

MUSIC PERFORMANCE ANALYSIS: A SURVEY

Alexander Lerch Claire Arthur Ashis Pati Siddharth Gururani

Center for Music Technology, Georgia Institute of Technology, Atlanta, USA

{alexander.lerch, claire.arthur, ashis.pati, siddgururani}@gatech.edu

ABSTRACT

Music Information Retrieval (MIR) tends to focus on the analysis of audio signals. Often, a single music recording is used as representative of a “song” even though different performances of the same song may reveal different properties. A performance is distinct in many ways from a (arguably more abstract) representation of a “song,” “piece,” or musical score. The characteristics of the (recorded) performance—as opposed to the score or musical idea—can have a major impact on how a listener perceives music. The analysis of music performance, however, has been traditionally only a peripheral topic for the MIR research community. This paper surveys the field of Music Performance Analysis (MPA) from various perspectives, discusses its significance to the field of MIR, and points out opportunities for future research in this field.

1. INTRODUCTION

Music, as a performing art, requires a performer or group of performers to render a musical score into an acoustic realization [38]. This is also true for non-classical music: for example, the ‘score’ might be a lead sheet or only a structured sequence of musical ideas, a ‘performer’ could also be a computer rendering audio, and the acoustic realization might be represented by a recording. The performance plays a major role in how listeners perceive a piece of music: even if the score content 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 actively or subconsciously interpreting, modifying, adding to, and dismissing score information.

Although the distinction between score and performance parameters is less obvious for other genres of western music, especially ones without clear separation between the composer and the performer, the concept of interpreting an underlying score 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, e.g., include the modification of content related to pitch, harmony, and rhythm.

Performance parameters can have a major impact on a listener’s perception of the music [13]. Formally, performance parameters can be structured in the same basic categories that we use to describe audio in general: tempo and timing, dynamics, pitch, timbre [51]. While the importance of different parameters might vary from genre to genre, the following list introduces some mostly genre-agnostic examples to clarify the 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 based on their plan for phrasing, tension, importance of certain parts of the score, and highlight specific events with accents.
- *Pitch* — the score usually defines the general pitches to play, but pitch-based performance parameters include expressive techniques such as vibrato as well as conscious or unconscious choices for intonation.
- *Timbre* — as the least specific category of musical sound, scores encode timbre parameters often only implicitly (e.g., instrumentation) while performers can, e.g., change playing techniques or the choice of specific instrument configurations (such as the choice of organ registers).

Note that usually the performance to be analyzed is a *recording* and not a live performance; every recording contains processing choices and interventions by sound engineer and editor with potential impact on expressivity—these modifications cannot be separated from the musicians’ creation and are thus an integral part of what is investigated [58].

The most 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. Early attempts at systematic and empirical MPA can be traced back to the 1930s with vibrato and singing analysis by Seashore [90] and the examination of piano rolls by Hartmann [37]. In the past two decades, MPA has greatly benefited from the advances in audio analysis made by members of the Music Infor-



mation Retrieval (MIR) community, significantly extending the volume of empirical data by simplifying access to a continuously growing heritage of commercial audio recordings. However, while advances in audio content analysis have had clear impact on MPA, the opposite is less true. While there have been publications on performance analysis at ISMIR, the major MIR conference, their absolute number remains comparably small (compare [3, 10, 18, 41, 53, 54, 73, 84, 85, 98, 104, 105] with a title referring to music performance out of approximately 1800 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 [5]; 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 researchers 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.

The remainder of this paper is structured as follows. The next Sect. 2 presents research on the objective description 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 (Sect. 3) or the listener (Sects. 4 and 5). We conclude our overview with a summary on applications of MPA and final remarks in Sect. 6. Note that while performance research has been inclusive of various musical genres, such as the Jingju music of the Beijing opera [33, 118], traditional Indian music [15, 34, 65] and jazz music [1], the vast majority of studies have been concerned with Western classical music. Therefore, the remainder of the paper focuses primarily on Western music.

2. PERFORMANCE MEASUREMENT

A large body of work focuses on a descriptive approach to analyzing performance recordings. Such studies typically extract characteristics such as the tempo curve [69, 75, 77] or loudness curve [82, 90] 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.

Several researchers observed a close relationship between musical phrase structure and deviations in tempo and timing [71, 75, 91]. For example, tempo changes in the form of *ritardandi* tend to occur at phrase boundaries [50, 69]. A similar co-occurrence was observed between dynamic patterns and timing [50, 81]. Additionally, Dalla Bella found the overall tempo influences the overall loudness of a performance [16]. There are also indications that loudness can

be linked to pitch height [81]. While the close relation of tempo and dynamics to structure has been repeatedly verified, Lerch did not succeed in finding similar relationships between structure and timbre properties in the case of string quartet recordings [50].

Pitch-based performance parameters have been analyzed mostly in the context of single-voiced instruments. Vibrato and its rate has, e.g., been studied for vocalists [17, 90] and violinists [22, 57]. Regarding intonation, Devaney et al. found significant differences between professional and non-professional vocalists in terms of the size of the interval between semi-tones [17].

Other studies use a multitude of performance parameters and aim to identify trends over time. For example, Ornoy and Cohen investigated violin performances of 19th century repertoire recorded in the past two decades [68]. They found a blend of stylistic approaches among violinists which questions the traditional distinction made between a historically informed and a mainstream performance.

The challenges in accessibility and interpretability of the extracted performance parameters have also led researchers to work on more intuitive or condensed forms of visualization that allow describing and comparing different performances beyond the traditional forms of visualization such as tempo curves [69, 75, 77] and scatter plots [50]. 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 [23, 48]. Sapp proposed so-called “Timescapes” and “Dynascapes” to visualize sub-sequence similarities of performance parameters [84, 85].

While most of the studies mentioned above make use of statistical methods to extract, visualize, and investigate patterns in the performance data, few studies make use of Machine Learning (ML) for MPA and performance modeling. However, ML-based approaches are useful for tasks such as composer classification, discovery of performance rules, or modeling performance characteristics. Widmer conducted studies that utilized ML to model expression in musical performance [109, 111, 112] and to learn simple rules from performance data with inductive learning [110]. He also applied ML to identify performers, showing that performer characteristics can be modeled by ML algorithms [112].

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 perceptually meaningful. Without this link, however, results can only provide limited insights into which parameters and parameter variations are ultimately important. Another typical challenge in such studies is the reliability of MIR algorithms for automatic annotations. While the accuracy of such algorithms has steadily improved over time, the fact that the majority of studies surveyed here continue to rely on manually-annotated data implies that the state-of-the-art algorithms for automatic annotation still

lack the required degree of accuracy for most tasks. As a result, most analyses are performed on small sample sizes possibly leading to poor generalizability of the studies.

3. PERFORMER

While most studies focus on the extraction of performance parameters or the mapping of these parameters to the listeners' reception (see Sects. 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 [14]. There is, as Palmer verified, a clear relation between reported intentions and objective parameters related to phrasing and timing of the performance [69]. Similar relations between the intended emotionality and loudness and timing measures were reported by Juslin [42] and Dillon [19–21]. 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. found that listeners were better able to perceive the presence of specific emotions in music when the performer played an expressive (versus mechanical) rendition of the composition [107]. 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 [44]. 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 [55, 86, 87]. Another interesting area of research is performer error. Repp 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 [40], suggesting that performers consciously or unconsciously avoid salient mistakes [80].

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 possibly excludes well-known and famous artists, resulting in a possible lack of generalizability. Depending on the experimental design, the separation of possible confounding variables (e.g., motor skills, random variations, and the influence of common performance rules) from the scrutinized performance data can be a challenge.

4. LISTENER

Every performance will ultimately be received and processed by a listener. The listener's meaningful interpretation of the incoming musical information relies on a sophisticated network of parameters. These include not only external, or semi-objective parameters such as score or performance-based features, but also "internal" ones such as those shaped by the culture, training, and history of the

listener. For this reason, listener-focused MPA remains one of the most challenging and elusive areas of research. However, to the extent that MPA research depends on purely perceptual information (e.g., expressiveness) or intends to deliver perceptually-relevant output (e.g., performance evaluation or reception, similarity), 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).

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 [89], listeners are very concerned with the notion of "expressiveness" which is a complex, multifaceted construct. Discovering which performance characteristics contribute to an expressive performance thus requires dissecting what listeners deem "expressive" as well as understanding the relation and potential differences between measured and perceived performance features. For instance, expressiveness is style-dependent, meaning that the perceived *appropriate* level of expression in a Baroque piece will be different from that of a Romantic piece — something that has been referred to as "stylishness" [25, 45]. In addition, there is the perceived *amount* of expressiveness, which is considered independent of stylishness [88]. Finally, Schubert and Fabian distinguish a third "layer" of expressiveness which arises from a performer's manipulation of various features specifically to alter or enhance emotion [89]. This is distinct from *musical* expressiveness 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 [89, p.293]), and research has demonstrated that there are interactions between them (e.g., [107]).

4.2 Expressive variation

Several scholars have made significant advances in our understanding of the role of timing, tempo, and dynamics on listeners' perception of music. As noted in Sect. 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, changes in timing and articulation within small structural units such as the measure, beat, or sub-beat appear to aid in the communication and emphasis of the metrical structure (e.g., [4, 29, 70, 93]), whereas changes across larger segments such as phrases, aid in the communication of formal structure. In fact, the communication of musical structure has been suggested as one of the primary roles or functions of a successful performer (see [43, 83]). For instance, an experiment by Sloboda found that listeners were better able to identify the meter of an ambiguous passage when performed by a more experienced performer [93]. 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 performer’s expressive variations align with a listener’s musical expectations 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 [66, 77, 78]. Likewise, Clarke [11] 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, the role of performance features such as timing and dynamics have also been studied extensively for their role in shaping “expressive” performance (see [12, 30]). For instance, a factor analysis in [89] 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.” Gingras et al. studied the relation between musical structure, expressive variation, and listeners’ ratings of musical tension. They found that variations in expressive timing were most predictive of listeners’ tension ratings [31]. While the role of expressive variation in timbre and intonation have 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 [96]). For instance, Sundberg found that a sharpened intonation at a phrase climax contributed to increased expressiveness and perceived excitement [97], and Siegwart and Scherer found that listener preferences were correlated with certain spectral components such as the relative strength of the fundamental and higher spectral centroid [92].

The reason why expressive variation is so enjoyable for listeners remains largely an open research question. As mentioned above, its role appears to go beyond bolstering the communication of musical structure. As pointed out by Repp, even a computerized or metronomic performance will contain grouping cues [83]. 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 [32, 49, 79, 102, 103] or human movements or gestures [8, 66]. For instance, Friberg and Sundberg suggested that the shape of final ritardandi matched the the velocity of runners coming to a stop [27], and Juslin includes “motion principles” in his model of performance expression [43].

4.3 Mapping and Predicting Listener Judgments

In order to isolate listeners’ perception of features that are strictly performance-related, several scholars have investi-

gated listeners’ judgments across multiple performances of the same excerpt of music (e.g., [24, 77]). 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., [11, 76, 83, 95]), and frequently making use of continuous data collection measures (e.g., [89]).

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., [26, 63]). 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 [39] showed there was no correlation between measured and perceived vibrato onset times. 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 a *perceived* “expressive” performance, especially if the changes in measured performance parameters are structurally normative, as discussed in Sect. 4.2.

Given a weak relation between a measured parameter and listeners’ perception of that parameter, an 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, who examined the representation of tempo and dynamics that best predicted listener judgments of musical similarity [101]. 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.

Clearly, the execution of several performance parameters are important for the perception of both fine-grained and large-grained musical structures, and appear 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 break down this high-level feature into meaningful collections of well-defined features that would be useful for MPA.

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 composition and performance, automatic accompaniment, virtual instrument design and control, or robotic instruments and HCI (see, e.g., the range of topics discussed in [46]). 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 (or conflate) the various definitions of “expressive.” 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 areas for future research.

5. PERFORMANCE ASSESSMENT

Assessment-focused MPA deals with modeling how we as humans assess a musical performance. While this is technically a subset of listener-focused MPA, its importance to MPA research and music education warrants a tailored review of research in this area. 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 [100, 108]. Work within Assessment-focused MPA seeks to increase the objectivity of performance assessments [62], and build accessible and reliable tools for automatic assessment.

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

Tools for performance assessment typically assess one or more performance parameters which are usually related to the accuracy of the performance in terms of pitch and timing [56, 72, 106, 113], or quality of sound (timbre) [47, 74]. 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 tools can also vary based on the granularity of assessments. Tools may simply classify a performance as ‘good’ or ‘bad’ [47, 64], or grade it on a scale, say from 1 to 10 [72]. Systems may provide fine-grained note-by-note

assessments [74] or analyze entire performances and report a single assessment score [64, 72].

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 cognitive model predicts the assessment. This approach requires availability of performance data (recordings) along-with human (expert) assessments for the rated parameters.

The level of sophistication of cognitive models was limited especially for early attempts; e.g., simple classifiers such as Support Vector Machines were used to predict human ratings. In this case, 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 [47]. In others, features aimed at capturing cognitive aspects of music perception were hand-designed using either musical intuition or expert knowledge [2, 52, 64, 74]. For instance, Nakano et al. used features measuring pitch stability and vibrato as inputs to a simple classifier to rate the quality of vocal performances [64]. Several studies also attempted to combine low-level audio features with hand-designed feature sets [56, 106, 113], as well as incorporating information from the musical score into feature computation [6, 7, 18, 106].

Recent methods, however, have transitioned towards using advanced ML techniques such as Sparse Coding [36, 114, 116] and Deep Learning [72] as a proxy to sophisticated cognitive models. Contrary to earlier methods which focused on extracting cognitively intuitive or 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. 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 [72, 114]. 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. 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 [35].

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 performance of these tools in modeling human judgments. Most of the presented systems either work well only for very select data [47] or have comparably low prediction accuracies [106, 113], rendering them unusable in most practical scenarios. 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. Second, the distribution of ground-truth (expert) ratings given by human judges is in many datasets skewed towards a particular class or value [35]. This makes it challenging to train unbiased models. Finally, the number of parameters required to adequately model a performance results in high dimensional data. 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. In many occasions, such assessments are either not available or are costly to obtain.

Given these data-related challenges, 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 [9]. In addition, the ability to interpret and understand the features learned by end-to-end models will play an important role in improving assessment tools. Interpretability of neural networks is still an active area of research in MIR, and performance assessment is an excellent test-bed 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. Furthermore, they enable 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 students' performance, provide a concise yet easily understandable analysis, give specific and understandable feedback on how to improve, and individualize the curriculum depending on

the students' 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 [59], Yousician [117], Music Prodigy [99], and SingStar [94].

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 been utilizing tempo and dynamics features successfully [28]. Generative models often require performance data to allow for the rendition of a convincing output. This obviously includes performance rendition systems that take a score and attempt to render a human-like output [60, 67], but it is also important for models of improvisation such as jazz solos as pitch information is part of the performance.

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 the sections above. 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 score 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, e.g., 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, lack of data continues to be a challenge for both, MIR core tasks and MPA; a focus on approaches for limited data [61], weakly labeled data, and unlabeled data [115] could help address this problem..

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.

7. REFERENCES

- [1] Jakob Abeßer, Klaus Frieler, Estefanía Cano, Martin Pfeleiderer, and Wolf-Georg Zaddach. Score-Informed Analysis of Tuning, Intonation, Pitch Modulation, and Dynamics in Jazz Solos. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 25(1):168–177, January 2017.
- [2] Jakob Abeßer, Johannes Hasselhorn, Christian Dittmar, Andreas Lehmann, and Sascha Grollmisch. Automatic Quality Assessment of Vocal and Instrumental Performances of Ninth-Grade and Tenth-Grade Pupils. In *International Symposium on Computer Music Multidisciplinary Research (CMMR)*, 2013.
- [3] Helena Bantula, Sergio Iván Giraldo, and Rafael Ramirez. Jazz Ensemble Expressive Performance Modeling. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, New York, 2016.
- [4] Klaus-Ernst Behne and Burkhard Wetekam. Musikpsychologische Interpretationsforschung: Individualität und Intention. *Musikpsychologie. Jahrbuch der Deutschen Gesellschaft für Musikpsychologie*, 10:24–37, 1993.
- [5] Emmanouil Benetos, Simon Dixon, Dimitrios Gianoulis, Holger Kirchhoff, and Anssi Klapuri. Automatic music transcription: challenges and future directions. *Journal of Intelligent Information Systems*, 41(3):407–434, Dec 2013.
- [6] Jordi Bonada, Alex Lascos, and Oscar Mayor. Performance Analysis and Scoring of the Singing Voice. In *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Audio Engineering Society, 2009.
- [7] Baris Bozkurt, Ozan Baysal, and Deniz Yüret. A Dataset and Baseline System for Singing Voice Assessment. In *International Symposium on Computer Music Multidisciplinary Research (CMMR)*, pages 430–438, 2017.
- [8] George John Broze III. *Animacy, anthropomimesis, and musical line*. PhD Thesis, The Ohio State University, 2013.
- [9] Keunwoo Choi, György Fazekas, Mark Sandler, and Kyunghyun Cho. Transfer Learning for Music Classification and Regression Tasks. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 141–149, Suzhou, 2017.
- [10] Ching-Hua Chuan and Elaine Chew. A Dynamic Programming Approach to the Extraction of Phrase Boundaries from Tempo Variations in Expressive Performances. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, 2007.
- [11] Eric F Clarke. Imitating and Evaluating Real and Transformed Musical Performances. *Music Perception*, 10:317–341, 1993.
- [12] Eric F Clarke. Rhythm and Timing in Music. In *The Psychology of Music*, pages 473–500. Academic Press, San Diego, 2nd edition, 1998.
- [13] Eric F Clarke. Listening to Performance. In John Rink, editor, *Musical Performance — A Guide to Understanding*. Cambridge University Press, Cambridge, 2002.
- [14] Eric F Clarke. Understanding the Psychology of Performance. In John Rink, editor, *Musical Performance – A Guide to Understanding*. Cambridge University Press, Cambridge, 2002.
- [15] Martin Clayton. *Time in Indian Music: Rhythm, Metre, and Form in North Indian Rag Performance*. Oxford University Press, August 2008. Google-Books-ID: pGOTwtQzxBYC.
- [16] Simone Dalla Bella and Caroline Palmer. Tempo and Dynamics in Piano Performance: The Role of Movement Amplitude. In *Proc. of the 8th International Conference on Music Perception & Cognition (ICMPC)*, Evanston, August 2004.
- [17] Johanna Devaney, Michael I. Mandel, Daniel P.W. Ellis, and Ichiro Fujinaga. Automatically Extracting Performance Data from Recordings of Trained Singers. *Psychomusicology: Music, Mind and Brain*, 21(1-2):108–136, 2011.
- [18] Johanna Devaney, Michael I Mandel, and Ichiro Fujinaga. A Study of Intonation in Three-Part Singing using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT). In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Porto, 2012.
- [19] Roberto Dillon. Extracting Audio Cues in Real Time to Understand Musical Expressiveness. In *Proc. of the MOSART workshop*, Barcelona, November 2001.
- [20] Roberto Dillon. A Statistical Approach to Expressive Intention Recognition in Violin Performances. In *Proc. of the Stockholm Music Acoustics Conference (SMAC)*, Stockholm, August 2003.
- [21] Roberto Dillon. *On the Recognition of Expressive Intentions in Music Playing: A Computational Approach with Experiments and Applications*. PhD Thesis, University of Genoa, Faculty of Engineering, Genoa, 2004.
- [22] Tomislav Dimov. *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, West Virginia, 2010.

- [23] Simon Dixon, Werner Goebel, and Gerhard Widmer. The Performance Worm: Real Time Visualisation of Expression based on Langner's Tempo Loudness Animation. In *Proc. of the International Computer Music Conference (ICMC)*, Göteborg, September 2002.
- [24] Dorottya Fabian and Emery Schubert. 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, July 2008.
- [25] Dorottya Fabian and Emery Schubert. Baroque Expressiveness and Stylishness in Three Recordings of the D minor Sarabanda for Solo Violin (BWV 1004) by JS Bach. *Music Performance Research*, 3:36–55, 2009.
- [26] Dorottya Fabian, Emery Schubert, and Richard Pulley. A Baroque Träumerei: The Performance and Perception of two Violin Renditions. *Musicology Australia*, 32(1):27–44, July 2010.
- [27] Anders Friberg and Johan Sundberg. Does Music Performance Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners. *Journal of the Acoustical Society of America*, 105(3):1469–1484, 1999.
- [28] Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang. A Survey of Audio-Based Music Classification and Annotation. *Trans. Multimedia*, 13(2):303–319, April 2011.
- [29] Alf Gabrielsson. Once Again: The Theme from Mozart's Piano Sonata in A Major (K. 331): A Comparison of Five Performances. In Alf Gabrielsson, editor, *Action and Perception in Rhythm and Music*, pages 81–103. Royal Swedish Academy of Music, No. 55., Stockholm, 1987.
- [30] Alf Gabrielsson. The Performance of Music. In Diana Deutsch, editor, *Psychology of Music*. Academic Press, San Diego, 2nd edition, 1998.
- [31] Bruno Gingras, Marcus T Pearce, Meghan Goodchild, Roger T Dean, Geraint Wiggins, and Stephen McAdams. Linking Melodic Expectation to Expressive Performance Timing and Perceived Musical Tension. *Journal of Experimental Psychology: Human Perception and Performance*, 42(4):594, 2016.
- [32] Robert O Gjerdingen. Shape and Motion in the Microstructure of Song. *Music Perception*, 6:35–64, 1988.
- [33] Rong Gong. *Automatic Assessment of Singing Voice Pronunciation: A Case Study with Jingju Music*. PhD Thesis, Universitat Pompeu Fabra, Barcelona, 2018.
- [34] Chitrallekha Gupta and Preeti Rao. Objective Assessment of Ornamentation in Indian Classical Singing. In Sølvi Ystad, Mitsuko Aramaki, Richard Kronland-Martinet, Kristoffer Jensen, and Sanghamitra Mohanty, editors, *Speech, Sound and Music Processing: Embracing Research in India*, Lecture Notes in Computer Science, pages 1–25. Springer Berlin Heidelberg, 2012.
- [35] Siddharth Gururani, K Ashis Pati, Chih-Wei Wu, and Alexander Lerch. Analysis of Objective Descriptors for Music Performance Assessment. In *Proc. of the International Conference on Music Perception and Cognition (ICMPC)*, Toronto, Ontario, Canada, 2018.
- [36] Yoonchang Han and Kyogu Lee. Hierarchical Approach to Detect Common Mistakes of Beginner Flute Players. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 77–82, Taipei, 2014.
- [37] Artur Hartmann. Untersuchungen über das metrische Verhalten in musikalischen Interpretationsvarianten. *Archiv für die gesamte Psychologie*, 84:103–192, 1932.
- [38] Peter Hill. From Score to Sound. In John Rink, editor, *Musical Performance -- A Guide to Understanding*. Cambridge University Press, Cambridge, 2002.
- [39] Patricia Howes, Jean Callaghan, Pamela Davis, Dianna Kenny, and William Thorpe. The Relationship between Measured Vibrato Characteristics and Perception in Western Operatic Singing. *Journal of Voice*, 18(2):216–230, June 2004.
- [40] David Huron. Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception: An Interdisciplinary Journal*, 19(1):1–64, 2001.
- [41] Luis Jure, Ernesto Lopez, Martin Rocamora, Pablo Canceleda, Haldo Sponton, and Ignacio Irigaray. Pitch Content Visualization Tools for Music Performance Analysis. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Porto, 2012.
- [42] Patrik N Juslin. Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception. *Journal of Experimental Psychology*, 26(6):1797–1813, 2000.
- [43] Patrik N Juslin. Five Facets of Musical Expression: A Psychologist's Perspective on Music Performance. *Psychology of Music*, 31(3):273–302, July 2003.
- [44] Patrik N Juslin and Petri Laukka. Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*, 129(5):770–814, September 2003.
- [45] Roger A Kendall and Edward C Carterette. The Communication of Musical Expression. *Music Perception*, 8(2):129–164, 1990.
- [46] Alexis Kirke and Eduardo R. Miranda, editors. *Guide to Computing for Expressive Music Performance*. Springer Science & Business Media, August 2012. Google-Books-ID: d9petWBUqHEC.

- [47] Trevor Knight, Finn Upham, and Ichiro Fujinaga. The Potential for Automatic Assessment of Trumpet Tone Quality. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, 2011.
- [48] Jörg Langner and Werner Goebel. Representing Expressive Performance in Tempo-Loudness Space. In *Proc. of the Conference of the European Society for the Cognitive Sciences of Music (ESCOM)*, Liege, April 2002.
- [49] Leon van Noorden and Dirk Moelants. Resonance in the Perception of Musical Pulse. *Journal of New Music Research*, 28(1):43–66, 1999.
- [50] Alexander Lerch. *Software-Based Extraction of Objective Parameters from Music Performances*. GRIN Verlag, München, 2009.
- [51] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Wiley-IEEE Press, Hoboken, 2012.
- [52] Pei-Ching Li, Li Su, Yi-Hsuan Yang, Alvin WY Su, and others. Analysis of Expressive Musical Terms in Violin Using Score-Informed and Expression-Based Audio Features. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 809–815, Malaga, 2015.
- [53] Cynthia CS Liem and Alan Hanjalic. Expressive Timing from Cross-Performance and Audio-based Alignment Patterns: An Extended Case Study. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, 2011.
- [54] Cynthia CS Liem and Alan Hanjalic. Comparative Analysis of Orchestral Performance Recordings: An Image-based Approach. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, 2015.
- [55] Paul Luizard, Erik Brauer, and Stefan Weinzierl. Singing in Physical and Virtual Environments: How Padapt to Room Acoustical Conditions. In *Proc. of the AES Conference on Immersive and Interactive Audio*, page 8, York, 2019. AES.
- [56] Yin-Jyun Luo, Li Su, Yi-Hsuan Yang, and Tai-Shih Chi. Detection of Common Mistakes in Novice Violin Playing. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, 2015.
- [57] Rebecca Bowman MacLeod. *Influences of Dynamic Level and Pitch Height on the Vibrato Rates and Widths of Violin and Viola Players*. PhD Thesis, Florida State University, Tallahassee, Florida, 2006.
- [58] Hans-Joachim Maempel. Musikaufnahmen als Datenquellen der Interpretationsanalyse. In Heinz von Lösch and Stefan Weinzierl, editors, *Gemessene Interpretation — Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen*, Klang und Begriff, pages 157–171. Schott, Mainz, 2011.
- [59] MakeMusic, Inc. SmartMusic, April 2019. <https://www.smartmusic.com>, last accessed 04/11/2019.
- [60] Iman Malik and Carl Henrik Ek. Neural Translation of Musical Style. In *Proc. of the NeurIPS Workshop on Machine Learning for Creativity and Design*, Long Beach, California, USA, 2017.
- [61] Brian McFee, Eric J Humphrey, and Juan Pablo Bello. A Software Framework for Musical Data Augmentation. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, 2015.
- [62] Gary E McPherson and William F Thompson. Assessing Music Performance: Issues and Influences. *Research Studies in Music Education*, 10(1):12–24, June 1998.
- [63] Toshie Nakamura. The Communication of Dynamics between Musicians and Listeners through Musical Performance. *Perception & Psychophysics*, 41(6):525–533, 1987.
- [64] Tomoyasu Nakano, Masataka Goto, and Yuzuru Hiraga. An Automatic Singing Skill Evaluation Method for Unknown Melodies using Pitch Interval Accuracy and Vibrato Features. In *Proc. of the International Conference on Spoken Language Processing (ICSLP)*, pages 1706–1709, 2006.
- [65] Krish Narang and Preeti Rao. Acoustic Features for Determining Goodness of Tabla Strokes. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 257–263, Suzhou, China, 2017.
- [66] Mitchell S Ohriner. Grouping Hierarchy and Trajectories of Pacing in Performances of Chopin’s Mazurkas. *Music Theory Online*, 18(1), 2012.
- [67] Sageev Oore, Ian Simon, Sander Dieleman, Douglas Eck, and Karen Simonyan. This Time with Feeling: Learning Expressive Musical Performance. *Neural Computing and Applications*, pages 1–13, 2018.
- [68] Eitan Ornoy and Shai Cohen. Analysis of Contemporary Violin Recordings of 19th Century Repertoire: Identifying Trends and Impacts. *Frontiers in Psychology*, 9:2233, 2018.
- [69] Caroline Palmer. Mapping Musical Thought to Musical Performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):331–346, 1989.
- [70] Caroline Palmer. *Timing in skilled music performance*. PhD Thesis, Cornell University, Ithaca, NY, 1989.
- [71] Caroline Palmer. Music Performance. *Annual Review of Psychology*, 48:115–138, 1997.

- [72] K Ashis Pati, Siddharth Gururani, and Alexander Lerch. Assessment of Student Music Performances Using Deep Neural Networks. *Applied Sciences*, 8(4):507, March 2018.
- [73] Jeroen Peperkamp, Klaus Hildebrandt, and Cynthia CS Liem. A Formalization of Relative Local Tempo Variations in Collections of Performances. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, 2017.
- [74] O Romani Picas, H Parra Rodriguez, Dara Dabiri, Hiroshi Tokuda, Wataru Hariya, Koji Oishi, and Xavier Serra. A Real-time System for Measuring Sound Goodness in Instrumental Sounds. In *Proc. of the Audio Engineering Society Convention*, volume 138, Warsaw, 2015.
- [75] Dirk-Jan Povel. Temporal Structure of Performed Music: Some Preliminary Observations. *Acta Psychologica*, 41(4):309–320, 1977.
- [76] Bruno H Repp. Expressive Microstructure in Music: A Preliminary Perceptual Assessment of Four Composers' "Pulses". *Music Perception*, 6(3):243–273, 1989.
- [77] Bruno H Repp. 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, August 1990.
- [78] Bruno H Repp. A Constraint on the Expressive Timing of a Melodic Gesture: Evidence from Performance and Aesthetic Judgment. *Music Perception*, 10:221–243, 1992.
- [79] Bruno H Repp. Music as Motion: A Synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik*. *Psychology of Music*, 21:48–72, 1993.
- [80] Bruno H Repp. The Art of Inaccuracy: Why Pianists' Errors are Difficult to Hear. *Music Perception*, 14(2):161–184, 1996.
- [81] Bruno H Repp. The Dynamics of Expressive Piano Performance: Schumann's 'Träumerei' Revisited. *Journal of the Acoustical Society of America (JASA)*, 100(1):641–650, 1996.
- [82] Bruno H Repp. 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, 1998.
- [83] Bruno H Repp. Obligatory "Expectations" of Expressive Timing Induced by Perception of Musical Structure. *Psychological Research*, 61(1):33–43, March 1998.
- [84] Craig S Sapp. Comparative Analysis of Multiple Musical Performances. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, September 2007.
- [85] Craig S Sapp. Hybrid Numeric/Rank Similarity Metrics for Musical Performance Analysis. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, September 2008.
- [86] Zora Schärer Kalkandjiev and Stefan Weinzierl. The Influence of Room Acoustics on Solo Music Performance: An Empirical Case Study. *Acta Acustica united with Acustica*, 99(3):433–441, May 2013.
- [87] Zora Schärer Kalkandjiev and Stefan Weinzierl. The Influence of Room Acoustics on Solo Music Performance. An Experimental Study. *Psychomusicology: Music, Mind, and Brain*, 25(3):195–207, 2015.
- [88] Emery Schubert and Dorottya Fabian. The Dimensions of Baroque Music Performance: A Semantic Differential Study. *Psychology of Music*, 34(4):573–587, 2006.
- [89] Emery Schubert and Dorottya Fabian. A Taxonomy of Listeners' Judgments of Expressiveness in Music Performance. In Dorottya Fabian, Renee Timmers, and Emery Schubert, editors, *Expressiveness in Music Performance: Empirical approaches across styles and cultures*. Oxford University Press, July 2014.
- [90] Carl E Seashore. *Psychology of Music*. McGraw-Hill, New York, 1938.
- [91] L Henry Shaffer. Timing in Solo and Duet Piano Performances. *The Quarterly Journal of Experimental Psychology*, 36A:577–595, 1984.
- [92] Hervine Siegwart and Klaus R Scherer. Acoustic Concomitants of Emotional Expression in Operatic Singing: The Case of Lucia in *Ardi gli incensi*. *Journal of Voice*, 9(3):249–260, 1995.
- [93] John A Sloboda. The Communication of Musical Metre in Piano Performance. *The Quarterly Journal of Experimental Psychology Section A*, 35(2):377–396, May 1983.
- [94] Sony Interactive Entertainment Europe. SingStar, April 2019. <http://www.singstar.com>, last accessed 04/11/2019.
- [95] Johan Sundberg. How can Music be Expressive? *Speech Communication*, 13(1):239–253, October 1993.
- [96] Johan Sundberg. The Singing Voice. In Sascha Frühholz and Pascal Belin, editors, *The Oxford Handbook of Voice Perception*. Oxford University Press, December 2018.
- [97] Johan Sundberg, Filipa MB Lã, and Evangelos Hionides. Intonation and Expressivity: A Single Case Study of Classical Western Singing. *Journal of Voice*, 27(3):391–e1, 2013.
- [98] Haruto Takeda, Takuya Nishimoto, and Shigeki Sagayama. Rhythm and Tempo Recognition of Music Performance from a Probabilistic Approach. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, 2004.

- [99] The Way of H, Inc. (dba Music Prodigy). Music Prodigy, April 2019. <http://www.musicprodigy.com>, last accessed 04/11/2019.
- [100] Sam Thompson and Aaron Williamon. Evaluating Evaluation: Musical Performance Assessment as a Research Tool. *Music Perception: An Interdisciplinary Journal*, 21(1):21–41, 2003.
- [101] Renee Timmers. Predicting the Similarity between Expressive Performances of Music from Measurements of Tempo and Dynamics. *Journal of the Acoustical Society of America (JASA)*, 117(1), 2005.
- [102] Neil P M Todd. The Dynamics of Dynamics: A Model of Musical Expression. *Journal of the Acoustical Society of America*, 91:3540–3550, 1992.
- [103] Neil P M Todd. The Kinematics of Musical Expression. *Journal of the Acoustical Society of America*, 97:1940–1949, 1995.
- [104] Ken’ichi Toyoda, Kenzi Noike, and Haruhiro Katayose. Utility System for Constructing Database of Performance Deviations. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Barcelona, 2004.
- [105] Sam Van Herwaarden, Maarten Grachten, and W Bas De Haas. Predicting Expressive Dynamics in Piano Performances Using Neural Networks. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, 2014.
- [106] Amruta Vidwans, Siddharth Gururani, Chih-Wei Wu, Vinod Subramanian, Rupak Vignesh Swaminathan, and Alexander Lerch. Objective Descriptors for the Assessment of Student Music Performances. In *Proc. of the AES Conference on Semantic Audio*, Erlangen, 2017. Audio Engineering Society (AES).
- [107] Sandrine Vieillard, Mathieu Roy, and Isabelle Peretz. Expressiveness in Musical Emotions. *Psychological Research*, 76(5):641–653, September 2012.
- [108] Brian C Wesolowski, Stefanie A Wind, and George Engelhard. 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, 2016.
- [109] Gerhard Widmer. Applications of Machine Learning to Music Research: Empirical Investigations into the Phenomenon of Musical Expression. In *Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications*. Wiley & Sons, 1997.
- [110] Gerhard Widmer. Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146(2):129–148, June 2003.
- [111] Gerhard Widmer and Werner Goebel. Computational Models of Expressive Music Performance: The State of the Art. *Journal of New Music Research*, 33(3):203–216, September 2004.
- [112] Gerhard Widmer and Patrick Zanon. Automatic Recognition of Famous Artists by Machine. In *Proc. of the 16th European Conference on Artificial Intelligence (ECAI)*, Valencia, August 2004.
- [113] Chih-Wei Wu, Siddharth Gururani, Christopher Laguna, K Ashis Pati, Amruta Vidwans, and Alexander Lerch. Towards the Objective Assessment of Music Performances. In *Proc. of the International Conference on Music Perception and Cognition (ICMPC)*, pages 99–103, San Francisco, 2016.
- [114] Chih-Wei Wu and Alexander Lerch. Assessment of Percussive Music Performances with Feature Learning. *International Journal of Semantic Computing*, 12(3):315–333, 2018.
- [115] Chih-Wei Wu and Alexander Lerch. From Labeled to Unlabeled Data – On the Data Challenge in Automatic Drum Transcription. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, Paris, 2018.
- [116] Chih-Wei Wu and Alexander Lerch. Learned Features for the Assessment of Percussive Music Performances. In *Proc. of the International Conference on Semantic Computing (ICSC)*, Laguna Hills, 2018. IEEE.
- [117] Yousician Oy. Yousician, April 2019. <https://www.yousician.com>, last accessed 04/11/2019.
- [118] Shuo Zhang, Rafael Caro Repetto, and Xavier Serra. 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, 2017.