

AUTOMATIC PRACTICE LOGGING: INTRODUCTION, DATASET & PRELIMINARY STUDY

R. Michael Winters, Siddharth Gururani, Alexander Lerch

Georgia Tech Center for Music Technology (GTCMT)

{mikewinters, siddgururani, alexander.lerch}@gatech.edu

ABSTRACT

Musicians spend countless hours practicing their instruments. To document and organize this time, musicians commonly use practice charts to log their practice. However, manual techniques require time, dedication, and experience to master, are prone to fallacy and omission, and ultimately can not describe the subtle variations in each repetition. This paper presents an alternative: by analyzing and classifying the audio recorded while practicing, logging could occur automatically, with levels of detail, accuracy, and ease that would not be possible otherwise. Towards this goal, we introduce the problem of Automatic Practice Logging (APL), including a discussion of the benefits and unique challenges it raises. We then describe a new dataset of over 600 annotated recordings of solo piano practice, which can be used to design and evaluate APL systems. After framing our approach to the problem, we present an algorithm designed to align short segments of practice audio with reference recordings using pitch chroma and dynamic time warping.

1. INTRODUCTION

Practice is a widespread and indispensable activity that is required of all musicians who wish to improve [5]. While a musical performance progresses through a score in linear-time and with few note-errors, practice is characterized by repetitions, pauses, mistakes, various tempi, and fragmentation. It can also take a variety of forms, including technique, improvisation, repertoire work, and sight-reading. It can occur with any musical instrument (often with many simultaneously), and can take place in a range of acoustic environments.

Within this context, we present the problem of Automatic Practice Logging (APL), which attempts to identify and characterize the content of musical practice from recorded audio during practice. For a given practice session, an APL system would output exactly what was practiced at all points in time, and describe how practice occurred.¹

¹ E.g., "Chopin's Raindrop Prelude, Op. 28, No. 15, mm. 1–26 was

By its nature, an APL system must be robust to wrong notes, pauses, repetitions, fragmentation, dynamic tempi, and other typical "errors" of practice. It should be able to operate in challenging acoustic environments, work with any instrument, and even with ensembles. Most importantly, it needs to identify *what* is being practiced and characterize *how* practice is occurring, so that it can describe and transcribe its content for a user.

In the following paper we elaborate on the subject of automatic practice logging (APL), including its benefits and challenges. We present precursors and relevant methods that have been developed in the MIR community, and which frame APL as a viable area of application. We then introduce a publicly available dataset of 34 hours of annotated piano practice including a typology for practice that informed our annotation. We conclude with a description of a preliminary algorithm capable of identifying the piece that is being practiced from short segments using pitch chroma and dynamic time warping.

2. MOTIVATION

At all skill levels, practice is key to learning music, advancing technique, and increasing expression [13]. Keeping track of the time spent practicing, or "practice logging" is an important component of practice, with many uses and benefits. Logging practice is a complex endeavor. For example, a description of practice might include amount of time spent practicing, specific pieces or repertoire that were practiced, specific sections or measure numbers, approaches to practicing, and types of practicing (e.g. technique exercises, sight-reading, improvisation, other instruments, or ensemble work). An even greater level of detail would describe how a particular section was practiced, and even the many nuances involved in each repetition. For performers, an APL system can offer unprecedented levels of detail, ease, and accuracy, not to mention additional advantages of digitization. The output of an APL system could help musicians to structure and organize the time spent practicing, to provide insight into personal improvement, and to engage in good practice habits (e.g., deliberate, goal-oriented practice [13]). For teachers and supporters, practice logs provide a window into a musician's private practice, which may foster a better understanding of improvements (or lack

practiced 11 times with a metronome gradually increasing tempo from 40–55 BPM. Mm. 19–26 were played slower on average and were characterized by fragmentation and pauses."



thereof), leading to more informed and thoughtful feedback. Researchers can benefit from detailed accounts of practice, gaining insights into performance and rehearsal strategies. For the field of Music Information Retrieval (MIR), APL offers a new and challenging area of application, which may culminate in valuable tools for researchers studying practice as well.

2.1 The Benefits of APL

Primary benefits of Automatic Practice Logging (APL) are increased levels of detail and ease of use. In repertoire practice, it is common for musicians to repeat sections of pieces many times, with the progression of these repetitions resulting in the musical development from error-ridden sight-reading to expressive performance. Marking and tallying these many repetitions manually would be impractical, and describing each repetition in terms of nuances (e.g., tempo changes, wrong/correct notes, expressive timing and intonation) would be even more so. However by using APL, repetitions could be identified and tallied automatically. Simply remembering to turn on the system and occasionally tagging audio could be the extent of user input. Once a section has been identified, a host of other MIR tools could be used to characterize and describe small variations in each repetition.

Another benefit of APL is accuracy. In addition to the relative dearth of detail that was mentioned previously, manual practice logging is plagued by the fallibility of human memory, resulting in omission and fallacy in logged practice [13]. Especially for students that are uncommitted to their instrument, manual logging may be prone to exaggeration and even deceit. By using the audio recorded directly from practice, an APL system could more accurately reflect the content of practice.

A host of other benefits would arise due to the digitization of the information. Using a digital format could lead to faster sharing of practice with teachers, who might be able to comment on practice remotely and provide support in a more continuous manner. Practice descriptions could be combined with ancillary information such as the day of the week, location of the practice, local weather, mood, and time of day, and lend itself to visualization through graphs and other data displays, assisting in review and decision making. Over time, this information might be combined and used by an intelligent “practice companion” that can encourage effective practice behaviors.

2.2 APL Challenges

Automatic practice logging, however, is not easy and a successful system must overcome a variety of challenges that are unique to audio recorded during practice. While live performances and studio recordings are almost flawless—including few (if any) wrong notes and unfolding linearly with respect to the score—the same can not be said about practice. Instead, practice is error-laden, characterized by fragmentation, wrong notes, pauses, short repetitions, erratic jumps (even to completely different pieces), and

slower, variable, and unsteady tempi. In polyphonic practice (e.g., a piano or ensemble), it is not uncommon to practice individual parts or hands separately.

Additional problems for APL arise from the fact that recordings made in a natural practice session will occur in an environment that is far from ideal. For example, metronomes, counting out-loud, humming, tapping, page-turning, and singing are common sound sources that do not arise directly from the instrument. Speech is also common in practice, and needs to be identified and removed from a search, but can also occur while the instrument is playing. Unlike recording studios and performance halls, practice environments are also subject to extraneous sound sources. These sources might include the sounds of other instruments and people, but also HVAC systems and a host of other environmental sounds. The microphone used to record practice might also be subject to bad practices such as poor placement, clipping, and sympathetic vibrations with the surface on which it was placed.

Last but not least, using APL for repertoire practice needs to address issues of audio-to-score alignment. Scores commonly include structural repetitions such as those marked explicit (e.g., repeat signs), and those occurring on a phrase level. At an even smaller time frame, it is not uncommon to have sequences of notes repeated in a row (e.g., ostinato), or short segments repeated at different parts of the piece (e.g., cadences). For a window that has many near-identical candidates in a given score, an APL system will have difficulties determining to which repeat the window belongs. This difficulty is compounded by the fact that practice is highly fragmented in time, so using longer time-frames for location cues may not be feasible.

3. RELATED WORK

Given the importance and prevalence of practice in the lives of musicians, the subject of practice has received considerable attention in the music research community [2, 13]. Important questions include the role of practice in attaining expertise [19], the effects of different types of practice [1, 6], and the best strategies for effective practice [8, 11]. However, to the best knowledge of the authors, automatically recognizing and characterizing musical practice has not specifically been addressed in MIR. It draws important parallels with many application spaces, but also offers its own unique challenges (see Sect. 2.2).

Perhaps its closest neighbor is the task of cover song detection [17], which in turn might derive methods from audio-to-audio or audio-to-score alignment and audio similarity [10]. Another possible area of interest is automatic transcription [12], and piano transcription [15] in particular for the presented dataset. In this section, techniques of cover song detection are described and compared with the unique requirements for an APL system. The cover song detection problem may be formulated as the following: Given a set of reference tracks and test tracks, identify tracks in the test set that are cover songs of a reference track. Ellis and Poliner derive a chroma-per-beat matrix representation and cross-correlate the reference and query track’s matrices

to search for sharp peaks in the correlation function that translate to a strong local alignment [7]. The chroma-per-beat helps with tempo-invariance and chroma-vectors can be circularly shifted to handle transpositions. Ravuri and Ellis make use of similar features to train a Support Vector Machine (SVM) classifier that classifies a reference/test song pair as a reference/cover song pair [16]. Serra et al. propose to extract harmonic pitch-class profile (HPCP) features from the reference and query track [18]. Dynamic Time Warping (DTW) is then used to compute the cost of alignment between the reference HPCP and query HPCP features. The DTW cost is representative of the degree to which a track is a cover of another. A system for large-scale cover-song detection is presented by Bertin-Mahieux and Ellis [4] as a modification of a landmark-based fingerprinting system [20]. The landmarks in this cover-song detection algorithm are pitch chroma-based instead of frequency-based as in the original fingerprinting algorithm. This makes the hashing key-invariant because it is possible to circular-shift the query chroma while searching for a match.

By analogy to cover song detection, repertoire practice consists of fragments of the practiced piece that should be independently identified as belonging to a particular track. Identifying the start and end times of a particular segment computationally is non-trivial, but must be the basis of a subsequence search algorithm (e.g., [9]). The subsequence search algorithm must furthermore be robust against practice artifacts such as pauses, various tempi, missed notes, short repetitions, and sporadic jumps. The cover-song detection methods described above take care of tempo invariance and algorithms for APL may leverage this for robustness against varying tempi.

Commercial products exist that focus on music practice and education, such as: SmartMusic,² Rocksmith³ and Yousician.⁴ SmartMusic is a music education software that enables teachers to enter lessons, track their students' progress and give feedback. Students also have access to pieces in the SmartMusic library. Rocksmith is an educational video game for guitar and bass that interfaces with a real instrument and helps users learn to play by choosing songs and exercises of a skill level that increases as a user progresses through the game. Yousician is a mobile application that teaches users how to play guitar, bass, ukulele and piano. It also employs tutorials to help users progress. In APL, the exercises are not predefined and an APL system should be able to detect and log a user's practice session without knowing what exercise or repertoire was practiced beforehand, making it less intrusive and more flexible.

4. THE APL DATASET

4.1 Considerations

Apart from the issues related to the recorded audio discussed in Sect. 2.2, APL needs to accommodate the many forms that practice might take. Although repertoire practice

using scores is common in the western art-music tradition, practice might also incorporate technique exercises, sight-reading, improvisation, and ensemble practice.

Bearing this framework in mind, the annotations for the dataset were informed by a typology of musical practice that frames the problem of APL in terms of two fundamental questions:

1. What type of practice occurred?
2. What was practiced?

The first question refers to the many types of practice that can occur, while the second question pertains to the actual content of practice. For a given type of practice (e.g., repertoire practice), question two can be addressed using two descriptors: *what* piece was practiced and *where* in the piece practice occurred.

To answer the first question, we organize the types of practice based upon the following basic categories: *technique*, *repertoire practice*, *sight-reading*, *improvisation*, and *ensemble work*.

'Technique' refers to the numerous fundamental repetitive patterns (e.g., scales and arpeggios) a performer would undertake. These have a pedagogical purpose, and typically involve basic musical elements, but would also include advanced technical and mental exercises like transposition and polymeters. 'Repertoire practice' refers to the repetitive practice of specific pieces of music for long-term musical goals such as public concerts and recordings. These repertoire pieces should be distinguishable from musical pieces that were practiced for a comparatively short amount of time (e.g., once or twice before moving on), which were labeled as 'sight-reading.' Although improvisation might be used as a type of technique or mental exercise, we choose to list it as a separate category given its importance in entire genres of music that is based only loosely upon a score if at all. The last category, 'ensemble work,' is meant to reflect the fact that the experience of practicing music is often shared by other performers, with their own unique instruments. However, it should be mentioned that the other items in this typology could be repeated in the ensemble work category.

4.2 Description

To begin working towards an APL system, we created a dataset of 34 hours of recorded piano practice including detailed annotations of the type of practice that was occurring, and the piece that was being played. These 34 hours of practice were chosen from a larger set of 250 hours of recordings made by one performer over the course of a year. They were targeted because they included repertoire practice that occurred in preparation for a studio-recording of a particular multi-movement piano piece: *Prokofiev's Piano Sonata No. 4 in C-minor, Op. 29*.

Recordings were made using a H4N Zoom recorder on a variety of Baby-Grand pianos in partially sound-isolated practice rooms. On each day of the recording, the microphone was placed upon the music rack of the piano, facing

² <http://www.smartmusic.com>, Date accessed: May 24, 2016.

³ <http://rocksmith.ubi.com/rocksmith/>, Date accessed: May 23, 2016.

⁴ <https://get.yousician.com/>, Date accessed: May 23, 2016.

the harp of the piano. The microphone input gain was adjusted to a level that was maximized to prevent clipping and adjusted only marginally if and only if clipping was discovered. To automatically remove silence from the recordings, an automatic recording process was used that triggered the start of a recording with signal level above a threshold SPL value. Similarly, recordings were automatically stopped when the SPL fell below a threshold, and stayed below the threshold for four seconds. This process created some tracks which were “empty” due to a false trigger. These were removed from the dataset. All recordings were made using the built-in stereo microphones. Recordings were made at 44.1kHz sampling rate and used the H4Ns built-in 96kbps MP3 encoder.

4.3 Annotation

Using this method of automatic recording, between 10 and 60 sound-files were recorded each day depending upon the length of practice, which ranged from approximately 30 minutes to 3 hours. The pieces were annotated by the performer, who by nature was the most familiar with the work and could identify and annotate their practice with the greatest speed and accuracy. The performer annotated them using Sennheiser CX 300 II earbuds at a comfortable listening volume in one to two hour-long chunks. Using VLC’s “short forward/back jump hot-key”, the performer made annotation of the piece being practiced in 10-second intervals. For each segment, the performer listened to enough audio to identify the piece being played and then skipped to the next section. In this way, if there were any changes in piece during a track, they could be identified efficiently.

Annotations were made on an online spreadsheet and exported to CSV and TSV format. The columns of the spreadsheet were titled as follows:

1. Track Name
2. Type of Practice
3. Descriptor #1 (e.g., Composer)
4. Descriptor #2 (e.g., Piece)
5. Start & End Time (if applicable)
6. Other (e.g., metronome, humming, distortion)

The track names were the auto-generated track names generated by the recorder, which include the date of the recording and the recording number. The type of practice was labeled as either repertoire, sight-reading, technique, or improvisation. The third category was used to list the composer for repertoire and sight-reading, or, for technique, was used to provide a general type (e.g., arpeggios, scales). For improvisation, this category and the next were not used. For repertoire and sight-reading, the next category was used to label the piece being played (e.g., Op. 29, Mvt. 1). For sight-reading, labeling this column was challenging as some pieces that had been played only once could not be identified by ear anymore.

Table 1. Number of files and length for major items in the APL dataset.

	# of Tracks	# of Minutes
Op. 29, Mvt. 1	74	115
Op. 29, Mvt. 2	50	61
Op. 29, Mvt. 3	215	250
Other Repertoire	94	80
Sight-Reading	45	73
Technique	106	177
Improvisation	14	17

The start and end times were used for cases when the track needed to be broken up due to the presence of other practice. In repertoire practice, this might occur when the performer suddenly switched pieces or movements without the necessary amount of silence to trigger a new recording. For these cases, a new annotation was created using the same track name as the original, but with different labels for composer and piece, and different start and end times. If the piece was kept constant throughout the track, the start and end times were not used. Last, the ‘Other’ category was used to provide annotations of atypical sounds that occurred such as humming, tapping, metronome use and practice of individual parts in an otherwise polyphonic texture. It was also used to denote tracks of special interest, such as when a score was played through without fragmentation as in a performance.

Table 1 presents the number of files and amount of time for major components of the dataset. The dataset, including the annotations and recordings have been made publicly available on Archiv.org.⁵ In the future, efforts will be directed towards extending the annotation scheme to accommodate more exact score-locations (e.g., measure numbers), adding a third question to the previous two: *How did practice occur?* Updated annotations will be kept with a version controlled repository. The database will also be expanded to include more instruments, and types of practice. Limiting factors to the growth are the creation of annotations, which require time and attention to annotate in detail. Those wishing to contribute to the database may contact the first author.

5. PRELIMINARY STUDY

5.1 Problem Formulation

As discussed in Sect. 4, we separate the APL task for repertoire practice into two primary components: 1) recognition of *which* repertoire piece is being practiced, and 2) recognition of *where* in the piece the practice is occurring. The former gives a general insight into the content of practice while the latter provides a more detailed view on the evolution of practice within the piece itself. Currently, we focus on the first component and present an algorithm that determines a matching reference track for each frame of the query track.

⁵ https://archive.org/details/Automatic_Practice_Logging. Date accessed, May 23, 2016.

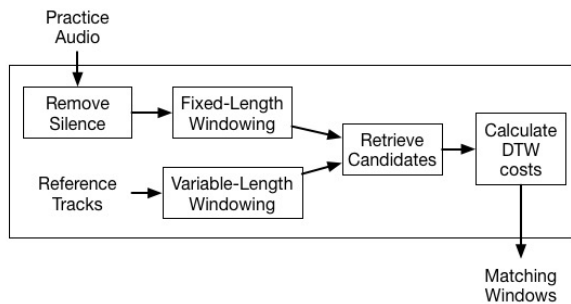


Figure 1. Block diagram of the presented system.

5.2 Overview

Although the task of automatically identifying practice audio is difficult, we present a simple approach that handles some of the major challenges of APL: pauses, fragmentation, and variable, unsteady tempi in the recorded audio.

A block diagram of the algorithm is provided in Fig. 1. We begin with a library of “reference tracks” that are full-length recordings of the repertoire being practiced. These reference tracks can, for example, be a commercial CD recording or a full recording of the student or teacher’s performance. After blocking these tracks, we compute a 12-dimensional pitch chroma vector per block. The pitch chroma captures the octave-independent pitch content of the block mapped across the 12 pitch classes [3]. We aggregate multiple pitch chromas by averaging them over larger texture windows with pre-defined lengths. Windows containing silence are dropped. The results of this computation are then one chroma vector per window, resulting in multiple chroma matrices for each of the reference tracks and window lengths.

Incoming query tracks are processed similarly. For each query texture window, a distance to all reference windows is calculated in order to select the candidates with the least distance. Subsequently, we compute the DTW cost between the selected reference texture window and the query texture window using the original (not aggregated) pitch chroma blocks. The DTW cost is the overall cost of warping the subsequence pitch chroma matrix from the query texture window to the reference pitch chroma matrix [14]. The reference track with the least DTW cost is chosen as the match for the query window.

5.3 Feature Extraction

The pitch chroma is extracted in blocks of length 4096 samples (app. 93 ms) with 50% overlap. The pitch chromas are then averaged into texture windows of 16 times the block length, with $\frac{7}{8}$ overlap between neighboring windows for the query audio. As a preprocessing step, silences are ignored. Windows containing more than 50% samples with magnitude less than a threshold are dropped and labeled as zero windows. The remaining windows are labeled non-zero windows and are used for search. The feature extraction for the reference tracks is identical, however, multiple

texture window lengths are used in order to account for different possible tempi. More specifically, lengths ranging from $N = 8, 10, 12, 14, 16, 18$ times the block size are used. Note that the length distribution is biased towards shorter windows as the query audio is more likely to be played slower than the reference. At the end of this step, we have an aggregated pitch chroma vector for the query audio and a set of aggregated pitch-chroma matrices for the reference tracks.

5.4 Candidate Track Selection

A match between query and reference is likely if the aggregated query pitch chroma matches one of the aggregated reference pitch chromas. We select a group of 15 likely track candidates for each reference track by computing the Euclidean distance between the query vector and all reference track vectors. At the end of this step, we have a pool of 15 candidates across all window lengths across for each of the reference tracks, making 45 matches total.

5.5 Track Identification

For the last step, we step back to the original short-time pitch chroma sequence. This means that our query track and reference tracks are now represented as a matrix of dimension $12 \times (2N - 1)$, where $N = 16$ for the query track and $N = \{8, 10, 12, 14, 16, 18\}$ for the reference tracks. The DTW cost is then computed for all 45 pairs of query matrix and reference matrices. For all pairs, the reference track with the texture window that has the lowest DTW cost relative to its path length and reference window size is chosen as the repertoire piece being practiced in that particular texture window of the query audio. Additional information such as the matching texture window length and matching frame are available, but not analyzed presently.

Using this sequence of steps, texture windows in the reference library will be chosen for each query texture window. These windows correspond to particular locations in the reference tracks, while the window sizes correspond to the best matching tempo. Figure 2 presents the results of running this algorithm on all of the non-zero windows one track of practiced audio, plotting the detected windows over the practiced windows. The correct track is plotted as asterisks.

6. RESULTS

To test our approach on a large body of practice audio, we ran our algorithm on 50,000 windows of practice from the APL dataset. As our approach is targeted towards repertoire practice, we chose recordings from a piece the performer was working towards at that time, namely Prokofiev’s *Piano Soanta No. 4 in C-Minor, Op. 29*. The piece is a three-movement work including sections of various tempi, note-densities, tonal strengths and key centers, and at various levels of completion and familiarity.

To create a roughly even distribution of query windows across the three reference tracks, particular days in the APL dataset were chosen for analysis. The APL dataset includes

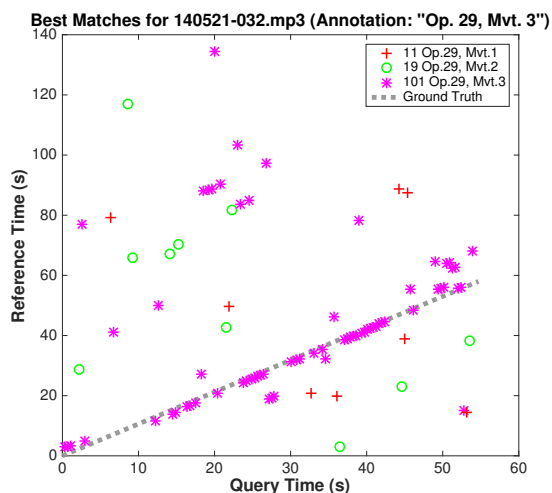


Figure 2. Detected piece across three reference tracks (see legend) and detected time in piece for all non-zero windows of a 60s query track of repertoire practice.

Table 2. Confusion matrix for the 50,000 windows belonging to either Mvt. 1, 2, or 3.

	Mvt. 1	Mvt. 2	Mvt. 3
Mvt. 1	0.55	0.29	0.15
Mvt. 2	0.17	0.66	0.17
Mvt. 3	0.13	0.21	0.66

a disproportionate amount of work on the third movement, so days were selected that included relatively more work on the first and second movements. These were May 5th, 7th, 11th, 14th, 15th, 21st and 22nd, 2014. Tracks annotated as ‘technique,’ ‘sight-reading,’ or ‘improvisation’ were not included. Furthermore, tracks that included annotations in the ‘Other’ category were not included as this category was used to indicate tracks with audio sources not from the instrument (e.g., metronome, humming, singing, counting, but also distortion). Last, tracks that included more than one piece being practiced, or more than one kind of practice were not included. A confusion matrix displaying the results of this test are displayed in Table 2.

7. DISCUSSION

The results demonstrate that an APL system based upon the pitch chroma of short windows of practice audio can be used to identify the piece being practiced. The results have targeted a broad level of description, specifically the correct identification of the piece being practiced. However, further levels of detail are provided by this approach: namely a specific location in the reference track, the window size corresponding to the match, and the amount of dissimilarity (cost) for that combination.

Although the present results are far from perfect, it is important to remember that APL by nature identifies audio that is error-laden. Pauses, short-repetitions, wrong-notes and general fragmentation make correct identification of

every window a hard challenge. Instead, it is more practical for APL to use some form of monotonicity constraint. In the example of the present algorithm, a single window that is identified as Op. 29, Mvt. 2 that is surrounded by windows that are classified as belonging to a particular section in Op. 29, Mvt. 1, likely belongs to Mvt. 1. One could also favor windows that are in a sequence in the reference tracks, or have the same window length (same relative tempo). It is interesting to note that for the present results, a simple majority vote for non-zero windows across each query track could be used to remove chosen candidates from minority identifications and replace them with candidates from the majority identification. Even this course interpolation would lead to dramatic improvements in the confusion matrix of Table 2.

It is also necessary to acknowledge the importance of reference tracks in APL. In the present case, we make use of “full versions” of the repertoire pieces played by the same performer in a similar recording environment as the practiced audio. However, in general, complete versions of repertoire pieces are not available until the performer has already practiced them significantly. Although one could choose to use studio recordings as reference, recording and production artifacts like microphone placement, SNR, spectral and temporal effects and reverberation may leave traces in the feature vector that can make correct identification more difficult. Furthermore, each performer and performance is subject to subtle timing deviations, which may create a systematic deviation when trying to match with those of the user. An alternative might be to use audio from a reference MIDI score, which would provide the highest amount of control and the additional benefit of measure numbers for matches. Generating “reference” material from the performer themselves however remains an interesting prospect for APL, which might have the most use when a score is not available (e.g., improvisation, new music).

8. CONCLUSION

This paper has presented current efforts towards Automatic Practice Logging (APL) including an annotated dataset, and a preliminary approach to identification. Practice is a ubiquitous component of music, and despite challenges, there are many benefits to logging its content automatically. Practice occurs in many forms, and for the purpose of annotating it, we presented a typology and annotation framework that can be generalized to many instruments, musicians and types of practice. We presented a preliminary approach that searches a reference library using pitch-chroma computed on very short segments, and uses dynamic-time warping as an additional step to find the best match from a collection of candidates. Incorporating additional local assumptions such as score-continuity and constant tempo might lead to increased performance in the future, but one should be mindful that practice is globally fragmented and variable in tempo. We hope that this work will encourage others to explore APL as an interesting and valuable topic for MIR.

9. REFERENCES

- [1] N. Barry. The effects of different practice techniques upon technical accuracy and musicality in student instrumental music performance. *Research Perspectives in Music Education*, 44(1):4–8, 1990.
- [2] N. Barry and S. Hallam. Practice. In R. Parncutt and G. E. McPherson, editors, *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*, pages 151–66. Oxford University Press, New York, NY, 2001.
- [3] M. A. Bartsch and G. H. Wakefield. To catch a chorus: Using chroma-based representations for audio thumbnailing. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 15–18, New Paltz, NY, October 2001.
- [4] T. Bertin-Mahieux and D. P. W. Ellis. Large-scale cover song recognition using hashed chroma landmarks. In *Proceedings of the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 117–20, New Paltz, NY, October 2011.
- [5] R. Chaffin and A. F. Lemieux. General perspectives on achieving musical excellence. In A. Williamson, editor, *Musical Excellence: Strategies and Techniques to Enhance Performance*, pages 19–40. Oxford University Press, New York, NY, 2004.
- [6] J. E. Driskell, C. Copper, and A. Moran. Does mental practice enhance performance? *Journal of Applied Psychology*, 79(4):481–92, 1994.
- [7] D. P. W. Ellis and G. E. Poliner. Identifying ‘cover songs’ with chroma features and dynamic programming beat tracking. In *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1429–32, Honolulu, HI, 2007.
- [8] K. A. Ericsson, R. T. Krampe, and C. Tesch-Rmer. The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3):363–406, 1993.
- [9] A. Guo and H. Siegelmann. Time-warped longest common subsequence algorithm for music retrieval. In *Proceedings of the 5th International Conference on Music Information Retrieval*, pages 258–61, Barcelona, Spain, 2004.
- [10] N. Hu, R. B. Dannenberg, and George Tzanetakis. Polyphonic audio matching and alignment for music retrieval. In *Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 185–8, New Paltz, NY, 2003.
- [11] H. Jørgensen. Strategies for individual practice. In Aaron Williamon, editor, *Musical Excellence: Strategies and Techniques to Enhance Performance*, pages 85–104. Oxford University Press, New York, NY, 2004.
- [12] A. P. Klapuri. Automatic music transcription as we know it today. *Journal of New Music Research*, 33(3):269–82, 2004.
- [13] A. C. Lehmann, J. A. Sloboda, and R. H. Woody. *Psychology for Musicians: Understanding and Acquiring the Skills*, chapter 4, pages 61–81. Oxford University Press, New York, NY, 2007.
- [14] M. Müller. *Information Retrieval for Music and Motion*, chapter 4, pages 69–84. Springer, Berlin, Germany, 2007.
- [15] C. Raphael. Automatic transcription of piano. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, pages 15–19, Paris, France, October 2002.
- [16] S. Ravuri and D. P. W. Ellis. Cover song detection: from high scores to general classification. In *Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing*, pages 65–8, Dallas, TX, 2010.
- [17] J. Serrà, E. Gómez, and P. Herrera. Audio cover song identification and similarity: Background, approaches, evaluation, and beyond. In Z. W. Raś and A. A. Wiczorkowska, editors, *Advances in Music Information Retrieval*, pages 307–32. Springer, Berlin, Germany, 2010.
- [18] J. Serrà, E. Gómez, P. Herrera, and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(6):1138–51, 2008.
- [19] J. A. Sloboda, J. W. Davidson, M. J. A. Howe, and D. G. Moore. The role of practice in the development of performing musicians. *British Journal of Psychology*, 87(2):287–309, 1996.
- [20] A. L.-C. Wang. An industrial strength audio search algorithm. In *Proceedings of the 4th International Conference on Music Information Retrieval*, pages 7–13, Baltimore, MD, 2003.