Grant number: HD-228966-15
Title of project: Automatic Music Performance Analysis and Comparison Toolkit (AMPACT)
Project Director: Johanna Devaney
Institution: The Ohio State University
Date submitted: January 31, 2018

Automatic Music Performance Analysis and Comparison Toolkit (AMPACT)
White Paper

Project Participants

Project Director: Johanna Devaney
Computer Science and Engineering Collaborator: Michael I. Mandel
Ohio State University Graduate Research Associates: Dana DeVlieger
Hubert Léveillé Gauvin
Jiawei Liu
Lindsay Warrenburg

Statement of problem

The overarching goal for this project is to create a suite of automatic software tools for empirically analyzing musical performances for which a corresponding musical score or transcription is available, entitled the Automatic Music Performance and Comparison Toolkit (AMPACT). The primary target audience is scholars who are interested in empirically examining musical performances but who do not have the specialist technical skill or time required to develop their own tools or fully utilize existing tools. Quantitative analyses of musical performances will allow scholars to study not just contemporary performance practices, but also the performance practices used in historical recordings. Empirical methods for analyzing musical scores have already had a large impact on music scholarship, and the examination of recorded performance promises to be just as important. The software tools developed in this project are useful to musicologists and ethnomusicologists who are interested in undertaking longitudinal studies of either the development of a single performer's practice over time or changes in a particular style across time or geographical distance. For general audiences in the humanities, the results of this type of longitudinal music performance research can be linked to other historical and/or geographic research in order to contextualize the similarities or differences observed across musical performances. For example, AMPACT would allow a scholar to make empirical measurements of the performance of lyrics in vocal performances, which would facilitate large-scale examinations of the ways in which specific lyrical content is emphasized or de-emphasized in different performance contexts. A concrete example of this would be a comparison of studio and live recordings of performances of the song "This Land is Your Land" (see Appendix B for an example workflow). The tools could be useful for music performance teachers and students as a way to track progress and for developmental psychologists working on the acquisition of musical skill. Another potential

application is to develop computer models of expert performers, which could be applied to novice or computer-generated musical performances. Though the focus of AMPACT is to use corresponding musical scores to guide performance parameter estimation, other rough approximations of events, such as manual annotations, could also be automatically refined with AMPACT. There are also potential applications in analyzing the audio products of performing arts other than music, such as performance art or theatre.

Outcomes

Methods for extracting loudness and pitch information from polyphonic music

The core research question undertaken in this project was how to reliably extract performance data from polyphonic musical performances. Throughout 2015 and 2016, the Project Director (PD) and Computer Science and Engineering Contributor (CSEC) compared three methods for score-guided estimation of loudness and pitch information from polyphonic audio. Preliminary results were presented at the International Society of Music Information Retrieval (ISMIR) late-breaking demo session in August 2016 (Devaney and Mandel, 2016) and the Society for Music Theory's Music Informatics group session in November 2016. The PD and CESC subsequently presented both the results and a performance optimization for the estimation task in an oral presentation at the 2017 International Conference on Acoustics, Speech and Signal Processing (ICASSP) in March 2017 (Devaney and Mandel, 2017). In this paper, we evaluated the relative performance of the estimation algorithms based on discrete Fourier transform, instantaneous frequency analysis, and high-resolution spectral analysis on the task of estimating frame-wise fundamental frequency and power values. These values are the building block of our score-informed system, where the timing information for each note is available in the alignment and note-wise loudness, tuning, and timbral parameters can be calculated from the frame-wise frequency and power estimates. We initially expected high-resolution spectral analysis method would perform best but found that it was prone to spurious estimates. The discrete Fourier transform and instantaneous frequency approaches performed comparably, and we are currently exploring whether there is any difference between them on the task of estimating note-level performance descriptors from the frame-wise estimates.

Increased accuracy aligning score and audio

In Summer and Fall 2015 a part-time School of Music graduate research associate (GRA) was hired to develop test sets for AMPACT. The GRA succeeded in finding a number of existing annotated multi-track recordings and also worked on annotating recordings, focusing on unfretted string instruments, with a particular emphasis on cello. These test sets facilitated an increase in the range of instruments that AMPACT's alignment algorithm works on.

Comparing Performances

In Fall 2015, the PD expanded a study on measuring inter- versus intra-performer differences and similarities using AMPACT into a journal article that was published in the *Journal of New Music Research* (Devaney, 2016).

Encoding

In January 2016, a second part-time School of Music GRA to work on defining best practices for encoding performance data and examining ways of implementing AMPACT data encoding in both Humdrum and Music Encoding Initiative (MEI) formats. The first step in this process was running an online survey about needs for the encoding of music performance data, in which more than 50 music scholars participated. We used the results of this survey when defining best practices for encoding performance data, from which we subsequently developed proposals for Humdrum- and MEI-specific implementations. The GRA presented the results of the survey at the Music Encoding Conference in May 2016 along with a comparative evaluation of the potential for extending Humdrum and MEI to encoding performance data. We subsequently submitted a manuscript detailing these topics to the Digital Libraries for Musicology (DLfM) workshop, which was presented and published in the proceedings in August 2016 (Devaney and Léveillé Gauvin, 2016). After the DLfM workshop, the PD and the GRA expanded the proceeding paper into a journal article that was published the *International Journal on Digital Libraries* (Devaney and Léveillé Gauvin, 2017).

Documentation

In Summer 2016, a third part-time School of Music GRA was hired part-time to write up documentation for AMPACT. Her work, combined with the documentation written up by the second GRA on the encoding formats and the PD during the Summer of 2017, is available on the ampact.org and https://github.com/jcdevaney/AMPACT websites and formed the basis for the workshop materials that the PD developed in Fall 2017.

Surveys

The start-up stage of the AMPACT project was to both develop the underlying technologies necessary for this analysis and to survey musicologists who may be interested in using the product. These surveys were done both formally, as in the web-based survey we ran regarding encoding needs, and informally at conferences and workshops for both musicologists (e.g., American Musicological Society and Society for Music Theory) and more general audiences (e.g., Digital Humanities conference and CUNY Graduate Center Digital Initiatives).

Evaluation

The core technology developed during the grant was evaluated through experimental methods with labelled data. Some of the labelled data was created during the course of this grant while other sourced from other researchers or the PD's previous projects. Publications related to these technological develops went through peer-review by experts in music information retrieval and signal process. The overall usability of the newly developed tools was self-assessed by the PD, CESC, and GRAs. We focused on internal evaluation at this stage because we still consider the toolkit to be in pre-release form. Going forward, we are planning to engage in-depth surveys with researchers who are interested in using the new technology integrated in AMPACT both in terms of the appropriateness of the range of newly included features and the accuracy with which AMPACT can estimate them on different types of instruments/performance settings. One of the main challenges that we dealt with during the grant was figuring out how to assess the point at which the technology we were developing was usable, which was not necessarily the same thing as 100% accurate in what it was trying to estimate. Our decisions about this were both informed by human perception (the computational estimates did not need to be more accurate than human perceptual thresholds) and developing an understanding of where users would be willing to make corrections if necessary (such as for alignments of a particularly complicated piece of music).

Lessons learned

Methods for extracting loudness and pitch information from polyphonic music

Our initial plan to work on more instruments than unfretted string, starting with woodwind and brass instruments. This was postponed until the next phase of the grant due to both the lack of available data and, more importantly, by the determination that developing best practices for encoding performance data was a more important use of resources in the Start-Up phase of the grant.

Port MATLAB research code to Python and integrate it with Music21

In Spring 2016, a part-time Computer Science GRA was hired to begin porting the existing and newly developed MATLAB AMPACT code into Python. Concurrently, the PD and CESC began working on design ideas for a Python-driven web-based version of AMPACT in order to ensure that the port from MATLAB to Python is done in a sufficiently extensible way. Unfortunately, the GRA had to stop working on the project in Summer 2016 due to graduating. Subsequently, the PD tried to recruit an additional Computer Science student worker, or a student in another department with

sufficient technical skills to work on the project, but was unsuccessful. The challenge of finding another student employee was in part because in September 2015, the CESC has left Ohio State and moved to Brooklyn College. Concurrent to this, we were also reconsidering the utility of porting our research code to Python at this point. Our initial rationale was that a Python port would facilitate integration with music21, the most commonly used computational musicological software, but came to realize that that can be supported through the MEI and Humdrum performance extensions to the existing MEI and Humdrum parsers in music21. Specifically, that having a clearly defined and workable encoding format for performance data is vital for sharing data amongst researchers and for porting performance data between applications. As such, we focused our attention on the encoding formats and writing documentation for the remainder of the grant, leaving the porting of our MATLAB research code to the next stage of the project where we plan to implement a web-based GUI for AMPACT.

Future plans

The start-up funding allowed AMPACT to reach a stage where it can start to be rolled out to a wider array of users, and their feedback will inform the research team's decisions about what features to add and how the toolkit could be expanded. The PDs current focus is on giving workshops on the current version of AMPACT to introduce it to researchers who are interested in using it in its current code-based form. These include the CUNY Graduate Center Digital Initiative workshop she held in November 2017 and the forthcoming workshop in the NYC Digital Humanities Week in February 2018.

One long-term goal for this project, is to develop a graphical-user interface for using AMPACT and to move beyond a collection of scripts towards a robust piece of software that will be usable by researchers who are not familiar with computer programming. To finance this, the research team plans to apply for an NEH Digital Humanities Advancement grant as well as an ACLS Digital Extension grant in order to develop the AMPACT. The researchers also plan to explore ways in which this project may be integrated with a larger-scale music recording analysis project with other researchers that would be suitable for the NEH Digging into Data initiative, specifically looking at ways of linking audio analysis techniques capable of extracting performance data from recordings with other large-scale music digitization efforts that engage with scanning and analyzing centuries of musical scores. Another long-term goal is to develop tools that do not rely on a corresponding musical score, a much more complex technical problem but one that will allow for the study of the full range of real-world polyphonic recordings made since the early days of sound recording.

References

Devaney, J. and H. Léveillé Gauvin. 2017. Encoding music performance data in Humdrum and MEI. *International Journal for Digital Libraries*.

Devaney, J. and M. I. Mandel. 2017. An evaluation of score-informed methods for estimating fundamental frequency and power from polyphonic audio. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*.

Devaney, J. and H. Léveillé Gauvin. 2016. Representing and Linking Music Performance Data with Score Information. In *Proceedings of the 3rd International workshop on Digital Libraries for Musicology*.

Devaney, J. and M. I. Mandel. 2016. Score-informed estimation of performance parameters from polyphonic audio using AMPACT. In *Proceedings of the Late-Breaking Demo Session, ISMIR*.

Devaney, J. 2016. Inter- versus intra-singer similarity and variation in vocal performances. *Journal of New Music Research*. 45(3): 252-64.