OVERVIEW ARTICLE

An Interdisciplinary Review of Music Performance Analysis

Alexander Lerch, Claire Arthur, Ashis Pati and Siddharth Gururani

A musical performance renders an acoustic realization of a musical score or other representation of a composition. Different performances of the same composition may vary in terms of performance parameters such as timing or dynamics, and these variations may have a major impact on how a listener perceives the music. The analysis of music performance has traditionally been a peripheral topic for the MIR research community, where often a single audio recording is used as representative of a musical work. This paper surveys the field of Music Performance Analysis (MPA) from several perspectives including the measurement of performance parameters, the relation of those parameters to the actions and intentions of a performer or perceptual effects on a listener, and finally the assessment of musical performance. This paper also discusses MPA as it relates to MIR, pointing out opportunities for collaboration and future research in both areas.

1. Introduction

Music, as a performing art, requires a performer or group of performers to render a musical "blueprint" into an acoustic realization (Hill, 2002; Clarke, 2002b). This musical blueprint can, for example, be a score as in the case of Western classical music, a lead sheet for jazz, or some other genre-dependent representation describing the compositional content of a piece of music. The performers are usually musicians but might also be, e.g., a computer rendering audio.

The performance plays a major role in how listeners perceive a piece of music: even if the blueprint is identical for different renditions, as is the case in Western classical music, listeners may prefer one performance over another and appreciate different 'interpretations' of the same piece of music. These differences are the result of the performers' intentionally or unintentionally interpreting, modifying, adding to, and dismissing information from the score or blueprint (for the sake of simplicity, the remainder of this text will use the terms score and blueprint synonymously). This constant re-interpretation of music is inherent to the art form and is a vital and expected component of music.

Formally, musical communication can be described as a chain as shown in **Figure 1**: the composer typically only communicates with the listener through the performer who renders the blueprint to convey musical ideas to the listener (Kendall and Carterette, 1990). The performer does this by varying musical parameters while leaving the compositional content untouched. The visualized model of the communication chain displays a one-way communication path of information transmission. There can be, however, flows of information in the other direction, influencing the performance itself. Such a feedback path might transport information such as the instrument's sound and vibration (Todd 1993), the room acoustics, (Luizard et al., 2020), and the audience reaction (Kawase, 2014). Although the recording studio lacks an audience, a performance can also evolve during a recording session. Katz (2004) points out that in such a session, performers will listen to the recording of themselves and adjust "aspects of style and interpretation." In addition, the producer might also have impact on the recorded performance (Maempel, 2011).

The large variety of performance scenarios makes it necessary to focus on the common core of all music performances: the audio signal. While a music performance can, for example, contain visual information such as gestures and facial expressions (Bergeron and Lopes, 2009; Platz and Kopiez, 2012; Tsay, 2013), not every music performance has these cues. A musical robot, for example, may or may not convey such cues. The acoustic rendition, however, is the integral part of a music performance that cannot be missing; simply put, there exists no music without sound. An audio recording is a fitting representation of the sound that allows for quantitative analysis. This audio-focused view should neither imply
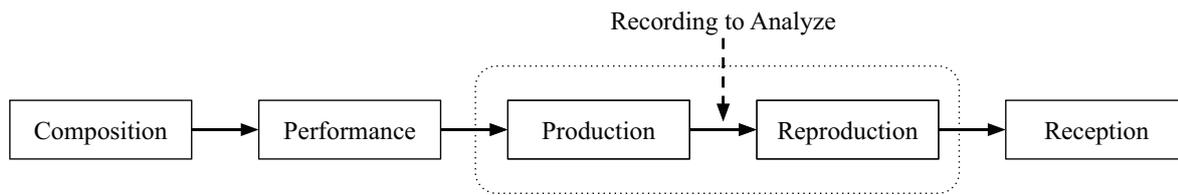
Center for Music Technology, Georgia Institute of Technology, US

Corresponding author: Alexander Lerch
(alexander.lerch@gatech.edu)

Recording to Analyze

Composition → Performance → Production → Reproduction → Reception

Figure 1: Chain of musical communication from composer to listener. The dotted box disappears for performances that are not recorded.

that non-audio information cannot be an important part of a performance nor that non-audio information cannot be analyzed. The focus on the *audio recording* of a performance makes it important to recognize that every recording contains processing choices and interventions by the production team with potential impact on the expressivity of the recording. Maempel (2011) discusses as main influences in the context of classical music the sound engineer (dynamics, timbre, panorama, depth) and the editor (splicing of different tracks, tempo and timing, pitch). These manipulations and any restrictions of the recording and distribution medium are part of the audio to be analyzed and cannot be separated anymore from the performers' creation. As the release of a recording usually has to be pre-approved by the performers it is assumed that the performers' intent has not been distorted.

Although the change of performance parameters can have a major impact on a listener's perception of the music (Clarke, 2002a, also Section 4), the nature of the performance parameter variation is often subtle, and must be evaluated in reference to something; typically either the same performer and piece at a different time, different performances of the same piece, or deviations from a perfectly quantized performance rendered from a symbolic score representation (e.g., MIDI) without tempo or dynamics variations. Notice that this reference problem makes the genre of music for performance analysis biased towards classical music, as there is commonly an abstract (i.e., score) representation that easily fills the model of the quantized version of a piece of music. In addition, the classical model of performance is based on a finite number of pre-composed musical works performed by numerous individuals over time, thus making it a rich resource for performance analysis. **Figure 2** visualizes two example piano performances of the beginning of Frédéric Chopin's *Fantasie in F Minor, Op. 49*, where variations in local tempo, dynamics, and pedaling can easily be identified. Even in musical traditions that also have scores, such as jazz, the role of the performer as 'interpreter' of the score is complicated by normative practices such as elaborate embellishment and improvisation (often to the point of highly obscuring the original material). Other genres might completely eliminate the concept of separable composition and performance. McNeil (2017) argues, for example, that the distinction between 'composition' and 'improvisation' cannot capture the essence of performance creativity in Indian Hindustani music, where the definition of fixed 'composed' material and improvised
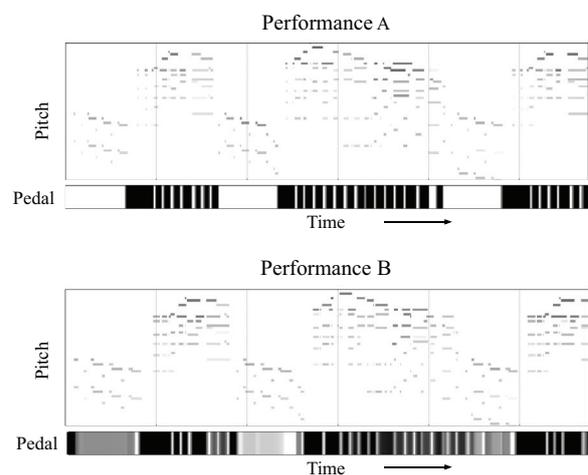

Figure 2: Variations in two different performances of Frédéric Chopin's Fantasie in F Minor, Op. 49 (taken from the Maestro dataset (Hawthorne et al., 2019). For each performance, timing and dynamics are shown using the piano rolls (darker color indicates higher velocity). Pedal control is shown below the piano roll (darker color indicates increasing usage of the pedal). Visualized excerpts correspond to the beginning of the piece.

material becomes hard and it might be more meaningful to refer to seed ideas which grow and expand throughout the performance.

Although the distinction between score and performance parameters is less obvious for non-classical genres of Western music, especially ones without clear separation between the composer and the performer, the concept and role of a performer as interpreter of a composition is still very much present, be it as a live interpretation of a studio recording or a cover version of another artist's song. In these cases, the freedom of the performers in modifying the score information is often much higher than it is for classical music — reinterpreting a jazz standard can, for example, include the modification of content related to pitch, harmony, and rhythm.

Formally, performance parameters can be structured in the same basic categories that we use to describe musical audio in general: tempo and timing, dynamics, pitch, and timbre (Lerch, 2012). While the importance of different parameters might vary from genre to genre, the following list introduces some mostly genre-agnostic examples to clarify these performance parameter categories:

- *Tempo and Timing*: the score specifies the general rhythmic content, just as it often contains a tempo indicator. While the rhythm is often only slightly modified by performers in terms of micro-timing, the tempo (both in terms of overall tempo as well as expressive tempo variation) is frequently seen only as a suggestion to the performer.
- *Dynamics*: in most cases, score information on dynamics is missing or only roughly defined. The performers will vary loudness and highlight specific events with accents based on their plan for phrasing and tension, and the importance of certain parts of the score.
- *Pitch*: the score usually defines the general pitches to play, but pitch-based performance parameters include expressive techniques such as vibrato as well as intentional or unintentional choices for intonation.
- *Timbre*: as the least specific category of musical parameters, scores encode timbre often only implicitly (e.g., instrumentation) while performers can, for example, change playing techniques or the choice of specific instrument configurations (such as the choice of organ registers).

There exist many performance parameters and playing techniques that either cannot be easily associated with one of the above categories or span multiple categories; examples of such parameters are articulation (*legato, staccato, pizzicato*) or forms of ornamentation in Baroque or jazz music.

One intuitive form of Music Performance Analysis (MPA) —discussing, criticizing, and assessing a performance after a concert— has arguably taken place since music was first performed. Traditionally, however, such reviews are qualitative and not empirical. While there are a multitude of approaches to music performance analysis, and numerous factors shape the insights and outcomes of such analyses, in this literature review we are only concerned with quantitative, systematic approaches to MPA scholarship. However, it is easily acknowledged that the design of any empirical study and the interpretation of results could be improved by a careful consideration of the musicological context (e.g., the choice of edition that a classical performance is rendered from — see Rink (2003)). Nevertheless, a detailed discussion of such contexts is beyond the scope of this article.

Early attempts at systematic and empirical MPA can be traced back to the 1930s with vibrato and singing analysis by Seashore (1938) and the examination of piano rolls by Hartmann (1932). In the past two decades, MPA has greatly benefited from the advances in audio analysis made by members of the Music Information Retrieval (MIR) community, significantly extending the volume of empirical data by simplifying access to a continuously growing heritage of commercial audio recordings. While advances in audio content analysis have had clear impact on MPA, the opposite is less true. An informal search reveals that while there have been publications on performance analysis at ISMIR, the major MIR conference, their absolute number remains comparably small

(compare Toyoda et al., 2004; Takeda et al., 2004; Chuan and Chew, 2007; Sapp, 2007, 2008; Hashida et al., 2008; Liem and Hanjalic, 2011; Okumura et al., 2011; Devaney et al., 2012; Jure et al., 2012; Van Herwaarden et al., 2014; Liem and Hanjalic, 2015; Arzt and Widmer, 2015; Page et al., 2015; Xia et al., 2015; Bantula et al., 2016; Peperkamp et al., 2017; Gadermaier and Widmer, 2019; Maezawa et al., 2019 with a title referring to "music performance" out of approximately 1,950 ISMIR papers overall).

Historically, MIR researchers often do not distinguish between score-like information and performance information even if the research deals with audio recordings of performances. For instance, the goal of music transcription, a very popular MIR task, is usually to transcribe all pitches with their onset times (Benetos et al., 2013); that means that a successful transcription system transcribes two renditions of the same piece of music differently, although the ultimate goal is to detect the same score (note that this is not necessarily true for all genres). Therefore, we can identify a disconnect between MIR research and performance research that impedes both the evolution of MPA approaches and robust MIR algorithms, slows gaining new insights into music aesthetics, and hampers the development of practical applications such as new educational tools for music practice and assessment. This paper aims at narrowing this gap by introducing and discussing MPA and its challenges from an MIR perspective. In pursuit of this goal, this paper complements previous review articles on music performance research (Sloboda, 1982; Palmer, 1997; Gabrielsson, 1999, 2003; Goebl et al., 2005) and expands on Lerch et al. (2019) by integrating non-classical and non-Western music genres, including a more extensive number of relevant publications, and clearly outlining the challenges music performance research is facing. While performance research has been inclusive of various musical genres, such as the Jingju music of the Beijing opera (Zhang et al., 2017; Gong, 2018), traditional Indian music (Clayton, 2008; Gupta and Rao, 2012; Narang and Rao, 2017) and jazz music (Abeßer et al., 2017), the vast majority of studies are concerned with Western classical music. The focus on Western music can also be observed in the field of MIR in general, despite efforts in diversifying the field (Serra, 2014; Tzanetakis, 2014). As mentioned, the reason for the focus on classical music within MPA may be due to the clear systematic differentiation between score and performance. This imbalance means that this overview article will necessarily emphasize Western classical music while referring to other musical styles wherever appropriate.

The remainder of this paper is structured as follows. The following Section 2 presents research on the objective description, modeling, and visualization of the performance itself, identifying commonalities and differences between performances. The subsequent sections focus on studies taking these objective performance parameters and relating them to either the performer (Section 3), the listener (Section 4), or the assessment of the performer from a listener's perspective

(Section 5). We conclude our overview with a summary on applications of MPA and final remarks in Section 6.

2. Performance Measurement

A large body of work focuses on an exploratory approach to analyzing performance recordings and describing performance characteristics. Such studies typically extract characteristics such as the tempo curve or histogram (Repp, 1990; Palmer, 1989; Povel, 1977; Srinivasamurthy et al., 2017) or loudness curve (Repp, 1998a; Seashore, 1938) from the audio and aim at either gaining general knowledge on performances or comparing attributes between different performances/performers based on trends observed in the extracted data. Additionally, there are also studies focusing on discovery of general patterns in performance parameters, which can be useful in identifying trends such as changes over eras (Ornoy and Cohen, 2018).

2.1 Tempo, timing, and dynamics

Deviations in tempo, timing, and dynamics are considered to be some of the most salient performance parameters and hence have been the focus of various studies in MPA. While these performance parameters have been studied in isolation in some instances, we present them together since there is substantial work aiming to understand their interrelation.

Close relationships were observed between musical phrase structure and deviations in tempo and timing (Povel, 1977; Shaffer, 1984; Palmer, 1997). For example, tempo changes in the form of *ritardandi* tend to occur at phrase boundaries (Palmer, 1989; Lerch, 2009). As a related structural cue, Chew (2016) proposed the concept of tipping points in the score, leading to a timing deviation with extreme pulse variability in the context of Western classical music performance.

Correlations were observed between timing and dynamics patterns (Repp, 1996b; Lerch, 2009). Dalla Bella and Palmer (2004) found that the overall tempo influences the overall loudness of a performance. There are also indications that loudness can be linked to pitch height (Repp, 1996b). Cheng and Chew (2008) analyzed global phrasing strategies for violin performance using loudness and tempo variation profiles and found dynamics to be more closely related to phrasing than tempo. While the close relation of tempo and dynamics to structure has been repeatedly verified, Lerch (2009) did not succeed in finding similar relationships between structure and timbre properties in the case of string quartet recordings.

In the context of jazz performances, Wesolowski (2016) found that both score and performance parameters such as underlying harmony, pitch interval size, articulation, and tempo had significant correlations with timing variations between successive eighth notes. He also found these parameters correlated with synchronicity between separate parts of a jazz ensemble. Several researchers conducted experiments studying swing style jazz (Ellis, 1991; Prögler 1995; Collier and Collier 2002; Friberg and

Sundström, 2002). Such studies focused on measuring or quantifying discrepancies and asynchrony of performers in order to study what characteristics of jazz performances made them 'swing' (Friberg and Sundström, 2002). Ellis (1991) found that asynchrony of jazz swing performers to the prevailing meter is positively correlated with the tempo and consists mainly of delaying attacks. Prögler (1995) noted that participatory asynchrony in swing is observed and measurable at a subsyntax level. Abeßer et al. (2014a) studied the relationship of note dynamics in jazz improvisation with other contextual information such as note duration, pitch, and position in the score. They used a score-informed music source separation algorithm to isolate the solo instrument and found that higher and longer notes tend to be louder, and structural accents are typically emphasized. Busse (2002) conducted an experiment to objectively measure deviations in terms of timing, articulation, and dynamics of jazz swing performers from mechanical regularity using MIDI-based 'groove quantization.' He created reference performer models using the measured performance parameter deviations and compared them against mechanical or quantized models. Experts were asked to rate the 'swing representativeness' of the different models. He found that reference performer models were rated to be representative of swing whilst similar mechanical models of performance were rated poorly. Ashley (2002) described timing in jazz ballad performance as melodic rhythm flexibility over a strict underlying beat pattern, which is a type of rubato. He found timing deviations to have strong relationships with musical structure. Collier and Collier (1994) studied tempo in corpora of jazz performances. They manually timed these recordings and found that tempo was normally (Gaussian) distributed when computed in terms of metronome markings but not when computed using note durations. They noted that while jazz performances are stable in terms of tempo, systematic patterns in timing variabilities tend to serve expressive functions. Iyer (2002) contrasted micro-timing in African-American music from techniques in Western classical music such as *rubato* and *ritardando*. Employing an embodied cognition framework, he argued that the differences are due to the emphasis of the human body in the cultural aesthetics of African-American musics.

Studies analyzing musical timing are not only limited to Western music. Srinivasamurthy et al. (2017) performed a large-scale computational analysis of rhythm in Hindustani classical percussion, confirming and quantifying tendencies pertaining to timing such as deviations of tempo within a metric cycle (also referred to as *tal*). The study demonstrated the value of using MIR techniques for rhythm analysis of large corpora of music. Bektaş (2005) confirmed the relationship between prosodic meter *arûz* and musical meter *usûl* in Turkish vocal music and found the existence of an even stronger concordance between a subset of prosodic patterns called *bahir* and *usûl*.

In addition to the studies already presented, there is a large body of work that focuses on modeling of

timing, tempo, and dynamics. Section 2.3 discusses such modeling approaches to performance measurement in detail.

2.2 Tonal analysis

Pitch-based performance parameters have been analyzed mostly in the context of single-voiced instruments. For instance, the vibrato range and rate has been studied for vocalists (Seashore, 1938; Devaney et al., 2011) and violinists (Bowman Macleod, 2006; Dimov, 2010). Regarding intonation, Devaney et al. (2011) found significant differences between professional and non-professional vocalists in terms of the size of the interval between semi-tones.

Studies have also been conducted on the relationship between pitch and meter in the context of bebop style jazz (Järvinen, 1995; and Toiviainen, 2000). These studies found that measurements of chorus-level tonal hierarchies match quite closely to rating profiles of chromatic pitches found in European art music and that metrical structure plays a role in determining which pitches are emphasized or de-emphasized. The studies also indicated that there is no effect of syncopation or polyrhythm on the use of certain pitches.

Franz (1998) demonstrated the utility of Markov chains in the analysis of jazz improvisation. He found Markov chains useful in quantifying the frequency of notes and patterns of notes which are in turn useful as comparison tools for musical scale analyses. Furthermore, he noted that this modeling technique can be useful for stylistic comparison as well as in developing metrics for style and creativity. Frieler et al. (2016) proposed an analysis framework for jazz improvisation based on so-called 'midlevel units' (MLU). These are musical units on the middle level between individual notes and larger form parts. They hypothesize that MLUs correspond to the improvisers' playing ideas and musical ideas and propose a taxonomy system for MLUs. The authors subsequently study the distribution of occurrences and durations of MLUs in a large corpus of jazz improvisations. They note that the most common MLUs used in improvisations belong to the *lick* and *line* categories. They also find that the distribution of MLU types differs between performers and styles.

Research on jazz solos has utilized MIR methods for source separation and pitch tracking to analyze intonation and other tonal features (Abeßer et al., 2014c, 2015). Abeßer et al. (2014c) proposed a score-informed pitch tracking algorithm for analysis. They analyze distributions of various pitch contour features to identify patterns between different artists and instruments. Several contextual parameters such as relative position of notes in a phrase, beat position in the bar, etc., are also extracted and correlations between these parameters and pitch features are studied. They found statistically significant correlations between the pitch contour features and contextual features but most of the correlations had small effect size. Abeßer et al. (2015) utilized score-based source separation and

pitch tracking to study intonation in jazz brass and woodwind solos to identify trends in tuning frequency over various decades, intonation for different artists, and properties of vibrato depending on context and performer.

A survey of computational methods utilized to study tonality in Turkish *Makam* music was conducted by Bozkurt et al. (2014), including research related to the analysis of tuning frequency and melodic phrases, transcription of performances, *Makam* recognition, as well as rhythmic and timbral analysis in *Makam* music. Atli et al. (2015) discuss the importance of tonic frequency or *karar* estimation for *Makam* music and devise a simple method for the task: they find that detecting the last note of a performance recording and estimating the frequency works well for *karar* estimation. Hakan et al. (2012) analyzed Turkish *ney* performances to identify key aspects of embellishments. They found that the rate of change of vibrato and 'pitch bump', which measures the deviation of pitch just before ascending or descending into the next note, were the key features useful for distinguishing performance styles across various performers.

Similarly, several researchers have worked on analysis of melody and tonality in Indian classical music. Ganguli and Rao (2017) modeled ungrammatical phrases, i.e., phrases straying away from the predetermined *raga*, in Hindustani music performance. They utilized computational techniques to model the tonal hierarchies and melodic shapes of different *ragas* toward that end. Viraraghavan et al. (2017) analyzed the use of ornamentation known as *gamakas* in *Carnatic* music performances. They find that the use of *gamakas* is vital in defining the *raga* during a performance.

Chen (2013) developed methods for analysis of intonation in Beijing opera or *Jingju* music. The methods, involving peak distribution analysis of pitch histograms, validated claims in literature that the fourth degree is higher, and the seventh degree is lower than the corresponding pitches in the equally tempered scale. Chen also found pitch histograms to be good features to distinguish role types in opera performances. Caro Repetto et al. (2015) utilized MIR techniques for pitch tracking and audio feature extraction to compare singing styles of two *Jingju* schools, namely the *Mei* and *Cheng* schools. Their experiments quantitatively support observations made in musicological texts about characteristics such as pitch register, vibrato, volume/dynamics, and timbre brightness. In addition, they find other properties not previously reported; for example, vibrato in *Mei* tends to be slower and wider on average than in *Cheng*. Yang et al. (2015) developed methods based on filter diagonalization and hidden Markov models to detect and model vibrato and portamento in performances of *erhu*, violin and *Jingju* opera vocals. There are studies aiming to understand the relationship between linguistic tone and melodic pitch contours in *Jingju* music by utilizing machine learning methods such as clustering (Zhang et al., 2015, 2014). *Jingju* music utilizes a two-dialect tone system making the tone-melody relationship complicated.

2.3 Modeling

While most of the studies mentioned above make use of statistical methods to extract, summarize, visualize, and investigate patterns in the performance data, researchers have also investigated modeling approaches to better understand performances. Several overview articles exist covering research on generating expressive musical performance (Cancino-Chacón et al., 2018; Kirke and Miranda, 2013; Widmer and Goebl, 2004). In this subsection, however, we primarily focus on methods that model performance parameters leading to useful insights, while ignoring the generative aspect.

Several researchers have attempted to model timing variations in performances. In early work, Todd (1995) modeled *ritardandi* and *accelerandi* using kinematics theory, concluding that tempo variation is analogous to velocity. Li et al. (2017) introduced an approach invariant to phrase length for analyzing expressive timing. They utilize Gaussian mixture models (GMMs) to model the polynomial regression coefficients for tempo curves instead of directly modeling expressive timing. Liem and Hanjalic (2011) proposed an entropy-based deviation measure for quantifying timing in piano performances and found it to be a good alternative to standard-deviation-based measures. Grachten et al. (2017) utilized recurrent neural networks to model timing in performances and demonstrated the benefits over static models that do not account for temporal dependencies between score features. Stowell and Chew (2013) introduced a Bayesian model of tempo modulation at various time-scales in performances.

For dynamics modeling, Kosta et al. (2015) applied and compared two change-point algorithms to detect dynamics changes in performed music, evaluating them using the corresponding dynamics markers in the score. In further research, Kosta et al. (2018) quantified the relationships between notated and performed dynamics using a corpus of performed Chopin Mazurkas. Kosta et al. (2016) applied various machine learning methods such as decision trees, support vector machines (SVM) and neural networks to understand the relationship between dynamics markings and performed loudness. They find that score-based features are more important than performer style features for predicting dynamics markings given performed loudness and vice versa. Similarly, Marchini et al. (2014) utilized decision trees, SVMs and k-nearest neighbor classifiers to model and predict performance features such as intensity, timing deviations, vibrato extent, and bowing speed of each note in string quartet performances. They found that inter-voice attributes played a strong role in models trained with ensemble recordings versus solo recordings. Grachten and Widmer (2012) introduced a so-called linear basis function model which encodes score information using weighted combinations of a set of basis functions. They utilize this model to predict and analyze dynamics in performance. Grachten et al. (2017) extended the framework with a recurrent model which better captures temporal relationships in order to improve modeling of

timing. Cancino-Chacón et al. (2017) performed a large-scale evaluation of linear, non-linear and temporal models for dynamics in piano and orchestral performances. These models utilize various features extracted from the score to encode structure in dynamics, pitch and rhythm.

An alternative direction in performance modeling research involves methods to discover rules for performance (Widmer, 2003). Widmer's ensemble learning method succeeds in finding simple, and in some cases novel, rules for music performance. An approach to modeling expressive jazz performance based on genetic algorithms was proposed by Ramirez et al. (2008), learning rules to generate sequences that are best able to fit the training data.

2.4 Visualization

Many traditional approaches to performance parameter visualization such as pitch contours (Gupta and Rao, 2012; Abeßer et al., 2014c), tempo curves (Repp, 1990; Palmer, 1989; Povel, 1977), and scatter plots (Lerch, 2009) are not necessarily interpretable or easily utilized for comparative studies. This led researchers to develop other, potentially more intuitive or condensed forms of visualization that allow describing and comparing different performances. The 'performance worm', for example, is a pseudo-3D visualization of the tempo-loudness space that allows the identification of specific gestures in that space (Langner and Goebl, 2002; Dixon et al., 2002). Sapp (2007, 2008) proposed the so-called 'timescapes' and 'dynascapes' to visualize subsequent similarities of performance parameters.

The 'Phenicx' project explored various ways of visualizing orchestral music information, including both score and performance information (Gasser et al., 2015). Dynagrams and tempograms, for example, are used to visualize various temporal levels of loudness and tempo variations, respectively. Dittmar et al. (2018) devised swingogram representations by analyzing and tracking the swing ratio implied by the ride cymbal in jazz swing performance. This visualization enables insights into jazz improvisation such as the interaction between a soloist and the drummer.

2.5 Challenges

The studies presented in this section often follow an exploratory approach; extracting various parameters in order to identify commonalities or differences between performances. While this is, of course, of considerable interest, one of the main challenges is the interpretability of these results. Just because there is a timing difference between two performances does not necessarily mean that this difference is musically or perceptually meaningful. Without this link, however, results can only provide limited insights into which parameters and parameter variations are ultimately important.

The acquisition of data for analysis is another challenge in MPA research. Goebl et al. (2005) discuss various methods that have been used to that end, including special instruments (such as Yamaha Disklaviers, piano

rolls), hand measurement of performance parameters, as well as automatic audio analysis tools. All these methods have potential downsides. For example, the use of special instruments and sensors excludes the analysis of performances not recorded on these specific devices and the associated formats (e.g., MIDI) may have difficulties representing special playing techniques. Manual annotation can be time consuming and tedious. The fact that the majority of studies surveyed here rely on manually annotated data implies that available algorithms for automatic performance parameter extraction lack the reliability and/or accuracy for practical MPA tasks. This is especially true for ensemble performances where the polyphonic and poly-timbral nature as well as timing fluctuations between individual voices complicate the analysis. As a result of these challenges, most studies are performed on manually annotated data with small sample sizes, possibly leading to poor generalizability of the results. The increasing number of datasets providing performance data as listed in **Table 1**, however, gives hope that this ceases to be an issue in the future.

3. Performer

While most studies focus on the extraction of performance parameters or the mapping of these parameters to the listeners' perception (see Sections 4 and 5), some investigate the capabilities, goals, and strategies of performers. A performance is usually based on an explicit or implicit performance plan with clear intentions (Clarke, 2002b). This seems to be the case also for improvised music: for instance, Dean et al. (2014) could verify clearly perceivable structural boundaries in free jazz piano improvisation. There is, as Palmer verified, a clear relation between reported intentions and objective parameters related to phrasing and timing of the performance (Palmer, 1989). Similar relations between the intended emotionality and loudness and timing measures were reported in multiple studies (Juslin, 2000; Dillon, 2001, 2003, 2004). For example, projected emotions such as anger and sadness show significant correlations with high and low tempo, and high and low overall sound level, respectively. Moreover, a performer's control of expressive variation has been shown to significantly improve the conveyance of emotion. For instance, a study by Vieillard et al. (2012) found that listeners were better able to perceive the presence of specific emotions in music when the performer played an 'expressive' (as opposed to a mechanical) rendition of the composition. This suggests that the performer plays a fairly large role in communicating an emotional 'message' above and beyond what is communicated through the score alone (Juslin and Laukka, 2003). In music performed from a score in particular, the score-based representation might be thought of as a set of instructions in the sense that the notational system itself is used to communicate basic structural information to the performer. However, as noted by Rink (2003), the performer is not simply a medium or vessel through which performance directions

are carried out, but "what performers do has the potential to impart meaning and create structural understanding."

Research by Friberg and Sundström (2002) set out to tackle the question of what makes music 'swing.' Their approach was to examine the variation in the 'swing ratio' between pairs of eighth notes in jazz music, and they found that it tends to vary as a function of tempo. This finding has interesting implications for MPA as well as perceptual experiments. Across performers there is a clear systematic relation between the stretching of the ratio at slower tempi and the compressing of the ratio at higher tempi, such that it approaches a 1:1 ratio at approximately 300 BPM. Interestingly, the duration of the second eighth note remained fairly constant at approximately 100 ms for medium to fast tempi, suggesting a practical limit on tone duration that, as the authors speculate, could be due to perceptual factors.

Another interesting area of research is performer error. Repp (1996a) analyzed performers' mistakes and found that errors were concentrated in mostly unimportant parts of the score (e.g., middle voices) where they are harder to recognize (Huron, 2001), suggesting that performers intentionally or unintentionally avoid salient mistakes.

3.1 Influences

In addition to the performance plan itself, there are other influences shaping the performance. Acoustic parameters of concert halls such as the early decay time have been shown to impact performance parameters such as tempo (Schärer Kalkandjiev and Weinzierl, 2013, 2015; Luizard et al., 2019, 2020). Related work by Repp showed that pedaling characteristics in piano performance are dependent on the overall tempo (Repp, 1996c, 1997b).

Other studies investigate the importance of the feedback of the music instrument to the performer (Sloboda, 1982); there have been studies reporting on the effect of deprivation of auditory feedback (Repp, 1999), investigating the performers' reaction to delayed or changed auditory feedback (Pfordresher and Palmer, 2002; Finney and Palmer, 2003; Pfordresher, 2005), or evaluating the role of tactile feedback in a piano performance (Goebl and Palmer, 2008). In summary, the different forms of feedback have been found to have small but significant impact on reproduction accuracy of performance parameters.

3.2 Challenges

There is a wealth of information about performances that can be learned from performers. The main challenge of this direction of inquiry is that such studies have to involve the performers themselves. This limits the amount of available data and usually excludes well-known and famous artists, resulting in a possible lack of generalizability. Depending on the experimental design, the separation of possible confounding variables (for example, motor skills, random variations, and the influence of common performance rules) from the scrutinized performance data can be a considerable challenge.

4. Listener

Every performance will ultimately be heard and processed by a listener. The listener's meaningful interpretation of the incoming musical information depends on a sophisticated network of parameters. These parameters include both objective (or, at least, measurable) features that can be estimated from a score or derived from a performance, as well as subjective and 'internal' ones such as factors shaped by the culture, training, and history of the listener. Presently, there remain many acoustic parameters related to music performance where the listener's response has not been measured, either in terms of perceptibility or aesthetic response or both. For this reason, listener-focused MPA remains one of the most challenging and elusive areas of research. However, to the extent that MPA research and its applications depend on perceptual information (e.g., perceived expressiveness), or intend to deliver perceptually-relevant output (e.g., performance evaluation or reception, similarity ratings), it is imperative to achieve a fuller understanding of the perceptual relevance of the manipulation and interaction of performance characteristics (e.g., tempo, dynamics, articulation). The subsequent paragraphs provide a brief overview of the relevant literature on music perception and MPA, along with some discussion of the relevance of this information for current and future work in both MPA and in MIR in general.

4.1 Musical expression

When it comes to listener judgments of a performance, it remains poorly understood which aspects are most important, salient, or pertinent for the listener's sense of satisfaction. According to Schubert and Fabian (2014), listeners are very concerned with the notion of 'expressiveness' which is a complex, multifaceted construct. Performance expression is commonly defined as "variations in musical parameters by a signal or instrumentalist" (Dibben, 2014). In other words, performance expression implies the intentional application of systematic variation on the part of the performer. On the other hand, expressive performance (or 'expressiveness') implies a judgment (either implicit or explicit) on the part of a listener.

As stated by Devaney (2016), however, not all variation is expressive: "The challenge [...] is determining which deviations are intentional, which are due to random variation, and which are due to specific physical constraints that a given performer faces, such as bio-mechanical limitations [...]. In regard to physical limitations, these deviations may be both systematic and observable in collected performance data, but may not be perceptible to listeners." Thus, identifying the variation in a performance that would be *intended as* expressive is only the first step. Discovering which performance characteristics contribute to an expressive performance requires dissecting what listeners deem 'expressive' as well as understanding the relation and potential differences between measured and perceived performance features.

Expressiveness is genre and style dependent, meaning that the perceived *appropriate* level and style of expression

in a pop ballad will be different from a jazz ballad, and that expression in a Baroque piece will be different from that of a Romantic piece —something that has been referred to as 'stylishness' (Fabian and Schubert, 2009; Kendall and Carterette, 1990). For example, the timing difference between the primary melody and the accompaniment tends to be wider in jazz than in classical music, and there is evidence that the direction of difference is reversed, i.e., the melody leads the accompaniment in classical piano music (Goebl, 2001; Palmer, 1996) while it follows in the case of jazz (Ashley, 2002). Similarly, syncopation created by anticipating the beat is normative in pop genres but appears to be reversed in jazz music where syncopation is created by delaying the onset of the melody (Dibben, 2014).

In addition to style-related expression, there is the perceived *amount* of expressiveness, which is considered independent of stylishness (Schubert and Fabian, 2006). Finally, Schubert and Fabian (2014) distinguish a third 'layer' of expressiveness, *emotional expressiveness*, which arises from a performer's manipulation of various features specifically to alter or enhance emotion. This is distinct from *musical* expressiveness, or expressive variation, which more generally refers to the manipulation of compositional elements by the performer in order to be 'expressive' without necessarily needing to express a specific emotion. Practically speaking, however, it may be difficult for listeners to separate these varieties of expressiveness (Schubert and Fabian, 2014, p.293), and research has demonstrated that there are interactions between them (e.g., Vieillard et al., 2012).

4.2 Expressive variation

Several scholars have made significant advances in our understanding of the role of timing, tempo, and dynamic variation on listeners' perception of music. As noted in Section 2, the subtle variations in tempo and dynamics executed by a performer have been shown to play a large role in highlighting and segmenting musical structure. For instance, the perception of metrical structure is largely mediated through changes in timing and articulation within small structural units such as the measure, beat, or sub-beat, whereas the perception of formal structures are largely communicated through changes across larger segments such as phrases (e.g., Sloboda, 1983; Gabrielsson, 1987; Palmer, 1996; Behne and Wetekam, 1993). An experiment by Sloboda (1983) found that listeners were better able to identify the meter of an ambiguous passage when performed by a more experienced performer. This suggests that even subtle changes in articulation and timing—more easily executed by an expert performer—play an important role in communicating structural information to the listener. Through measuring the differences in the performers' expressive variations, Sloboda identified dynamics and articulation —in particular, a *tenuto* articulation— as the most important features for communicating which notes were accented.

The extent to which a listener's musical expectations align with a performer's expressive variations appears an important consideration. For example, because of the predictable relation between timing and structural segmentation, it has been demonstrated that listeners find it difficult to detect timing (and duration) deviations from a 'metronomic' performance when the pattern and placement of those deviations are stylistically typical (Repp, 1990, 1992; Ohriner, 2012). Likewise, Clarke (1993) found pianists able to more accurately reproduce a performance when the timing profile was 'normative' with regards to the musical structure, and also found listeners' aesthetic judgments to be highest for those performances with the original timing profiles compared with those that were inverted or altered.

In addition to communicating structural information to the listener, performance features such as timing and dynamics have also been studied extensively for their role in contributing to a perceived 'expressive' performance (see Clarke, 1998; Gabrielsson, 1999). For instance, a factor analysis by Schubert and Fabian (2014) examined the features and qualities that may be related to perceived expressiveness, finding that dynamics had the highest impact on the factor labeled 'emotional expressiveness.' Recent work by Battcock and Schutz (2019) showed attack rate to be the most important predictor of intensity (or "arousal" in terms of two-dimensional models of emotion). While this work was not strictly performance analysis since the authors measured elements that correspond to fixed directions from a score (e.g., mode; pitch height), the authors do analyze attack rate, which is related to timing. Specifically, the authors point out that understanding the role of timing is confounded by the fact that it encompasses several distinct musical properties such as tempo and rhythm. Although the authors do not attempt to segregate these phenomena (tempo and rhythm) in their perceptual experiments, it is clear that for a performer, adjusting the tempo (globally or locally) would influence the attack rate, and therefore have an impact on perceived intensity.

The relation between changes in various expressive parameters and their effect on perceived tension ratings has been fairly well studied but with conflicting results. Krumhansl (1996) found that in an experiment comparing an original performance to versions with flat dynamics, flat tempo, or both, listeners' continuous tension ratings were not affected, implying that tension was primarily conveyed by the melodic, harmonic, and durational elements central to the composition (rather than the performance). A similar result was reported by Farbood and Upham (2013) where repetitions of the same verse across a single performance —as well as a harmonic reduction of it— were found to produce strongly correlated tension ratings. However, Gingras et al. (2016) studied the relation between musical structure, expressive variation, and listeners' ratings of musical tension, and found that variations in expressive timing were most predictive of listeners' tension ratings.

It is equally important to empirically test assumptions about the perceptual effects of expressive variation. For instance, some aspects of so-called 'micro-timing' variation —defined as small, systematic, intentional deviations in timing— have been debated with regard to their perceptual effects. In particular, micro-timing has been suggested as one of the principle contributors to the perception of 'groove' (Iyer, 2002; Roholt, 2014). In fact, there is a sizable portion of literature dedicated to this phenomenon, and the role of micro-timing in generating embodied cognitive responses (Dibben, 2014). However, Davies et al. (2013) parametrically varied the amount of micro-timing in certain jazz, funk, and samba rhythm patterns, and, contrary to popular belief, found that systematic micro-timing generally led to decreased ratings of perceived groove, naturalness, and liking. Similarly, Frühauf et al. (2013) found that the highest ratings of perceived 'groove quality' were given to drum patterns that were perfectly quantized, and that increasing systematic micro-timing (by shifting either forwards or backwards), resulted in lower quality ratings.

While the role of expressive variation in timbre and intonation has generally been less studied, there has been substantial attention given to the expressive qualities of the singing voice, where these parameters are especially relevant (see Sundberg, 2018). For instance, Sundberg et al. (2013), found that a sharpened intonation at a phrase climax contributed to increased perception of expressiveness and excitement, and Siegwart and Scherer (1995) found that listener preferences were correlated with certain spectral components such as the relative strength of the fundamental and the value of the spectral centroid. Similarly, the role of ornamentation in contributing to perceived expression, skill, or overall quality, has been largely overlooked, especially as it relates to music outside of the classical canon. Some exceptions include research showing subjective preferences for an idealized pitch contour and timing profile of the Indian classical music ornament *Gamak* (Gupta and Rao, 2012), and, in pop music, the expressive and emotional effects of portamento (or pitch 'slides'), as well as the so-called 'noisy' sounds of the voice, have been theorized to be of strong importance in generating an emotional response (Dibben, 2014). In the latter case, no actual perceptual experiments have been conducted to investigate this claim, however, it is consistent with ethological research on the role of vocalizations and sub-vocalizations in affective communication (Huron, 2015).

The reason why expressive variation is so enjoyable for listeners remains largely an open research question. Expressive variation is assumed to be the most important cue to a listener that they are hearing a uniquely *human* performance and is regularly hailed as the key component in communicating an aesthetically pleasing performance. As mentioned above, its role appears to go beyond bolstering the communication of musical structure. And, as pointed out by Repp, even a computerized or metronomic performance will contain grouping cues (Repp, 1998b). However, one prominent theory suggests that systematic

performance deviations (such as tempo) may generate aesthetically pleasing expressive performances in part due to their exhibiting characteristics that mimic natural motion in the physical world (Gjerdingen, 1988; Todd, 1992; Repp, 1993; Todd, 1995; van Noorden and Moelants, 1999) or human movements or gestures (Ohriner, 2012; Broze III, 2013). For instance, Friberg and Sundberg (1999), suggested that the shape of final *ritardandi* matched the velocity of runners coming to a stop and Juslin (2003) includes 'motion principles' in his model of performance expression.

4.3 Mapping and Predicting Listener Judgments

In order to isolate listeners' perception of parameters that are strictly performance-related, several scholars have investigated listeners' judgments across multiple performances of the same excerpt of music (e.g., Repp, 1990; Fabian and Schubert, 2008). A less-common technique relies on synthesized constructions or manipulations of performances, typically using some kind of rule-based system to manipulate certain musical parameters (e.g., Repp, 1989; Sundberg, 1993; Clarke, 1993; Repp, 1998b), and frequently making use of continuous data collection measures (e.g., Schubert and Fabian, 2014).

From these studies, it appears that listeners (especially 'trained' listeners) are capable not only of identifying performance characteristics such as phrasing, articulation, and vibrato, but that they are frequently able to identify them in a manner that is aligned with the performer's intentions (e.g., Nakamura, 1987; Fabian and Schubert, 2009). However, while listeners may be able to identify performers' intentions, they may not have the perceptual acuity to identify certain features with the same precision allowed by acoustic measures. For instance, a study by Howes et al. (2004) showed there was no correlation between measured and perceived vibrato onset times. Similarly, Geringer (1995) found that listeners consistently identified increases in intensity (crescendos) with a greater perceived magnitude of contrast than the decreases in intensity (decrescendos) regardless of the actual magnitude of change. This suggests that there are some measurable performance parameters that may not map well to human perception. For example, an objectively measurable difference between a 'deadpan' and 'expressive' performance does not necessarily translate to *perceived expressivity*, especially if the changes in measured performance parameters are structurally normative, as discussed in Section 4.2. Two related papers, by Li et al. (2015) and Sulem et al. (2019), describe research attempting to better understand the communication chain from score interpretation to performance and performance to perception, respectively. The former attempted to match quantitative acoustic measures with expressive musical terms (commonly used in score directions as the principal means of communicating expressive instruction), while the latter asked performers to match the same expressive musical terms in terms of

their perceived emotion along a common dimensional model of emotion (i.e., Russell, 1980). This work lays the foundation for future research to empirically examine the full chain of communication; in attempting to manipulate the same acoustic measurements it may be possible to predict perceived musical and emotional correlates.

An important but rarely discussed consideration is the relation between observed differences in a model and the perceptual evaluation of those differences by a listener. For instance, Dixon et al. (2006) experimented with various methods for extracting perceived tempo information in relation to expressively-performed excerpts with an emphasis on some of the assumptions of beat-tracking algorithms. They discuss the presumption that what is desirable in a beat-tracking model is typically to accurately mark what was performed rather than what was perceived, even though the two may differ. In particular, they note that the perceived beat is smoother than the performance data would indicate. Busse (2002) evaluated expert listener judgments of the optimal 'swing style' of performed jazz piano melodies that were either unmodified, or modified according to one of four 'derived' models. The first derived model altered parameters (durations, onsets, and velocities) according to performer averages, whereas the other three derived 'mechanical' models had the same parameters fixed by simple ratio relationships. Unsurprisingly, the unaltered and derived models were generally preferred to the mechanical models. (It is well known that *some* randomness is required in order for a performance to sound convincingly human, and various jitter functions have been implemented in computer music software for this reason since the 1980s.) Despite that one might predict human preference for one of the original (performed) melodies, several of the derived models were not rated statistically different from the original melodies, suggesting that the averaged parameter values created a realistic model. However, the parameters of the unmodified originals were not reported nor compared against each other or those of the derived models, making it impossible to examine any difference thresholds across the measured parameters in terms of their impact on the swing ratings. Devaney (2016) also compared model classification against human classification using a singer-identification task to explore differences between inter-singer variability and intra-singer similarity across different performance parameters. In general, listeners performed the singer identification task better than chance but far below the abilities of the computational model. However, there were some similarities between the model parameters and the features reported by listeners as important determinants of their classification (e.g., vibrato, pitch stability, timbre, breathiness, intonation). Furthermore, the same pair of singers that 'confused' the model were the same two conflated by listeners. These experiments represent excellent examples from a scarce pool of research attempting to bridge MIR and cognitive approaches to performance research.

However, only a comparison of systematic manipulations between contrived stimuli will allow sufficient control over the individual parameters necessary to come to definitive conclusions about the perceptibility of variations in performance and their aesthetic value for performance.

Given a weak relation between a measured parameter and listeners' perception of that parameter, another important question arises: is the parameter itself not useful in modeling human perception, or is the metric simply inappropriate? For example, there are many aspects of music perception that are known to be categorical (e.g., pitch) in which case a continuous metric would not work well in a model designed to predict human ratings.

Similarly, there is the consideration of the role of the *representation* and transformation of a measured parameter for predicting perceptual ratings. This question was raised by Timmers (2005), who examined the representation of tempo and dynamics that best predicted listener judgments of musical similarity. This study found that, while most existing models rely on normalized variations of tempo and dynamics, the absolute tempo and the interaction of tempo and loudness were better predictors.

Finally, there are performance features that are either not captured in the audio signal or else not represented in a music performance analysis that may well contribute to a listener's perception. For instance, if judgments of perception are made in a live setting, then many visual cues —such as performer movement, facial expression, or attire— will be capable of altering the listener's perception (Huang and Krumhansl, 2011; Juchniewicz, 2008; Wapnick et al., 2009; Livingstone et al., 2009; Silvey, 2012). Importantly, visual information such as performer gesture and movement may contribute to embodied sensorimotor engagement, which is thought to be an essential component of music perception (e.g., Leman and Maes, 2014; Bishop and Goebl, 2018), and could therefore be influential on ratings of performance aesthetics and/or musical expression.

Clearly, the execution of multiple performance parameters is important for the perception of both small-scale and large-scale musical structures, and appears to have a large influence over listeners' perception and experience of the emotional and expressive aspects of a performance. Since the latter appears to carry great significance for both MPA and music perception research, it suggests that future work ought to focus on disentangling the relative weighting of the various features controlled by performers that contribute to an expressive performance. Since it is frequently alluded to that a performer's manipulation of musical tension is one of the strongest contributors to an expressive performance, further empirical research must attempt to systematically break down the concept of tension as a high-level feature into meaningful collections of smaller, well-defined features that would be useful for MPA.

4.4 Challenges

The research surveyed in this section highlights the importance of human perception in MPA research, especially as it pertains to the communication of emotion, musical structure, and creating an aesthetically pleasing performance. In fact, the successful modeling of perceptually relevant performance attributes, such as those that mark 'expressiveness', could have a large impact not only for MPA but for many other areas of MIR research, such as computer-generated performance, automatic accompaniment, virtual instrument design and control, or robotic instruments and HCI (see, for example, the range of topics discussed by Kirke and Miranda (2013)). A major obstacle impeding research in this area is the inability to successfully isolate (and therefore understand) the various performance characteristics that contribute to a so-called 'expressive' performance from a listener's perspective. Existing literature reviews on the topic of MPA have not been able to shed much light on this problem, in part because researchers frequently disagree on (or conflate) the various definitions of 'expressive,' or else findings appear inconsistent across the research, likely as a result of different methodologies, types of comparisons, or data. As noted by Devaney (2016), combining computational and listening experiments could lead to a better understanding of *which* aspects of variation are important to observe and model. Careful experimental design and/or meta-analyses across both MPA and cognition research, as well as cross-collaboration between MIR and music cognition researchers, may therefore prove fruitful endeavors for future research.

5. Performance Assessment

Assessment of musical performances deals with providing a rating of a music performance with regard to specific aspects of the performance such as accuracy, expressivity, and virtuosity. Performance assessment is a critical and ubiquitous aspect of music pedagogy: students rely on regular feedback from teachers to learn and improve skills, recitals are used to monitor progress, and selection into ensembles is managed through competitive auditions. The performance parameters on which these assessments are based are not only subjective but also ill-defined, leading to large differences in subjective opinion among music educators (Thompson and Williamon, 2003; Wesolowski et al., 2016). However, other studies have shown that humans tend to rate prototypical (average) performances higher than individual performances (Repp, 1997a; Wolf et al., 2018). This might indicate that performances are rated based on some form of perceived distance from an 'ideal' performance. Apart from music education, assessment of performances is also an important area of focus for the evaluation of computer-generated music performances (Bresin and Friberg, 2013) where researchers have primarily focused on listening studies to understand the effect of musical knowledge and biases on rating performances (De Poli et al., 2014) and the degree to which computer generated performances stack up against those by humans (Schubert et al., 2017).
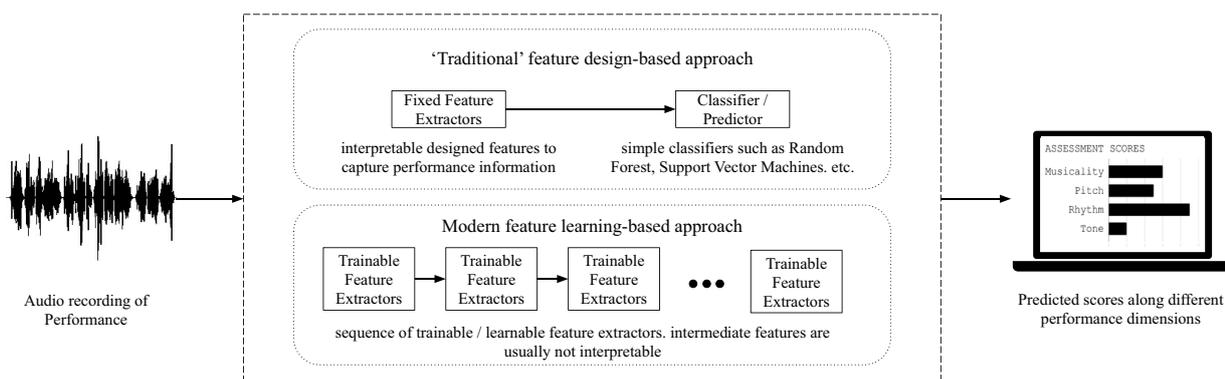
Figure 3: Schematic showing the comparison between different approaches for music performance assessment.

Work within assessment-focused MPA deals with modeling how humans assess a musical performance. The goal is to increase the objectivity of performance assessments (McPherson and Thompson, 1998) and to build accessible and reliable tools for automatic assessment. While this might be considered a subset of listener-focused MPA, its importance to MPA research and music education warrants a tailored review of research in this area.

Over the last decade, several researchers have worked towards developing tools capable of automatic music performance assessment. These can be loosely categorized based on (i) the parameters of the performance that are assessed, and (ii) the technique/method used to design these systems.

5.1 Assessment parameters

Tools for performance assessment evaluate one or more performance parameters typically related to the accuracy of the performance in terms of pitch and timing (Wu et al., 2016; Vidwans et al., 2017; Pati et al., 2018; Luo, 2015), or quality of sound (timbre) (Knight et al., 2011; Romani Picas et al., 2015; Narang and Rao, 2017). In building an assessment tool, the choice of parameters may depend on the proficiency level of the performer being assessed. For example, beginners will benefit more from feedback in terms of low-level parameters such as pitch or rhythmic accuracy as opposed to feedback on higher-level parameters such as articulation or expression. Assessment parameters can also be specific to culture or the musical style under consideration, for example, in the case of Indian classical music the nature of pitch transitions or *gamakas* plays an important role (Gupta and Rao, 2012), while correct pronunciation of syllables is a strict requirement for Chinese *Jingju* music (Gong, 2018).

Assessment tools can also vary based on the granularity of assessments. Tools may simply classify a performance as 'good' or 'bad' (Knight et al., 2011; Nakano et al., 2006), or grade it on a scale, e.g., from 1 to 10 (Pati et al., 2018). Systems may provide fine-grained note-by-note assessments (Romani Picas et al., 2015; Schramm et al., 2015) or analyze entire performances and report a single assessment score (Nakano et al., 2006; Pati et al., 2018; Huang and Lerch, 2019).

5.2 Assessment methods

While different methods have been used to create performance assessment tools, the common approach has been to use descriptive features extracted from the audio recording of a performance, based on which a classifier predicts the assessment. This approach requires availability of performance data (recordings) along with human (expert) assessments for the rated performance parameters.

The level of sophistication of classifiers was limited especially for early attempts, in which classifiers such as Support Vector Machines were used to predict human ratings. In these systems, the descriptive features became an important aspect of the system design. In some approaches, standard spectral and temporal features such as spectral centroid, spectral flux, and zero-crossing rate were used (Knight et al., 2011). In others, features aimed at capturing certain aspects of music perception were hand-designed using either musical intuition or expert knowledge (Nakano et al., 2006; Abeßer et al., 2014b; Romani Picas et al., 2015; Li et al., 2015). For instance, Nakano et al. (2006) used features measuring pitch stability and vibrato as inputs to a simple classifier to rate the quality of vocal performances. Several studies also attempted to combine low-level audio features with hand-designed feature sets (Luo, 2015; Wu et al., 2016; Vidwans et al., 2017), as well as incorporating information from the musical score or reference performance recordings into the feature computation process (Devaney et al., 2012; Mayor et al., 2009; Vidwans et al., 2017; Bozkurt et al., 2017; Molina et al., 2013; Falcao et al., 2019).

Recent methods, however, have transitioned towards using advanced machine learning techniques such as sparse coding (Han and Lee, 2014; Wu and Lerch, 2018c,a) and deep learning (Pati et al., 2018). Contrary to earlier methods which focused on hand-designing musically important features, these techniques input raw data (usually in the form of pitch contours or spectrograms) and train the models to automatically learn meaningful features so as to accurately predict the assessment ratings.

In some ways, this evolution in methodology has mirrored that of other MIR tasks: there has been a gradual transition from feature design to feature learning (compare **Figure 3**). Feature design and feature learning

have an inherent trade-off. Learned features extract relevant information from data which might not be represented in the hand-crafted feature set. This is evident from their superior performance at assessment modeling tasks (Wu and Lerch, 2018a; Pati et al., 2018). However, this superior performance comes at the cost of low interpretability. Learned features tend to be abstract and cannot be easily understood. Custom-designed features, on the other hand, typically either measure a simple low-level characteristic of the audio signal or link to high-level semantic concepts such as pitch or rhythm which are intuitively interpretable. Thus, such models allow analysis that can aid in the interpretation of semantic concepts for music performance assessment. For instance, Gururani et al. (2018) analyzed the impact of different features on an assessment prediction task and found that features measuring tempo variations were particularly critical, and that score-aligned features performed better than score-independent features.

5.3 Challenges
In spite of several attempts across varied performance parameters using different methods, the important features for assessing music performances remain unclear. This is evident from the average accuracy of these tools in modeling human judgments. Most of the presented models either work well only for very select data (Knight et al., 2011) or have comparably low prediction accuracies (Vidwans et al., 2017; Wu et al., 2016), rendering them unusable in most practical scenarios (Eremenko et al., 2020). While this may be partially attributed to the subjective nature of the task itself, there are several other factors which have limited the improvement of these tools. First, most of the models are trained on small task-specific or instrument-specific datasets that might not reflect noisy real-world data. This reduces the generalizability of these models. The problem becomes more serious for data-hungry methods such as deep learning which require large amounts of data for training. The larger datasets (>3000 performances) based on real-world data are either not publicly available (for example, the FBA dataset (Pati et al., 2018)) or only provide intermediate representations such as pitch contours (for example, the MAST melody dataset (Bozkurt et al., 2017)). Thus, more efforts are needed towards creating and releasing larger performance datasets for the research community. Second, the distribution of ground-truth (expert) ratings given by human judges, in many datasets, skewed towards a particular class or value (Gururani et al., 2018). This makes it challenging to train unbiased models. Finally, the performance parameters required to adequately model a performance are not well understood. While the typical approach is to train different models for different parameters, this approach necessitates availability of performance data along with expert assessments for all these parameters. On many occasions, such assessments are either not available or are costly to obtain. For instance, while the MAST rhythm dataset (Falcao et al., 2019) contains performance recordings (and pass/fail assessment ratings) for around 1000 students, the finely annotated (on a 4-point scale) version of the same dataset contains only 80 performances. An interesting direction for future research might consider leveraging models which are successful at assessing a few parameters (and/or instruments) to improve the performance of models for other parameters (and/or instruments). This approach, usually referred to as transfer learning, has been found to be successful in other MIR tasks (Choi et al., 2017).

In addition to the data-related challenges, there are several other challenging problems for MIR researchers interested in this domain. Better techniques need to be developed to factor the score (or reference) information into the assessments. So far, this has been accomplished by either using dynamic time warping (DTW) based methods (Vidwans et al., 2017; Bozkurt et al., 2017; Molina et al., 2013) to compute distance-based features between the reference and the performance or by computing vector similarity between features extracted from the performance and the reference (Falcao et al., 2019). However, expressive performances are supposed to deviate from the score and simple distance-based features may fail to adequately capture the nuances. The problem of how to incorporate this information into the assessment computation process remains an open problem.

Another area which requires attention from researchers lies in improving the ability to interpret and understand the features learned by end-to-end models. This will play an important role in improving assessment tools. Interpretability of neural networks is still an active area of research, and performance assessment is an excellent testbed for developing such methods.

6. Conclusion
The previous sections outlined insights gained by MPA at the intersection of audio content analysis, empirical musicology, and music perception research. These insights are of importance for better understanding the process of making music as well as affective user reactions to music.

6.1 Applications
The better understanding of music performance enables a considerable range of applications spanning a multitude of different areas including systematic musicology, music education, MIR, and computational creativity, leading to a new generation of music discovery and recommendation systems, and generative music systems.

The most obvious application example connecting MPA and MIR is music tutoring software. Such software aims at supplementing teachers by providing students with insights and interactive feedback by analyzing and assessing the audio of practice sessions. The ultimate goals of an interactive music tutor are to highlight problematic parts of the student's performance, provide a concise yet easily understandable analysis, give specific and understandable feedback on how to improve, and

individualize the curriculum depending on the student's mistakes and general progress. Various (commercial) solutions are already available, exhibiting a similar set of goals. These systems adopt different approaches, ranging from traditional music classroom settings to games targeting a playful learning experience. Examples for tutoring applications are SmartMusic,[1] Yousician,[2] Music Prodigy,[3] and SingStar.[4] However, many of these tools are not reliable enough to be used in educational settings. More studies are needed to properly evaluate the usability of performance assessment systems in real classroom environments (Eremenko et al., 2020).

Performance parameters have a long history being either explicitly or implicitly part of MIR systems. For instance, core MIR tasks such as music genre classification and music recommendation systems have a long history of utilizing tempo and dynamics features successfully (Fu et al., 2011).

Another area which has relied extensively on using performance data is the field of generative modeling. Much of the recent research has been on generating expressive performances with or without a musical score as input. While the vast majority of this body of work has focused on piano performances (Cancino-Chacón and Grachten, 2016; Malik and Ek, 2017; Jeong et al., 2019; Jeong et al., 2019b,a; Oore et al,. 2020; Maezawa et al., 2019), there are a few studies focused on other instruments such as violin and flute (Wang and Yang, 2019). The common thread across these approaches is that they use end-to-end data-driven techniques to generate the performance (either predict note-wise performance features such as timing, tempo and dynamics, or directly generate the audio) given the score as input. While these methods have achieved some success, they mostly operate as black boxes, and hence, lack in their ability to either provide deeper insights regarding the performance generation process or exert any form of explicit control over different performance parameters. There have been some attempts to alleviate these limitations. For instance, Maezawa et al. (2019) tried to learn an abstract representation capturing the musical interpretation of the performer. This could allow generation of different performances of the same piece with varying interpretations. More studies like this would allow better modeling of musical performances and improving the quality and usability of performance generation systems.

6.2 Challenges

Despite such practical applications, there are still many open topics and challenges that need to be addressed. The main challenges of MPA have been summarized at the end of each of the previous sections. The related challenges to the MIR community, however, are multi-faceted as well. First, the fact that the majority of the presented studies use manual annotations instead of automated methods should encourage the MIR community to re-evaluate the measures of success of their proposed systems if, as it appears to be, the outputs lack the robustness or accuracy required for a detailed analysis even for tasks

considered to be 'solved.' Second, the missing separation of composition and performance parameters when framing research questions or problem definitions can impact not only interpretability and reusability of insights but might also reduce algorithm performance. If, for example, a music emotion recognition system does not differentiate between the impact of core musical ideas and performance characteristics, it will have a harder time differentiating relevant and irrelevant information. Thus, it is essential for MIR systems to not only differentiate between score and performance parameters in the system design phase but also analyze their respective contributions during evaluation. Third, when examining phenomena that are complex and at times ambiguous — such as 'expressiveness'— it is imperative to fully define the scope of the associated terminology. Inconsistently used or poorly defined terms can obfuscate results making it more challenging to build on prior work or to propagate knowledge across disciplines. Fourth, a greater flow of communication between MIR and music perception communities would bolster research in both areas. However, differing methodologies, tools, terminology, and approaches have often created a barrier to such an exchange (Aucouturier and Bigand, 2012). One way of facilitating this communication between disciplines is to maximize the interpretability and reusability of results. In particular, acknowledging or addressing the perceptual relevance of predictor variables or results, or even explicitly pointing to a possible gap in the perceptual literature, can aid knowledge transfer by pointing to 'meaningful' or perceptually-relevant features to focus subsequent empirical work. In addition, it would be prudent to ensure that any underlying assumptions of perceptual validity (linked to methods or results) are made overt and, where possible, supported with empirical results. Fifth, lack of data continues to be a challenge for both MIR core tasks and MPA; a focus on approaches for limited data (McFee et al., 2015), weakly labeled data, and unlabeled data (Wu and Lerch 2018b) could help address this problem. There is, however, a slow but steady growth in the number of datasets available for performance analysis, indicating growing awareness and interest in this topic. **Table 1** lists the most relevant currently available datasets for music performance research. Note that 22 of the 30 datasets listed have been released in the last 5 years.

In conclusion, the fields of MIR and MPA each depend on the advances in the other field. In addition, music perception and cognition, while not a traditional topic within MIR, can be seen as an important linchpin for the advancement of MIR systems that depend on reliable and diverse perceptual data. Cross-disciplinary approaches to MPA bridging methodologies and data from music cognition and MIR are likely to be most influential for future research. Empirical, descriptive research driven by advanced audio analysis is necessary to extend our understanding of music and its perception, which in turn will allow us to create better systems for music analysis, music understanding, and music creation.

Table 1: A list of datasets useful for various tasks within MPA.

name	reference	instruments	genre	data	size	performance parameters
APL	Winters et al. (2016)	piano	classical	audio	621 recordings	piano practice
CBFdataset	Wang et al. (2019)	bamboo flute	chinese	audio	1GB	playing techniques
CrestMusePEDB	Hashida et al. (2008)	piano	classical	xml	121 performances	timing, dynamics
CSD	Cuesta et al. (2018)	vocals	classical	audio, f0 series	48 recordings	intonation
DAMP	–	vocals	popular	audio	24874 recordings (14 songs)	singing
DrumPT	Wu and Lerch (2016)	drums	popular	audio	30 recordings	playing techniques
Duet	Xia and Dannenberg (2015)	piano	classical	MIDI	105 performances	timing, dynamics
EEP	Marchini et al. (2014)	string quartet	classical	audio	23 recordings	timing, gestures, bowing techniques
Erkomaishvili	Rosenzweiget al. (2020)	vocals	Georgian	audio, f0 series, MusicXML	116 recordings	timing, pitch
Groove MIDI	Gillick et al. (2019)	drums	popular	MIDI	13.6 hours	drum timing
GPT	Su et al. (2014)	guitar	popular	audio	6580 recordings	playing techniques
IDMT-SMT-Bass	Abeßer et al. (2010)	bass	popular	audio	3.6 hours	playing techniques
IDMT-SMT-Guitar	Kehling et al. (2014)	guitar	popular	audio	4700 note events	playing techniques
Intonation	Wager et al. (2019)	vocals	popular	audio, f0 series	4702 performances	singing
Jingju-Pitch	Gong et al. (2016)	vocals	Beijing Opera	f0 series	13MB	intonation
JKU-ScoFo	Henkel et al. (2019)	piano	classical	audio, MIDI	16 performances	timing, dynamics
Kara1k	Bayle et al. (2017)	vocals	popular	audio	1000 songs	singing
Maestro	Hawthorne et al. (2019)	piano	classical	audio, MIDI	200 hours	timing, dynamics
MASTmelody	Bozkurt et al. (2017)	vocals	–	f0 series	1018 recordings	pass/fail ratings
MASTrhythm	Falcao et al. (2019)	percussion	–	audio	3721 recordings	pass/fail ratings
Mazurka	Sapp (2007)	piano	classical	beat markers	2732 recordings	tempo, dynamics
PGD	Sarasúa et al. (2017)	piano	classical	audio, video, MIDI	210 recordings	gestures, intentions
QUARTET	Papiotis (2016)	string quartet	classical	audio, video	96 recordings	timing, gestures, bowing techniques
SMD	Müller et al. (2011)	piano	classical	audio, MIDI	50 performances	timing, dynamics
SUPRA	Shi et al. (2019)	piano	classical	piano rolls, MIDI	478 performances	gestures, timing, dynamics
URMP	Li et al. (2019)	multi	classical	audio, video	44 pieces	timing, dynamics
VGD	Sarasúa et al. (2017)	violin	classical	audio, EMG, IMU	960 recordings	position data, playing techniques
Vienna 4x22	Goebl (1999)	piano	classical	audio, MIDI	4 pieces, 22 pianists	timing, dynamics
VocalSet	Wilkins et al. (2018)	vocals	popular	audio	6GB	singing techniques
WJazzD	Pfleiderer et al. (2017)	wind instruments	jazz	MIDI	456 solos	timing, pitch

Notes

[1] MakeMusic, Inc., www.smartmusic.com, last accessed 04/11/2019.
[2] Yousician Oy, www.yousician.com, last accessed 04/11/2019.
[3] The Way of H, Inc., www.musicprodigy.com, last accessed 04/11/2019.
[4] Sony Interactive Entertainment Europe, www.singstar.com, last accessed 04/11/2019.

Competing Interests

The authors have no competing interests to declare.

References

Abeßer, J., Cano, E., Frieler, K., & **Pfleiderer, M.** (2014a). Dynamics in jazz improvisation: Scoreinformed estimation and contextual analysis of tone intensities in trumpet and saxophone solos. In *9th Conference on Interdisciplinary Musicology (CIM14)*, Berlin, Germany.

Abeßer, J., Cano, E., Frieler, K., & **Zaddach, W.-G.** (2015). Score-informed analysis of intonation and pitch modulation in jazz solos. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, Spain.

Abeßer, J., Frieler, K., Cano, E., Pfleiderer, M., & **Zaddach, W.-G.** (2017). Score-informed analysis of tuning, intonation, pitch modulation, and dynamics in jazz solos. *IEEE Transactions on Audio, Speech and Language Processing*, *25*(1), 168–177. DOI: https://doi.org/10.1109/TASLP.2016.2627186

Abeßer, J., Hasselhorn, J., Grollmisch, S., Dittmar, C., & **Lehmann, A.** (2014b). Automatic competency assessment of rhythm performances of ninth-grade and tenth-grade pupils. In *Proceedings of the International Computer Music Conference (ICMC)*, Athens, Greece.

Abeßer, J., Lukashevich, H., & **Schuller, G.** (2010). Feature-based extraction of plucking and expression styles of the electric bass guitar. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2290–2293, Dallas. DOI: https://doi.org/10.1109/ICASSP.2010.5495945

Abeßer, J., Pfleiderer, M., Frieler, K., & **Zaddach, W.-G.** (2014c). Score-informed tracking and contextual analysis of fundamental frequency contours in trumpet and saxophone jazz solos. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Erlangen, Germany.

Arzt, A., & **Widmer, G.** (2015). Real-time music tracking using multiple performances as a reference. In Müller, M., & Wiering, F., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 357–363, Malaga, Spain.

Ashley, R. (2002). Do[n't] change a hair for me: The art of jazz rubato. *Music Perception: An Interdisciplinary Journal*, *19*(3), 311–332. DOI: https://doi.org/10.1525/mp.2002.19.3.311

Atli, H. S., Bozkurt, B., & **Sentürk, S.** (2015). A method for tonic frequency identification of Turkish makam music recordings. In *Proceedings of the 5th International Workshop on Folk Music Analysis (FMA)*, Paris, France. Association Dirac. DOI: https://doi.org/10.1109/SIU.2015.7130148

Aucouturier, J.-J., & **Bigand, E.** (2012). Mel Cepstrum & Ann Ova: The difficult dialog between MIR and music cognition. In *ISMIR*, pages 397–402.

Bantula, H., Giraldo, S. I., & **Ramirez, R.** (2016). Jazz ensemble expressive performance modeling. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, New York.

Battcock, A., & **Schutz, M.** (2019). Acoustically expressing affect. *Music Perception*, *37*(1), 66–91. DOI: https://doi.org/10.1525/mp.2019.37.1.66

Bayle, Y., Maršík, L., Rusek, M., Robine, M., Hanna, P., Slaninová, K., Martinovic, J., & **Pokorný, J.** (2017). Kara1k: A karaoke dataset for cover song identification and singing voice analysis. In *International Symposium on Multimedia (ISM)*, pages 177–184, Taichung, Taiwan. IEEE. DOI: https://doi.org/10.1109/ISM.2017.32

Behne, K.-E., & **Wetekam, B.** (1993). Musikpsychologische Interpretationsforschung: Individualität und Intention. *Musikpsychologie. Jahrbuch der Deutschen Gesellschaft für Musikpsychologie*, *10*, 24–37.

Bektaş, T. (2005). Relationships between prosodic and musical meters in the beste form of classical Turkish music. *Asian Music*, *36*(1), 1–26. DOI: https://doi.org/10.1353/amu.2005.0003

Benetos, E., Dixon, S., Giannoulis, D., Kirchhoff, H., & **Klapuri, A.** (2013). Automatic music transcription: Challenges and future directions. *Journal of Intelligent Information Systems*, *41*(3), 407–434. DOI: https://doi.org/10.1007/s10844-013-0258-3

Bergeron, V., & **Lopes, D. M.** (2009). Hearing and seeing musical expression. *Philosophy and Phenomenological Research*, *78*(1), 1–16. DOI: https://doi.org/10.1111/j.1933-1592.2008.00230.x

Bishop, L., & **Goebl, W.** (2018). Performers and an active audience: Movement in music production and perception. *Jahrbuch Musikpsychologie*, *28*, 1–17. DOI: https://doi.org/10.5964/jbdgm.2018v28.19

Bowman Macleod, R. (2006). *Influences of Dynamic Level and Pitch Height on the Vibrato Rates and Widths of Violin and Viola Players*. Dissertation, Florida State University, College of Music, Tallahassee, FL.

Bozkurt, B., Ayangil, R., & **Holzapfel, A.** (2014). Computational analysis of Turkish makam music: Review of state-of-the-art and challenges. *Journal of New Music Research*, *43*(1), 3–23. DOI: https://doi.org/10.1080/09298215.2013.865760

Bozkurt, B., Baysal, O., & **Yüret, D.** (2017). A dataset and baseline system for singing voice assessment. In *Proceedings of the International Symposium on Computer Music Modeling and Retrieval (CMMR)*, Matosinhos.

Bresin, R., & **Friberg, A.** (2013). Evaluation of computer systems for expressive music performance. In Kirke, A., & Miranda, E. R., editors, *Guide to Computing for Expressive Music Performance*, pages 181–203. Springer, London. DOI: https://doi.org/10.1007/978-1-4471-4123-5_7

Broze III, G. J. (2013). *Animacy, Anthropomimesis, and Musical Line*. PhD Thesis, The Ohio State University.

Busse, W. G. (2002). Toward objective measurement and evaluation of jazz piano performance via MIDIbased groove quantize templates. *Music Perception: An Interdisciplinary Journal, 19*(3), 443–461. DOI: https://doi.org/10.1525/mp.2002.19.3.443

Cancino-Chacón, C. E., Gadermaier, T., Widmer, G., & Grachten, M. (2017). An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonic music. *Machine Learning, 106*(6), 887–909. DOI: https://doi.org/10.1007/s10994-017-5631-y

Cancino-Chacón, C. E., & Grachten, M. (2016). The Basis Mixer: A computational Romantic pianist. In *Late Breaking Demo (Extended Abstract), International Society for Music Information Retrieval Conference (ISMIR)*, New York.

Cancino-Chacón, C. E., Grachten, M., Goebl, W., & Widmer, G. (2018). Computational models of expressive music performance: A comprehensive and critical review. *Frontiers in Digital Humanities, 5*. DOI: https://doi.org/10.3389/fdigh.2018.00025

Caro Repetto, R., Gong, R., Kroher, N., & Serra, X. (2015). Comparision of the singing style of two Jingju schools. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, Spain.

Chen, K. (2013). Characterization of pitch intonation of Beijing opera. Master's thesis, Universitat Pompeu Fabra, Barcelona.

Cheng, E., & Chew, E. (2008). Quantitative analysis of phrasing strategies in expressive performance: Computational methods and analysis of performances of unaccompanied Bach for solo violin. *Journal of New Music Research, 37*(4), 325–338. DOI: https://doi.org/10.1080/09298210802711660

Chew, E. (2016). Playing with the edge: Tipping points and the role of tonality. *Music Perception, 33*(3), 344–366. DOI: https://doi.org/10.1525/mp.2016.33.3.344

Choi, K., Fazekas, G., Sandler, M. B., & Cho, K. (2017). Transfer learning for music classification and regression tasks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 141–149, Suzhou, China.

Chuan, C.-H., & Chew, E. (2007). A dynamic programming approach to the extraction of phrase boundaries from tempo variations in expressive performances. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria.

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception: An Interdisciplinary Journal, 10*(3), 317–341. DOI: https://doi.org/10.2307/40285573

Clarke, E. F. (1998). Rhythm and timing in music. In *The Psychology of Music*, pages 473–500. Academic Press, San Diego, 2nd edition. DOI: https://doi.org/10.1016/B978-012213564-4/50014-7

Clarke, E. F. (2002a). Listening to performance. In Rink, J., editor, *Musical Performance – A Guide to Understanding*.

Cambridge University Press, Cambridge. DOI: https://doi.org/10.1017/CBO9780511811739.014

Clarke, E. F. (2002b). Understanding the psychology of performance. In Rink, J., editor, *Musical Performance – A Guide to Understanding*. Cambridge University Press, Cambridge. DOI: https://doi.org/10.1017/CBO9780511811739.005

Clayton, M. (2008). *Time in Indian Music: Rhythm, Metre, and Form in North Indian Rag Performance*. Oxford University Press, New York.

Collier, G. L., & Collier, J. L. (1994). An exploration of the use of tempo in jazz. *Music Perception: An Interdisciplinary Journal, 11*(3), 219–242. DOI: https://doi.org/10.2307/40285621

Collier, G. L., & Collier, J. L. (2002). A study of timing in two Louis Armstrong solos. *Music Perception: An Interdisciplinary Journal, 19*(3), 463–483. DOI: https://doi.org/10.1525/mp.2002.19.3.463

Cuesta, H., Gómez Gutiérrez, E., Martorell Domínguez, A., & Loáiciga, F. (2018). Analysis of intonation in unison choir singing. In *Proceedings of the International Conference on Music Perception and Cognition (ICMPC)*, Graz, Austria.

Dalla Bella, S., & Palmer, C. (2004). Tempo and dynamics in piano performance: The role of movement amplitude. In *Proceedings of the 8th International Conference on Music Perception & Cognition (ICMPC)*, Evanston.

Davies, M., Madison, G., Silva, P., & Gouyon, F. (2013). The effect of microtiming deviations on the perception of groove in short rhythms. *Music Perception: An Interdisciplinary Journal, 30*(5), 497–510. DOI: https://doi.org/10.1525/mp.2013.30.5.497

De Poli, G., Canazza, S., Rodà, A., & Schubert, E. (2014). The role of individual difference in judging expressiveness of computer-assisted music performances by experts. *ACM Transactions on Applied Perception, 11*(4), 22:1–22:20. DOI: https://doi.org/10.1145/2668124

Dean, R. T., Bailes, F., & Drummond, J. (2014). Generative structures in improvisation: Computational segmentation of keyboard performances. *Journal of New Music Research, 43*(2), 224–236. DOI: https://doi.org/10.1080/09298215.2013.859710

Devaney, J. (2016). Inter- versus intra-singer similarity and variation in vocal performances. *Journal of New Music Research, 45*(3), 252–264. DOI: https://doi.org/10.1080/09298215.2016.1205631

Devaney, J., Mandel, M. I., Ellis, D. P. W., & Fujinaga, I. (2011). Automatically extracting performance data from recordings of trained singers. *Psychomusicology: Music, Mind and Brain, 21*(1–2), 108–136. DOI: https://doi.org/10.1037/h0094008

Devaney, J., Mandel, M. I., & Fujinaga, I. (2012). A study of intonation in three-part singing using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT). In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal.

Dibben, N. (2014). Understanding performance expression in popular music recordings. In Fabian, D., Timmers, R., & Schubert, E., editors, *Expressiveness in*

Music Performance: Empirical Approaches Across Styles and Cultures. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199659647.003.0007

Dillon, R. (2001). Extracting audio cues in real time to understand musical expressiveness. In *Proceedings of the MOSART workshop*, Barcelona.

Dillon, R. (2003). A statistical approach to expressive intention recognition in violin performances. In *Proceedings of the Stockholm Music Acoustics Conference (SMAC)*, Stockholm.

Dillon, R. (2004). *On the Recognition of Expressive Intentions in Music Playing: A Computational Approach with Experiments and Applications.* Dissertation, University of Genoa, Faculty of Engineering, Genoa.

Dimov, T. (2010). *Short Historical Overview and Comparison of the Pitch Width and Speed Rates of the Vibrato Used in Sonatas and Partitas for Solo Violin by Johann Sebastian Bach as Found in Recordings of Famous Violinists of the Twentieth and the Twenty- First Centuries.* D.M.A., West Virginia University, United States.

Dittmar, C., Pfleiderer, M., Balke, S., & **Müller, M.** (2018). A swingogram representation for tracking micro-rhythmic variation in jazz performances. *Journal of New Music Research, 47*(2), 97–113. DOI: https://doi.org/10.1080/09298215.2017.1367405

Dixon, S., Goebl, W., & **Cambouropoulos, E.** (2006). Perceptual smoothness of tempo in expressively performed music. *Music Perception, 23*(3), 195–214. DOI: https://doi.org/10.1525/mp.2006.23.3.195

Dixon, S., Goebl, W., & **Widmer, G.** (2002). The Performance Worm: Real time visualisation of expression based on Langner's tempo loudness animation. In *Proceedings of the International Computer Music Conference (ICMC)*, Göteborg.

Ellis, M. C. (1991). An analysis of "swing" subdivision and asynchronization in three jazz saxophonists. *Perceptual and Motor Skills, 73*(3), 707–713. DOI: https://doi.org/10.2466/pms.1991.73.3.707

Eremenko, V., Morsi, A., Narang, J., & **Serra, X.** (2020). Performance assessment technologies for the support of musical instrument learning. In *Proceedings of the International Conference on Computer Supported Education (CSEDU)*, Prague. DOI: https://doi.org/10.5220/0009817006290640

Fabian, D., & **Schubert, E.** (2008). Musical character and the performance and perception of dotting, articulation and tempo in 34 recordings of Variation 7 from J.S. Bach's Goldberg Variations (BWV 988). *Musicae Scientiae, 12*(2), 177–206. DOI: https://doi.org/10.1177/102986490801200201

Fabian, D., & **Schubert, E.** (2009). Baroque expressiveness and stylishness in three recordings of the D minor Sarabanda for solo violin (BWV 1004) by J. S. Bach. *Music Performance Research, 3*, 36–55.

Falcao, F., Bozkurt, B., Serra, X., Andrade, N., & **Baysal, O.** (2019). A dataset of rhythmic pattern reproductions and baseline automatic assessment system. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands.

Farbood, M. M., & **Upham, F.** (2013). Interpreting expressive performance through listener judgments of musical tension. *Frontiers in Psychology*, 4. DOI: https://doi.org/10.3389/fpsyg.2013.00998

Finney, S. A., & **Palmer, C.** (2003). Auditory feedback and memory for music performance: Some evidence for an encoding effect. *Memory & Cognition, 31*(1), 51–64. DOI: https://doi.org/10.3758/BF03196082

Franz, D. M. (1998). Markov chains as tools for jazz improvisation analysis. Master's thesis, Virginia Tech.

Friberg, A., & **Sundberg, J.** (1999). 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, 105*(3), 1469–1484. DOI: https://doi.org/10.1121/1.426687

Friberg, A., & **Sundström, A.** (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception: An Interdisciplinary Journal, 19*(3), 333–349. DOI: https://doi.org/10.1525/mp.2002.19.3.333

Frieler, K., Pfleiderer, M., Zaddach, W.-G., & **Abeßer, J.** (2016). Midlevel analysis of monophonic jazz solos: A new approach to the study of improvisation. *Musicae Scientiae, 20*(2), 143–162. DOI: https://doi.org/10.1177/1029864916636440

Frühauf, J., Kopiez, R., & **Platz, F.** (2013). Music on the timing grid: The influence of microtiming on the perceived groove quality of a simple drum pattern performance. *Musicae Scientiae, 17*(2), 246–260. DOI: https://doi.org/10.1177/1029864913486793

Fu, Z., Lu, G., Ting, K. M., & **Zhang, D.** (2011). A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia, 13*(2), 303–319. DOI: https://doi.org/10.1109/TMM.2010.2098858

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A Major (K. 331): A comparison of five performances. In Gabrielsson, A., editor, *Action and Perception in Rhythm and Music*, pages 81–103. Royal Swedish Academy of Music, No. 55, Stockholm.

Gabrielsson, A. (1999). The performance of music. In Deutsch, D., editor, *The Psychology of Music*. Academic Press, San Diego, 2nd edition. DOI: https://doi.org/10.1016/B978-012213564-4/50015-9

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of Music, 31*(3), 221–272. DOI: https://doi.org/10.1177/03057356030313002

Gadermaier, T., & **Widmer, G.** (2019). A study of annotation and alignment accuracy for performance comparison in complex orchestral music. In Flexer, A., Peeters, G., Urbano, J., & Volk, A., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 769–775, Delft, Netherlands.

Ganguli, K. K., & **Rao, P.** (2017). Towards computational modeling of the ungrammatical in a raga performance. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China.

Gasser, M., Arzt, A., Gadermaier, T., Grachten, M., & **Widmer, G.** (2015). Classical music on the web: User interfaces and data representations. In Müller, M., & Wiering, F., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 571–577, Malaga, Spain.

Geringer, J. M. (1995). Continuous loudness judgments of dynamics in recorded music excerpts. *Journal of Research in Music Education, 43*(1), 22–35. DOI: https://doi.org/10.2307/3345789

Gillick, J., Roberts, A., Engel, J., Eck, D., & **Bamman, D.** (2019). Learning to groove with inverse sequence transformations. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 2269–2279.

Gingras, B., Pearce, M. T., Goodchild, M., Dean, R. T., Wiggins, G., & **McAdams, S.** (2016). Linking melodic expectation to expressive performance timing and perceived musical tension. *Journal of Experimental Psychology: Human Perception and Performance, 42*(4), 594. DOI: https://doi.org/10.1037/xhp0000141

Gjerdingen, R. O. (1988). Shape and motion in the microstructure of song. *Music Perception: An Interdisciplinary Journal, 6*(1), 35–64. DOI: https://doi.org/10.2307/40285415

Goebl, W. (1999). *Numerisch-klassifikatorische Interpretationsanalyse mit dem 'Bösendorfer Computerflügel'.* Diploma Thesis, Universität Wien, Vienna, Austria.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *The Journal of the Acoustical Society of America, 110*(1), 563–572. DOI: https://doi.org/10.1121/1.1376133

Goebl, W., Dixon, S., De Poli, G., Friberg, A., Bresin, R., & **Widmer, G.** (2005). 'Sense' in expressive music performance: Data acquisition, computational studies, and models. In Leman, M., & Cirotteau, D., editors, *Sound to Sense, Sense to Sound: A State of- the-Art.* Logos, Berlin.

Goebl, W., & **Palmer, C.** (2008). Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research, 186*(3), 471–479. DOI: https://doi.org/10.1007/s00221-007-1252-1

Gong, R. (2018). *Automatic Assessment of Singing Voice Pronunciation: A Case Study with Jingju Music.* PhD Thesis, Universitat Pompeu Fabra, Barcelona, Spain.

Gong, R., Yang, Y., & **Serra, X.** (2016). Pitch contour segmentation for computer-aided Jingju singing training. In *Proceedings of the Sound and Music Computing Conference (SMC)*, Hamburg, Germany.

Grachten, M., Cancino-Chacón, C. E., & **Chacón, C. E. C.** (2017). Temporal dependencies in the expressive timing of classical piano performances. In *The Routledge Companion to Embodied Music Interaction*, pages 360–369. Routledge, New York. DOI: https://doi.org/10.4324/9781315621364-40

Grachten, M., & **Widmer, G.** (2012). Linear basis models for prediction and analysis of musical expression. *Journal of New Music Research, 41*(4), 311–322. DOI: https://doi.org/10.1080/09298215.2012.731071

Gupta, C., & **Rao, P.** (2012). Objective assessment of ornamentation in Indian classical singing. In Ystad, S., Aramaki, M., Kronland-Martinet, R., Jensen, K., & Mohanty, S., editors, *Speech, Sound and Music Processing: Embracing Research in India*, Lecture Notes in Computer Science, pages 1–25. Springer Berlin Heidelberg.

Gururani, S., Pati, K. A., Wu, C.-W., & **Lerch, A.** (2018). Analysis of objective descriptors for music performance assessment. In *Proceedings of the International Conference on Music Perception and Cognition (ICMPC)*, Toronto, Ontario, Canada.

Hakan, T., Serra, X., & **Arcos, J. L.** (2012). Characterization of embellishments in ney performances of makam music in Turkey. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal.

Han, Y., & **Lee, K.** (2014). Hierarchical approach to detect common mistakes of beginner flute players. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 77–82, Taipei, Taiwan.

Hartmann, A. (1932). Untersuchungen über das metrische Verhalten in musikalischen Interpretationsvarianten. *Archiv für die gesamte Psychologie, 84*, 103–192.

Hashida, M., Matsui, T., & **Katayose, H.** (2008). A new music database describing deviation information of performance expressions. In Bello, J. P., Chew, E., & Turnbull, D., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 489–494, Philadelphia, PA.

Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.-Z. A., Dieleman, S., Elsen, E., Engel, J., & **Eck, D.** (2019). Enabling factorized piano music modeling and generation with the MAESTRO Dataset. In *Proceedings of the International Conference on Learning Representations (ICLR)*. arXiv: 1810.12247.

Henkel, F., Balke, S., Dorfer, M., & **Widmer, G.** (2019). Score following as a multi-modal reinforcement learning problem. *Transactions of the International Society for Music Information Retrieval, 2*(1), 67–81. DOI: https://doi.org/10.5334/tismir.31

Hill, P. (2002). From score to sound. In Rink, J., editor, *Musical Performance − A Guide to Understanding*. Cambridge University Press, Cambridge. DOI: https://doi.org/10.1017/CBO9780511811739.010

Howes, P., Callaghan, J., Davis, P., Kenny, D., & **Thorpe, W.** (2004). The relationship between measured vibrato characteristics and perception in Western operatic singing. *Journal of Voice, 18*(2), 216–230. DOI: https://doi.org/10.1016/j.jvoice.2003.09.003

Huang, J., & **Krumhansl, C. L.** (2011). What does seeing the performer add? It depends on musical style, amount of stage behavior, and audience expertise. *Musicae Scientiae, 15*(3), 343–364. DOI: https://doi.org/10.1177/1029864911414172

Huang, J., & **Lerch, A.** (2019). Automatic assessment of sight-reading exercises. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, Netherlands.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, *19*(1), 1–64. DOI: https://doi.org/10.1525/mp.2001.19.1.1

Huron, D. (2015). Affect induction through musical sounds: An ethological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140098. DOI: https://doi.org/10.1098/rstb.2014.0098

Iyer, V. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception: An Interdisciplinary Journal*, *19*(3), 387–414. DOI: https://doi.org/10.1525/mp.2002.19.3.387

Järvinen, T. (1995). Tonal hierarchies in jazz improvisation. *Music Perception: An Interdisciplinary Journal*, *12*(4), 415–437. DOI: https://doi.org/10.2307/40285675

Järvinen, T., & **Toiviainen, P.** (2000). The effect of metre on the use of tones in jazz improvisation. *Musicae Scientiae*, *4*(1), 55–74. DOI: https://doi.org/10.1177/102986490000400103

Jeong, D., Kwon, T., Kim, Y., Lee, K., & **Nam, J.** (2019a). VirtuosoNet: A hierarchical RNN-based system for modeling expressive piano performance. In Flexer, A., Peeters, G., Urbano, J., & Volk, A., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 908–915, Delft, Netherlands.

Jeong, D., Kwon, T., Kim, Y., & **Nam, J.** (2019b). Graph neural network for music score data and modeling expressive piano performance. In *International Conference on Machine Learning*, pages 3060–3070.

Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychology of Music*, *36*(4), 417–427. DOI: https://doi.org/10.1177/0305735607086046

Jure, L., Lopez, E., Rocamora, M., Cancela, P., Sponton, H., & **Irigaray, I.** (2012). Pitch content visualization tools for music performance analysis. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Porto.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology*, *26*(6), 1797–1813. DOI: https://doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N. (2003). Five myths about expressivity in music performance and what to do about them. In *Proceedings of the International Conference on Arts and Humanities*, Honolulu, Hawaii.

Juslin, P. N., & **Laukka, P.** (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. DOI: https://doi.org/10.1037/0033-2909.129.5.770

Katz, M. (2004). *Capturing Sound: How Technology has Changed Music*. University of California Press, Berkeley and Los Angeles.

Kawase, S. (2014). Importance of communication cues in music performance according to performers and audience. *International Journal of Psychological Studies*, *6*(2), 49. DOI: https://doi.org/10.5539/ijps.v6n2p49

Kehling, C., Abeßer, J., Dittmar, C., & **Schuller, G.** (2014). Automatic tablature transcription of electric guitar recordings by estimation of score- and instrument-related parameters. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Erlangen, Germany.

Kendall, R. A., & **Carterette, E. C.** (1990). The communication of musical expression. *Music Perception*, *8*(2), 129–164. DOI: https://doi.org/10.2307/40285493

Kirke, A., & **Miranda, E. R.,** editors (2013). *Guide to Computing for Expressive Music Performance*. Springer Science & Business Media, London. DOI: https://doi.org/10.1007/978-1-4471-4123-5

Knight, T., Upham, F., & **Fujinaga, I.** (2011). The potential for automatic assessment of trumpet tone quality. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 573–578, Miami, FL.

Kosta, K., Bandtlow, O. F., & **Chew, E.** (2015). A change-point approach towards representing musical dynamics. In Collins, T., Meredith, D., & Volk, A., editors, *Mathematics and Computation in Music*, Lecture Notes in Computer Science, pages 179–184, Cham. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-20603-5_18

Kosta, K., Bandtlow, O. F., & **Chew, E.** (2018). Dynamics and relativity: Practical implications of dynamic markings in the score. *Journal of New Music Research*, *47*(5), 438–461. DOI: https://doi.org/10.1080/09298215.2018.1486430

Kosta, K., Ramirez, R., Bandtlow, O., & **Chew, E.** (2016). Mapping between dynamic markings and performed loudness: A machine learning approach. *Journal of Mathematics and Music*, *10*(2), 149–172. DOI: https://doi.org/10.1080/17459737.2016.1193237

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception: An Interdisciplinary Journal*, *13*(3), 401–432. DOI: https://doi.org/10.2307/40286177

Langner, J., & **Goebl, W.** (2002). Representing expressive performance in tempo-loudness space. In *Proceedings of the Conference European Society for the Cognitive Sciences of Music (ESCOM)*, Liege.

Leman, M., & **Maes, P.-J.** (2014). The role of embodiment in the perception of music. *Empirical Musicology Review*, *9*(3–4), 236–246. DOI: https://doi.org/10.18061/emr.v9i3-4.4498

Lerch, A. (2009). *Software-Based Extraction of Objective Parameters from Music Performances*. GRIN Verlag, München.

Lerch, A. (2012). *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Wiley-IEEE Press, Hoboken. DOI: https://doi.org/10.1002/9781118393550

Lerch, A., Arthur, C., Pati, A., & **Gururani, S.** (2019). Music performance analysis: A survey. In *Proceedings*

of the International Society for Music Information Retrieval Conference (ISMIR), Delft.

Li, B., Liu, X., Dinesh, K., Duan, Z., & **Sharma, G.** (2019). Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. *IEEE Transactions on Multimedia*, *21*(2), 522–535. DOI: https://doi.org/10.1109/TMM.2018.2856090

Li, P.-C., Su, L., Yang, Y.-H., & **Su, A. W. Y.** (2015). Analysis of expressive musical terms in violin using score-informed and expression-based audio features. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 809–815, Malaga, Spain.

Li, S., Dixon, S., & **Plumbley, M. D.** (2017). Clustering expressive timing with regressed polynomial coefficients demonstrated by a model selection test. In Cunningham, S. J., Duan, Z., Hu, X., & Turnbull, D., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 457–463, Suzhou, China.

Liem, C. C., & **Hanjalic, A.** (2011). Expressive timing from cross-performance and audio-based alignment patterns: An extended case study. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, FL.

Liem, C. C., & **Hanjalic, A.** (2015). Comparative analysis of orchestral performance recordings: An imagebased approach. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, Spain.

Livingstone, S. R., Thompson, W. F., & **Russo, F. A.** (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography. *Music Perception*, *26*(5), 475–488. DOI: https://doi.org/10.1525/mp.2009.26.5.475

Luizard, P., Brauer, E., & **Weinzierl, S.** (2019). Singing in physical and virtual environments: How performers adapt to room acoustical conditions. In *Proceedings of the AES Conference on Immersive and Interactive Audio*, York. AES.

Luizard, P., Steffens, J., & **Weinzierl, S.** (2020). Singing in different rooms: Common or individual adaptation patterns to the acoustic conditions? *The Journal of the Acoustical Society of America*, *147*(2), EL132–EL137. DOI: https://doi.org/10.1121/10.0000715

Luo, Y.-J. (2015). Detection of common mistakes in novice violin playing. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 316–322, Malaga, Spain.

Maempel, H.-J. (2011). Musikaufnahmen als Datenquellen der Interpretationsanalyse. In von Lösch, H., & Weinzierl, S., editors, *Gemessene Interpretation – Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen*, Klang und Begriff, pages 157–171. Schott, Mainz.

Maezawa, A., Yamamoto, K., & **Fujishima, T.** (2019). Rendering music performance with interpretation variations using conditional variational RNN. In Flexer, A., Peeters, G., Urbano, J., & Volk, A., editors,

Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), pages 855–861, Delft, Netherlands.

Malik, I., & **Ek, C. H.** (2017). Neural translation of musical style. *arXiv:1708.03535 [cs]*.

Marchini, M., Ramirez, R., Papiotis, P., & **Maestre, E.** (2014). The sense of ensemble: A machine learning approach to expressive performance modeling in string quartets. *Journal of New Music Research*, *43*(3), 303–317. DOI: https://doi.org/10.1080/09298215.2014.922999

Mayor, O., Bonada, J., & **Loscos, A.** (2009). Performance analysis and scoring of the singing voice. In *Proceedings of the Audio Engineering Society Convention*, pages 1–7.

McFee, B., Humphrey, E. J., & **Bello, J. P.** (2015). A software framework for musical data augmentation. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, Spain.

McNeil, A. (2017). Seed ideas and creativity in Hindustani Raga music: Beyond the compositionimprovisation dialectic. *Ethnomusicology Forum*, *26*(1), 116–132. DOI: https://doi.org/10.1080/17411912.2017.1304230

McPherson, G. E., & **Thompson, W. F.** (1998). Assessing music performance: Issues and influences. *Research Studies in Music Education*, *10*(1), 12–24. DOI: https://doi.org/10.1177/1321103X9801000102

Molina, E., Barbancho, I., Gómez, E., Barbancho, A. M., & **Tardón, L. J.** (2013). Fundamental frequency alignment vs. note-based melodic similarity for singing voice assessment. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 744–748, Vancouver, Canada. DOI: https://doi.org/10.1109/ICASSP.2013.6637747

Müller, M., Konz, V., Bogler, W., & **Arifi-Müller, V.** (2011). Saarland Music Data (SMD). In *Late Breaking Demo (Extended Abstract), International Society for Music Information Retrieval Conference (ISMIR)*, Miami, FL.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics*, *41*(6), 525–533. DOI: https://doi.org/10.3758/BF03210487

Nakano, T., Goto, M., & **Hiraga, Y.** (2006). An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features. In *Proceedings of the International Conference on Spoken Langaunge Processing (INTERSPEECH)*, volume 12, Pittsburgh, PA.

Narang, K., & **Rao, P.** (2017). Acoustic features for determining goodness of tabla strokes. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 257–263, Suzhou, China.

Ohriner, M. S. (2012). Grouping hierarchy and trajectories of pacing in performances of Chopin's Mazurkas. *Music Theory Online*, *18*(1). DOI: https://doi.org/10.30535/mto.18.1.6

Okumura, K., Sako, S., & **Kitamura, T.** (2011). Stochastic modeling of a musical performance with expressive

representations from the musical score. In Klapuri, A., & Leider, C., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 531–536, Miami, FL. University of Miami.

Oore, S., Simon, I., Dieleman, S., Eck, D., & **Simonyan, K.** (2020). This time with feeling: Learning expressive musical performance. *Neural Computing and Applications*, *32*(4), 955–967. DOI: https://doi.org/10.1007/s00521-018-3758-9

Ornoy, E., & **Cohen, S.** (2018). Analysis of contemporary violin recordings of 19th century repertoire: Identifying trends and impacts. *Frontiers in Psychology*, 9. DOI: https://doi.org/10.3389/fpsyg.2018.02233

Page, K. R., Nurmikko-Fuller, T., Rindfleisch, C., Weigl, D. M., Lewis, R., Dreyfus, L., & **De Roure, D.** (2015). A toolkit for live annotation of opera performance: Experiences capturing Wagner's Ring Cycle. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, Spain.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 331–346. DOI: https://doi.org/10.1037/0096-1523.15.2.331

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception: An Interdisciplinary Journal*, *14*(1), 23–56. DOI: https://doi.org/10.2307/40285708

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48*, 115–138. DOI: https://doi.org/10.1146/annurev.psych.48.1.115

Papiotis, P. (2016). *A Computational Approach to Studying Interdependence in String Quartet Performance*. PhD thesis, Univesitat Pompeu Fabra, Barcelona, Spain.

Pati, K. A., Gururani, S., & **Lerch, A.** (2018). Assessment of student music performances using deep neural networks. *Applied Sciences*, *8*(4), 507. DOI: https://doi.org/10.3390/app8040507

Peperkamp, J., Hildebrandt, K., & **Liem, C. C.** (2017). A formalization of relative local tempo variations in collections of performances. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou.

Pfleiderer, M., Frieler, K., Abeßer, J., Zaddach, W.-G., & **Burkhart, B.,** editors (2017). *Inside the Jazzomat: New Perspectives for Jazz Research*. Schott Campus.

Pfordresher, P. Q. (2005). Auditory feedback in music performance: The role of melodic structure and musical skill. *Journal of Experimental Psychology*, *31*(6), 1331–1345. DOI: https://doi.org/10.1037/0096-1523.31.6.1331

Pfordresher, P. Q., & **Palmer, C.** (2002). Effects of delayed auditory feedback on timing of music performance. *Psychological Research*, *16*, 71–79. DOI: https://doi.org/10.1007/s004260100075

Platz, F., & **Kopiez, R.** (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception*, *30*(1), 71–83. DOI: https://doi.org/10.1525/mp.2012.30.1.71

Povel, D.-J. (1977). Temporal structure of performed music: Some preliminary observations. In *Acta Psychologica*, volume 41, pages 309–320. DOI: https://doi.org/10.1016/0001-6918(77)90024-5

Prögler, J. A. (1995). Searching for swing: Participatory discrepancies in the jazz rhythm section. *Ethnomusicology*, *39*(1), 21–54. DOI: https://doi.org/10.2307/852199

Ramirez, R., Hazan, A., Maestre, E., & **Serra, X.** (2008). A genetic rule-based model of expressive performance for jazz saxophone. *Computer Music Journal*, *32*(1), 38–50. DOI: https://doi.org/10.1162/comj.2008.32.1.38

Repp, B. H. (1989). Expressive microstructure in music: A preliminary perceptual assessment of four composers' "pulses". *Music Perception: An Interdisciplinary Journal*, *6*(3), 243–273. DOI: https://doi.org/10.2307/40285589

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America (JASA)*, *88*(2), 622–641. DOI: https://doi.org/10.1121/1.399766

Repp, B. H. (1992). A constraint on the expressive timing of a melodic gesture: Evidence from performance and aesthetic judgment. *Music Perception: An Interdisciplinary Journal*, *10*(2), 221–241. DOI: https://doi.org/10.2307/40285608

Repp, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) Gestaltung und Bewegung in der Musik. *Psychology of Music*, *21*(1), 48–72. DOI: https://doi.org/10.1177/030573569302100104

Repp, B. H. (1996a). The art of inaccuracy: Why pianists' errors are difficult to hear. *Music Perception*, *14*(2), 161–184. DOI: https://doi.org/10.2307/40285716

Repp, B. H. (1996b). The dynamics of expressive piano performance: Schumann's 'Träumerei' revisited. *Journal of the Acoustical Society of America (JASA)*, *100*(1), 641–650. DOI: https://doi.org/10.1121/1.415889

Repp, B. H. (1996c). Pedal timing and tempo in expressive piano performance: A preliminary investigation. *Psychology of Music*, *24*(2), 199–221. DOI: https://doi.org/10.1177/0305735696242011

Repp, B. H. (1997a). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception*, *14*(4), 419–444. DOI: https://doi.org/10.2307/40285732

Repp, B. H. (1997b). The effect of tempo on pedal timing in piano performance. *Psychological Research*, *60*(3), 164–172. DOI: https://doi.org/10.1007/BF00419764

Repp, B. H. (1998a). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America (JASA)*, *104*(2), 1085–1100. DOI: https://doi.org/10.1121/1.423325

Repp, B. H. (1998b). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research*, *61*(1), 33–43. DOI: https://doi.org/10.1007/s004260050011

Repp, B. H. (1999). Effects of auditory feedback deprivation on expressive piano performance.

Music Perception, 16(4), 409–438. DOI: https://doi.org/10.2307/40285802

Rink, J. (2003). In respect of performance: The view from musicology. *Psychology of Music, 31*(3), 303–323. DOI: https://doi.org/10.1177/03057356030313004

Roholt, T. C. (2014). *Groove: A Phenomenology of Rhythmic Nuance*. Bloomsbury Publishing USA. Google-Books-ID: qvJzBAAAQBAJ.

Romani Picas, O., Rodriguez, H. P., Dabiri, D., Tokuda, H., Hariya, W., Oishi, K., & **Serra, X.** (2015). A real-time system for measuring sound goodness in instrumental sounds. In *Proceedings of the Audio Engineering Society Convention*, volume 138, Warsaw.

Rosenzweig, S., Scherbaum, F., Shugliashvili, D., Arifi- Müller, V., & **Müller, M.** (2020). Erkomaishvili Dataset: A curated corpus of traditional Georgian vocal music for computational musicology. *Transactions of the International Society for Music Information Retrieval, 3*(1), 31–41. DOI: https://doi.org/10.5334/tismir.44

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. DOI: https://doi.org/10.1037/h0077714

Sapp, C. S. (2007). Comparative analysis of multiple musical performances. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria.

Sapp, C. S. (2008). Hybrid numeric/rank similarity metrics for musical performance analysis. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, PA.

Sarasúa, A., Caramiaux, B., Tanaka, A., & **Ortiz, M.** (2017). Datasets for the analysis of expressive musical gestures. In *Proceedings of the International Conference on Movement Computing (MOCO)*, pages 1–4, London, UK. Association for Computing Machinery. DOI: https://doi.org/10.1145/3077981.3078032

Schärer Kalkandjiev, Z., & **Weinzierl, S.** (2013). The influence of room acoustics on solo music performance: An empirical case study. *Acta Acustica united with Acustica, 99*(3), 433–441. DOI: https://doi.org/10.3813/AAA.918624

Schärer Kalkandjiev, Z., & **Weinzierl, S.** (2015). The influence of room acoustics on solo music performance: An experimental study. *Psychomusicology: Music, Mind, and Brain, 25*(3), 195–207. DOI: https://doi.org/10.1037/pmu0000065

Schramm, R., de Souza Nunes, H., & **Jung, C. R.** (2015). Automatic Solfège assessment. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 183–189, Malaga, Spain.

Schubert, E., Canazza, S., De Poli, G., & **Rodà, A.** (2017). Algorithms can mimic human piano performance: The deep blues of music. *Journal of New Music Research, 46*(2), 175–186. DOI: https://doi.org/10.1080/09298215.2016.1264976

Schubert, E., & **Fabian, D.** (2006). The dimensions of Baroque music performance: A semantic differential study. *Psychology of Music, 34*(4), 573–587. DOI: https://doi.org/10.1177/0305735606068105

Schubert, E., & **Fabian, D.** (2014). A taxonomy of listeners' judgments of expressiveness in music performance. In Fabian, D., Timmers, R., & Schubert, E., editors, *Expressiveness in Music Performance: Empirical Approaches Across Styles and Cultures*. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199659647.003.0016

Seashore, C. E. (1938). *Psychology of Music*. McGraw-Hill, New York. DOI: https://doi.org/10.2307/3385515

Serra, X. (2014). Creating research corpora for the computational study of music: The case of the CompMusic Project. In *Proceedings of the AES International Conference on Semantic Audio*, pages 1–9, London, UK. AES.

Shaffer, L. H. (1984). Timing in solo and duet piano performances. *The Quarterly Journal of Experimental Psychology, 36A*, 577–595. DOI: https://doi.org/10.1080/14640748408402180

Shi, Z., Sapp, C. S., Arul, K., McBride, J., & **Smith, J. O.** (2019). SUPRA: Digitizing the Stanford University Piano Roll Archive. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 517–523, Delft, The Netherlands.

Siegwart, H., & **Scherer, K. R.** (1995). Acoustic concomitants of emotional expression in operatic singing: The case of Lucia in Ardi gli incensi. *Journal of Voice, 9*(3), 249–260. DOI: https://doi.org/10.1016/S0892-1997(05)80232-2

Silvey, B. A. (2012). The role of conductor facial expression in students' evaluation of ensemble expressivity. *Journal of Research in Music Education, 60*(4), 419–429. DOI: https://doi.org/10.1177/0022429412462580

Sloboda, J. A. (1982). Music Performance. In Deutsch, D., editor, *The Psychology of Music*. Academic Press, New York. DOI: https://doi.org/10.1016/B978-0-12-213562-0.50020-6

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *The Quarterly Journal of Experimental Psychology Section A, 35*(2), 377–396. DOI: https://doi.org/10.1080/14640748308402140

Srinivasamurthy, A., Holzapfel, A., Ganguli, K. K., & **Serra, X.** (2017). Aspects of tempo and rhythmic elaboration in Hindustani music: A corpus study. *Frontiers in Digital Humanities, 4*. DOI: https://doi.org/10.3389/fdigh.2017.00020

Stowell, D., & **Chew, E.** (2013). Maximum a posteriori estimation of piecewise arcs in tempo time-series. In Aramaki, M., Barthet, M., Kronland-Martinet, R., & Ystad, S., editors, *From Sounds to Music and Emotions*, Lecture Notes in Computer Science, pages 387–399, Berlin, Heidelberg. Springer. DOI: https://doi.org/10.1007/978-3-642-41248-6_22

Su, L., Yu, L.-F., & **Yang, Y.-H.** (2014). Sparse cepstral, phase codes for guitar playing technique classification. In Wang, H.-M., Yang, Y.-H., & Lee, J. H., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 9–14.

Sulem, A., Bodner, E., & **Amir, N.** (2019). Perceptionbased classification of expressive musical terms: Toward a parameterization of musical expressiveness.

Music Perception, 37(2), 147–164. DOI: https://doi.org/10.1525/mp.2019.37.2.147

Sundberg, J. (1993). How can music be expressive? *Speech Communication, 13*(1), 239–253. DOI: https://doi.org/10.1016/0167-6393(93)90075-V

Sundberg, J. (2018). The singing voice. In Frühholz, S., & Belin, P., editors, *The Oxford Handbook of Voice Perception*. Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198743187.013.6

Sundberg, J., Lã, F. M. B., & **Himonides, E.** (2013). Intonation and expressivity: A single case study of classical Western singing. *Journal of Voice, 27*(3), 391.e1–391.e8. DOI: https://doi.org/10.1016/j.jvoice.2012.11.009

Takeda, H., Nishimoto, T., & **Sagayama, S.** (2004). Rhythm and tempo recognition of music performance from a probabilistic approach. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain.

Thompson, S., & **Williamon, A.** (2003). Evaluating evaluation: Musical performance assessment as a research tool. *Music Perception: An Interdisciplinary Journal, 21*(1), 21–41. DOI: https://doi.org/10.1525/mp.2003.21.1.21

Timmers, R. (2005). Predicting the similarity between expressive performances of music from measurements of tempo and dynamics. *The Journal of the Acoustical Society of America, 117*(1), 391–399. DOI: https://doi.org/10.1121/1.1835504

Todd, N. P. M. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91*, 3540–3550. DOI: https://doi.org/10.1121/1.402843

Todd, N. P. M. (1993). Vestibular feedback in musical performance. *Music Perception, 10*(3), 379–382. DOI: https://doi.org/10.2307/40285575

Todd, N. P. M. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America (JASA), 97*(3), 1940–1949. DOI: https://doi.org/10.1121/1.412067

Toyoda, K., Noike, K., & **Katayose, H.** (2004). Utility system for constructing database of performance deviations. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain.

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences, 110*(36), 14580–14585. DOI: https://doi.org/10.1073/pnas.1221454110

Tzanetakis, G. (2014). Computational ethnomusicology: A music information retrieval perspective. In *Proceedings of the Joint International Computer Music/Sound and Music Computing Conference*, Athens.

Van Herwaarden, S., Grachten, M., & **De Haas, W. B.** (2014). Predicting expressive dynamics in piano performances using neural networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, Taiwan.

van Noorden, L., & **Moelants, D.** (1999). Resonance in the perception of musical pulse. *Journal of New Music Research, 28*(1), 43–66. DOI: https://doi.org/10.1076/jnmr.28.1.43.3122

Vidwans, A., Gururani, S., Wu, C.-W., Subramanian, V., Swaminathan, R. V., & **Lerch, A.** (2017). Objective descriptors for the assessment of student music performances. In *Proceedings of the AES Conference on Semantic Audio*, Erlangen. Audio Engineering Society (AES).

Vieillard, S., Roy, M., & **Peretz, I.** (2012). Expressiveness in musical emotions. *Psychological Research, 76*(5), 641–653. DOI: https://doi.org/10.1007/s00426-011-0361-4

Viraraghavan, V. S., Aravind, R., & **Murthy, H. A.** (2017). A statistical analysis of gamakas in Carnatic music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China.

Wager, S., Tzanetakis, G., Sullivan, S., Wang, C.-i., Shimmin, J., Kim, M., & **Cook, P.** (2019). Intonation: A dataset of quality vocal performances refined by spectral clustering on pitch congruence. In *Proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 476–480, Brighton, UK. IEEE. DOI: https://doi.org/10.1109/ICASSP.2019.8683554

Wang, B., & **Yang, Y.-H.** (2019). PerformanceNet: Score-to-audio music generation with multi-band convolutional residual network. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(01), 1174–1181. DOI: https://doi.org/10.1609/aaai.v33i01.33011174

Wang, C., Benetos, E., Lostanlen, V., & **Chew, E.** (2019). Adaptive time-frequency scattering for periodic modulation recognition in music signals. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, Netherlands.

Wapnick, J., Campbell, L., Siddell-Strebel, J., & **Darrow, A.-A.** (2009). Effects of non-musical attributes and excerpt duration on ratings of high-level piano performances. *Musicae Scientiae, 13*(1), 35–54. DOI: https://doi.org/10.1177/1029864909013001002

Wesolowski, B. C. (2016). Timing deviations in jazz performance: The relationships of selected musical variables on horizontal and vertical timing relations: A case study. *Psychology of Music, 44*(1), 75–94. DOI: https://doi.org/10.1177/0305735614555790

Wesolowski, B. C., Wind, S. A., & **Engelhard, G.** (2016). Examining rater precision in music performance assessment: An analysis of rating scale structure using the multifaceted Rasch partial credit model. *Music Perception: An Interdisciplinary Journal, 33*(5), 662–678. DOI: https://doi.org/10.1525/mp.2016.33.5.662

Widmer, G. (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence, 146*(2), 129–148. DOI: https://doi.org/10.1016/S0004-3702(03)00016-X

Widmer, G., & **Goebl, W.** (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research, 33*(3), 203–216. DOI: https://doi.org/10.1080/0929821042000317804

Wilkins, J., Seetharaman, P., Wahl, A., & **Pardo, B.** (2018). VocalSet: A singing voice dataset. In Gómez, E., Hu, X., Humphrey, E., & Benetos, E., editors, *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 468–474, Paris, France.

Winters, R. M., Gururani, S., & **Lerch, A.** (2016). Automatic practice logging: Introduction, dataset & preliminary study. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, New York.

Wolf, A., Kopiez, R., Platz, F., Lin, H.-R., & **Mütze, H.** (2018). Tendency towards the average? The aesthetic evaluation of a quantitatively average music performancea successful replication of Repp's (1997) study. *Music Perception, 36*(1), 98–108. DOI: https://doi.org/10.1525/mp.2018.36.1.98

Wu, C.-W., Gururani, S., Laguna, C., Pati, A., Vidwans, A., & **Lerch, A.** (2016). Towards the objective assessment of music performances. In *Proceedings of the International Conference on Music Perception and Cognition (ICMPC)*, pages 99–103, San Francisco.

Wu, C.-W., & **Lerch, A.** (2016). On drum playing technique detection in polyphonic mixtures. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, New York. ISMIR.

Wu, C.-W., & **Lerch, A.** (2018a). Assessment of percussive music performances with feature learning. *International Journal of Semantic Computing (IJSC), 12*(3), 315–333. DOI: https://doi.org/10.1142/S1793351X18400147

Wu, C.-W., & **Lerch, A.** (2018b). From labeled to unlabeled data – on the data challenge in automatic drum transcription. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Paris.

Wu, C.-W., & **Lerch, A.** (2018c). Learned features for the assessment of percussive music performances. In *Proceedings of the International Conference on Semantic Computing (ICSC)*, Laguna Hills. IEEE. DOI: https://doi.org/10.1109/ICSC.2018.00022

Xia, G., & **Dannenberg, R.** (2015). Duet interaction: Learning musicianship for automatic accompaniment. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME 2015, pages 259–264, Baton Rouge, Louisiana, USA. The School of Music and the Center for Computation and Technology (CCT), Louisiana State University.

Xia, G., Wang, Y., Dannenberg, R. B., & **Gordon, G.** (2015). Spectral learning for expressive interactive ensemble music performance. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 816–822, Malaga, Spain.

Yang, L., Tian, M., & **Chew, E.** (2015). Vibrato characteristics and frequency histogram envelopes in Beijing opera singing. In *Proceedings of the Fifth International Workshop on Folk Music Analysis (FMA)*, Paris, France.

Zhang, S., Caro Repetto, R., & **Serra, X.** (2014). Study of the similarity between linguistic tones and melodic pitch contours in Beijing opera singing. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, Taiwan.

Zhang, S., Caro Repetto, R., & **Serra, X.** (2015). Predicting pairwise pitch contour relations based on linguistic tone information in Beijing opera singing. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, Spain.

Zhang, S., Caro Repetto, R., & **Serra, X.** (2017). Understanding the expressive functions of Jingju metrical patterns through lyrics text mining. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China.