Review

Research in computational expressive music performance and popular music production: a potential field of application?

Pierluigi Bontempi [1], Filippo Carnovalini [2], Antonio Rodà [3] and Sergio Canazza [4]

1 University of Padua, Department of Information Engineering; pierluigi.bontempi@phd.unipd.it
2 University of Padua, Department of Information Engineering; filippo.carnovalini@dei.unipd.it
3 University of Padua, Department of Information Engineering; roda@dei.unipd.it
4 University of Padua, Department of Information Engineering; canazza@dei.unipd.it
* Correspondence: canazza@dei.unipd.it

Abstract: In music, the role of the interpreter is to play her/his part manipulating the performance parameters in order to offer a sonic rendition of the piece capable of conveying specific expressive intentions. Since the 1980s there has been a growing interest in computational expressive music performance (EMP). This research field has two fundamental objectives: the understanding of the phenomenon of human musical interpretation and the automatic generation of expressive performances. Rule based, statistical, machine and deep learning approaches have been proposed, most of them devoted to the classical repertoire, in particular to piano pieces. On the contrary, we present an introduction to the role of expressive performance within popular music and to the contemporary ecology of pop music production, based on the use of Digital Audio Workstations (DAWs) and virtual instruments. After an analysis of the tools related to expressiveness commonly available to modern producers we propose a detailed survey of research into the computational EMP field, highlighting the potential and limits of what is present in literature with respect to the context of popular music, which by its nature cannot be completely superimposed on the classical one. In the concluding discussion we suggest possible lines of future research in the field of computational expressiveness applied to pop music.

Keywords: computational music expressive performance, popular music, music production, Digital Audio Workstation, virtual instruments

check for updates

Citation: Bontempi, P.; Carnovalini, F.; Rodà, A.; Canazza, S.; Title. *Preprints* **2022**, *1*, 0. https://doi.org/

1. Introduction

When a skilled musician plays a piece of music, s/he usually does not do it mechanically, keeping a perfectly steady timing without any concession to loudness or timbre variations, or possibly to the use of embellishment techniques. Actually, expert musicians commonly manipulate at least some of the performance parameters to produce expressive sonic renditions of the pieces:

> *[...] performers are able to use systematic variations in performance parameters to convey emotion and structure to listeners in a musically sensitive manner* [1, pg. 64].

With regard to western classical music, traditionally expressiveness has been defined in terms of deviations from what is prescribed in the score:

> *When playing a piece, expert performers shape various parameters (tempo, timing, dynamics, intonation, articulation, etc.) in ways that are not prescribed by the notated score, in this way producing an expressive rendition that brings out dramatic, affective, and emotional qualities that may engage and affect the listeners* [2, pg. 1].

On the other hand, in the context of popular music the traditional separation between composer, author of the score, and performer, who interprets and gives a sonic rendition of the same, is only rarely fully appropriate. Musical parts, for example, can be improvised, or can be based on more or less detailed lead sheets (an essential form of musical notation

 doi:10.20944/preprints202212.0494.v1

that describes the fundamental elements of a song), or can be created and memorized by the musician without the need of a formal notation. Moreover, with the spread of computer assisted music production, popular music creation has shifted more and more towards a model in which the traditional composer/performer distinction further loses meaning, in favor of a deep integration of activities associated with music composition, engineering, production, and performance [3].

This does not mean that the expressiveness of the performance is not relevant in popular music, although with some specific exceptions: in electro-pop "deadpan" performances, commonly considered inexpressive and not desirable in classical music, can fall within sought after aesthetic intentions, as well as the construction of musical parts that can only be played by machines and not by humans [4]. Apart from these specific cases, human-like expressivity in popular music performance plays generally a crucial role.

Within computer assisted music production the producer often creates - through mouse and keyboard editing and/or use of dedicated hardware controllers - MIDI parts associated with virtual instruments [5]. Clearly, simply using the mouse or the keyboard to insert into the rhythm grid of the Digital Audio Workstation (DAW) the MIDI notes that the virtual instrument will have to sonorize is not enough to produce something that can be perceived as expressive, nor is it easy to obtain realistic and expressive sounding parts using a MIDI controller, because of the possible technical limitations of the controller and/or the skills required from the performer. Manually refining parts so that they result more expressive is only partially possible, and it is undoubtedly tiring and time consuming. The commercially available tools (the ones commonly used by pop producers, see Section 2 for some examples) cover only partially the needs related to expressive performance.

Innovative solutions capable of automatically or semi-automatically conferring expressiveness to the MIDI parts produced would be highly valuable, and it is our opinion that existing academic research on Expressive Music Performance (EMP) could help improve the tools available or suggest directions to develop new ones.

The following part of this contribution is organized as follows. In Section 2, an analysis of the strengths and limitations of what is commercially available in relation to musical expressiveness in computer-assisted popular production is offered. In Section 3, existing EMP literature is reviewed; the contributions are organized in thematic subsections such as relation between expressiveness and structure, local expressiveness, relationship between emotional intention and expressive parameters and so on. At the end of each subsection a comment about the possible applications or limits of the EMP literature presented in relation to popular music computer assisted production is provided. Section 4 is devoted to general discussion and conclusions.

2. Commercial products

Digital Audio Workstations, since the early 2000s, have become central to the popular music production workflow, and are now a fundamental working tool for professionals as well as for "bedroom" producers [6]. Among the benefits of DAWs there is the

> *ability to build up complex musical arrangements using realistic-sounding virtual instruments* [7, pg. 78].

Virtual instruments can be based on audio samples of real instruments as well as on sound synthesis techniques [8]. The Virtual Studio Technology (VST) standard is probably the most widespread and used to build virtual instruments and audio effects [9].

When a popular music producer uses a virtual instrument in an arrangement, s/he can program a MIDI part asynchronously, or play the part in real-time using a MIDI controller (usually, but not necessary, a keyboard-shaped one [10]). In both cases, making the part sound expressive is not a trivial task, despite the extensive tonal and dynamic potential of the modern virtual instruments.

 doi:10.20944/preprints202212.0494.v1

To obtain an updated list of tools and practices commonly used in popular music production and concerning expressiveness, we entered the following keywords in the search fields of the websites of the most accredited sector magazines: Sound on Sound (https://www.soundonsound.com), Tape OP (https://tapeop.com/), MusicRadar (https://www.musicradar.com/) – which includes the magazines Future Music, Computer Music, Music Tech:

- expressiveness;
- expressive performance;
- virtual performance;
- virtual performer.

Up to sixty results were consulted (when available) per search key per portal. Once the results not relevant for the purposes of this contribution have been filtered, examples have been selected from the rest that are useful for providing a general overview of the sector. In the event of the presence of distinct commercial products, but similar in terms of operating principles and construction/development logic, not all of them have necessarily been mentioned, as the objective of this contribution is not to offer a complete overview of what the market offers, including each individual product, but to account for the main tools associated with expressiveness generically available to contemporary music producers. The results of the research are categorized in thematic areas and presented in the following subsections of Section 2.

2.1. Common tools

Among practically any recent virtual instrument there are some basic parameters that can be controlled using MIDI messages and/or DAW automations, namely equalisation (EQ), high pass filter (HPF) - possibly with settable Resonance, and, sometimes, compression. More advanced sound control potential may be present.

It's common for sampled instruments to have multiple samples of each note, usually selected and loaded based on the MIDI velocity of the note played. Sample based virtual instruments frequently also implement the round-robin technique (when multiple consecutive requests of the same note are received, the instrument loads different samples, in random or predetermined order, to avoid the so-called "shotgun effect", the repetition of a series of identical sounds, an event perceived as unnatural in the musical field).

Lastly, DAWs can usually randomize MIDI data, in compliance with the range of relative values set by the user. This is sometimes done to alleviate the feeling of unnaturalness that quantized, fixed velocity parts transmit [11].

All these expressive tools and tricks need to be triggered (in a static or – when possible and appropriate - dynamic fashion) using offline editing or MIDI controllers [12]. That translates into a wide range of possible sound nuances, that anyway must be controlled manually by the producer, being it in real time or not.

2.2. Triggerable instrument-specific patterns

Especially among virtual instruments that model physical ones, the presence of patterns that can be recalled by the user is frequent. These are normally pre-recorded or pre-programmed performances of short parts, usually structured so that they can be repeated in a seamless loop. Sometimes, but not necessarily, these patterns can be customized by the user, directly within the virtual instrument or using the MIDI editing tools of the DAW.

Some of the products that include this functionality - among many others - are:

- Steinberg Groove Agent 5;
- UJAM collection (Virtual Guitarist, Virtual Bassist, Virtual Drummer, Virtual Pianist);
- Native instruments Action Strings and Emotive Strings.

This approach - that could be considered somehow an evolution of traditional arpeggiators - can be useful among producers that are not able or interested in writing the instrument parts note by note. Moreover, some degree of expressivity can be included in the patterns, what kind exactly depending on how the pattern was realized and on the sonic capabilities of the virtual instrument. Sometimes the patterns can be modified in real-time setting parameters such as Complexity and Intensity in Steinberg's percussive oriented sampler Groove Agent 5 (Figure 1): there Complexity (x-axis) loads richer, more nuanced patterns as you move to the right, while Intensity (y-axis) affects the MIDI velocity of the strokes.

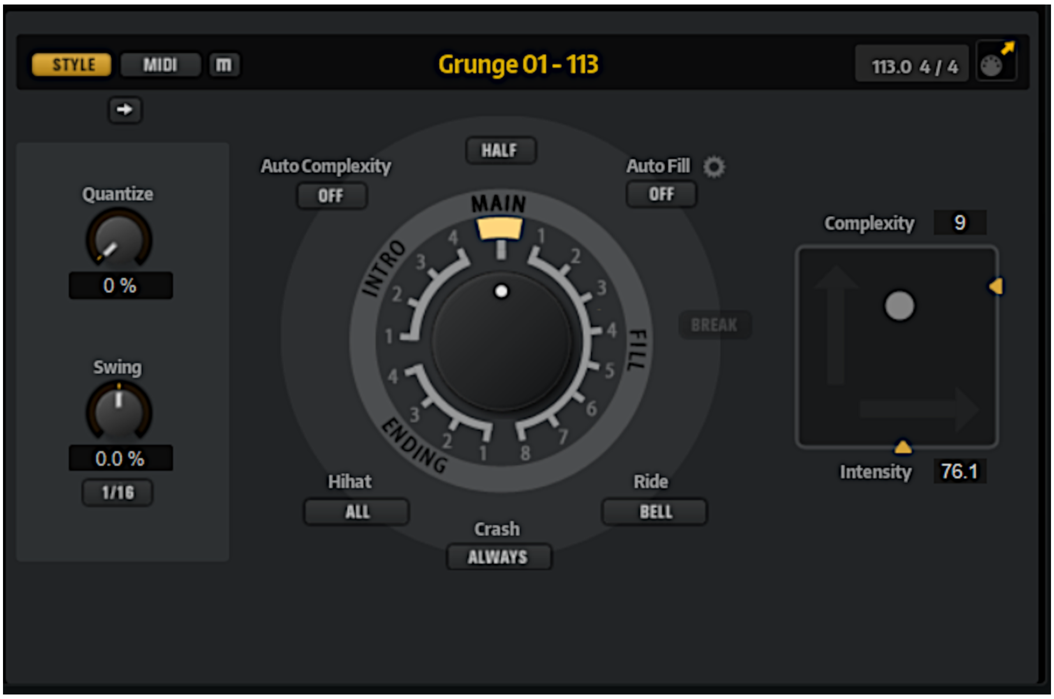

Figure 1. The pattern management area in Steinberg Groove Agent 5. On the right it can be seen the two-dimensional control surface (that works also in real time) dedicated to the Intensity and Complexity parameters. Courtesy of Steinberg Media Technologies GmbH.

Significantly, the review of the UJAM Virtual Bassist Collection published on Sound on Sound [13, pg. 130] says:

> *[. . .] my completely virtual band of session musicians just needing pointing in the right direction. Add in some 'human' with a few guitar overdubs and some vocals, and a song idea can be fleshed out very quickly. What's more, the virtual band sounds very polished indeed.*

This quotation highlights two common limitations of the pre-made patterns approach: they can hardly be completely "tuned" to the expressive intention of the specific piece of music under processing (making a human contribution necessary, in the form of human played added parts or manual editing of the patterns), and they usually sound very polished, maybe even too polished to be perceived as truly "real". Moreover, having pre-established patterns, albeit sometimes customizable, places this type of resource between expressiveness and automatic/assisted music generation, touching territories outside the boundaries of this contribution.

2.3. Triggerable instrument articulations

In the jargon of contemporary music production, *articulations* mean different timbres or performative techniques that can be associated with the virtual instrument. For example, the same notes can be played by a virtual violin (among other possibilities) with *detaché*, *staccato*, or *col legno* articulations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

This approach is very common among sample based virtual instruments belonging to the classical orchestra category (strings, brasses, woodwinds), but can potentially be found in about any kind of virtual instrument.

Commercial products falling into this category are offered by manufacturers such as EastWest, IK Multimedia, Vienna Symphonic Library, Steinberg, Native Instruments, Spitfire Audio. Similarly to what has been said in the Common Tools section 2.1, also in articulations triggering a manual intervention of the producer is always needed.

2.4. Advanced hardware timbre and expressivity control

On the controller front, many attempts to produce innovative and powerful MIDI devices, sometimes called hyper-instruments (term used in particular in the case of traditional instruments modified to act as MIDI controllers [14]), capable of manipulating multiple parameters in real time, can be cited:

- the pioneering Kurzweil XM1 Expression Mate, that dates to 2000;
- ROLI Seaboard family, whose operation is traditionally based on the MPE (MIDI Polyphonic Expression) protocol (Figure 2);
- Expressive E Touché and Osmose;
- Keith McMillen SoftStep 2;
- Erae Touch (which makes use of MIDI 2.0).

Figure 2. The MIDI controller ROLI Seaboard Rise. Courtesy of ROLI Ltd.

These tools can be touch and/or velocity-sensitive, and often allow control of pitch bend, polyphonic aftertouch, microtonal slides, and in general MIDI Control Changes/Continuous Controllers (CCs) through gestures.

In academia, that can anyway have direct impact on commercial products, the International Conference on New Interfaces for Musical Expression (NIME - https://www.nime.org/), held since 2001, is of particular relevance [15].

Once again, these tools are designed to allow the producer direct control of the parameters of the virtual instrument; they are not able to generate autonomously musical expressiveness, nor they are usually aimed at that.

2.5. Automatic analysis of the harmonic structure and generation of new musical parts

In some cases, not particularly widespread, the virtual instrument is able to automatically generate parts on the basis of the data obtained from the analysis of the parts of other instruments or of the harmonic progression.

One of the best representatives of this approach is Toontrack EZ Bass (Figure 3), that can analyse MIDI or audio parts and generate matched bass lines, highly expressive thanks to the use of different articulations, embellishments, velocity variations, etc. Similarly to what has been said about pattern based virtual instruments, also in this case it is questionable whether this kind of capability should be included in the automatic composition/arrangement area, or in the expressive performance one. Nonetheless the bass lines generated by the software present expressiveness, and – even more important – are context-aware, at least with regard to some parameters/musical dimensions, things that are of prime relevance in the field under consideration.

Figure 3. Toontrack EZ Bass. Courtesy of Toontrack Music AB.

2.6. The missing link

From what has been observed in the previous subsections it is evident that, in the current panorama, extremely sonically versatile virtual instruments with rich expressive potential are available on a commercial level, which must however be controlled in real time (there is no shortage of hardware tools for sophisticated real-time control) or through offline MIDI sequencing by the producer. Many of them provide pre-established - sometimes customizable - patterns, and some can generate context aware ones. However, these last features, more than in EMP, fall within an area on the border between expressiveness and automatic or computer aided composition. What is missing here are tools capable of processing musical parts provided by the producer - potentially time-quantized and fixed-velocity ones - making them expressive acting automatically on tone parameters of the virtual instrument, time deviations and - possibly - on the automatic selection of the appropriate articulations or embellishments. To be completely effective, these potentials should then be sensitive to the context, and possibly to high-level indications (for example, to emotional intention) provided by the producer. Could academic research in computational expressive music performance help? Considering also that historically EMP has addressed mostly classical music for solo instruments (see below), which research lines could find direct application in popular music? Which adaptations, if any, would be needed?

 doi:10.20944/preprints202212.0494.v1

3. Research products

3.1. A multifaceted field of research

Computational Expressive Music Performance (EMP) is a particularly rich and multifaceted field of research, which sits at the intersection of different approaches and areas of expertise. Several academic disciplines are involved (computer science, musicology, psychology). It is therefore not surprising that academic research has generated diverse approaches with respect to the type(s) of technology employed and the specific aspect(s) of computational EMP addressed.

The first scientific contributions to the topic date back at least to the 80s, and more than one review work has already been published [2,16–20]. There are two main purposes associated with EMP computational models: they can be used as an analytical tool for understanding how humans perform music, or to generate new performances of musical pieces, in many different contexts [2]. Actually, the two things are connected: to develop models capable of credible virtual performances it is necessary to understand what makes human performances worthy of interest and expressive. Also for what concerns the technologies involved in the existing models we can identify, generally speaking, two areas: data driven and rule based models [2]. The former category relies on machine learning, probabilistic, and Artificial Neural Networks (ANNs) approaches (and consequently on large collections of data), the latter on manually designed rules, based on musical hypothesis. Another commonly used nomenclature is the one that contrasts *analysis by synthesis* to *analysis by measurement*: the former indicates the implementation of rules obtained from the dialogue with human experts, the latter the use of real performances parameters measurements to extract rules or other significant regularities.

Below, the main research lines that have been followed by scholars will be presented, at a medium to high level of abstraction. They will be divided into the following thematic areas:

- visual representation of expressive performances features;
- relation between expressiveness and structure of the musical piece;
- local expressiveness;
- score markings interpretation;
- relationship between emotional intention and expressive parameters;
- relationship between sensorial experiences and expressive parameters;
- identification and modelling of the performative styles of real musicians;
- identification of physical and psychological limits of the performer;
- ensemble music modelling;
- conductor systems.

We believe that this subdivision into thematic macro areas could prove to be particularly functional to a clear understanding of the computational EMP field, despite the fact that the musical parameters investigated may be common to all areas (timing, loudness, timbre, etc.). What changes is the logic that leads an approach to study or modify in a certain way specific parameters. Moreover, some of the research lines reviewed here address more than one area at the same time, and can therefore appear several times in the course of the treatment, in different subsections.

The main computational EMP systems under examination will then be recalled in Section 3.12, where for each of them a summary of the main characteristics will be provided in table form, specifically with regard to:

- technologies involved;
- user interaction;
- main goal(s).

A color based association of the systems in relation to the thematic macro areas will also be included.

Before getting to the heart of the discussion we consider relevant offering some general insights obtained through a meta reading of the recent literature review published by C.E.

Cancino-Chacón and co-authors [2], regarding the broad characteristics of the existing research about computational EMP. The following data are based on the summary tables presented there.

- Total number of reported models: 18;
- Reported expressive parameters: metronomic tempo, timing (notes onset deviations), dynamics (most of the times linked to MIDI velocity), articulation. Only once the ornamentation parameter is present [21] ;
- Total number of datasets reported: 12;
- Music genres of the datasets: classical (10), popular (2);
- Sources of the datasets: computer-controlled piano (8), audio recordings (3), audio recordings and computer-controlled piano (1) - the fact that the piano is the instrument of choice for research on musical expressiveness can be traced back to technical reasons: being based on the percussion of the strings and not allowing a continuous control of the timbre, as it happens for example in the violin, it is relatively simple to build functional expressive models taking into account a minimum number of parameters (timing and dynamics/velocity) [22,23]. Moreover, hybrid acoustic/digital instruments such as the Yamaha Disklavier allow for easy recording of MIDI data from human performances.

By no means the above summary should be considered exhaustive, but it gives a first rough idea of the field. What this entails on the subject of this contribution will be discussed later.

3.2. Visual representation of expressive performances features

Although a visual representation of expressive performance features is not a computational model for expressiveness, its relevance in the understanding of the human performing dynamics and subsequent construction of models must not be underestimated. For this reason we considered justified the inclusion of a dedicated subsection, comprising at least some of the most relevant experiences in the field.

Despite the fact that there are many examples of technical solutions for the detection and representation of single performative parameters, the overall perception of the performance is linked to the interaction of several parameters rather than to decontextualised, individual ones. A system for the real-time representation of performances in a two-dimensional graph, with tempo (bpm) on the x-axis and loudness on the y-axis was proposed by J. Langner and W. Goebl [24]. Along the two axis, a dot moves in synchrony with sound. The materials analyzed are piano performances played on a Bösendorfer computer-controlled grand piano (SE290). Timing information is extracted from MIDI data, while loudness data is extracted from the audio files of the performances. The trajectory of the dot (that gradually fades away, leaving behind a visible tail) is a visual description of the two most important parameters of the performance, tempo and loudness.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

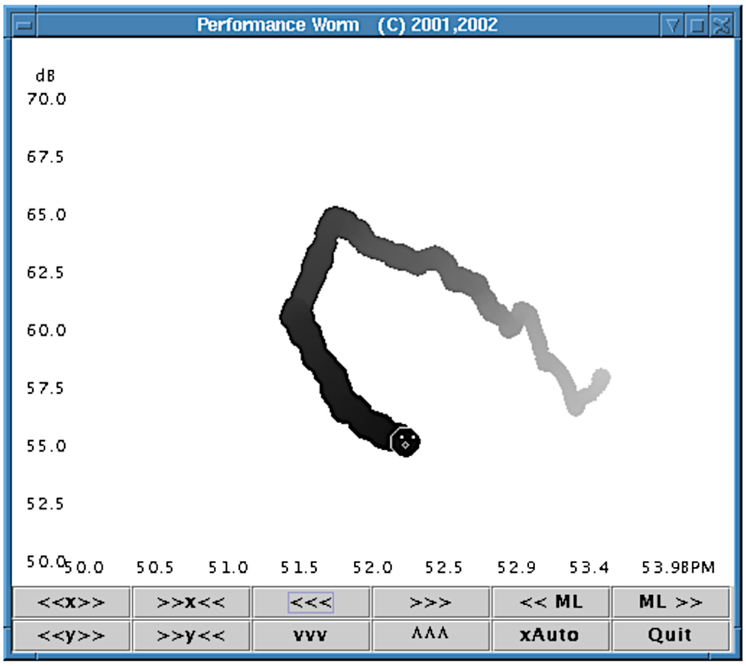

Figure 4. A screenshot of the performance worm. On the x-axis it is shown the tempo in beats per minute; on the y-axis the loudness in decibel [25].

This kind of visual representation has been subsequently taken up and evolved, with the extraction of time information directly from audio, and not from MIDI data, and the coinage of the term *performance worm* [26], which has also been used in other scientific publications [25] (see Figure 4).

Non-standard music transcription techniques, indicated for repertoires handed down orally or to report dimensions that escape standard Western notation, are for obvious reasons of primary interest in the field of ethnomusicology [27]. With the necessary differences (symbolic representation instead of direct *analogical* representation of the phenomenon [28]), it is also in this case a question of moving performative sound dimensions into the visual sphere. It does not seem unreasonable to us to think that reflections of this kind could have positive repercussions also in the EMP field, although the goals are clearly different in the two cases. A way of representing unambiguously the desired prosodic interpretation of melodies using a dedicated small alphabet $A = \{l^-, l^x, l^+, l, l, l^*\}$ and a deterministic mapping from the prosodically labeled score to sound synthesisis has been proposed by Christopher Raphael [22] (see Figure 5).

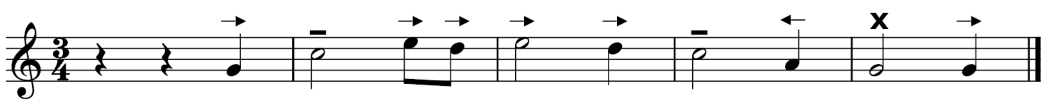

Figure 5. The melody of the popular tune Amazing Grace showing a custom note-level prosodic labeling (reconstruction of part of an example present in [22]).

Integrated graphical representations of expressive parameters such as the ones related to the *performance worm* [24–26] could be relevant in the implementation of popular music expressive models not only because they offer an intuitive and immediate tool for understanding expression dynamics that may be hard to catch otherwise by a human observer, but also because they could simplify the intuition and understanding of the analogies and differences between musical expressiveness in the historical/stylistic/instrumental fields to which academic research has traditionally turned and what happens in a popular context.

Also the reflection on the notation of dimensions of the musical performance not included in the Western standard writing can offer useful insights for the development of EMP models. In particular, we imagine possible applications in the field of conductor systems (see Section 3.11).

3.3. Relation between expressiveness and structure of the musical piece

Pieces of music can be described in terms of form or structure at different levels of abstraction. At the highest level we have the sections of the piece, as in the classic sonata form, divided into introduction, exposition, development, recapitulation and coda sections. At a lower - and probably more relevant for computational music expression - level we can identify structural elements such as motifs, phrases, and periods. These are defined in Grove Music Online (based on the renowned Oxford Dictionary of Music and Oxford Companion to Music) [29–31] respectively as:

A short musical idea, melodic, harmonic, rhythmic, or any combination of these three. A motif may be of any size, and is most commonly regarded as the shortest subdivision of a theme or phrase that still maintains its identity as an idea.

A term adopted from linguistic syntax and used for short musical units of various lengths; a phrase is generally regarded as longer than a Motif but shorter than a Period. It carries a melodic connotation, insofar as the term 'phrasing' is usually applied to the subdivision of a melodic line.

[...] a musical statement terminated by a cadence or built of complementary members, each generally two to eight bars long and respectively called 'antecedent' and 'consequent'.

Generally speaking, quite often the clarification of the piece structure is considered one of the main aims of expressive performance [32]. A radical (non-computational) approach, that substantially brings the performative expressiveness back to the structure of the piece, is that of E.F. Clarke [33], who bases the discussion on generative principles [34], in turn influenced by Schenkerian analysis.

Many attempts at modeling expressive parameters based on structure analysis are present in the literature. Tempo changes related to phrase structure in tonal music have been at the center of the seminal research of N.P.M. Todd [35]. The author's approach is based again on Lerdahl and Jackendoff's generative theory [34]. The piece is divided into nested hierarchical organized time spans. The model reflects the structure of the piece slowing at structural endings, in a more or less pronounced way depending on the hierarchical importance of the syntactic break (see Figure 6). Subsequently, Todd expanded his approach to include the computational modeling of rubato in relation to music structure [36] and the dynamics, based on the assumption that there is a correspondence between speed and intensity (the faster, the louder) [37]. Although subsequent publications have shown the relevant limitations of Todd's model [38,39], this remains a cardinal point in the evolution of the research field of computational musical expressiveness.

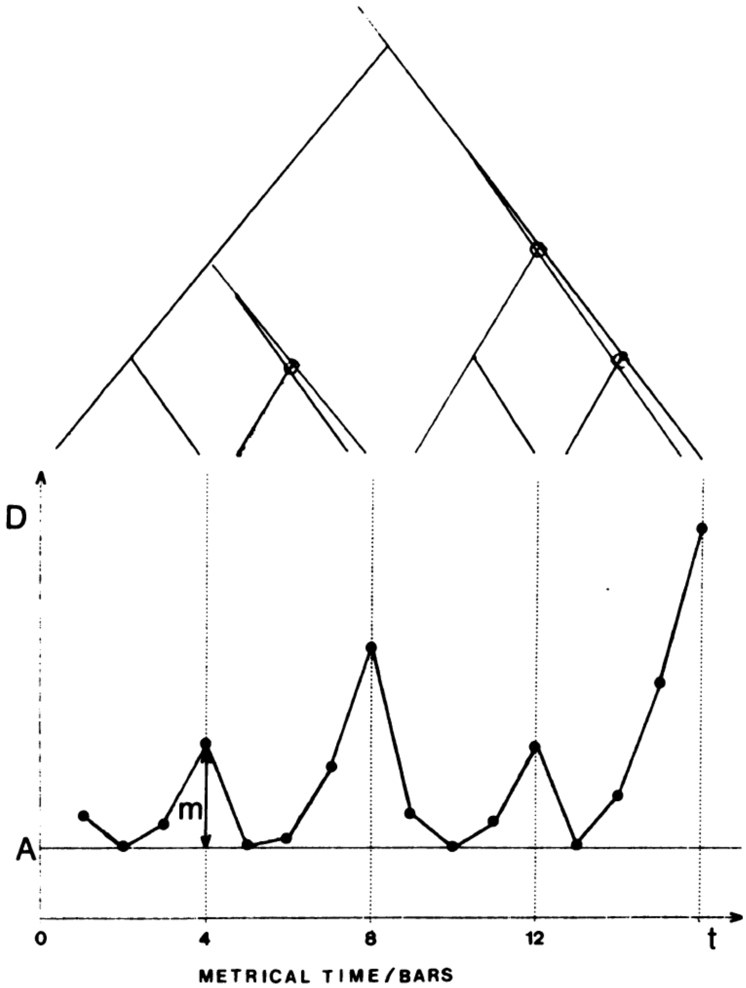

Figure 6. A tree diagram representing the hierarchical structure of the piece and, below, the corresponding tempo variations generated by the model (the higher the value on the y-axis, the more pronounced the slowdown) [35].

G. Grindlay and D. Helmbold proposed a hierarchical hidden Markov model (HHMM) to extract statistical data about the relations between score structure and associated performances (in particular with regard to time deviations) [40]. The model is based on a two-level hierarchical structure, with the top-level representing the musical phrase context, while the lower one is associated with note-level contexts. Once trained, the model is able to generate expressive performances, but can also be used to recognize individual performers.

A similar approach, based on the distinction and relation between a phrase level and a note level, can also be declined in a rule-based system, as shown by G. Widmer and A. Tobudic [41,42]. Note-level and phrase-level expressive patterns are combined there to generate predictions about complex composite expression curves for new pieces.

One of the most relevant EMP lines of research is that of the KTH rule system for musical performance [43]. Substantially based on the *analysis-by-synthesis* approach [44] (simplifiable into: selection and analysis of the performances and of the variables to study tentative synthesis of varying versions through the definition of rules human judgement of the performance and iterative return to the previous phase to improve the model), the system also provides a complementary *analysis-by-measure* approach (the rules are designed based on objective data derived from real performances). The rules affect various parameters such as timing, dynamics and articulation, and they are weighted. They are organized in the categories Phrasing, Micro-level timing, Metrical patterns and grooves,

 doi:10.20944/preprints202212.0494.v1

Articulation, Tonal tension, Intonation, Ensemble timing, and Performance noise; therefore they cover the piece of music at different levels. Specifically, in relation to the structural dimension, the KTH system suggests rules for the creation of phrase arch-like tempo and loudness contours, and for the insertion of a *ritardando* in the end of the piece.

F. Carnovlalini and A. Rodà proposed a system capable of generating brief melodies together with their expressive performance, all based on a multilayered hierarchical subdivision of the notes reminiscent of the Schenkerian approach [45]. The length and loudness of the notes are automatically set according to their hierarchical relevance, within a three levels model. The KTH phrase-arc rule is also implemented.

The structure of a piece can also often be analysed in terms of construction and resolution of tensions (a task usually easier within tonal music, associated with a more general logic of construction and resolution of expectations [46,47]). Computational expression models can refer to those tension patterns.

A three dimensional spiral representation of pitch classes, chords and keys to compute variations in tension during the piece of music was introduced by D. Herremans and E. Chew [48]; based on that, a computational study of the role of tonal tension in the prediction of expressive performances of classical piano music was presented by C. Cancino-Chacón and M. Grachten [49]. A computational model for the calculation of tempo and dynamic variations is there implemented using a bidirectional LSTM Recurrent Neural Network (which processes both backwards and forwards information). It is noted that the use of tonal tension as defined by Herremas and Chew is useful to predict expressive changes in tempo and dynamics, but not to predict specific values for those parameters. Tonal tension seems to be an additional, but not self-sufficient, resource in the creation of models for musical expressiveness.

The relation between music performance expressiveness and piece structure could be as relevant in popular as it is in classical music, but the subject needs deep analysis and investigation before being able to draw any conclusions, since the two contexts are only partially superimposable from a structural perspective. Moreover, the tonality related dynamics enhanced in classical music by some researchers could be relevant only in specific popular contexts, but would probably be out of focus and of little help in many others [50]. For this reason, the applicability of computational solutions for the automatic identification of the structure of the piece (aimed at the subsequent use of expressive models) based on tonal music theories [51,52], may not be ideal in many cases in popular music. The melodic analysis model suggested by Carnovalini and Rodà [45], based on the previous works of theirs and N. Orio and F. Simonetta [53,54], seems to be much more promising with regard to popular music, since it represents a more theory-agnostic approach. It classifies the notes of the piece within a three levels hierarchical structure based on metric position, relevance of the underlying chord with regard to tonality (but this could be easily adapted also to the modern modal approach to popular music composition), and relevance of the melodic note within the underlying chord.

3.4. Local expressiveness

Musical expressiveness can also be observed and modelled locally, note-wise or in relation to small groupings of notes.

Fundamental in this perspective were the seminal studies of A. Gabrielsson - among many others [44,55–57]. Gabrielsson's approach is measure-based: an analysis of real performances, looking for relevant performance variables, makes it possible to detect systematic variations (SYVARs) related to some type of norm. These variations, or deviations from the mechanical regularity, may vary according to the specific context (type of music, performer etc.). Gabrielsson investigated in particular rhythmic micro structures (e.g. a half note followed by a quarter one, or a dotted chrome followed by a sixteenth note, etc.), as well as deviations at half and full measure level, finding significant regularities, at least given the same context. Gabrielsson observed that:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

It seems safe to assume that such differences in performance are not primarily made in order to affect the perceived structure, but rather to contribute to a proper motional-emotional character of the music in question. [44, pg. 79]

Generating systematically varying sound sequences based on the findings of analysis and subjecting them to a human evaluation allows the testing of the validity of the detected SYVARs and of the relationship between them and the experiential psychological variables associated with the listener.

The expert system approach has also been investigated. In M.L. Johnson's model for the expressive rendition of Bach's fugues [58] the rules, based on the expertise of two professional performers, affect tempo and articulation, and are associated with specific rhythmic patterns.

Many of the rules of the KTH system [43] also address the *local EMP area*. For example, the Duration contrast rule does "*Shorten relatively short notes and lengthen relatively long notes*", the Double duration one does "*Decrease duration ratio for two notes with a nominal value of 2:1*". A combination of the rule based approach with Artificial Neural Networks and user interaction has also been tried [59].

Generally speaking, it seems that structure based expressiveness (see previous section) can be more easily appreciated by following an analysis by synthesis approach, while local expressiveness is more prone to data driven approaches, at least taking into consideration current technologies and the studies already carried out. This is confirmed in S. Oore and co-authors' work [23], where an LSTM network used to generate both piano solo musical parts (automatic composition) and their expressive performance shows better results on a local basis, more than in long term structure. In K. Teramura and co-authors' proposal [60] a Machine Learning Gaussian Process Regression predominantly based on local input data (durations and pitches of the notes concerned and of those belonging to the previous and subsequent measures are analyzed, as well as more general indicators such as meter and belonging to the melodic line) is used to render expressive performances of music scores. Similar parameters are considered in the statistical model YQX [61], together with principles taken from E. Narmour's implication-realization model [47]. Binary tree based clustering has been explored by K. Okumura and co-authors to classify the local context of the notes; to each context is then applied a specific stochastic model (see Figure 7) [62,63]. Also the Maximum Entropy model proposed by S. Moulieras and F. Pachet [64] is explicitly based on the assumption that musical expression refers to local texture, rather than long-range correlations. In this case the reference repertoires are jazz, pop, and latin jazz.

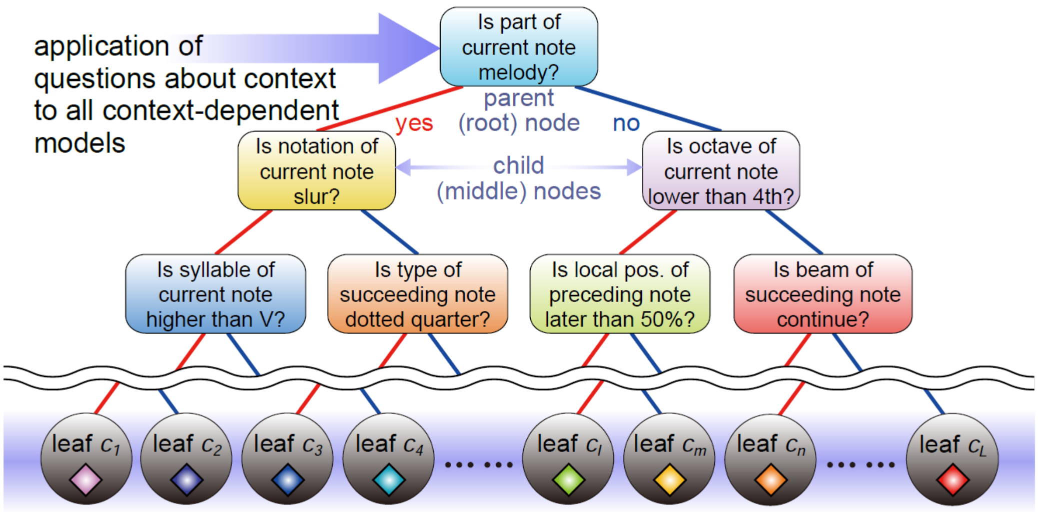

Figure 7. Binary tree based clustering of the local context [63].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

Although in-depth studies are necessary before reaching any conclusion, it is our opinion - also based on preliminary informal pop performances analyses - that in the field of popular music the study of local expressiveness could prove to be of extreme relevance. Most of the research cited in this subsection could find direct application or could be adapted to work also in a popular music context.

3.5. Score markings interpretation

Quite peculiarly, M. Grachten and G. Widmer focused their attention on a very specific and limited area of performative musical expressiveness: the interpretation of dynamic score markings [65]. Their model (which must not be considered limited to dynamics alone, but can also be applied to other musical expressive categories) is focused on the explicit expressive markings written in the score (e.g. *crescendo*/*diminuendo* signs, accents, dynamic markings such as *p*, *f*, *mf*, directly linked to basis functions - numeric descriptors that represent specific components of the score), and investigates how the combination of these markings/*basis functions* influences specific target parameters. In other words, expressive parameters (e.g note dynamics) are modelled after a weighted linear combination of score information, plus noise. In a subsequent work a Bayesian probabilistic alternative to the original approach to weights estimation (based on least squares regression) was proposed, together with a new set of *basis functions* and the contextualization of gradual loudness annotations (*crescendo*/*diminuendo*) in relation with the preceding and following notated loudness levels (e.g. *p crescendo f*) [66]. An ANN based approach (Non Linear Basis Model - NBM) was then investigated [67,68]. In these last contributions, instead of using a simple weighted linear combination of the *basis functions* (LBM - Linear Basis Model), Feed Forward Neural Networks (FFNNs), bidirectional Recurrent Neural Networks (RNNs - see [69]), a combination of FFNN and RNN, and a Long Short Term Memory network are tried on piano solo and symphonic repertoires, showing better prediction accuracy than the original LBM. Another development of the model therefore took place with the addition of the relationship between the formation of musical expectations and the corresponding musical performances to the analysis of score features based on basis functions, with significantly positive results [70].

The KTH rule system [43] provides articulation rules, and in particular it takes into account the markings of legato and staccato present in the score.

While attention to indications present in the score similar or such as those mentioned above could prove useful in popular music production, it must be noted that the MIDI protocol does not natively provide the tools to describe the score markings, being it a performance description language more than a symbolic prescriptive language (like the score). In any case, the score features could be included in MIDI in the form of CCs, or out-of-range notes, as commonly happens in virtual instruments with multiple articulations. MusicXML or similar approaches may be difficult to integrate in popular music producers' workflow, but innovative ways to harmonize them with MIDI could produce positive results.

3.6. Relationship between emotional intention and expressive parameters

Much of the interest shown by people in music is due to the emotional dimension [71]. The research field that investigates music and emotion has seen a relevant and growing interest in the last decades. An excellent general introduction to the topic was offered by P.N. Juslin and P. Laukka [72]. A relatively recent but deep review of the state of the art was presented by T. Eerola and J.K. Vuoskoski [73].

Before delving into more computational oriented approaches that focus on the relation between music expressiveness and emotion, it seems appropriate to outline the essential characteristics of general reflection on music and emotions.

 doi:10.20944/preprints202212.0494.v1

A first distinction to be made is the one between emotions expressed by the music and identified by the listener, and music-inducted emotions (felt by the listener). A. Gabrielsson observed that there is not necessarily a positive correlation between perceived and felt emotions in music [74]. P. Evans and E. Schubert further investigated the possible relationships between the emotional quality attributable to musical materials (expressed emotion, called *external locus of emotion*) and the subjective emotional response to music (felt emotion, called *internal locus*), describing the simple hypothesis of equality between the two as overly simplistic [75]. Moreover, P.N. Juslin observed that emotions perceived and expressed in music may be different from each other, and that not always does music arouse an emotional response in the listener [71].

Of fundamental importance in any approach to the description of the relationship between emotion and music is the emotion model adopted. In [73] four models are proposed: discrete model (all emotions can be derived from a limited set of basic emotions, usually fear, anger, disgust, sadness, and happiness); dimensional model (frequently traceable back to J. Russell's circumplex model [76], still relevant today, where emotions can be represented as a mixture of the core dimensions of valence and arousal, in a bidimensional space); miscellaneous models (based on a collection of concepts such as preference, similarity, tension); music specific models (that focus on emotions that are directly relevant for music, while the other approaches are more general-purpose and may not be fully suited to the music field).

Numerous (configurations of) music features have been linked in past studies to the expression of discrete emotions (for example, fast tempo, major mode, simple and consonant harmony, and ascending pitch are positively correlated with happiness) [71]. A. Gabrielsson published some of the the seminal papers concerning the relation between parameters such as timing, dynamics, intonation and expressed emotion [77,78].

S.R. Livingstone and co-authors proposed a rule system based on the previous studies about music parameters and expressed emotions, capable of modifying not only performance parameters but also score indications according to the desired emotion [79].

R. Bresin and A. Friberg presented a synthesis approach to the topic: 20 performers were asked to intervene on seven musical variables (tempo, sound level, articulation, phrasing, register, timbre, and attack speed) simultaneously - and not, as it often happens, one at a time - for communicating five different emotional intentions (neutral, happy, scary, peaceful, sad). For each of the five emotions the mean values and ranges of the musical variables are detected (see Figure 8). It is noted that these expressive parameters are not dependent on the score to be played [80].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

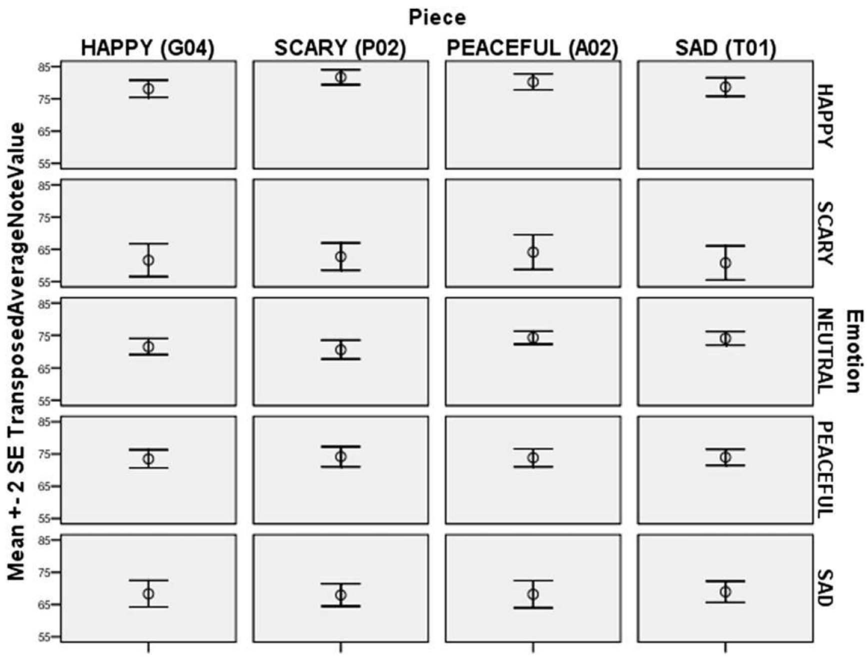

Figure 8. Relation between register and expressed emotion (mean values and range) [80].

The ability of the KTH rule system and of Director Musices to produce virtual performances that can be associated with different emotional states was investigated by R. Bresin and A. Friberg [81].

Several musical cues were analyzed through systematic parameters variations with respect to the emotion expressed by T. Eerola and co-authors: tempo, register, dynamics, articulation, and timbre. The detected relevance of each cue corresponds to the order in which they are listed above. Another finding is that the musical cues seem not to have significant interactions; their contribution to the overall expressed emotion appears to be based on simple linear combination [82].

The origins of these associations may be traced back to analogies with emotional speech [83], to human movement [84], to personal and/or cultural associations.

For what concerns the induction of emotions, one of the seminal approaches is the one of L.B. Meyer [46], which bases it on the creation and subsequent confirmation or disruption of expectations, an approach later deepened and extended in E. Narmour's work [47]. P.N. Juslin proposed a unified framework called BRECVEMA, that takes into account eight emotion induction mechanisms, to be added to the cognitive one: *brain stem reflex* (an emotion is evoked when one or more music parameters exceed a specific threshold), *rhythmic entrainment* (the rhythm of the music influences some internal bodily rhythm of the listener, e.g. the heart rate), *evaluative conditioning* (specific traits of the music are associated with emotions because they were heard many times in specific contexts), *emotional contagion* (the brain responds to specific music stimuli as if they were coming from a human voice that expresses emotion, and mimics that emotion), *visual imagery* (music stimulates imagery of bodily experiences with which it has something in common), *episodic memory* (there's a connection between the music and personal memories of the listener), *music expectancy* (see above the Meyer approach), and *aesthetic judgement* (the evaluation of the aesthetic value of the piece makes the listener feel an emotion) [85].

Probably, most of the research described above could find direct application in popular music oriented EMP models or could be adapted for this purpose. Popular music seems to be able to evoke and express emotions that are equally powerful to high-art (classical) music, at least when only liked or loved pieces are taken into consideration [86]. Anyway, although some research from different perspectives has been conducted on emotion and popular music (see for example Y. Song and co-authors' research [87]), the commitment is certainly much lower than what it has been done in the field of classical music [86]. Much research is needed, but it could definitely be worth it.

3.7. Relationship between sensorial experiences and expressive parameters

The expressive intention of a performance can be traceable back also to the desire of expressing sensorial perceptions. S. Canazza and co-authors asked a professional clarinet player to play multiple times the same excerpt taken from the W.A. Mozart's *Clarinet Concert in E Major* (K622), once in a scholastic, *normal* way and the other times trying to express the adjectives *light*, *heavy*, *soft*, *hard*, *bright*, *dark*. The recordings were analysed with regard to time (correlated to amplitude and duration) and frequency (timbre) domains. The data collected were then used to synthesize virtual performances aimed at expressing the above adjectives. A panel of musicians correctly recognized the expressive intention of the computational renditions [88]. In this work we can already see the bases on which the CaRo system [18,89–91] would later be developed. CaRo will be addressed in detail below, in Section 3.11.

In the pDM system, a real time sequencer integrated with the KTH rule system - see Section 3.11, a virtual performance can be manipulated through a set of mappers that translate high-level indications into rule parameters. Among them there are descriptive adjectives such as hard, light, heavy or soft [92].

A. Friberg and J. Sundberg compared the stopping of running and the final *ritardando* that marks the termination of a piece of music, noting that they present significant similarities [93].

The association between sensorial or physical experiences and performance parameters could be of prime relevance in the development of popular music oriented EMP systems. Within pop music production teams this kind of linguistic parallels seems to be often used [94], and it could be more easily interpretable by producers and better convey expressive intentions than other terminological families. More in-depth research on the subject, whose surface has only been scratched by the academy, would certainly be valuable to clarify the real scope of this approach to computational expressiveness applied to the popular context.

3.8. Identification and modelling of the performative styles of real musicians

The study of a specific performer's style can serve multiple purposes: automatic artist recognition, quantitative analysis of the individual style of his/hers, creation of real human-based expressive models.

Visualization and analysis of the performance style of famous artists is one of the explicit objectives mentioned in the work of S. Dixon and co-authors [26].

Different weight distributions among the rules of the KTH system can be used to represent different performers' styles [43]. Director Musices, a software available for GNU/Linux, Macintosh and Windows that implements most of the KTH rule system, was used to try and reproduce a specific pianist's expressive timing, with good but not optimal results [95]. From the study it emerges that rule combinations have to change between sections in order to better match the pianist's actual deviations. In the above mentioned hybridation of KTH rules and ANN [59], after a first step in which the ANN is trained to emulate a selection of the KTH rules, a more complex version of the system, aware of the local context ($n-1$, $n+1$ and $n+2$ notes parameters contribute to the definition of the parameters of the current note n), is proposed to learn the playing style of a specific pianist. In informal listening tests, better judgment were obtained by the ANN trained with the real pianist, compared to the one trained with the KTH rules.

S.I. Giraldo and R. Ramirez worked on a machine learning system for performative rules discovering and modeling of expressive jazz guitar performances [21,96]. The ML approach to feature selection, rules discovering and performance modelling is suitable for research applied to specific individual musicians.

 doi:10.20944/preprints202212.0494.v1

In C. Saunders and co-authors' research tempo and loudness deviations of multiple performers playing the same piece of music are translated into performance worms; from there general performance alphabets can be derived, and the performances can be represented as strings [97].

Generally speaking, Machine Learning techniques lend themselves well to the identification of performers. That is the case, for example, of the contributions of E. Stamatatos and G. Widmer [98] and of R. Ramirez and co-authors [99], respectively dedicated to piano and saxophone performers automatic identification.

The attempt to imitate specific musicians, potentially very different from each other by their nature, imposes an approach to the management of performative parameters that tends to be more agnostic than what occurs in the general modeling of expressiveness in specific stylistic or historical contexts. Therefore, we see no reason why the research already carried out in the field of recognition and modeling of specific performers should not be applicable also in the field of popular music, possibly with some minor adaptations.

3.9. Identification of physical and psychological limits of the performer

While on the one hand a lot of attention (see previous paragraphs) has been paid to the observation of the deviations of the performers attributable to structural high level or to local musical logics, to emotional rendering intentions, and to sensory references, on the other hand little research was devoted to understanding the implications of biomechanical constraints and internal processes of the musician in the implementation of expressive music performance models, an approach - the latter - that has been called *Performer-Based Modelling* [100]. L.L. Costalonga and co-authors observed that

> *physical manipulation of an instrument by the performer is often neglected in previous research* [100, pg. 332]

and that the data and evidences collected by the authors in this perspective

> *provide important insights for the development of Expressive Music Performance Models, specifically for guitar performance. In our point of view, such comprehension is essential for better proposing Digital Musical Instruments and EMP systems [...].*

In fact, many studies have been carried out to better understand - among other performance characteristics - motor control, hand dexterity, and timing precision in musicians [101], but it is very rare to see the fallout from these studies on EMP modelling research. The most common objectives in this kind of studies are, for example, the prevention of injuries [102] or the study of optimal fingerings from the biomechanical point of view [103,104].

One excellent exception is the above mentioned work of L.L. Costalonga and co-authors [100], where biomechanical constraints, errors, noise generated, muscle strength, speed and endurance are deeply investigated in relation to the guitar, with the declared intention of producing data potentially useful for the development of EMP models.

Another work that we think could be particularly inspiring in an EMP development perspective (although its declared objectives are not related to that) is the one by P. Visentin and co-authors about the biomechanics of left-hand position changes (shifting) in violin performance [105]. There, violinists' left hand position shifts are investigated, with specific regard to EST (end of shift) and DOS (duration of shift) parameters, measured in milliseconds. It is demonstrated that each performer tends to develop a more or less fixed left hand shifting time (given the space covered by the move is the same), independent of the metronomic tempo and the specific context. This seems to partially contradict B.H. Repp's research [106], in which two pianists are analysed. The results suggest here that while major (cognitively controlled) temporal and dynamic parameters of a performance

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

change substantially in proportion with tempo, minor features tend to be determined by tempo-independent motor constraints.

It is our opinion that *Performer-Based Modelling* could play a key-role in the future development of EMP models, in particular if they were oriented to attempts at creating computational models of real musicians. This could be particularly true in the popular music EMP field, since from certain points of view the degree of freedom of the performer is more limited there than what happens in western classical music, if only for the fact that generally the timing of the performance is bound to a fixed metronomic tempo, or in any case to drums and percussion parts. Some of the previous approaches to expressive timing may not be fully applicable, resulting in the need to follow other strategies to investigate and model performative expressiveness, and *Performer-Based Modelling* seems particularly promising in this perspective.

3.10. Ensemble music modelling

As seen above, most of the computational EMP publications deals with solo - usually piano - music. Only few steps regarding expressive ensemble performance have been taken [107]. Among the very few approaches to the topic it can be cited the one of J. Sundberg and co-authors [108], where an embryonic version of the soon to be KTH rule system is proposed, together with a hardware/software system capable of receiving in input a music score complemented with phrasing boundaries, harmonies, ties, etc., and returning to output (through the Rulle software) an expressive rendition of the piece. Ensemble synchronization and intonation issues are here taken into account. Perfect synchronization between the notes belonging to different instruments but contemporaneous in the score was better judged by a panel of musician than the competing synchronization solution, based on the freedom of the various instruments, that anyway had to be perfectly in sync once for each bar. The former approach provides that all the musical parts have to synchronize each time to the most significant one (the one that executes the shortest note) for each note. For what concerns the tuning, a general preference for the equal temperament over other solutions is detected.

The modest attention paid by research to expressiveness in the context of ensemble performance represents a strong limitation in the field of contemporary popular music production, where generally more instruments and voices are present at the same time. Any approach to computational expressiveness must be context-sensitive here, and the deviations produced cannot be thought of as if the instrument of interest operated alone.

3.11. Conductor systems

Solutions that involve meaningful user interaction are categorized as conductor systems. In conductor systems the user can apply, either asynchronously or in real-time, changes to the expressive rendering of the musical piece. Two aspects are frequently highlighted in conductor systems research: the capability for users without specific musical training to interact with music, focusing on the creative dimension [91], and the possibility for the user of concentrating on the expressive component of the music, without having to worry about the technical difficulties inherent in playing musical instruments [109].

The association between conductor systems and traditional orchestra conducting inspired some of the earlier research in this area. In M.V. Mathews and J. Lawson work a radio baton capable of controlling expressive parameters of a synthesizer (timing and dynamics) through the dedicated software Conductor is presented [109–111]. Another example of virtual orchestra conduction system is E. Lee and co-authors' one [112], based on the recognition of baton gestures and associated time-stretching on the audiovisual recording of a performing orchestra (see Figure 9). A more recent implementation of virtual orchestral conduction was proposed by T. Baba and co-authors [113].

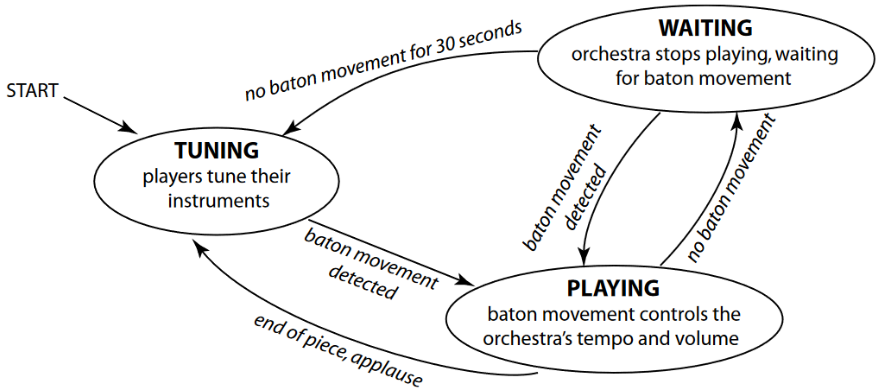

Figure 9. State machine for *You're the conductor* [112].

In [114] M.V. Mathews and co-authors suggested an integration between Conductor software and Director Musices. A real time extension in Pure Data language of Director Musices is presented by A. Friberg and co-authors in [115]. Also correlated to the KTH rule system is the already mentioned work of R. Bresin [59], where the user can interact with the system directly providing input data to the ANN or acting on some of the rules involved in the model. R. Bresin also proposed a system for the real time control of pDM parameters [92]. The system provides for the visual recognition of human gestures, used to control music parameters at three possible abstraction levels: listener level - the controlling activity is based on basic emotions (happy, sad, angry); simple conductor level - basic overall musical features are controlled using the energy-kinematics space or similar solutions; advanced conductor level - level 1 and 2 are combined with the explicit control of each beat. In [116] S. Canazza and co-authors proposed an integration between the KTH rule system and the expressiveness model and two-dimensional real-time control space developed at the CSC (Centro di Sonologia Computazionale) of the University of Padua (see below).

Of particular relevance is the CaRo system [18,89,90]. The system is based on the idea that there are two main sources for musical expression: structure of the piece (see subsection 3.3) and expressive intention (e.g. bright, dark, hard, see subsection 3.7). Asking professional musicians to play the same melody in a neutral, scholastic way, and with a definite set of expressive intentions, it is possible to estimate how performing parameters (e.g. intensity, legato, attack duration, brightness) are affected by the expressive intention of the musician. For each expressive parameter and each expressive intention two parameters are extracted: k (associated with the mean value) and m (the range of the values, affecting the variance). The factor loadings obtained from factor analysis (with a two dimension solution) are then used as coordinates capable of describing the expressive parameters of the performances in a two-dimensional space. Reversing the process, the two-dimensional space can be used to control the expressive intention of neutral performances (see Figure 10). The CaRo system can operate directly on audio signals, but also on MIDI data. Years after the initial release, the CaRo system was adapted to work in a Web 2.0 environment [91].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

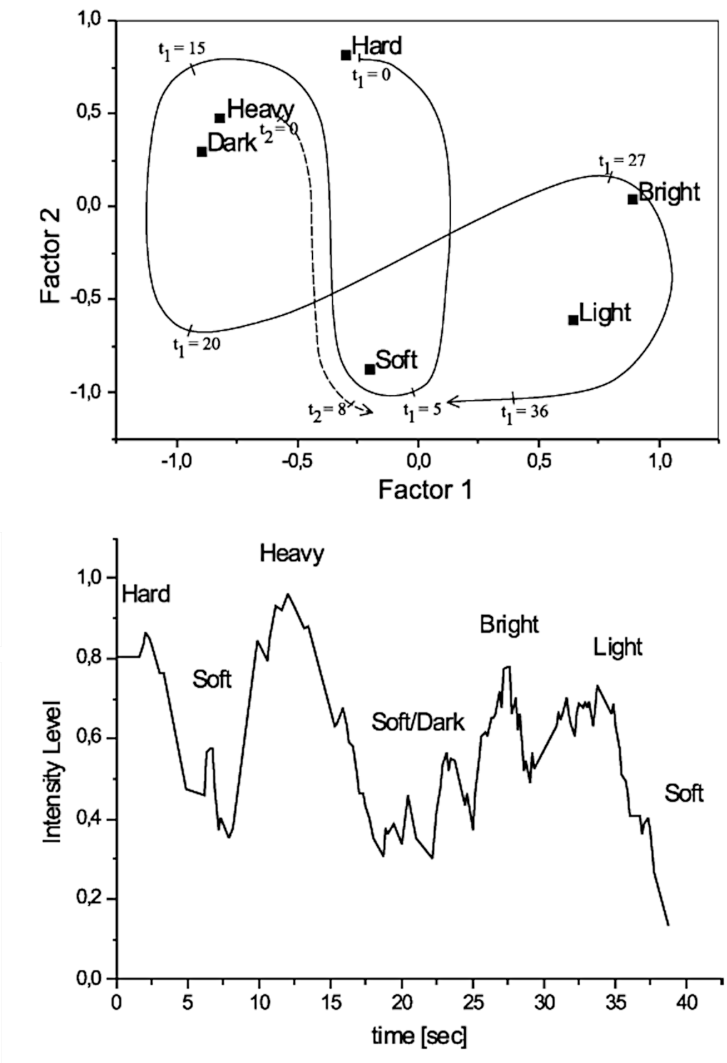

Figure 10. The control space movements and the corresponding intensity parameter trend in CaRo [90].

In [117], S. Dixon and co-authors resumed the metaphor of the performance worm [25,26], inverting it, so that tempo and loudness can be controlled in real time by the user. This can be done using hand movements and a digital theremin (Air Worm), or the computer mouse (Mouse Worm). In the same contribution the Air Tapper and Mouse Tapper systems are also presented. These can be traced back to the more traditional control of the metronomic tempo, through hand gestures or use of the mouse.

Generally speaking, conductor systems seem to be of great interest from the point of view of the possible applications in the field of popular music production. Given that in popular music the metronomic tempo is generally tied to a steady beat, or to drums and percussions parts, direct control of the same (like in virtual orchestra conduction solution) appears to be of little relevance here. On the other hand, the control of the loudness parameter, or of the high level expressive intention of the performance (sensorial or emotional terms were used in scientific literature, but their appropriateness in the pop world should be verified) should be more relevant.

3.12. Table summary

Within the following table the main features of the most relevant EMP systems reviewed are summarized. In the first cell on the left, after the recall of the research line and

 doi:10.20944/preprints202212.0494.v1

the references, a color coded indication of the thematic areas covered is provided. The correspondences are the following:

- Visual representation: Vis
- Structure based expressiveness: Str
- Local based expressiveness: Loc
- Score markings: Sco
- Emotion: Emo
- Sensorial experience: Sen
- Identification and modelling of performers: Per
- Physical and psychological limits: Lim
- Ensemble modelling: Ens
- Conductor systems: Con

doi:10.20944/preprints202212.0494.v1

References	Technologies	User interaction	Main goal(s)
KTH rule system [43,59,81,95,108,115,118] Str Emo Loc Sen Per Con	Rules, ANN [59]	Interaction with the ANN [59], Director Musices and pDM applications [95,115,118]	Expressive music performances generation, modelling of real performers or music performing styles
LNM and NBM [65–68] Loc	Linear weighted combination of parameters [65,66], FFNN RNN LSTM [67,68,70]	/	Modelling of the influence of explicit score markings on expressive parameters. Music expectations consedered in [70]
Performance worm [24–26,117] Vis	MIR through analysis of MIDI or audio data	Tempo and loudness control using hand gestures or PC mouse [117]	Real time graphical representation of tempo and loudness, user control of tempo and loudness
Tonal tension in expressive piano performance [49] Str Loc	RNN LSTM	/	Expressive music performances generation based on the analysis of tonal tensions
CaRo and CaRo 2.0 [18,89–91] Con Loc Sen	Statistical analysis (principal component)	Real-time interaction through an abstract two-dimensional control space	Graphical description of performances, generation of expressive music performances starting from neutral ones, through user interaction
SYVARs [44,55–57] Loc	Statistical research of regularities	/	Find and validate systematic expressive variations in specific contexts
This time with feeling [23] Loc	LSTM	/	Generation of solo piano musical parts and their expressive performance at the same time
Expert system [58] Loc	Analysis by synthesis rule based expert system	/	Rendition of expressive performances of Bach's fugues
ML approach to jazz guitar solos [21,96,119] Loc Lim	ML ANN, decision trees, SVM, feature selection	/	Discover of rules for expressive performance in jazz guitar and expressive models creation through ML techniques
Rule system for modifying score and performance to express emotions [79] Loc	Rules	/	Express emotions modifying not only performance parameters but also score ones
ESP [40] Str Loc	Hierarchical Hidden Markov Model (HHMM)	/	Expressive performance generation based on the score structure
Gaussian process regression [60] Loc	Gaussian process ML	/	Expressive performance generation
YQX [47] Loc	Bayesian networks	/	Expressive performance generation
Laminae [62,63] Loc	Tree-based clustering, Gaussian distributions	/	Expressive performance generation
Maximum entropy [64] Loc	Maximum entropy	/	Expressive performance generation given a specific musical style

 doi:10.20944/preprints202212.0494.v1

4. Conclusions

In this paper we briefly presented the tools available today to modern popular music producers that can help in building an expressive virtual music performance. We noticed that, although powerful tools for synthesis and sound sampling are commercially available, and although there is the possibility to freely vary many of the most relevant expressive parameters, what is missing is the ability to automatically process musical parts provided by the producer - potentially time-quantized and fixed-velocity ones - making them expressive, acting automatically on tone parameters of the virtual instrument, time and loudness deviations, articulation, embellishments, or even errors and noise traceable back to performer's constraints, all of that possibly in a context-aware manner. In the second part of the paper we conducted a reasoned review of the scientific literature dedicated to the EMP sector, evaluated from the point of view of its potential impact on contemporary popular music production. What emerged is that more than one research line could prove useful in helping modern producers. Graphical representations of expressive parameters can help in the understanding of expression dynamics that may be hard to catch otherwise by a human observer, and of analogies and differences between musical expressiveness in popular music and in other contexts. Much attention has been paid in EMP research to the relation between structure and expressiveness. Part of this research could probably be adapted and applied to popular music, but more investigation is needed before being able to draw any conclusions. Automatic structure identification solutions not based on classical tonality seem to be promising with regard to pop music. Much more relevant in view of possible applications to pop music is the local expressiveness approach. Generally speaking, the data driven solutions introduced could all find applications in pop music. The same can be told about identification and modelling of the performative styles of real musicians. The emotional intention and sensorial parallelisms related research could find positive applications in popular music, but first of all it would be necessary to better clarify the role of these associations in pop, also from the production practise, psychological, cultural, linguistic and anthropological points of view. The identification of physical and psychological limits of the performer is a topic not particularly well-trodden in EMP research. What has already been done can be of great significance in pop music applications, and more research in the field would be highly desirable. The same can be said about expressive ensemble modelling. Conductor systems are of prime relevance here, because they offer several examples of how it is possible to decline the control of performative expressiveness. The score markings interpretation topic could be included in conductor systems, as a kind of offline conduction tool.

The main reasons because of which past EMP studies outcomes often cannot be directly applied, or at least not before being rethought and adapted, to the typical workflow of a modern music producer, are:

- most of the times the reference repertoire is the classical eurocultural one, which presumably may be subject to different rules and practices than popular music. This seems to be confirmed by the fact that similarities and differences between machine learning induced rules in expressive jazz guitar and rules reported in the literature were found [21];
- the role of music scores is also profoundly different between the classical and popular contexts. While in classical music the performer offers an interpretation of the composer's intention (given through the score written by the latter), in popular music there can be many different situations. Musical parts can be improvised, or there can be only chord charts or lead sheets, which respectively show only the harmonic skeleton or the reference melody, which can be embellished and modified even in depth during the performance. Anyway, if we take into account the declared objective of this contribution, that is to understand which research products could be useful to the contemporary producer in conferring expressiveness to pre-defined parts, and not

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

in the automatic generation of new parts, it is possible to leave out the specific case of free improvisation. As for the relationship and alignment between real performance and lead sheet, the topic was dealt with in relation to jazz guitar by S.I. Giraldo and R. Ramirez [96,119];

- while musical expression is partly instrument-specific, most of the past research deals with classical solo piano [22];
- in art music and particularly in classical music artists have a lot of freedom to express their individuality [26], while in popular music production there tend to be more constraints, if only for the fact that often the pieces are recorded with constant metronome tempo;
- while usually in modern popular music there are many instruments and vocal parts played or sung at the same time, many past studies on EMP deal with solo instruments (notably solo piano), that for obvious reasons have greater freedom of expression, in particular with respect to tempo and timing.

In conclusion, it is our hope that in the future studies on musical expression will involve more the field of popular music. Part of what has been done in computational EMP can be directly applied or adapted to this specific context, but much remains to be done. This need is justified not only by an expansion of the academic understanding of the phenomenon of musical expressiveness, but also by the possible positive effects in the daily work of popular producers and by potential commercial applications.

Author Contributions: conceptualization, Pierluigi Bontempi and Sergio Canazza; methodology, Pierluigi Bontempi, Filippo Carnovalini, and Sergio Canazza; investigation, Pierluigi Bontempi; writing—original draft preparation, Pierluigi Bontempi; writing—review and editing, Pierluigi Bontempi, Filippo Carnovalini, Antonio Rodà, and Sergio Canazza; supervision, Pierluigi Bontempi and Sergio Canazza.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
CC	Continuous Controller/Control Change
DAW	Digital Audio Workstation
DOS	Duration Of Shift)
EMP	Expressive Music Performance
EQ	EQualization
EST	End of ShifT
FFNN	Feed Forward Neural Network
HHMM	Hierarchical Hidden Markov model
HPF	Hi Pass Filter
LBM	Linear Basis Model
LSTM	Long Short Term Memory
MIR	Music Information Retrieval
ML	Machine Learning
MPE	Midi Polyphonic Expression
NBM	Non linear Basis Model
RNN	Recurrent Neural Network
SVM	Support Vector Machine
SYVAR	SYstematic VARiations

References

1. Juslin, P.N.; Friberg, A.; Bresin, R. Toward a computational model of expression in music performance: the GERM model. *Musicae Scientiae* **2001**, *5*, 63–122. https://doi.org/10.1177/10298649020050S104.

2. Cancino-Chacón, C.E.; Grachten, M.; Goebl, W.; Widmer, G. Computational models of expressive music performance: a comprehensive and critical review. *Frontiers in Digital Humanities* **2018**, *5*. https://doi.org/10.3389/fdigh.2018.00025.
3. Moir, Z.; Medbøe, H. Reframing popular music composition as performance-centred practice. *Journal of Music, Technology & Education* **2015**, *8*, 147–161. https://doi.org/10.1386/jmte.8.2.147_1.
4. Dibben, N. Understanding performance expression in popular music recordings. *Expressiveness in music performance: Empirical approaches across styles and cultures* **2014**, pp. 117–132. https://doi.org/10.1093/acprof:oso/9780199659647.003.0007.
5. Collins, M. *A professional guide to audio plug-ins and virtual instruments*; Routledge, 2012.
6. Owsinski, B.; Hal Leonard, 2016.
7. Marrington, M.; et al. Composing with the digital audio workstation. *The singer-songwriter handbook* **2017**, pp. 77–89.
8. Yun, Y.; Cha, S.H. Designing virtual instruments for computer music. *International Journal of Multimedia and Ubiquitous Engineering* **2013**, *8*, 173–178. https://doi.org/http://dx.doi.org/10.14257/ijmue.2013.8.5.16.
9. Tanev, G.; Božinovski, A. Virtual Studio Technology inside music production. In Proceedings of the ICT Innovations 2013; Trajkovik, V.; Anastas, M., Eds.; Springer: Heidelberg, 2014; pp. 231–241. https://doi.org/10.1007/978-3-319-01466-1_22.
10. Moog, B. MIDI. *Journal of Audio Engineering Society* **1986**, *34*.
11. Hennig, H.; Fleischmann, R.; Fredebohm, A.; Hagmayer, Y.; Nagler, J.; Witt, A.; Theis, F.J.; Geisel, T. The nature and perception of fluctuations in human musical rhythms. *PloS one* **2011**, *6*, e26457. https://doi.org/10.1371/journal.pone.0026457.
12. Organic mixdowns. *Computer Music* **2015**.
13. Walden, J. UJAM Virtual Bassist: Royal, Mellow & Rowdy. *Sound On Sound* **2019**, *34*, 128–130.
14. Tzanetakis, G. Natural human-computer interaction with musical instruments. In *Digital Tools for Computer Music Production and Distribution*; Politis, D.; Miltiadis, T.; Ioannis, I., Eds.; IGI Global: Hershey, PA, 2016; chapter 6, pp. 116–136. https://doi.org/10.4018/978-1-5225-0264-7.ch006.
15. Fasciani, S.; Goode, J. 20 NIMEs: twenty years of new interfaces for musical expression. In Proceedings of the Proceedings of the International Conference on New Interfaces for Musical Expression (NIME); , 2021. https://doi.org/10.21428/92fbeb44.b368bcd5.
16. Gabrielsson, A. Music performance research at the millennium. *Psychology of Music* **2003**, *31*, 221–272. https://doi.org/10.1177/03057356030313002.
17. Widmer, G.; Goebl, W. Computational models of expressive music performance: the state of the art. *Journal of New Music Research* **2004**, *33*, 203–216. https://doi.org/10.1080/0929821042000317804.
18. De Poli, G. Expressiveness in music performance. In *Algorithms for Sound and Music Computing*; 2018.
19. Kirke, A.; Miranda, E.R. A survey of computer systems for expressive music performance. *ACM Computing Surveys* **2009**, *42*, 1–41. https://doi.org/10.1145/1592451.1592454.
20. Delgado, M.; Fajardo, W.; Molina-Solana, M. A state of the art on computational music performance. *Expert Systems with Applications* **2011**, *38*, 155–160. https://doi.org/https://doi.org/10.1016/j.eswa.2010.06.033.
21. Giraldo, S.; Ramirez, R. A machine learning approach to ornamentation modeling and synthesis in jazz guitar. *Journal of Mathematics and Music* **2016**, *10*, 107–126. https://doi.org/10.1080/17459737.2016.1207814.
22. Raphael, C. Symbolic and structural representation of melodic expression. *Journal of New Music Research* **2010**, *39*, 245–251. https://doi.org/10.1080/09298215.2010.512978.
23. Oore, S.; Simon, I.; Dieleman, S.; Eck, D.; Simonyan, K. This time with feeling: learning expressive musical performance. *Neural Computing and Applications* **2020**, *32*, 955–967. https://doi.org/10.1007/s00521-018-3758-9.
24. Langner, J.; Goebl, W. Representing expressive performance in tempo-loudness space. In Proceedings of the ESCOM Conference on Musical Creativity, 2002, pp. 5–8.
25. Dixon, S.; Goebl, W.; Widmer, G. The performance worm: real time visualisation of expression based on Langner's tempo-loudness animation. In Proceedings of the Proceedings of the International Computer Music Conference (ICMC). Michigan Publishing, 2002.
26. Dixon, S.; Goebl, W.; Widmer, G. Real time tracking and visualisation of musical expression. In Proceedings of the International Conference on Music and Artificial Intelligence (ICMAI). Springer, 2002, pp. 58–68. https://doi.org/10.1007/3-540-45722-4_7.
27. Stanyek, J. Forum on transcription. *Twentieth-Century Music* **2014**, *11*, 101–161. https://doi.org/10.1017/S1478572214000024.
28. Dannenberg, R.B. Music representation issues, techniques, and systems. *Computer Music Journal* **1993**, *17*, 20–30. https://doi.org/10.2307/3680940.
29. Drabkin, W. Motif [motive]. https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000019221#omo-9781561592630-e-0000019221, 2001. Accessed: 2022-06-05.
30. Phrase. https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000021599, 2001. Accessed: 2022-06-05.
31. Ratner, L.G. Period. https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000021337#omo-9781561592630-e-0000021337, 2001. Accessed: 2022-06-05.
32. Palmer, C. Music performance. *Annual Review of Psychology* **1997**, *48*, 115–138. https://doi.org/10.1146/annurev.psych.48.1.115.
33. Clarke, E.F. Generative principles in music performance. In *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*; Sloboda, J., Ed.; Clarendon Press/Oxford University Press, 2001; pp. 1–26. https://doi.org/10.1093/acprof:oso/9780198508465.003.0001.
34. Lerdahl, F.; Jackendoff, R. *A generative theory of tonal music*; The MIT Press, 1983.

35. Todd, N. A model of expressive timing in tonal music. *Music Perception: an Interdisciplinary Journal* **1985**, *3*, 33–57. https://doi.org/10.2307/40285321.
36. Todd, N. A computational model of rubato. *Contemporary Music Review* **1989**, *3*, 69–88. https://doi.org/10.1080/07494468900640061.
37. McAngus Todd, N.P. The dynamics of dynamics: a model of musical expression. *The Journal of the Acoustical Society of America* **1992**, *91*, 3540–3550. https://doi.org/10.1121/1.402843.
38. Windsor, W.L.; Clarke, E.F. Expressive timing and dynamics in real and artificial musical performances: using an algorithm as an analytical tool. *Music Perception* **1997**, *15*, 127–152. https://doi.org/10.2307/40285746.
39. Clarke, E.F.; Windsor, W.L. Real and simulated expression: a listening study. *Music Perception* **2000**, *17*, 277–313. https://doi.org/10.2307/40285819.
40. Grindlay, G.; Helmbold, D. Modeling, analyzing, and synthesizing expressive piano performance with graphical models. *Machine learning* **2006**, *65*, 361–387. https://doi.org/10.1007/s10994-006-8751-3.
41. Widmer, G.; Tobudic, A. Playing Mozart by analogy: learning multi-level timing and dynamics strategies. *Journal of New Music Research* **2003**, *32*, 259–268. https://doi.org/10.1076/jnmr.32.3.259.16860.
42. Widmer, G. Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence* **2003**, *146*, 129–148. https://doi.org/10.1016/S0004-3702(03)00016-X.
43. Friberg, A.; Bresin, R.; Sundberg, J. Overview of the KTH rule system for musical performance. *Advances in cognitive psychology* **2006**, *2*, 145–161. https://doi.org/10.2478/v10053-008-0052-x.
44. Gabrielsson, A. Interplay between analysis and synthesis in studies of music performance and music experience. *Music Perception* **1985**, *3*, 59–86. https://doi.org/10.2307/40285322.
45. Carnovalini, F.; Rodà, A. A multilayered approach to automatic music generation and expressive performance. In Proceedings of the International Workshop on Multilayer Music Representation and Processing (MMRP), 2019, pp. 41–48. https://doi.org/10.1109/MMRP.2019.00016.
46. Meyer, L.B. *Emotion and meaning in music*; University of Chicago Press, 1956.
47. Narmour, E. *The analysis and cognition of basic melodic structures: the implication-realization model.*; University of Chicago Press, 1990.
48. Herremans, D.; Chew, E. Tension ribbons: quantifying and visualising tonal tension. In Proceedings of the Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR); Hoadley, R.; Nash, C.; Fober, D., Eds.; Anglia Ruskin University: Cambridge, UK, 2016; pp. 8–18.
49. Cancino-Chacón, C.; Grachten, M. A computational study of the role of tonal tension in expressive piano performance. *arXiv preprint* **2018**. https://doi.org/10.48550/arXiv.1807.01080.
50. Tagg, P. *Everyday Tonality II — Towards a tonal theory of what most people hear*; Mass Media Scholars Press, 2018.
51. Marsden, A. Schenkerian analysis by computer: a proof of concept. *Journal of New Music Research* **2010**, *39*, 269–289. https://doi.org/10.1080/09298215.2010.503898.
52. Hamanaka, M.; Hirata, K.; Tojo, S. Implementing "A generative theory of tonal music". *Journal of New Music Research* **2006**, *35*, 249–277. https://doi.org/10.1080/09298210701563238.
53. Orio, N.; Rodà, A. A measure of melodic similarity based on a graph representation of the music structure. In Proceedings of the Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 2009, pp. 543–548. https://doi.org/10.5281/zenodo.1415010.
54. Simonetta, F.; Carnovalini, F.; Orio, N.; Rodà, A. Symbolic music similarity through a graph-based representation. In Proceedings of the Proceedings of the Audio Mostly Conference on Sound in Immersion and Emotion, 2018, pp. 1–7. https://doi.org/10.1145/3243274.3243301.
55. Gabrielsson, A. Performance of rhythm patterns. *Scandinavian Journal of Psychology* **1974**, *15*, 63–72. https://doi.org/10.1111/j.1467-9450.1974.tb00557.x.
56. Gabrielsson, A. Performance and training of musical rhythm. *Psychology of Music - Special Issue* **1982**, pp. 42–46.
57. Gabrielsson, A.; Bengtsson, I.; Gabrielsson, B. Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology* **1983**, *24*, 193–213. https://doi.org/10.1111/j.1467-9450.1983.tb00491.x.
58. Johnson, M.L. Toward an expert system for expressive musical performance. *Computer* **1991**, *24*, 30–34. https://doi.org/10.1109/2.84832.
59. Bresin, R. Artificial neural networks based models for automatic performance of musical scores. *Journal of New Music Research* **1998**, *27*, 239–270. https://doi.org/10.1080/09298219808570748.
60. Teramura, K.; Okuma, H.; Taniguchi, Y.; Makimoto, S.; Maeda, S.i. Gaussian process regression for rendering music performance. *Proceedings of International Conference on Music Perception and Cognition (ICMPC)* **2008**, pp. 167–172.
61. Flossmann, S.; Grachten, M.; Widmer, G. Expressive performance rendering with probabilistic models. In *Guide to Computing for Expressive Music Performance*; Springer, 2013; pp. 75–98. https://doi.org/10.1007/978-1-4471-4123-5_3.
62. Okumura, K.; Sako, S.; Kitamura, T. Stochastic modeling of a musical performance with expressive representations from the musical score. In Proceedings of the Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 2011, pp. 531–536.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 December 2022 doi:10.20944/preprints202212.0494.v1

63. Okumura, K.; Sako, S.; Kitamura, T. Laminae: a stochastic modeling-based autonomous performance rendering system that elucidates performer characteristics. In Proceedings of the Proceedings of the International Computer Music Conference (ICMC), 2014.
64. Moulieras, S.; Pachet, F. Maximum entropy models for generation of expressive music. *arXiv preprint arXiv:1610.03606* **2016**. https://doi.org/10.48550/arXiv.1610.03606.
65. Grachten, M.; Widmer, G. Linear basis models for prediction and analysis of musical expression. *Journal of New Music Research* **2012**, *41*, 311–322. https://doi.org/10.1080/09298215.2012.731071.
66. Grachten, M.; Cancino Chacón, C.E.; Widmer, G. Analysis and prediction of expressive dynamics using Bayesian linear models. In Proceedings of the Proceedings of the 1st International Workshop on Computer and Robotic Systems for Automatic Music Performance, 2014, pp. 545–552.
67. Cancino Chacón, C.E.; Grachten, M. An evaluation of score descriptors combined with non-linear models of expressive dynamics in music. In Proceedings of the Proceedings of the 18th International Conference on Discovery Science (DS); Springer: Banff, Canada, 2015. https://doi.org/10.1007/978-3-319-24282-8_6.
68. Cancino-Chacón, C.E.; Gadermaier, T.; Widmer, G.; Grachten, M. An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonic music. *Machine Learning* **2017**, *106*, 887–909. https://doi.org/10.1007/s10994-017-5631-y.
69. Graves, A. Generating sequences with recurrent neural networks, 2013. https://doi.org/10.48550/arXiv.1308.0850.
70. Cancino-Chacón, C.; Grachten, M.; Sears, D.R.; Widmer, G. What were you expecting? Using expectancy features to predict expressive performances of classical piano music. *arXiv preprint* **2017**. https://doi.org/10.48550/arXiv.1709.03629.
71. Juslin, P.N.; Laukka, P. Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *Journal of New Music Research* **2004**, *33*, 217–238. https://doi.org/10.1080/0929821042000317813.
72. Juslin, P.N.; Sloboda, J. *Handbook of music and emotion: theory, research, applications*; Oxford University Press, 2011. https://doi.org/10.1093/acprof:oso/9780199230143.001.0001.
73. Eerola, T.; Vuoskoski, J.K. A review of music and emotion studies: approaches, emotion models, and stimuli. *Music Perception* **2013**, *30*, 307–340. https://doi.org/10.1525/mp.2012.30.3.307.
74. Gabrielsson, A. Emotion perceived and emotion felt: same or different? *Musicae Scientiae* **2001**, *5*, 123–147. https://doi.org/10.1177/10298649020050S105.
75. Evans, P.; Schubert, E. Relationships between expressed and felt emotions in music. *Musicae Scientiae* **2008**, *12*, 75–99. https://doi.org/10.1177/102986490801200105.
76. Russell, J.A. A circumplex model of affect. *Journal of Personality and Social Psychology* **1980**, *39*, 1161. https://doi.org/10.1037/h0077714.
77. Gabrielsson, A. Intention and emotional expression in music performance. In Proceedings of the Proceedings of the Stockholm Music Acoustics Conference (SMAC), 1993, pp. 108–111.
78. Gabrielsson, A. Expressive intention and performance. In *Music and the Mind Machine*; Springer, 1995; pp. 35–47. https://doi.org/10.1007/978-3-642-79327-1_4.
79. Livingstone, S.R.; Muhlberger, R.; Brown, A.R.; Thompson, W.F. Changing musical emotion: a computational rule system for modifying score and performance. *Computer Music Journal* **2010**, *34*, 41–64. https://doi.org/10.1162/comj.2010.34.1.41.
80. Bresin, R.; Friberg, A. Emotion rendering in music: range and characteristic values of seven musical variables. *Cortex* **2011**, *47*, 1068–1081. https://doi.org/10.1016/j.cortex.2011.05.009.
81. Bresin, R.; Friberg, A. Emotional coloring of computer-controlled music performances. *Computer Music Journal* **2000**, *24*, 44–63. https://doi.org/10.1162/014892600559515.
82. Eerola, T.; Friberg, A.; Bresin, R. Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in pPsychology* **2013**, *4*. https://doi.org/10.3389/fpsyg.2013.00487.
83. Bhatara, A.; Laukka, P.; Levitin, D.J. Expression of emotion in music and vocal communication: Introduction to the research topic, 2014. https://doi.org/10.3389/fpsyg.2014.00399.
84. Sievers, B.; Polansky, L.; Casey, M.; Wheatley, T. Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences* **2013**, *110*, 70–75. https://doi.org/10.1073/pnas.1209023110.
85. Juslin, P.N. Emotional reactions to music. *The Oxford handbook of music psychology* **2014**, pp. 197–213. https://doi.org/10.1093/oxfordhb/9780198722946.013.17.
86. Schubert, E. Emotion in popular music: a psychological perspective. *Volume! La revue des musiques populaires* **2013**, *1*, 265–266. https://doi.org/10.4000/volume.3626.
87. Song, Y.; Dixon, S.; Pearce, M.T.; Halpern, A.R. Perceived and induced emotion responses to popular music: categorical and dimensional models. *Music Perception: an Interdisciplinary Journal* **2016**, *33*, 472–492. https://doi.org/10.1525/mp.2016.33.4.472.
88. Canazza, S.; Poli, G.D.; Rinaldin, S.; Vidolin, A. Sonological analysis of clarinet expressivity. In Proceedings of the Joint International Conference on Cognitive and Systematic Musicology. Springer, 1996, pp. 431–440. https://doi.org/10.1007/BFb0034131.
89. Canazza, S.; Poli, G.; Rodà, A.; Vidolin, A. An abstract control space for communication of sensory expressive intentions in music performance. *Journal of New Music Research* **2003**, *32*, 281–294. https://doi.org/10.1076/jnmr.32.3.281.16862.

90. Canazza, S.; De Poli, G.; Drioli, C.; Roda, A.; Vidolin, A. Modeling and control of expressiveness in music performance. *Proceedings of the IEEE* **2004**, *92*, 686–701. https://doi.org/10.1109/JPROC.2004.825889.
91. Canazza, S.; De Poli, G.; Rodà, A. Caro 2.0: an interactive system for expressive music rendering. *Advances in Human-Computer Interaction* **2015**, *2015*. https://doi.org/10.1155/2015/850474.
92. Friberg, A. Home conducting-control the overall musical expression with gestures. In Proceedings of the Proceedings of the International Computer Music Conference (ICMC). Michigan Publishing, 2005.
93. Friberg, A.; Sundberg, J. Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *The Journal of the Acoustical Society of America* **1999**, *105*, 1469–1484. https://doi.org/10.1121/1.426687.
94. Porcello, T. Speaking of sound: language and the professionalization of sound-recording engineers. *Social Studies of Science* **2004**, *34*, 733–758. https://doi.org/10.1177/0306312704047328.
95. Sundberg, J.; Friberg, A.; Bresin, R. Attempts to reproduce a pianist's expressive timing with director musices performance rules. *Journal of New Music Research* **2003**, *32*, 317–325. https://doi.org/10.1076/jnmr.32.3.317.16867.
96. Giraldo, S.I.; Ramirez, R. A machine learning approach to discover rules for expressive performance actions in jazz guitar music. *Frontiers in psychology* **2016**, *7*. https://doi.org/10.3389/fpsyg.2016.01965.
97. Saunders, C.; Hardoon, D.R.; Shawe-Taylor, J.; Widmer, G. Using string kernels to identify famous performers from their playing style. In Proceedings of the Machine Learning: ECML 2004; Boulicaut, J.F.; Esposito, F.; Giannotti, F.; Pedreschi, D., Eds.; Springer: Berlin, Heidelberg, 2004; pp. 384–395. https://doi.org/10.1007/978-3-540-30115-8_36.
98. Stamatatos, E.; Widmer, G. Automatic identification of music performers with learning ensembles. *Artificial Intelligence* **2005**, *165*, 37–56. https://doi.org/10.1016/j.artint.2005.01.007.
99. Ramirez, R.; Maestre, E.; Pertusa, A.; Gomez, E.; Serra, X. Performance-based interpreter identification in saxophone audio recordings. *IEEE Transactions on Circuits and Systems for Video Technology* **2007**, *17*, 356–364. https://doi.org/10.1109/TCSVT.2007.890862.
100. Costalonga, L.L.; Pimenta, M.S.; Miranda, E.R. Understanding biomechanical constraints for modelling expressive performance: A guitar case study. *Journal of New Music Research* **2019**, *48*, 331–351. https://doi.org/10.1080/09298215.2019.1643892.
101. Metcalf, C.D.; Irvine, T.A.; Sims, J.L.; Wang, Y.L.; Su, A.W.; Norris, D.O. Complex hand dexterity: a review of biomechanical methods for measuring musical performance. *Frontiers in psychology* **2014**, *5*. https://doi.org/10.3389/fpsyg.2014.00414.
102. Wristen, B.G. Avoiding piano-related injury: a proposed theoretical procedure for biomechanical analysis of piano technique. *Medical Problems of Performing Artists* **2000**, *15*, 55–64. https://doi.org/10.21091/mppa.2000.2012.
103. Parncutt, R.; Sloboda, J.A.; Clarke, E.F.; Raekallio, M.; Desain, P. An ergonomic model of keyboard fingering for melodic fragments. *Music Perception* **1997**, *14*, 341–382. https://doi.org/10.2307/40285730.
104. Jacobs, J.P. Refinements to the ergonomic model for keyboard fingering of Parncutt, Sloboda, Clarke, Raekallio, and Desain. *Music Perception* **2001**, *18*, 505–511. https://doi.org/10.1525/mp.2001.18.4.505.
105. Visentin, P.; Li, S.; Tardif, G.; Shan, G. Unraveling mysteries of personal performance style; biomechanics of left-hand position changes (shifting) in violin performance. *PeerJ* **2015**, *3*. https://doi.org/10.7717/peerj.1299.
106. Repp, B.H. Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological research* **1994**, *56*, 269–284. https://doi.org/10.1007/BF00419657.
107. Marchini, M.; Ramirez, R.; Papiotis, P.; Maestre, E. The sense of ensemble: a machine learning approach to expressive performance modelling in string quartets. *Journal of New Music Research* **2014**, *43*, 303–317. https://doi.org/10.1080/09298215.2014.922999.
108. Sundberg, J.; Friberg, A.; Frydén, L. Rules for automated performance of ensemble music. *Contemporary Music Review* **1989**, *3*, 89–109. https://doi.org/10.1080/07494468900640071.
109. Mathews, M.V. The radio baton and conductor program, or: pitch, the most important and least expressive part of music. *Computer Music Journal* **1991**, *15*, 37–46. https://doi.org/10.2307/3681070.
110. Lawson, J.; Mathews, M.V. Computer program to control a digital real-time sound synthesizer. *Computer Music Journal* **1977**, pp. 16–21.
111. Mathews, M.V. The conductor program and mechanical baton. In Proceedings of the Proccedings of 1989 International Symposium on Music and Information Science, 1989, pp. 58–70.
112. Lee, E.; Nakra, T.M.; Borchers, J. You're the conductor: a realistic interactive conducting system for children. In Proceedings of the Proceedings of the International Conference on New Interfaces for Musical Expression (NIME), 2004, pp. 68–73. https://doi.org/10.5281/zenodo.1176629.
113. Baba, T.; Hashida, M.; Katayose, H. "VirtualPhilharmony": a conducting system with heuristics of conducting an orchestra. In Proceedings of the Proceedings of the International Conference on New Interfaces for Musical Expression (NIME), 2010, Vol. 2010, pp. 263–270. https://doi.org/10.5281/zenodo.1177715.
114. Mathews, M.V.; Friberg, A.; Bennett, G.; Sapp, C.; Sundberg, J. A marriage of the Director Musices program and the conductor program. In Proceedings of the Proceedings of the Stockholm Music Acoustics Conference (SMAC), 2003, Vol. 1, pp. 13–16.
115. Friberg, A. pDM: an expressive sequencer with real-time control of the KTH music-performance rules. *Computer Music Journal* **2006**, *30*, 37–48. https://doi.org/10.1162/comj.2006.30.1.37.
116. Canazza, S.; Friberg, A.; Rodà, A.; Zanon, P. Expressive Director: a system for the real-time control of music performance synthesis. In Proceedings of the Proceedings of Stockholm Music Acoustics Conference (SMAC), 2003, Vol. 2, pp. 521–524.

117. Dixon, S.; Goebl, W.; Widmer, G. The "air worm": an interface for real-time manipulation of expressive music performance. In Proceedings of the Proceedings of the International Computer Music Conference (ICMC). Michigan Publishing, 2005.
118. Friberg, A.; Colombo, V.; Frydén, L.; Sundberg, J. Generating musical performances with director musices. *Computer Music Journal* **2000**, *24*, 23–29. https://doi.org/10.1162/014892600559407.
119. Giraldo, S.; Ramírez, R. Performance to score sequence matching for automatic ornament detection in jazz music. In Proceedings of the International Conference of New Music Concepts (ICMNC), 2015, Vol. 8.