical representation are the duration of a measure and the relative inter-onset interval (IOI), i.e., the time difference between the next event and the actual event divided by the symbolic (score) duration. Notice that the inter-onset interval is not the physical duration of a note; notes can be played staccato or legato, greatly affecting their expressive character.

While tempo and timing both refers to time values, they tend to be perceived somewhat independently by listeners. Thus, timing models should take into account both aspects trying to separate them (Honing, 2001). A study trying to address that problem has been presented by Cambouropoulos et al. (2001). Often expressive timing is considered as describing the timing deviations in a performance (e.g., accentuating notes by lengthening them slightly, or playing notes after the beat). In addition, timing might be perceived independently of any changing tempo (tempo rubato). So it could be argued that expressive timing and expressive tempo possibly co-exist as two, relatively independent, perceptible aspects of a performance (Honing, 2002).

The musical parameters used in modelling can be represented as values or attributes of discrete time instants (musical events) such as notes or structural units. Alternatively they can be represented as profiles, i.e., as functions of continuous (performance or score) time. An example of discrete time representation is the articulation of timing of individual notes or the micropauses between melodic units. An example of continuous time representation is the vibrato of a note or a crescendo curve. The first representation is more related to the symbolic level, while the second relates to the physical level. The choice depends on the aim of a model, on availability of data, and on their ability to explain. Sometimes models combine both kinds of representations or are able to transform data from one to the other representation, e.g., by interpolation or sampling. For example, a crescendo is a discrete parameter for the piano, but not for other instruments, e.g., the violin. Moreover it can be interpreted as continuous curve sampled at the note onsets.

Another aspect of the representation is the granularity. When possible, the information is represented as numerical values. Sometimes absolute values, e.g., time interval in milliseconds, sometimes relative values, e.g., relative inter-onset interval, are used. In this case the inter-onset intervals are represented as normalized to their score duration (see, e.g., Repp (1992)), or at a certain metrical level, most often the beat level (see, e.g., Shaffer et al. (1985)) or the bar level (see, e.g., Todd (1985); Repp (1992)). In this last way, the timing pattern becomes a local tempo indicator. In other situations the information is categorical, describing one choice among few alternatives, e.g., staccato versus legato, shortening versus lengthening. Even for granularity, the effectiveness of the representation depends on the problem we are dealing with and on the musical context (see, e.g., Timmers and Honing (2002)). However, in symbolic representation of music often the concepts are not easily expressible as numbers or as precisely defined categories. A possibility of using effectively vague definitions is offered by the techniques of soft computing such as fuzzy sets (Bresin et al., 1995a,b; Weyde and Dalinghaus, 2003).

Music is an organization of events in time and often a hierarchical time structure can be envisaged. Therefore, models are developed for representing performance aspects at different time scales (Lerdahl and Jackendorff, 1983; Friberg, 1995; Todd, 1992, 1995; De Poli et al., 1998). We may have models at note scale, e.g., for attack time or vibrato, at local scale considering only few notes, e.g., articulation of a melodic gesture, or at a more global scale, e.g., for phrase crescendo. The most complete models deal with the different time scales by using distinct but coordinated strategies.

3.3 Expressive deviations

Most studies of performance expressiveness aims at understanding the systematic presence of *deviations* from the musical notation as a communication means between musician and listener. Deviations introduced by technical constraints (such as fingering) or by imperfect performer skill, are not normally considered part of expression communication and thus are often filtered out as noise. Deviations considered in models normally refers to the expressive performance parameters as discussed above.

The analysis of these systematic deviations has led to the formulation of several models that try to describe their structure, with the aim to explain where, how and why a performer modifies, sometimes unconsciously, what is indicated by the notation in the score. It should be noticed that, although deviations are only the external surface of something deeper and often not directly accessible, they are quite easily measurable, and thus widely used to develop computational models in scientific research and generative models for musical applications.

When we talk of deviation, it is important to define which is the reference used for computing deviation. Very often the score is taken as *reference*, both for theoretical (the score represents the music structure) and practical (it is easily available). However, the use of a score as reference has some drawbacks for the interpretation of how listeners judge expressiveness. Alternative approaches are the *intrinsic definitions* of expression (expressive deviations defined in terms of the performance itself) (Gabrielsson, 1974; Desain and Honing, 1991) or non-structural approaches relating expression to motion, emotion, etc. (see Clarke (1995) for a general discussion). The idea is that, from the structural description of a music piece, we can in-

dividuate units which can act as a reference at that level. Its subunits will act as atomic parts whose internal detail will be ignored. Then expression is defined as the deviation from the norm as given by a higher level unit. For example. the expressive variations of the durations of beats are expressed with reference to the bar duration (as ratio). Using this *intrinsic definition,* expression can be extracted from the performance data itself, taking more global measurements as reference for local ones. (Desain and Honing, 1991).

However the choice depends on the problem we are dealing with. When we studied how a performer plays a piece according to different expressive intentions, we found that a clearer interpretation and best results in simulation are obtainable by using a *neutral* performance as reference (Canazza et al., 2004). We intend neutral in the sense of a human performance without any specific expressive intention. By neutral we intend a human performance without any specific expressive intention. In other studies the mean performance (i.e., the mathematical mean across different performances by the same or many performers) was taken as the reference when stylistic choices and preferences were investigated (e.g., Repp (1992, 1997)).

4. MODELS OF/FOR MUSIC PERFORMANCE

Models are developed with different aims. A basic difference is between models *of* music performance, i.e. models *for* understanding (also called *analysis* models), and models for music performance, i.e. models able to produce music performances (also called *synthesis* models). In the following sections, the main paradigms will be presented and discussed.

4.1 Model structures

In developing and using models it is often convenient to break the problem into simpler parts, each one described and modelled by a proper strategy, and then combine everything into a larger unit. In the following, the principal way used to combine rules or models will be discussed.

The first, and frequently used, strategy assumes that the partial results computed by sub-models can be *added* to obtain the final result. For example, the deviations computed by the KTH rule system are obtained by a weighted sum of the deviations computed by the single rules (Sundberg et al., 1983, 1989; Friberg et al., 1991, 1998). Another application is when the final result is obtained as sum of profiles at different time scale, e.g. the crescendo and accelerando curves computed for phrases and sub-phrases by Todd (Todd, 1992, 1995). An application of this strategy in analysis is when the principal component analysis (PCA) of measured deviations on a musical passage is used to highlight differences among performing styles of different pianists (Repp, 1992). In fact PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components*. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The original data are thus expressed as a linear combination of (few) significant and independent variations around their mean values.

The additivity hypothesis is attractive from both a mathematical and a practical point of view; it allows the use of many computational tools and it is easily interpretable. However, it may result in over-simplifying and tends to hide the interrelation of different aspects of performance. A partially different strategy for combining numerical values consists in *multiplying* the partial results. It is often used when relative values are employed. Of course taking the logarithms will transform it in an additive strategy.

More complex is the *nonlinear combination* of the sources $y = f(x1, x2, ..., xn)$. In this way the interrelations of inputs can taken into account. An example is the use of feed-forward neural networks as general approximators of observed performance deviations (Bresin, 1998).

Models are sometimes combined using the output of a model as input to a second one, i.e. by *functional composition*, as in cascade model that compute $y = f[g(x)]$. A typical example is timing function composition as discussed by Honing (2001).

A more general approach is in *hierarchical models* when they operate at different abstraction level. The information is processed and combined at the proper level. An example is the distinction of rules and metarules in the KTH system, where the metarules choose the proper setting of basic rules to express, for example, different emotion (Bresin and Friberg, 2000).

From another point of view we may identify *local models*, that acts at note level and try to explain the observed facts in a local context (see, e.g., Friberg et al. (1991); Widmer (2002)). A different perspective is assumed by *phrasing models* (see, e.g., Todd (1992, 1995); Battel and Fimbianti (1999); Widmer and Tobudic (2003)) that take into account higher levels of the musical structure or more abstract expression patterns. The two approaches often require different modelling strategies and structures. In certain cases it is possible to devise a combination of both approaches with the purpose being to obtain better results. The *composed models* are built by several components, each one aiming to represent the different sources of expression. However, a good combination of the different parts is still quite challenging.

4.2 Comparing performances

A problem that normally arises in performance research is how performances can be compared. In subjective comparisons often a supposed ideal performance is taken as reference by the evaluator. In other cases, an actual reference performance can be assumed. Of course subjects with different backgrounds can have dissimilar preferences that are not easily made explicit.

However when we consider computational models, objective numerical comparisons would be very appealing. In this case, performances are represented by a set of values. Sometimes the adopted strategies compare absolute

or relative values. As measure of distance the mean of the absolute differences can be considered, or the Euclidean distance (square root of difference squares) or maximum distance (i.e. take the maximal difference component). It is not clear how to weight the components, nor which distance formulation is more effective. Different researchers employ different measures. More basically it is not clear how to combine time and loudness distances for a comprehensive performance comparison. For instance as already discussed, the emphasis of a note can be obtained by lengthening, dynamic accent, time shift, and timbre variation. Moreover, it is not clear how perception can be taken into account, nor how to model subjective preferences. How are subjective and objective comparisons related? The availability of good and agreed methods for performance comparison would be very welcome in performance research. A subjective assessment of objective comparison is needed. More research effort on this direction is advisable.

4.3 Models for understanding

We may distinguish some strategies in developing the structure of the model and in finding its parameters. The most prevalent ones are analysis-by-measurement and analysis-by-synthesis. Recently some methods from artificial intelligence started being developed: machine learning and case based reasoning.

4.3.1 Analysis by measurements

The first strategy, analysis-by-measurement, is based on the analysis of deviations measured in recorded human performances. The analysis aims at recognizing regularities in the deviation patterns and to describe them by means of a mathematical model, relating score to expressive values (Gabrielsson, 1999). The method consists in different stages:

1. *Selection of performances.* The choice of good and/or typical performances of the musical excerpt to study is important. Often a rather small set of carefully selected performances are used. While the performer normally is left free to play according to his own taste, for experimental purposes he may sometimes be asked to play according to specific instructions, e.g. to convey a specific emotion.

2. *Measurement of the performance parameters of every note.* The physical variations of the performance are several: duration, intensity, frequency, envelope, note vibrato; which and how many variables to study depends on the aims and working hypothesis, on the technical possibility of the instrument and on the considered instruments.

3. *Reliability control and classification of performances.* It is necessary to verify the reliability and consistence of the data obtained from the physical variable measurement, classifying the performance in different categories, with different characteristics, taking into account the collected data.

4. *Selection and analysis of the most relevant variables.* This stage depends on the two previous ones and it ends temporarily the analytic part of the scheme to give space to the judgment by the listeners, in the following stages.

5. *Statistical analysis and development of mathematical interpretation model of the data.* The analysis of the selected variables is often carried out on different time scale representations.

The most frequently used approaches are statistical models (Repp, 1992) and mathematical models (Todd, 1992, 1995). Sometimes multidimensional analysis is applied to performance profiles (e.g. principal component analysis) in order to extract independent pattern (Repp, 1992). Often the hypothesis, that deviations deriving from different patterns or hierarchical levels can be separated and then added, is implicitly assumed. This hypothesis helps the modeling phase, but may be oversimplified.

Several methodologies of approximation of human performances were developed using neural network techniques (Bresin, 1998)or fuzzy logic approach (Bresin et al., 1995a,b) or using a multiple regression analysis algorithm (Ishikawa et al., 2000) or linear vector space theory (Zanon and De Poli, 2003a,b). In these cases, the researcher devises a parametric model and then estimates the parameters that best approximate a set of given performances.

In alternative to this method that analyses actual music performances, some researchers are performing controlled experiments in collecting and studying performances. The idea is that by manipulating one parameter in a performance (e.g. the instruction to play at a different tempo), the measurements may reveal something of the underlying mechanisms (Desain et al., 2001).

4.3.2 Analysis by synthesis

The analysis-by-synthesis paradigm (Gabrielsson, 1985) takes into account the performance-perception and it starts from the results of the previous stages continuing with the following stages.

6. *Synthesis of performances with systematic variations.* At this stage the researcher produces different versions of the piece in order to have performances in which the physical variables to be studied (duration, intensity, etc.) systematically vary.

7. *Judgment of synthesized versions, paying particular attention to the different experimental aspects selected.* Knowledge of relevant experimental variables and the designation of useful evaluations scales are required.

8. *Control of the reliability of the judgments followed by classifications of the listeners.* We need to use adequate methods to control the listeners? reliability and their judgments, possibly classifying them in different classes.

9. *Study of relation between performance and experimental variables.* At this point, it is possible to observe the relations between performances with manipulated physical variations and the selected variables asking questions such as: are the listeners sensitive to the manipulations made? If yes, in which way? Are there general effects or interactions among different variables? Which are the most important variables? Can we eliminate some of them?

10. *Repetition of the procedure (steps 3-9) until the results converge.* In relation to the results of stages 3-9, the process should be continued in an interactive manner until the relations of the selected variables of the performance converge to the experimental variables.

The scheme here described can be modified and extended, but the main concept remains the following: the analysis of the real performances produces hypothesis to be tested through the systematic variations introduced in the synthetic versions. With regard to such variations, it should be noticed that factors must be modified one by one keeping the rest constant. The best method to generate them should be, for instance, to produce simplified versions where only one variable is modified, while imposing constant values to the others. The product will sound rather different from a real performance where all the physical variables change continuously. In order to obtain data about the effect of the other variables and their interaction, we must proceed with further experiments, in a long series of working sessions.

This strategy derives models, which are described with a collection of rules, using an *analysis-by-synthesis* method. The most important is the KTH rule system (Friberg et al., 1991, 1998, 2000; Sundberg et al., 1983, 1989, 1991). Other rules were developed by De Poli et al. (1990). Dannenberg and Derenyi (1998) employed this methodology for developing a performance model that generates continuous control information to synthesize trumpet performances from a symbolic score input. In the KTH system, the rules describe quantitatively the deviations to be applied to a musical score in order to produce a more attractive and human-like performance than the mechanical rendering that results from a literal playing of the score. Every rule tries to predict (and to explain with musical or psychoacoustic principles) some deviations that a human performer is likely to insert. At first, rules are obtained based on the indications of professional musicians, using *knowledge engineering* paradigms. Then, the performances, produced by applying the rules, are evaluated by listeners, allowing further tuning and development of the rules. The rules can be grouped according to the purposes that they apparently have in music communication. *Differentiation rules* appear to facilitate categorization of pitch and duration, whereas *grouping rules* appear to facilitate grouping of notes, both at micro and macro level. As an example of such rules, let us consider the Duration Contrast rule; it shortens and decreases the amplitude of notes with durations between 30 and 600 ms, depending on their duration according to a suitable function. The value computed by the rule is then weighted by a quantity parameter k.

4.3.3 Machine learning

In the traditional way of developing models, the researchers normally makes some hypothesis on the performance aspects they want to model and then he tries to establish the empirical validity of the model by testing it on real data or on synthetic performances. A different approach, pursued by Widmer and coworkers (Widmer, 1995a,b, 1996, 2003; Widmer and Zanon, 2004), instead tries to extract new and potentially interesting regularities and performance principles from many performance examples, by using machine learning and data mining algorithms. Aim of these methods is to search for and discover complex dependencies in very large data sets, without any preliminary hypothesis. The advantage is the possibility of discover new (and possibly interesting) knowledge, avoiding any musical expectation or assumption. Moreover, these algorithms normally allow describing discoveries in intelligible terms. The main criteria for acceptance of the results are generality, accuracy, and simplicity.

4.3.4 Case based reasoning

An alternative approach, much closer to the observation-imitation-experimentation process observed in humans, is that of directly using the knowledge implicit in human performance samples. Case-based reasoning (CBR) is based on the idea of solving new problems by using (often with some kind of adaptation) similar previously solved problems. Two basic mechanisms are used; retrieval of solved problems (called cases) using suitable criteria, and adaptation of solutions used in previous cases to the actual problem. The assumption is that similar problems have similar solutions. CBR is appropriate for problems where many examples of solved problems can be obtained and a large part of the knowledge involved in the solution of problems is tacit, difficult to verbalize and generalize. Moreover, a new problem solution can be checked by the user and then memorized. Thus, the system learns from experience. An important achievement in this direction is the SaxEx system for expressive performance of jazz ballads (Arcos et al., 1998; Arcos and de Mántaras, 2001) and the Suzuki and Tokunaga (1999) system. The success of this approach greatly depends on the availability of a large amount of well-distributed previously solved problems, not easy to collect. These are not easy to collect.

4.3.5 Expression recognition models

The methods described in the previous sections aim at explaining how expression is conveyed by the performer and how it is related to the musical structure. Recently this accumulated research results started giving rise to models that aim to extract and recognize expression from a performance. For example Dannenberg et al. (1997) developed a system to classify improvisational performance style among different alternatives and Friberg et al. (2002) recognizes the basic emotion in music performance. Zanon

addressed the question whether it is possible for a machine to learn to identify famous performers (pianists) based on their style of playing (Zanon and Widmer, 2003; Widmer and Zanon, 2004)

4.4 Models for music production

4.4.1 Performance synthesis models

While the models above described were developed mainly for analysis and understanding purpose, they also are often used for synthesis purpose. Starting from expressiveness models, several software systems for the computer automatic generation of musical performances were developed. Some examples are: POCO (Honing, 1990), the RUBATO system, based on performance vector field (Mazzola and Zahorka, 1994), Director Musices (Friberg et al., 2000) Super Conductor (Clynes, 1998) and CaRo (Canazza et al., 2000). Moreover, many sequencers now implement functions, called humanizer, that add deviations to the score, computed in a random way or according to specific criteria. The typical scheme is represented in figure 3.

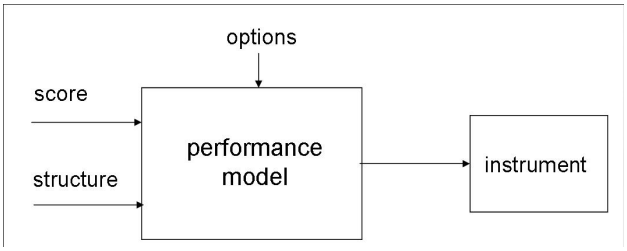

Figure 3. Typical structure of a performance synthesis model.

The model defined at Centro di Sonologia Computazionale (CSC), University of Padova, was developed using the results of perceptual and sonological analyses made on professional performances (Canazza et al., 2004). Different applications based on this model were developed. Music performance is an activity that is well suited as a target for multimodal concepts. Music is a nonverbal form of communication that requires both logical precision and intuitive expression. Our research in the domain of creative arts has focused on musical mapping of gestural input. Since the control space works at an abstract level, it can be used as an interface between transmodal signals. In particular, we developed an application allowing control of the expressive content of a pre-recorded music performance by means of dancers movement as captured by a camera. Then expressive features extracted by dancers movements are used as input for the abstract space. In the entertainment area, we built the application Once upon the time (released as an applet) for the enjoyment of fairy-tales in a remote multimedia environment (Canazza et al., 2000). IIn this software, an expressive identity can be assigned to each character in the tale and to the different multimedia objects of the virtual environment. Starting from the storyboard of the tale, the different expressive intentions are located in synthetic control spaces defined for the specific contexts of the tale. The expressive content of the audio is gradually modified with respect to the position and movements of the mouse pointer, using the abstract control space described above.

4.4.2 Discussion on synthesis models

The idea of automatic expressive music performance, especially when it is applied to the performance of classical music, is questionable. We can remark that classical music was not written for this purpose. Even if the models could be very accurate (and they still are not), some very important artistic aspects of this kind of music will be omitted. When we listen to a recording of a classic music performance, we are aware that it is just a reproduction of an event and not an experience of the music as it was conceived at its time. On the other hand, the possibility to fully model and render the artistic creativity implied in the performance is still to be demonstrated. For the moment, at best, we can expect a reproduction of a specific performance, without a real new creative contribution that would make listening interesting. Or we can expect the rendering of some, hopefully relevant, aspects of a musically acceptable performance, but not sufficient for a full artistic appreciation.

Performers are particularly sensitive to these aspects and usually look at performance synthesis in a very suspicious manner. An instinctive fear of a possible danger for their competence and even their job can be guessed to contribute, but the cultural motivations are definitely true. On the other hand if we think of music applications where a real artistic value is not necessary (even if useful as in many multimedia applications), and where the alternative is a mechanic performance of the score (as in many sequencers), automatic performance can be acceptable. From this point of view such models can be used for entertainment applications (Bresin and Friberg, 2001) or when it is not necessary to preserve the exact artistic environment of the composition, as in popular music. However, in many occasions a human performer is not available and should be substituted in a certain way. Performance models or processing of MIDI recorded performances could be a solution. Notice that the quality of performance processing is much higher when it is based on performance models and knowledge.

Another important application of performance models, even of classical music, is in *education*. The knowledge embodied in performance models may help teachers to increase their students' awareness of certain performance strategies and to better convey their teaching goals.

4.4.3 Models for multimedia application

Representing, modeling and processing expressive information is useful not only for automatic music performance. In fact a user can interact with the model during the performance. We can thus consider interactive performance models where expression is conveyed by a joint action of the user and of the model. This paradigm of human machine interaction for expression communication is not only fruitful in music applications, but it can be extended to many other fields where non-verbal content can be very

relevant. We may distinguish two main classes of possible interfaces for the human-machine communication:

- Graphic panel dedicated to the control, where the control variables are directly displayed on the panel and the user should learn how to use it.
- Multimodal, where the user interacts freely through movements and non-verbal communication. The task of the interface is to analyze and to identify human intentions correctly.

Expressiveness control is a relevant aspect in multimodal systems. The current state-of-the-art allows for a growing number of applications, from advanced human-computer interfaces in multimedia systems to new kinds of interactive multimodal systems. An explosion of human interface technologies involving ecological interface design, agents, virtual immersive workspaces, decision support systems, avatars, distributed architectures, and computer-supported cooperative work, are appearing on the scene as means to address these complex problems.

Multimodal interfaces have the potential to offer users more expressive power and flexibility, as well as better tools for controlling sophisticated visualization and multimedia output capabilities. As these interfaces develop, research will be needed on how to design complete multimodal-multimedia systems that are capable of highly robust functioning. To achieve this goal, a better expressive content analysis and processing ability will be essential. The computer science community is just beginning to understand how to design innovative, well-integrated, and robust multimodal systems. Most multimodal systems remain bimodal, and recognition technologies related to several human senses (e.g., haptics, smell, taste) are not well represented in multimodal interfaces. This means that it is very important to consider performance models for non-verbal communication in a successful design of multimodal systems.

4.5 Models for artistic creation

The situation is different when music is created expressively, bearing in mind the use of technology. We are in the era of the information society and artists are more frequently using technology in their artworks. Since the beginning of the last century, some musicians started to think how to enlarge the sound palette by using unconventional instruments. The availability of new electronic and computer-generated sounds gave rise to a new kind of music. Artists exploited and innovated greatly the methods of producing and performing music. In the first period of computer music, a lot of research effort was dedicated to sound synthesis and modeling. New synthesis algorithms were discovered, such as frequency modulation, and new paradigms were developed for musical sound generation, such as spectral and physical models. On the other side, models for music representation and algorithmic composition were developed.

Less attention was being paid to the performance aspects. The music was automatically generated from the score as it was written by the composer, or generated by a composition program. The composer had to take into account all the nuances often implicit in the score to communicate the expressive content of the music. In this situation, the composer must explicitly preview what the performer normally handles. The composer is also a performer and needs to formalize the performance process. A different approach, to overcome the limitations of computer-generated music, was taken by music for live electronics. Here the performer interacts with technology on the stage, transforming the sound produced by traditional or synthetic instruments in real time.

In both cases, a central challenge is the control of the sound synthesis or processing engines (systems, algorithms, etc.). This problem is a typical performance topic and it refers to the need of establishing and computing the relations between musical and compositional aspects and sound parameters according to the expressive aim of the musician. The inputs are discrete events, as described in the score or generated by computer, and continuous signals, e.g., performer gestures. The most widespread computer environment for realizing live electronic music are the Max paradigm based languages Max/MSP, jmax, and Pd (Puckette, 1991, 1997, 2002) and the Eyesweb platform (Camurri et al., 2000, 2004). Many new input devices for music expression have been proposed (see, e.g., Paradiso (1997); Wanderley and Orio (2002)). The inputs should be coordinated and merged to produce and process sound events. In music technology, the concept of mapping strategies which describe these relations is of great importance. The conventional (and simplest) approach refers to specific relations; for example, how to convert pitch and loudness information into proper spectral and micro-timing values of a synthetic note. Nevertheless, the work strategies tend to include other possible choices and sources of information as phrasing, musical character, mood of the performer, and stylistic alternatives. All these aspects are typical music performance issues, and suitable music performance models are very desirable.

Figure 4 the typical situation of music performance with digital instruments where the electronic instrument performer controls the sound synthesis with gestures and suitable processes. A performance model lies between the symbolic and the audio control level. The performer receives audio feedback from the instrument as with traditional instruments. In live electronics, the scheme is different (see Figure 5). Here the live electronics performer processes the sound produced by the instrument performer, acting on his computer. In the live electronics box, we still have score processes and gestures controlling the sound processing devices via a performance model. However, in this case the input is sounding music, already performed. In a certain sense, we have a combined effect of performances (e.g. deviations of deviations) that the models should take into account. The performer receives audio feedback from both the instrument and the sound processing (Vidolin, 1997).

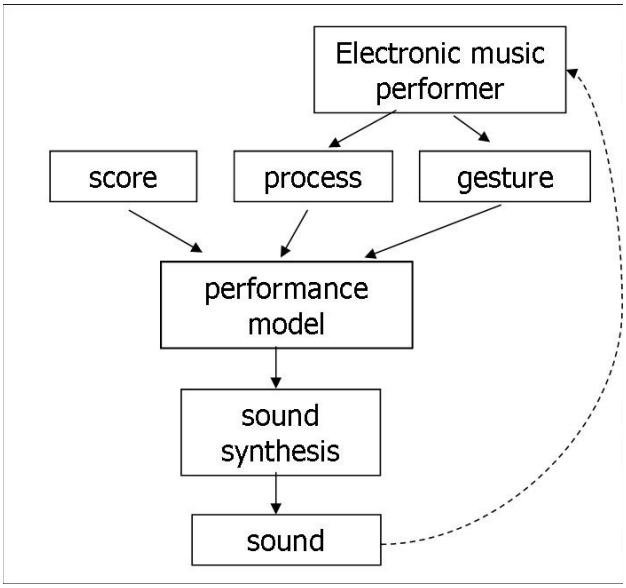

Figure 4. Scheme of music performance with digital instruments where the electronic instrument performer controls the sound synthesis with gestures and suitable processes. A performance model lies between the symbolic and the audio control level. The performer receives an audio feedback from the instrument as with traditional instruments.

5. CONCLUSIONS

IRecently music performance researchers are becoming more aware of the need of a well-founded approach based on strong scientific knowledge. This aim can be faced from two complementary directions. One way is to start from the knowledge gained in classical music performance studies and formalized in performance models, then generalize their results and apply them to the performance of new music creations. The other direction starts from the practical knowledge of new music creators (often embodied in their music performance systems) in order to extract possible suggestions and proposals of new performance models. From the joint effort of scientists and musicians valid results can be expected, and real new tools can be developed, not only inspired by problems and solutions of past times.

It can be noticed that music performance is an interesting topic for scientific investigation and for technology research; it involves human non-verbal communication, has artistic-creative finality, and requires strong cooperation between art and science - technology. Probably still more important is the fact that music is an immaterial art that has a strong tradition of symbolic representation and abstract thinking. This attitude may explain why musicians were the most enthusiastic and successful in promoting and contributing to the joint development of art and science since the beginning of computer science. In other arts, this collaboration started much later and very often it is restricted to the use of technology rather than a real contribution to a joint development of knowledge and tools.

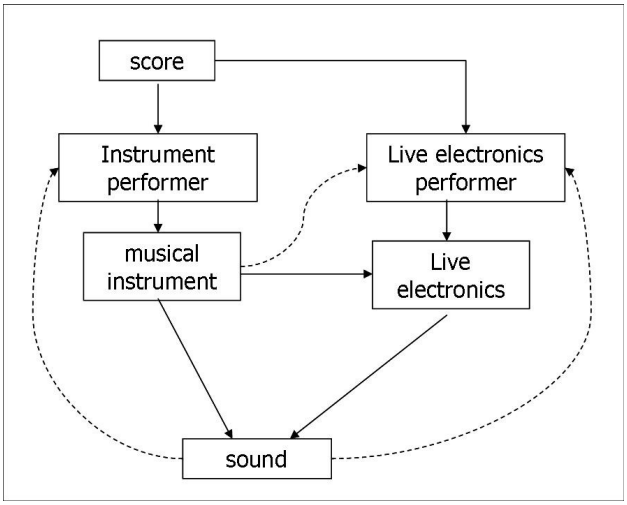

Figure 5. Scheme of live electronic music performance. The live electronics performer processes the sound produced by the instrument performer, acting on his computer. The live electronics box, merging score processes and gestures, controls the sound processing devices via a performance model. The performer receives audio feedback from both the instrument and the sound processing.

Acknowledgement

This overview was prepared partly thanks to funding from the Enactive Interfaces Network of Excellence under the 6th Framework program of the European Commission, and of the National PRIN03 project Sound and control codesign.

References

J. L. Arcos and R. L. de Mántaras. An interactive case-based reasoning approach for generating expressive music.applied intelligence. *Applied Intelligence*, 14(1): 115–129, 2001.

J. L. Arcos, R. L. de Mántaras, and Xavier Serra. Saxex: A case-based reasoning system for generating expressive musical performances. *Journal of New Music Research*, pages 194–210, September 1998.

G.U. Battel and R. Fimbianti. Expressive intentions in five pianists' performances. *General Psychology*, 3:277–295, 1999.

R. Bresin. Artificial neural networks based models for automatic performance of musical scores. *Journal of New Music Research*, 27(3):239–270, 1998.

R. Bresin and A. Friberg. Emotional coloring of computer controlled music performance. *Computer Music Journal*, 24(4):44–62, 2000.

R. Bresin and A. Friberg. Expressive musical icons. In *Proceedings of the 2001 International Conference on Auditory Display (ICAD'2001)*, pages 141–143, Espoo, Finland, 2001.

R. Bresin, G. De Poli, and R. Ghetta. A fuzzy approach to performance rules. In *Proceedings of XI CIM Colloquium on Musical Informatics*, pages 163–168, Bologna, 1995a.

R. Bresin, G. De Poli, and R. Ghetta. A fuzzy formulation of kth performance rule system. In *Proceedings of. 2nd Int. Conference on Acoustics and Musical Research CIARM 95*, pages 433–438, Ferrara, 1995b. CIARM.

E. Cambouropoulos. Musical rhythm: a formal model for determining local boundaries, accents and meter in a melodic surface. In M. Leman, editor, *Music, Gestalt and Computing - Studies in Systematic and Cognitive Musicology*, pages 277–293. Springer Verlag, 1997.

E. Cambouropoulos, S.E. Dixon, W. Goebl, and G. Widmer. Human preferences for tempo smoothness. In H. Lappalainen, editor, *Proceedings of the Seventh International Symposium on Systematic and Comparative Musicology, Third International Conference on Cognitive Musicology*, pages 18–26, Jyväskylä, Finland:, 2001. University of Jyväskylä.

A. Camurri, S. Hashimoto, M. Ricchetti, R. Trocca, K. Suzuki, and G. Volpe. Eyesweb - toward gesture and affect recognition in interactive dance and music systems. *Computer Music Journal*, 24(1):57–69, 2000.

A. Camurri, B. Mazzarino, and G. Volpe. Analysis of expressive gesture: the eyesweb expressive gesture process- ing library. In A. Camurri and G. Volpe, editors, *Gesture-based Communication in Human-Computer Interaction*, LNAI 2915, pages 20–39. Springer Verlag, 2004.

S. Canazza, G. De Poli, C. Drioli, A. Rodà, and A. Vidolin. Audio morphing different expressive intentions for multimedia systems. *IEEE Multimedia*, 7(3):79–83, 2000.

S. Canazza, G. De Poli, A. Rodà, and A. Vidolin Vidolin. An abstract control space for communication of sensory expressive intentions in music performance. *Journal of New Music Research*, 32(3):281–294, 2003.

S. Canazza, G. De Poli, C. Drioli, A. Rodà, and A. Vidolin. Modeling and control of expressiveness in music performance. *The Proceedings of the IEEE*, 92(4):686–701, 2004.

E.F. Clarke. Expression in performance: generativity, perception and semiosis. In J. Rink, editor, *The Practice of Performance: Studies in Musical Interpretation*, pages 21–54. Cambridge University Press, Cambridge, 1995.

M. Clynes. Superconductor: The global music interpretation and performance program. Available on-line at: http://www.microsoundmusic.com/clynes/superc.htm, 1998.

R.B. Dannenberg and I. Derenyi. Combining instrument and performance models for high-quality music synthesis. *Journal of New Music Research*, 27(3):211–238, 1998.

R.B. Dannenberg, B. Thom, and D. Watson. A machine learning approach to musical style recognition. In *Proceedings of the 1997 International Computer Music Conference*, pages 344–347, San Francisco, CA, 1997. International Computer Music Association.

G. De Poli, L. Irone, and A. Vidolin. Music score interpretation using a multilevel knowledge base. *Journal of New Music Research*, 19(2-3):137–146, 1990.

G. De Poli, A. Rodà, and A. Vidolin. Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance. *Journal of New Music Research*, 27(3):293–321, 1998.

P. Desain and H. Honing. Towards a calculus for expressive timing in music. *Computers in Music Research*, 3:43–120, 1991.

P. Desain, H. Honing, and R. Timmers. Music performance panel: NICI / MMM position statement. Available on-line at http://www.nici.kun.nl/mmm/papers/NICI-position.pdf, 2001.

A. Friberg. Matching the rule parameters of phrase arch to performances of 'Träumerei': A preliminary study. In *Proceedings of the KTH symposium on Grammars for music performance*, pages 37–44, Stockholm, 1995. A. Friberg J. Sundberg (ed.).

A. Friberg, L. Frydén, L. G. Bodin, and J. Sundberg. Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal*, 15(2):49–55, 1991.

A. Friberg, R. Bresin, L. Frydén, and J. Sundberg. Musical punctuation on the microlevel: automatic identification and performance of small melodic units. *Journal of New Music Research*, 27(3):271–292, 1998.

A. Friberg, V. Colombo, L. Frydén, and J. Sundberg. Generating musical performances with director musices. *Computer Music Journal*, 24(3):23–29, 2000.

A. Friberg, E. Schoonderwaldt, P. Juslin, and R. Bresin. Automatic real-time extraction of musical expression. In *Proceedings of the 2002 International Computer Music Conference*, pages 365–367, San Francisco, CA, 2002. International Computer Music Association.

A. Gabrielsson. Performance of rhythm patterns. *Scandinavian Journal of Psychology*, 15(1):63–72, 1974.

A. Gabrielsson. Interplay between analysis and synthesis in studies of music performance and music experience. *Music Perception*, 3(1):59–86, 1985.

A. Gabrielsson. The performance of music. In D. Deutsch, editor, *The Psychology of Music*, pages 201–602. Academic Press, San Diego, CA, 2nd edition, 1999.

A. Gabrielsson and P. Juslin. Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychology of Music*, 24(1): 68–91, 1996.

H. Honing. POCO: An environment for analysing, modifying, and generating expression in music. In *Proceedings of the 1990 International Computer Music Conference*, pages 364–368, San Francisco, CA, 1990. International Computer Music Association.

H. Honing. From time to time: The representation of timing and tempo. *Computer Music Journal*, 25(3):50–61, 2001.

H. Honing. Structure and interpretation of rhythm and timing. *Tijdschrift voor Muziektheorie*, 7:227–232, 2002.

O. Ishikawa, Y. Aono, H. Katayose, and S. Inokuchi. Extraction of musical performance rule using a modified algorithm of multiple regression analysis. In *Proceedings of the 2000 International Computer Music Conference*, pages 348–351, San Francisco, CA, 2000. International Computer Music Association.

P. N. Juslin. Communicating emotion in music performance: a review and a theoretical framework. In P. Juslin and J. Sloboda, editors, *Music and Emotion: Theory and Research*. Oxford University Press, Oxford, 2001.

F. Lerdahl and R. Jackendorff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, MA, 1983.

M.V. Matthews and F.R. Moore. GROOVE – a program to compose, store, and edit functions of time. *Communications of the ACM*, 13(12):715–721, 1970.

G. Mazzola and O. Zahorka. The RUBATO performance workstation on NEXTSTEP. In *Proceedings of the 1994 International Computer Music Conference*, pages 102–108, San Francisco, CA, 1994. International Computer Music Association.

E. Narmour. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. The University of Chicago Press, Chicago, IL, 1900.

C. Palmer. Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):331–346, 1989.

C. Palmer. Music performance. *Annual Review Psychology*, 48:115–138, 1997.

J. Paradiso. New ways to play: electronic music interfaces. *IEEE Spectrum*, 34(12):18–30, 1997.

R.W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, 1997.

M. Puckette. Combining event and signal processing in the MAX graphical programming environment. *Computer Music Journal*, 15(3):68–77, 1991.

M. Puckette. Pure data. In *Proceedings of the 1997 International Computer Music Conference*, pages 43–46, San Francisco, CA, 1997. International Computer Music Association.

M. Puckette. Max at seventeen. *Computer Music Journal*, 26(4):31–43, 2002.

B. H. Repp. Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's 'Träumerei'. *Journal of the Acoustical Society of America*, 92:2546–2568, 1992.

B. H. Repp. The aesthetic quality of a quantitatively average music performance: two preliminary experiments. *Music Perception*, 14(4):419–444, 1997.

K. Scherer. Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, 33(3): 247–259, 2004.

E. Schubert. Continuous measurement of self-report emotional response to music. In P. Juslin and J. Sloboda, editors, *Music and Emotion: Theory and Research*, pages 393–414. Oxford University Press, Oxford, 2001.

C.E. Seasore, editor. *Objective Analysis of Music Performance*. University of Iowa Press, Iowa City, IA, 1936.

I. H. Shaffer, E.F. Clarke, and N. P. Todd. Meter and rhythm in piano playing. *Cognition*, 20(1):61–77, 1985.

J. Sundberg, A. Askenfelt, and L. Frydén. Musical performance: a synthesis-by-rule approach. *Computer Music Journal*, 7(1):37–43, 1983.

J. Sundberg, A. Friberg, and L. Frydén. Rules for automated performance of ensemble music. *Contemporary Music Review*, 3(1):89–109, 1989.

J. Sundberg, A. Friberg, and L. Frydén. Threshold and preference quantities of rules for music performance. *Music Perception*, 9(1):71–92, 1991.

K. Suzuki and S. Hashimoto. Robotic interface for embodied interaction via dance and musical performance. *The Proceedings of the IEEE*, 92(4):656–671, 2004.

T. Suzuki and H. Tokunaga, T. andTanaka. A case based approach to the generation of musical expression. In *Proceedings of 1999 IJCAI*, pages 642–648, 1999.

R. Timmers and H. Honing. On music performance, theories, measurement and diversity. *Cognitive Processing (International Quarterly of Cognitive Sciences)*, 1/2:1–19, 2002.

N. P. Todd. A model of expressive timing in tonal music. *Music Perception*, 3(1):33–58, 1985.

N. P. Todd. The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, 91(6):3540–3550, 1992.

N. P. Todd. The kinematics of musical expression. *Journal of the Acoustical Society of America*, 97(3):1940–1949, 1995.

A. Vidolin. Musical interpretation and signal processing. In C. Roads, S.T. Pope, A. Piccialli, and G. De Poli, editors, *Musical Signal Processing*, pages 439–459. Swets & Zeitlinger, Lisse, 1997.

M. Wanderley and N. Orio. Evaluation of input devices for musical expression: borrowing tools from HCI. *Computer Music Journal*, 26(3):62–76, 2002.

T Weyde and K. Dalinghaus. Design and optimization of neuro-fuzzy-based recognition of musical rhythm patterns. *International Journal of Smart Engineering System Design*, 5(2):67–79, 2003.

G. Widmer. A machine learning analysis of expressive timing in pianists' performances of Schumann's "Träumerei". In A. Friberg and J. Sundberg, editors, *Proceedings of the KTH Symposion on Grammars for Music Performance*, pages 69–81, Stockholm, 1995a. Department of Speech Communication and Music Acoustics.

G. Widmer. Modeling rational basis for musical expression. *Computer Music Journal*, 19(2):76–96, 1995b.

G. Widmer. Learning expressive performance: The structure-level approach. *Journal of New Music Research*, 25(2):179–205, 1996.

G. Widmer. Machine discoveries: a few simple, robust local expression principles. *Journal of New Music Research*, 31(1):37–50, 2002.

G. Widmer. Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146(2):129–148, 2003.

G. Widmer and A. Tobudic. Playing Mozart by analogy: learning multi-level timing and dynamics strategies. *Journal of New Music Research*, 32(3):259–268, 2003.

G. Widmer and P. Zanon. Automatic recognition of famous artists by machine. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI'2004)*, pages 1109–1110, Valencia, Spain, 2004.

P. Zanon and G. De Poli. Estimation of parameters in rule system for expressive rendering of musical performance. *Computer Music Journal*, 27(1):29–46, 2003a.

P. Zanon and G. De Poli. Time-varying estimation of parameters in rule systems for music performance. *Journal of New Music Research*, 32(3):295–315, 2003b.

P. Zanon and G. Widmer. Learning to recognize famous pianists with machine learning techniques. In R. Bresin, editor, *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03)*, volume 2, pages 581–584, Stockholm, 2003. Department of Speech, Music, and Hearing, Royal Institute of Technology.