Broader Perspectives in the Understanding of Musical Expression

Jeff Gregorio
Matthew Prockup
Brandon G. Morton
Youngmoo E. Kim
Drexel University
Philadelphia, PA 19104, USA
jgregorio@drexel.edu
mprockup@drexel.edu
bmorton@drexel.edu
ykim@drexel.edu

CHI'16, May 7–12, 2016, San Jose, CA, USA.

DOI string from ACM form confirmation

Abstract
Designing interfaces for musical expressivity is unusual in that expression often takes on an abstract, highly individualized, creative character, placing it among the most nebulous problems addressed by the HCI community. For these reasons, expression in Music Interaction literature has been characterized in subtly different ways depending on the goals of a study, tools available, and modalities under consideration. We argue that there may never emerge a single dominant definition of musical expression, yet Music Interaction and HCI researchers may benefit from a synthesis of a wide range of perspectives used to approach various facets of expression and related problems. This may inform the design and evaluation of a wide range of creative tools.

Author Keywords
musical expression; music information retrieval

ACM Classification Keywords
H.5.5 [Sound and Music Computing]: Methodologies and techniques

Background
From a purely creative perspective, the observation that musical expression continues to elude a universal definition can be liberating, yet the HCI perspective dictates an awareness of the implications that attempting to charac-

terize expression has on the applicability of any model to specific instruments, styles, and modalities. Prevailing models of expression often center on performer's *intention* and the listener's *perception* at the conceptual level, and embed relationships to *musical structure* and measure *timbre modulation* at the quantitative level.

Intention vs. Perception

Intention of the performer is often assumed to play the most salient role in identifying expression in quantitative studies. Typically, participants are asked to perform a piece of music with several expressive intentions. A supervised machine learning model is then trained to predict those intentions [1] [2] [10]. In [8], Juslin notes that modeling expression as strictly intended fails to account for variations in listener perception, while expression as strictly perceived potentially validates any arbitrary association. Intention may indeed be seen as primary from an HCI perspective, since recognition of intention has utility regardless of whether it is successfully communicated. However, addition of perceptual evaluation to work previously focused on intention could have strong design implications, particularly if an intention-perception gap is identified.

Musical Structure and Timbre

One may choose to define expression as distinct from the composition, that is, as systematic variations of the *manner* in which a piece of music is played. This amounts to rule-based variation of performance parameters affecting timbre such as timing, tempo, dynamics, and articulation. These score-*dependent* models typically aim to synthesize rule-based expressive deviations from a nominal score or 'unexpressive' performance [3] [4] [8]. This assumes that 'appropriate' expression exists and serves to clarify the written musical structure. This can be problematic in that it precludes expressivity in improvisation and many contemporary styles [7]. Strictly score-*independent* expression has been studied [1] [11] by constraining the performance to single notes, scales, or simple melodies. This allows precise conclusions to be drawn regarding low-level timbral parameters, but sacrifices ecological validity. Score-*agnostic* expression studies acknowledge that expression is inextricable from musical structure, but model expression without using score information. These are also limited in scope, typically employing only a few performers and musical examples. Therefore, conclusions drawn are only applicable in their self-curated contexts and may vary considerably if extended to a larger set of performers and musical styles. In each of the previous methods, the sole focus on only audio signals leads to shortcomings that also reside in certain instances of the aforementioned intention-perception gaps, such as a pianist's percussive key press being found indistinguishable from a non-percussive press [5]. Further, recent work has indicated the significance of visual modalities on perception of expertise [14]. We believe this indicates potential utility in incorporation of multi-modal data sources, including audio, symbolic, visual, and high-resolution sensor data related to the modes of sound production.

Potential Insights in Related Work

Music interaction is a highly interdisciplinary field which progresses by synthesizing elements from a wide range of perspectives and insights gained on related problems. At Drexel's Music & Entertainment Technology Lab (MET-lab), our work spans the domains of music information retrieval, quantitative performance analysis, and humanoid robotics. In this section, we provide a brief overview of recent work that may be of interest to music interaction researchers.

New Sources of Sensor Data

Many past studies of expressive piano performance utilizing symbolic performance data were limited to the MIDI proto-

col, which reduces a rich variety of piano key articulations to simple timing and velocity information, thereby obscuring details that could potentially illuminate gaps in our knowledge of expression. In [10], McPherson developed a tool which enabled the collection of continuous key position profiles using a modular platform and demonstrated that these low-level position profiles could be used to sense five separate mid-level parameters of key motion, with potential utility extending to a wide range of non-musical button interfaces. Subsequent work [9] expanded on the theme of continuous control in keyboard instruments via capacitive multi-touch sensing key surfaces, which offer three degrees of freedom (key position and contact area). We believe these new sources of high-resolution sensor data provide rich, underutilized potential for a window into the physical parameters of sound production on keyboard instruments.

Expression and Music Information Retrieval
Much of the work by signal processing engineers at MET-Lab has been situated in the domain of Music Information Retrieval (MIR), often concerned with problems applicable to music recommendation systems including automatic classification of genre and emotion. These and many other aspects of music under study in MIR are closely related to expression. Recorded instrument mixtures offer a larger source of data. However, because they are less controlled, they are often overlooked in attempting to answer general questions regarding expression. Previous work in genre classification has studied the general recognition of styles in music audio signals, but few efforts have focused on the construction of its foundational components, throughout which expression plays a large part. In [13], Prockup chose to first focus on modeling rhythm-related attributes of meter and 'feel' (e.g., 'swing') in music by designing targeted acoustic features to accurately represent these attributes. Subsequent work has expanded beyond rhythm to characteristics of instrumentation (i.e., guitar distortion, vocal timbre) and shows that understanding similarities and differences in these individual aspects of performance can lead to a better understanding of certain musical styles.

Avoiding Encoded Biases with Feature Learning
Previous studies of musical expression have encoded implicit biases as to what constitutes expression, whether it be in a stimulus condition (e.g. "play with a happy tone"), or in reducing high-dimensional raw performance data (MIDI, piano key presses, audio spectra, etc.) by computing features informed by musicological assumptions. The shortcomings of the latter assumption has led some in the MIR community toward the adoption of feature learning methods for the discovery of useful descriptors that might be overlooked in hand-designed feature sets. These methods avoid bias by using minimally-processed data as inputs to computational methods for finding inherent structure, making no assumptions about what is informative. In MET-Lab's MIR work, Schmidt [15] applied several variations on feature learning methods and deep neural networks toward predicting human-annotated emotion labels. In [12], Morton used feature learning methods based on non-negative matrix factorization and deep belief networks to study musical artist influence. We believe this work indicates that there is much unexplored potential utility in both prediction and analysis of expression using feature learning methods.

Alternative Methods of Analysis-by-Synthesis
The success of analysis-by-synthesis methods has revealed perceptually salient components of musical expression, where parameterized expression is used to modify symbolic data prior to audio synthesis, yet synthesized audio has questionable ecological validity for many musical styles. Robotic platforms have the potential to synthesize precise expression profiles from symbolic data while pro-

ducing an audio signal complete with the acoustic complexity missing from many synthesized instruments. Moreover, Grunberg has utilized several humanoid robot platforms in perceptual evaluations of synthesized motion profiles for dancing in response to emotional content in an audio signal [6]. We believe this indicates some potential for humanoid platforms in synthesizing both the aural and visual aspects of performance expression missing in synthesized audio.

Conclusions

It is apparent that no single, all-encompassing characterization of expression has emerged due to the constraints imposed by specific instruments, modalities, methods, and targets. While a large degree of success has been achieved in the recognition and synthesis of expression in controlled situations, it could be beneficial to Music Interaction researchers to be aware of the relative strengths and weaknesses of each characterization, model, and modality. Further, recent developments and successes within and outside the Music Interaction community indicate underutilized potential, not only in informing the next generation of interactive musical interfaces, but in answering fundamental questions regarding the nature of musical expression.

References

[1] F. B. Baraldi, G. De Poli, and A. Rodà. 2006. Communicating expressive intentions with a single piano note. *Journal of New Music Research* 35, 3 (2006), 197–210.

[2] M Bernays and T. Caroline. 2013. Expressive Production of Piano Timbre: Touch and Playing Techniques for Timbre Control in Piano Performance. In *Proc. Sound and Music Computing Conference*, R. Bresin (Ed.). KTH Royal Institute of Technology, Logos Verlag, Berlin, 341–346.

[3] A. Bhatara, A. K. Tirovolas, L. M. Duan, B. Levy, and D. J. Levitin. 2011. Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance* 37, 3 (June 2011), 921–934.

[4] S. Canazza, G. De Poli, C. Drioli, A. Roda, and A. Vidolin. 2004. Modeling and control of expressiveness in music performance. *Proc. of the IEEE* 92, 4 (Apr 2004), 686–701.

[5] W. Goebl, R. Bresin, and E. Galembo. 2004. Once Again: The Perception of Piano Touch and Tone. Can Touch Audibly Change Piano Sound Independently of Intensity. In *Proc. International Symposion on Music Acoustics*. 332–335.

[6] D. K. Grunberg, A. M. Batula, E. M. Schmidt, and Y. E. Kim. 2012. Affective Gesturing with Music Mood Recognition. In *Proc. International Conference on Humanoid Robotics*.

[7] M. Gurevich and J. Treviño. 2007. Expression and Its Discontents: Toward an Ecology of Musical Creation. In *Proc. of the 7th Intl. Conference on New Interfaces for Musical Expression*. ACM, New York, NY, USA, 106–111.

[8] P. N. Juslin. 2003. Five Facets of Musical Expression: A Psychologist's Perspective on Music Performance. *Psychology of Music* 31, 3 (2003), 273–302.

[9] A. McPherson and Y. E. Kim. 2011a. Design and applications of a multi-touch musical keyboard. In *Proc. Sound and Music Computing Conference*. Padova, Italy.

[10] A. McPherson and Y. E. Kim. 2011b. Multidimesional gesture sensing at the piano keyboard. In *Proc. ACM Human Factors in Computing Systems Conference*. Association for Computing Machinery (ACM), Vancouver, BC, 2789–2798.

[11] L. Mion and G. De Poli. 2008. Score-Independent Audio Features for Description of Music Expression. *Audio, Speech, and Language Processing, IEEE Trans.* 16, 2 (Feb 2008).

[12] B. G. Morton and Y. E. Kim. 2015. Acoustic Features For Recognizing Musical Artist Influence. In *Proc. International Conference on Machine Learning and Applications*.

[13] M. Prockup, A. F. Ehmann, F. Gouyon, E. M. Schmidt, and Y. E. Kim. 2015. Modeling Musical Rhythm at Scale with the Music Genome Project. In *IEEE WASPAA*.

[14] M. W. M. Rodger, C. M. Craig, and S. O'Modhrain. 2012. Expertise is perceived from both sound and body movement in musical performance. *Human Movement Science* 31, 5 (2012), 1137 – 1150.

[15] E. M. Schmidt and Y. E. Kim. 2013. Learning rhythm and melody features with deep belief networks. In *Proc. International Society for Music Information Retrieval Conference*.