COMPARATIVE ANALYSIS OF ORCHESTRAL PERFORMANCE RECORDINGS: AN IMAGE-BASED APPROACH

Cynthia C. S. Liem
Delft University of Technology
Multimedia Computing Group
c.c.s.liem@tudelft.nl

Alan Hanjalic
Delft University of Technology
Multimedia Computing Group
a.hanjalic@tudelft.nl

ABSTRACT

Traditionally, the computer-assisted comparison of multiple performances of the same piece focused on performances on single instruments. Due to data availability, there also has been a strong bias towards analyzing piano performances, in which local timing, dynamics and articulation are important expressive performance features. In this paper, we consider the problem of analyzing multiple performances of the same symphonic piece, performed by different orchestras and different conductors. While differences between interpretations in this genre may include commonly studied features on timing, dynamics and articulation, the timbre of the orchestra and choices of balance within the ensemble are other important aspects distinguishing different orchestral interpretations from one another. While it is hard to model these higher-level aspects as explicit audio features, they can usually be noted visually in spectrogram plots. We therefore propose a method to compare orchestra performances by examining visual spectrogram characteristics. Inspired by eigenfaces in human face recognition, we apply Principal Components Analysis on synchronized performance fragments to localize areas of cross-performance variation in time and frequency. We discuss how this information can be used to examine performer differences, and how beyond pairwise comparison, relative differences can be studied between multiple performances in a corpus at once.

1. INTRODUCTION

A written notation is not the final, ultimate representation of music. As Babbitt proposed, music can be represented in the acoustic (physical), auditory (perceived) and graphemic (notated) domain, and as Wiggins noted, in each of these, projections are observed of the abstract and intangible concept of 'music' [29]. In classical music, composers usually write down a notated score. Subsequently, in performance, multiple different musicians will present their own artistic reading and interpretation of it.

Nowadays, increasing amounts of digital music recordings become available. As a consequence, for musical pieces, an increasing amount of (different) recorded performances can be found. Therefore, in terms of data availability, increasing opportunities emerge to study and compare different recordings of the same piece. Beyond the Music Information Retrieval (Music-IR) domain, this can serve long-term interests in psychology and cognition on processes and manifestations of expressive playing (e.g. [6, 21, 26]), while the analysis of performance styles and schools also is of interest to musicologists [5, 16].

In this paper, we mostly are interested in the analysis of multiple performances of the same piece from a search engine and archive exploration perspective. If one is looking for a piece and is confronted with multiple alternative performances, how can technology assist in giving overviews of main differences between available performances? Given a corpus, are certain performances very similar or dissimilar to one another?

In contrast to common approaches in automated analysis of multiple performances, we will not depart from explicit modeling of performance parameters from a signal. Instead, we take a more holistic approach, proposing to consider spectrogram images. This choice has two reasons: first of all, we are particularly interested in finding methods for comparative analysis of orchestra recordings. We conjecture that the richness of orchestra sounds is better captured in spectrogram images than in mid-level audio features. Secondly, as we will demonstrate in this paper, we believe spectrogram images offer interpretable insights into performance nuances.

After discussing the state of the art in performance analysis in Section 2, in Section 3, we will further motivate our choice to compare performances through visual comparison of spectrogram images. Subsequently, Section 4 details our chosen comparison method, after which we present the experimental setup for this paper in Section 5. We will then illustrate our approach and its outcomes through a case study in Section 6, with a detailed discussion of selected musically meaningful examples. This is followed by a discussion on how our method can assist corpus-wide clustering of performances in Section 7, after which the Conclusion will be presented.

2. STATE-OF-THE-ART REVIEW

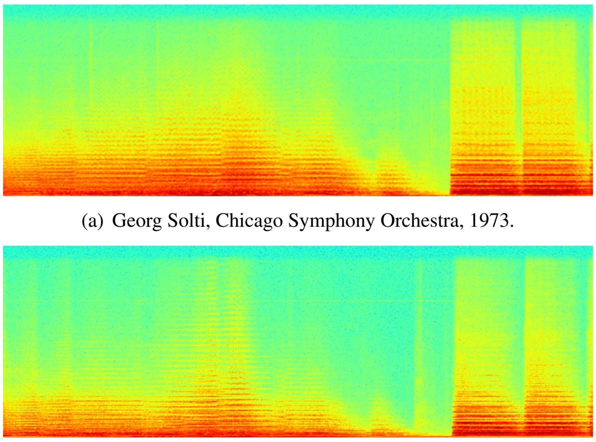
A lot of work exists on analyzing musical performance expressivity. In several cases, establishing models for computer-rendered expressive performances was the ultimate goal (e.g. see [10,11]). Other works focused on identifying reasons behind performance expressivity, including lower-level perceptual processes [21]; varying score editions, individual treatments of ornamentation and pedaling, and music-theoretic notions of expectation and tension-relaxation [20]; generative rules, emotional expression, random variability, motion principles and stylistic unexpectedness [14]; and musical structure [9, 13, 20]. Historically, the analysis of musical performance strongly focused on expressivity in piano playing (e.g. [6, 20–22]). The few exceptions to this rule focused on violin performance (e.g. [4]), movement in clarinet players (e.g. [8]), and performance of trained and untrained singers (e.g. [7], inspired by [26]), but to the best of our knowledge, no systematic comparative studies have been performed considering larger ensembles.

A reason for the general bias towards piano performance may be that digital player pianos (e.g. the Yamaha Disklavier) allow a very precise recording of mechanical performance parameters. When such parameters are available, inter-onset-intervals (IOIs), expressing the time between subsequent onsets, are frequently studied. Otherwise, performance parameters have to be extracted or annotated from the audio signal. As a piano has a discrete pitch set and percussive mechanics, expressive possibilities for a pianist are restricted to timing, dynamics and articulation. As a consequence, audio-based performance analysis methods usually focus on local timing and dynamics. Since it is not trivial to find a suitable time unit for which these parameters should be extracted, supervised or semi-supervised methods often have been applied to obtain this, e.g. by departing from manually annotating beat labels (e.g. [24, 25]). However, it is hard (if not infeasible) to realize such a (semi-)supervised approach at scale. Therefore, while a very large corpus of recorded Chopin Mazurkas exists, in practice only the Mazurkas for which annotated beat information exists have been studied in further depth (e.g. [15, 19, 24, 25]).

Alternatively, in [17, 18] an unsupervised approach for comparing Mazurka recordings was proposed which does not rely on explicitly modeled higher-level performance parameters or semantic temporal units, but rather on alignment patterns from low-level short-time frame analyses. As such, this approach would be scalable to a larger corpus. Furthermore, while the choice of not adopting explicit performance parameters makes evaluation of a clear-cut ground truth less trivial, at the same time it allows for any salient variations to emerge automatically from the analysis. The work of this paper follows a similar philosophy.

3. MOTIVATION FOR SPECTROGRAM IMAGES

In this paper, we focus on the comparative analysis of orchestra recordings. An orchestra involves a mix of many instruments. Hence, the overall orchestral sound is richer than that of a piano, although individual beat placings and note onsets will be much smoother. Given the multitude of involved players, an orchestra needs guidance by a conductor. Due to this coordinated setup, there is less room for individual freedom in both local dynamics and tempo than in Romantic piano music repertoire. Thus, while local tempo deviations still occur in orchestral recordings, one cannot expect these to reflect performer individuality as strongly as for example in the case of Chopin Mazurkas.



(a) Georg Solti, Chicago Symphony Orchestra, 1973.

(b) Nikolaus Harnoncourt, Chamber Orchestra of Europe, 1990.

Figure 1. Beethoven's Eroica symphony, 2nd movement, spectrogram of bars 56-60 for two different interpretations.

At the same time, in terms of timbre, balance and phrasing articulation, a conductor has a much richer palette than isolated instruments can offer. These aspects are not trivial to explicitly model or interpret from audio signals. However, relevant information may be reflected in recording spectrograms, as illustrated in Figure 1. While it is hard to point out individual instruments, a spectrogram can visually reveal how rich the overall sound is, where signal energy is concentrated, and if there are any salient sound quality developments over time, such as vibrato notes.

Indeed, spectrograms are commonly used in audio editing tools for visualization, navigation and analysis purposes. In an ethnographic study of musicologists studying historical recordings, it further was shown that examination of the spectrogram helped musicologists in discovering and listening to performance nuances [1]. Therefore, regarding potential end users of performance analysis and exploration tools, spectrogram images may be more familiar and interpretable than reduced mid-level representations such as chroma.

4. METHOD

Our proposed analysis method for spectrogram images is inspired by the eigenfaces method of Turk and Pentland [27], which was originally proposed in the context of human face recognition. Since human faces share many common features, by applying Principal Components Analysis (PCA) on a dataset of aligned facial images, a set of basis images ('eigenfaces') can be found, explaining most of the variability found in the face dataset. While PCA has previously been applied as a tool in musical performance analysis [23], this analysis was performed on annotation-intensive IOI data. In contrast, our analysis considers information which only requires alignment of different fragments (as will described in Section 5), but no further manual annotation effort.

We apply the same principle to a set of N spectrogram images for a time-aligned music fragment, as represented by N different recordings. Each spectrogram image $\mathbf{x}$ is $(i \cdot j)$ pixels in size. We treat each pixel in the image as a feature; as such, $\mathbf{x}$ is a vector of length $i \cdot j$. We collect all spectrogram images in an $(N \times (i \cdot j))$ matrix $\mathbf{X}$.

By applying PCA, we decompose $\mathbf{X}$ into an $(N \times N)$ matrix of principal component loadings $\mathbf{W}$ and an $((i \cdot j) \times N)$ matrix of principal components scores $\mathbf{T}$. $\mathbf{X}$ can be reconstructed by performing $\mathbf{X} = \mathbf{T} \cdot \mathbf{W}^T$.

Since the PCA is constructed such that principal components are ordered in descending order of variance, dimension reduction can be applied by not using the full $\mathbf{T}$ and $\mathbf{W}$, but only the first L columns of both.

The component scores in $\mathbf{T}$ can now be interpreted and visualized as basis images, each representing a linear component explaining part of the variability in the dataset.

5. EXPERIMENTAL SETUP

Unfortunately, no standardized corpora on multiple performances of the same orchestra piece exist. [1] Furthermore, no clear-cut ground truth exists of performance similarity. We therefore consider a dataset collected for the PHENICX [2] project, consisting of 24 full-length recordings of Beethoven's Eroica symphony, as well as 7 recordings of the Alpensinfonie by Richard Strauss. In the Beethoven dataset, 18 different conductors and 10 orchestras are featured (with a major role for the recording catalogue of the Royal Concertgebouw Orchestra (RCO)), meaning that the same conductor may conduct multiple orchestras, or even the same orchestra at different recording moments. While metadata and audio content are not fully identical, in two cases in the dataset (Harnoncourt, Chamber Orchestra of Europe (COE) 1990 and 1991; Haitink, London Symphony Orchestra (LSO) 2005 ($\times$ 2)), there are suspicions that these near-duplicates pairs consider the same original recording. In the Strauss dataset, 6 conductors and 6 orchestras are featured: Haitink conducts both the RCO and LSO, and the RCO is represented once more with Mariss Jansons as conductor. The oldest (Mengelberg, RCO, 1940) and newest (Fischer, RCO, 2013) recordings are both featured in the Beethoven dataset.

We will demonstrate insights from the PCA spectrogram analysis in two ways: (1) by highlighting several analysis examples in detail in Section 6, based on manual selection of musically relevant fragments and (2) by discussing generalization opportunities in Section 7, based on aggregation of 4-bar analysis frames.

[1] While a dataset of orchestral recordings with multiple renditions of the same piece was used in [2], these recordings are not publicly available.

[2] http://phenicx.upf.edu

Figure 2. Eroica 1st movement, score bars 3-10.

In both cases, a similar strategy is taken: first, a musical fragment is designated, for which all recordings of the piece should be aligned. Alignment is performed automatically using the method described in [12]. Then, the audio fragments, which are all sampled at Fs = 44.1 kHz, are analyzed using a Hann window of 1024 samples and a hop size of 512, and the corresponding magnitude spectrum is computed using the Essentia framework [3]. Combining the spectra for all frames results in a spectrogram image. To ensure that all images have equal dimensions, a constant heigth of 500 pixels is imposed, and the longest fragment in terms of time determines a fixed width of the image, to which all other spectrograms are scaled accordingly. While all recordings are offered at 44.1 kHz, the original recordings sometimes were performed at a lower sampling rate (particularly in more historical recordings). Therefore, a sharp energy cut-off may exist in the higher frequency zones, and for analysis, we try to avoid this as much as possible by only considering the lower 90% of the image. In general, by using raw spectrogram images, a risk is that recording quality is reflected in this spectrum; nonetheless, in the next sections we will discuss how musically relevant information can still be inferred.

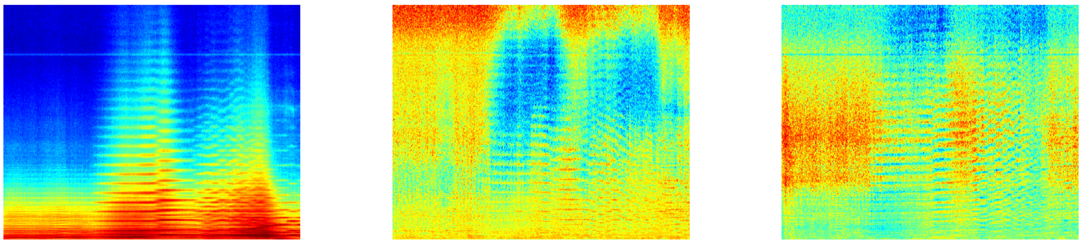
6. CASE STUDY

In this case study, to illustrate the information revealed by PCA analysis, we will look in detail at information obtained on two selected fragments: the start of the first movement of the Eroica symphony, first theme (bars 3-15), and the 'maggiore' part of the Eroica symphony, second movement (bars 69-104).

6.1 Eroica first movement, bars 3-15

A score fragment for bars 3-10 of the first movement of the Eroica is given in Figure 2. In our case, we consider the full phrase up to bar 15 in our analysis.

The first three basis images (component scores) resulting from PCA analysis are shown in Figure 3. The first component of the PCA analysis gives a smoothed 'basic' performance version of the fragment. For this very general component, it is rather hard to truly contrast performances. However, a more interesting mapping can be done in higher-order components. As an example, Figure 4 dis-



(a) First component (b) Second component (c) Third component

Figure 3. Eroica, 1st movement, 1st theme start (bars 3-15); first three principal component basis images.

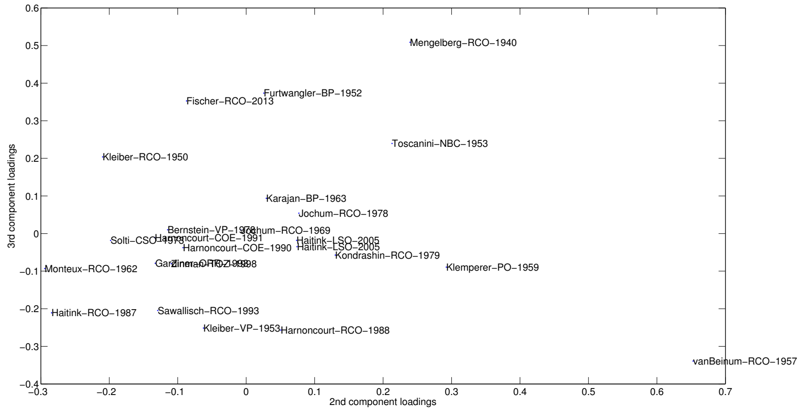
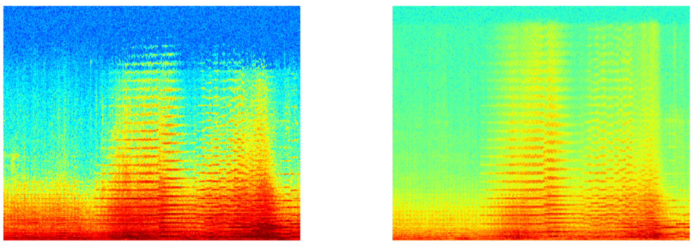


Figure 4. 2nd and 3rd PCA component scatter plot for Eroica 1st movement, bars 3-15.



(a) Fisher, RCO, 2013 (b) Haitink, RCO, 1987

Figure 5. Spectrogram image examples for Fisher and Haitink interpretations of Eroica 1st movement, bars 3-15.

plays a scatter plot of the second and third principal component loadings for this fragment.

While as expected, several historical (and acoustically noisy) recordings cause outliers, by comparing the component scores and loadings to corresponding data samples, we still note interpretable differences. For example, the RCO recordings of Fischer and Haitink, of which respective spectrogram images for the excerpt are shown in Figure 5, have contrasting loadings on the third PCA component. Judging from the principal component image in Figure 3, this component indicates variability at the start of the fragment (when the celli play), and in between the fragments highlighted by the second component; more specifically, a variability hotspot occurs at the sforzato in bar 10. When contrasting two opposite examplars in terms of scores, such as Fischer and Haitink, it can be heard that in the opening, Haitink emphasizes the lower strings more strongly than Fischer, while at the sforzato, Haitink strongly emphasizes the high strings, and lets the sound develop over the a-flat played by violin 1 in bar 10. Fischer maintains a 'tighter' sound over this sforzato.

6.2 Eroica second movement, maggiore

To illustrate findings on another manually selected (and slightly longer) relevant fragment, we also consider the

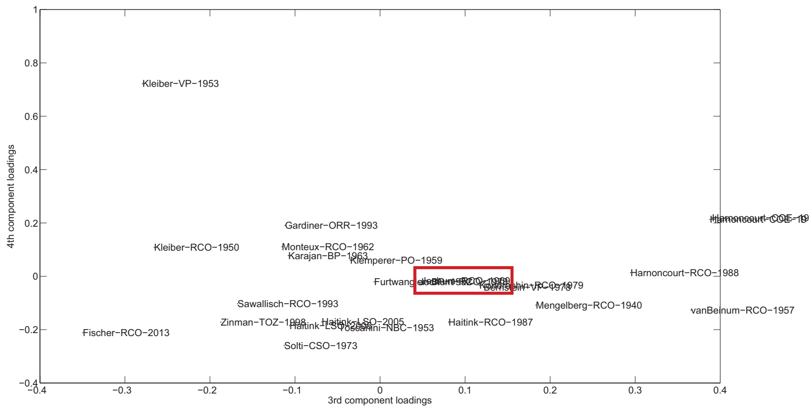


Figure 6. 3rd and 4th PCA component scatter plot for Eroica 2nd movement, maggiore. Jochum's 1969 and 1978 recordings occur within the marked rectangular border.

'maggiore' part of the second movement of the Eroica. Analyses of scatter plots and component images show that the second principal component is affected by historical recording artefacts. However, this is less so for the third and fourth component, of which the scatter plot is displayed in Figure 6. It can be seen that the suspected near-duplicates of Harnoncourt's two COE recordings have near-identical loadings on these components. Next to this, another strong similarity is noted between the recordings of Jochum with the RCO in 1969 and 1978. While these both recordings acoustically are clearly different and also seem to be explicitly different interpretations, there still are consistencies in Jochum's work with the same orchestra for these two recordings.

7. CORPUS-WIDE CLUSTERING

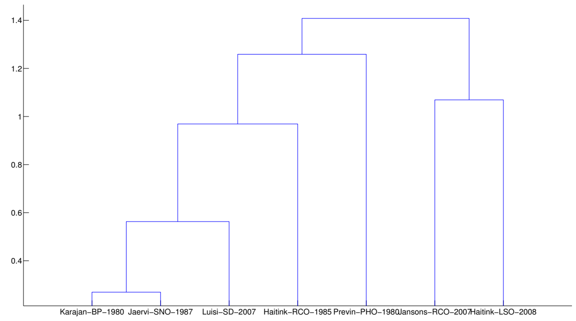
As demonstrated in the previous section, PCA analysis can be used as an exploratory tool to reveal differences between selected fragments in recordings. However, selecting incidental manual examples will not yet allow for scalable analysis of information over the full timeline of a piece. To do this, instead of pre-selecting designated fragments, we perform a 4-bar sliding window PCA analysis on full synchronized recordings, where bar boundaries are obtained through the score-to-performance mapping obtained in the alignment procedure. Instead of examining individual component images, in each 4-bar analysis frame, we consider vectors of component loadings for the minimum amount of components required to explain 95% of the variance observed. From these component loading vectors, we compute the Euclidean distance between recordings within a frame, and aggregate these at the recording track level. [3]

[3] Note that component loadings obtained for different frames cannot be directly averaged, as the components are different per frame. However, observed distances between recordings still remain valid and can be aggregated.

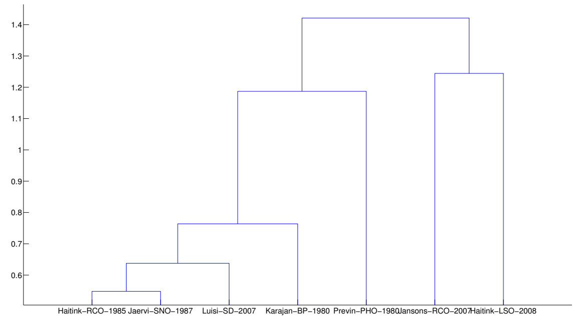
Based on distances found between performances, clustering can be performed. This reveals whether stable performer clusters can found for different movements within a piece, and to what extent clusterings found in local fragments match those found for a full piece.

Regarding the first question, for each of the Eroica movements, we calculated the average between-performer distances per movement, and then made 5 clusters of performers based on Ward's linkage method [28]. While space does not allow a full cluster result report, several clusters co-occur consistently:

- The two Harnoncourt COE recordings consistently form a separate cluster. These are highly likely to be duplicate recordings.
- Haitink's two LSO recordings also consistently co-occur, and like Harnoncourt are highly likely to be duplicate recordings. However, Bernstein's 1978 Vienna Philharmonic recording co-occurs with these two Haitink recordings in the first three Eroica movements, and thus may be similar in terms of interpretation. It is striking that Haitink's 1987 recording with the RCO never co-occurs in this cluster.
- In the first three movements, a consistent cluster occurs with recordings by Klemperer (Philharmonia Orchestra, 1959), Toscanini (NBC Symphony Orchestra, 1953) and Van Beinum (RCO, 1957). While this may be due to recording artefacts, other historical recordings (e.g. Kleiber, RCO 1950 / Vienna Philharmonic 1953) do not co-occur.
- Surprisingly, Gardiner's historically informed recording with the Orchestre Révolutionaire et Romantique (1993) clusters with Kleiber's 1950 RCO recording for the first and last movement of the Eroica. Upon closer listening, Gardiner's choice of concert pitch matches the pitch of Kleiber's recording, and the sound qualities of the orchestras



(a) ‘Sonnenaufgang’ fragment (bars 46-63).



(b) Average over full Alpensinfonie.

Figure 7. Dendrogram images for performer distances in the Alpensinfonie.

are indeed similar (although in case of Kleiber, this is caused by recording artefacts).

- The 1969 and 1978 Jochum recordings with the RCO always co-occur, though in the largest cluster of recordings. As such, they are similar, but no clear outlier pair compared to the rest of the corpus.

Regarding consistent clusterings over the course of a piece, we further illustrate an interesting finding from the Alpensinfonie, in which we compare a clustering obtained on 18 bars from the ‘Sonnenaufgang’ movement to the clustering obtained for average distances over the full piece, as visualized in the form of dendrograms in Figure 7. As can be noted, the clusterings are very close, with the only difference that within the ‘Sonnenaufgang’ movement, Karajan’s interpretation is unusually close to Järvi’s interpretation, while Haitink’s interpretation is unusually different.

8. CONCLUSION

In this paper, we proposed to analyze differences between orchestral performance recordings through PCA analysis of spectrogram images. As we showed, PCA analysis is capable of visualizing areas of spectral variation between recordings. It can be applied in a sliding window setup to assess differences between performers over the timeline of a piece, and findings can be aggregated over interpretations of multiple movements. While spectrograms inevitably have sensitivity to recording artefacts, we showed that near-duplicate recordings in the corpus could be identified, and historical recordings in the corpus do not consistently form outliers in the different analyses.

While certain interesting co-occurrences were found between recordings, no conclusive evidence was found regarding consistent clustering of the same conductor with different orchestras, or the same orchestra with different conductors. This can either be due to interference from artefacts and different recording setups, but at the same time may suggest that different conductors work differently with different orchestras.

Several directions of future work can be identified. First of all, further refinement regarding the generation and analysis of the spectrogram images should be performed. At the moment, given the linear way of plotting and high sample rate, the plain spectrogram may be biased towards higher-frequency components, and risks to be influenced by sharp frequency cut-offs from lower original recording sample rates.

Furthermore, it would be interesting to study more deeply if visual inspection of spectrograms can indeed assist people in becoming more actively aware of performance differences. While the spectrogram images are expected to already be understandable to potential end-users, appropriate techniques should still be found for visualizing differences between multiple performers in a corpus. In the current paper, this was done with scatter plots and dendrograms, but for non-technical end-users, more intuitive and less mathematically-looking visualizations may be more appropriate.

One concern that may come up with respect to our work, is that it may be hard to fully associate our reported findings to expressive performance. As indicated, recording artefacts are superimposed on the signal, and effects of different halls and choices of orchestra instruments and concert pitch may further influence acoustic characteristics, which will in turn influence our analysis. Furthermore, since we are dealing with commercial recordings, we are dealing with produced end results which may have been formed out of multiple takes, and as such do not reflect ‘spontaneous’ performance.

However, our main interest is not in analyzing performance expressivity per se, but in providing novel ways for archive and search engine exploration, and making general sense of larger volumes of unannotated performance recordings. In such settings, the data under study will mostly be produced recordings with the above characteristics. For this, we believe our approach is useful and appropriate, offering interesting application opportunities.

Acknowledgements: The research leading to these results has received funding from the European Union Seventh Framework Programme FP7 / 2007–2013 through the PHENICX project under Grant Agreement no. 601166.

9. REFERENCES

[1] M. Barthet and S. Dixon. Ethnographic observations of musicologists at the British Library: implications for Music Information Retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, USA, 2011.

[2] J. P. Bello. Measuring structural similarity in music. *IEEE Transactions on Audio, Speech and Language Processing*, 19(7):2013–2025, 2011.

[3] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra. ESSENTIA: an Audio Analysis Library for Music Information Retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 493–498, 2013.

[4] E. Cheng and E. Chew. Quantitative Analysis of Phrasing Strategies in Expressive Performance: Computational Methods and Analysis of Performances of Unaccompanied Bach for Solo Violin. *Journal of New Music Research*, 37:325–338, December 2008.

[5] N. Cook. Towards the compleat musicologist? In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR) [invited talk]*, London, UK, 2005.

[6] P. Desain and H. Honing. Does expressive timing in music performance scale proportionally with tempo? *Psychological Research*, 56(4):285–292, July 1994.

[7] J. Devaney, M. I. Mandel, D. P. W. Ellis, and I. Fujinaga. Automatically extracting performance data from recordings of trained singers. *Psychomusicology: Music, Mind & Brain*, 21:108–136, 2011.

[8] M. M. Wanderley E. C. F. Teixeira, M. A. Loureiro and H. C. Yehia. Motion Analysis of Clarinet Performers. *Journal of New Music Research*, July 2014.

[9] A. Friberg and J. Sundberg. Does music performance allude to locomotion? A model of final *ritardandi* derived from measurements of stopping runners. *Journal of the Acoustic Society of America*, 105(3):1469–1484, March 1999.

[10] W. Goebl, S. Dixon, G. De Poli, A. Friberg, R. Bresin, and G. Widmer. “Sense” in expressive music performance: Data acquisition, computational studies, and models. In P. Polotti and D. Rocchesso, editors, *Sound to sense, sense to sound: a state of the art in sound and music computing*. Logos Verlag, 2007.

[11] W. Goebl and G. Widmer. On the use of computational methods for expressive music performance. In T. T. Crawford and L. Gibson, editors, *Modern Methods for Musicology: Prospects, Proposals and Realities*, Digital Research in the Arts and Humanities, pages 93–113. Ashgate, 2009.

[12] M. Grachten, M. Gasser, A. Arzt, and G. Widmer. Automatic Alignment of Music Performances with Structural Differences. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 607–612, 2013.

[13] M. Grachten and G. Widmer. Who is who in the end? Recognizing pianists by their final ritardandi. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, October 2009.

[14] P. N. Juslin. Five facets of musical expression: a psychologist’s perspective on music performance. *Psychology of Music*, 31(3):273–302, July 2003.

[15] K. Kosta, O. F. Bandtlow, and E. Chew. Practical implications of dynamic markings in the score: Is piano always piano? In *Proceedings of the 53rd International AES Conference on Semantic Audio*, London, UK, January 2014.

[16] E. Liebman, E. Ornoy, and B. Chor. A Phylogenetic Approach to Music Performance Analysis. *Journal of New Music Research*, 41:195–222, June 2012.

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

[18] C. C. S. Liem, A. Hanjalic, and C. S. Sapp. Expressivity in musical timing in relation to musical structure and interpretation: A cross-performance, audio-based approach. In *Proceedings of the 42nd International AES Conference on Semantic Audio*, pages 255–264, Ilmenau, Germany, July 2011.

[19] M. Müller, P. Grosche, and C. S. Sapp. What makes beat tracking difficult? a case study on chopin mazurkas. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, The Netherlands, August 2010.

[20] C. Palmer. Anatomy of a performance: Sources of musical expression. *Music Perception*, 13:433–453, Spring 1996.

[21] A. Penel and X. Drake. Sources of timing variations in music performance: a psychological segmentation model. *Psychological Research*, 61(1):12–32, March 1998.

[22] B. Repp. A microcosm of musical expression. I. Quantitative analysis of pianist’s timing in the initial measures of Chopin’s Etude in E major. *Journal of the Acoustic Society of America*, 104(2):1085–1100, August 1998.

[23] B. Repp. A microcosm of musical expression. I. Quantitative analysis of pianist’s timing in the initial measures of Chopin’s Etude in E major. *Journal of the Acoustic Society of America*, 104(2):1085–1100, August 1998.

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

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

[26] C. Seashore. *Psychology of music*. University of Iowa Press, Iowa City, 1938.

[27] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Maui, Hawaii, USA, June 1991.

[28] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[29] G. A. Wiggins. Computer-representation of music in the research environment. In T. T. Crawford and L. Gibson, editors, *Modern Methods for Musicology: Prospects, Proposals and Realities*, Digital Research in the Arts and Humanities, pages 7–22. Ashgate, Aldershot, UK, 2009.