EXPLORING GESTURALITY IN MUSIC PERFORMANCE

Jan C. Schacher
Institute for Computer Music
and Sound Technology ICST
Zurich University of the Arts
jan.schacher@zhdk.c

Daniel Bisig
Institute for Computer Music
and Sound Technology ICST
Zurich University of the Arts
daniel.bisig@zhdk.ch

Patrick Neff
University of Zurich
Department of Psychology
patrick.neff@uhz.ch

ABSTRACT

Perception of gesturality in music performance is a multimodal phenomenon and is carried by the differentiation of salient features in movement as well as sound. In a mix of quantitative and qualitative methods we collect sound and motion data, Laban effort qualifiers, and in a survey with selected participants subjective ratings and categorisations. The analysis aims at uncovering correspondences in the multi-modal information, using comparative processes to find similarity/differences in movement, sound as well categorical data. The resulting insights aim primarily at developing tools for automated gestural analysis that can be used both for musical research and to control interactive systems in live performance.

1. INTRODUCTION

By taking instrumental performance of a canonical contemporary piece as study object, this investigation aims at understanding which elements contribute to the multimodal nature of music perception and how these elements are interrelated.

This investigation is carried out in the context of a larger research project which aims to develop analytical methods for the identification of gestures in composition and performance. The project complements a strong focus on artistic practice with a cross-disciplinary approach that integrates three academic disciplines. Psychological research explores gesture categories that inform music-perception. Music Technology uses motion data to recognise and categorise gestures in an automated way. Music Analysis builds a framework for gestures classification in composition and performance.

In the context of the cross-disciplinary research project the necessary skills for the multi-methodological approach are present and in the dialogue, a perspective is developed that can bridge between academic investigations and artistic practice, in particular with the terminology and the tools that are refined throughout the process.

2. BACKGROUND

Designing research methods that bridge between objective, data-driven methods and subjective, perceptual reporting is a standard approach in social sciences, but much less so in music research. Since music is a social, cultural as much as as physical phenomenon, blending the two perspectives [1] provides an overlapping field, that potentially describes the impact and import of music in a more appropriate manner, than using exclusively either one of the methods.

By triangulating between the four positions of the theories used to structure the work-flow, the methods used to carry out the investigation, the type and nature of the data collected and the changing roles of the investigators, the validity of results increases [2] and produce the essential effects of convergence, inconsistency, and contradiction [3]. For a more in-depth overview on mixed-method research please refer to [4] and in relation to motion analysis in music refer to [5].

Music analysis methods in the domains of empirical research comprise the rich set of music information retrieval (MIR) methodologies, that originate from the need to search and identify music pieces and have a set of metrics and descriptors that are unique to a given piece [6–8]. Similarly, movement research dates back at least to the nineteenth century with the chrono-photographies by Muybridge [9], and has numeric tools for extracting significant features from, for example, motion capture data [10].

On the qualitative side, *movement analysis* is an established topic in dance-research [11, 11], but also robotics [12] and physiology and rehabilitation [13]. Here it is particularly interesting to observe that perceived movement qualities, i.e., what makes movement expressive for our perception, has been formalised and is now usable in mathematical models as well as descriptive analysis, as we will discuss further on.

An additional pole in our configuration is a *terminological* and categorisations investigation about *musical gesture* [14, 15]. This domain is informed as much by body-related dimensions, body-instrument relationships as by musical categories such as phrase, chunk, segment, or semantic unit [16]. The question of musical content and the size of units to investigate is directly related to the temporal frames that are perceived as a musical unit, albeit this highly dependent on stylistic and other contextual elements of the music investigated.

2.1 Modelling the Methods

In mixed-method research the question of the balance between the qualitative information and the quantitative data is critical. Even when deploying mainly data-driven methodologies, the decisions about what data to treat in

what way are to a certain extent subjective. In this study, we explore a mixed method work-flow that is cyclical, and fluctuates between purely quantitative data-driven analysis with mathematical methods and subjective, perceptual qualitative interpretation, based on reports and observations.

From a methodological point of view, this should not represent a problem, provided there is clear an unambiguous declaration of which element is situated in which domain. Therefore a step by step description and assignment in categories between the dimensions at hand can shed light on the validity of the methods and the extracted insights.

In this complex work-flow that straddles the divide between objective, data-driven measurements and subjective, perceptual and self-reported impressions, the relationships between the elements are not merely cause-and-effect driven, which makes understanding the interrelation between aspects more complex.

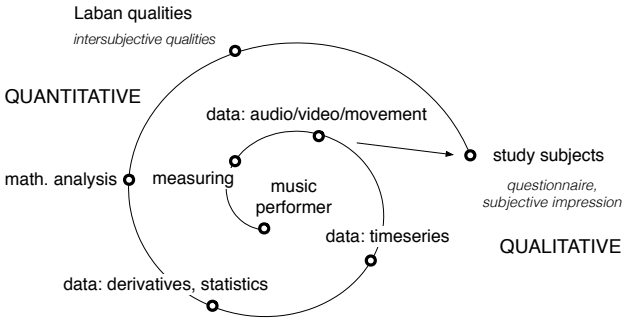

Figure 1: Cyclical model of mixed quantitative and qualitative methods. The axes stretch between qualitative and quantitative methods, and between tools and processes.

In an attempt to order and map the stages of the process and to assign the different outcomes and results to the main categories, the following cyclical map is proposed (see Fig. 1). Even if this arrangement neatly covers the steps in the work-flow, it is important to stress that this representation, as with each and every map, only represents one point of view on the configuration.

Reading the map is done by going from the centre to the periphery, from the musical performance, to measurement using technologies such as Motion Capture (MoCap) and audio video-recordings. Building on the collected data in time-series, a mathematical analysis leads to computable qualitative outcomes on the one hand and by evaluation through participants of a video of the performance to qualitative outcomes on the other hand.

3. THE CASE STUDY

A motion capture and audio-video recording session of a violoncello player performing the canonical piece 'Pression' by Helmut Lachenmann [17] was carried out (see Fig. 2). This composition is particularly interesting for a motion study on perceived effort, because the score prescribes the movements pertaining to specific playing technique rather than the sounding results. In addition, it focuses, as the name suggests, on extended playing techniques for the cello that have to do with rubbing, scratching and pressing in various ways with the hands and the bow on the strings and other parts of the cello.

Figure 2: The Musician wearing passive reflective markers used for kinematic motion capture. Note the markers on the back of the hands (RFIN/LFIN), the tip of the bow (CLTP), the elbows (RELB/LELB), and the forehead (HDFR)

3.1 Multi-modal Approach

The multi-modal, blended nature of music perception that is occurring as much on the auditory, the visual as on the kinaesthetic sensory channels, means that investigation perceptual salience or 'gesturality' needs to occur in more than one modality in order to be meaningful. The modalities that are measurable from outside the performer are more practical to use than those relating to her body's physiological data, although they may be more telling about physicality and exertion during playing. In this study we use kinematic motion data from Motion Capture in combination with audio-data for the quantitative analysis, and for the qualitative investigation we use audio and video recordings evaluated by participants.

The point of departure for this investigation is perceived performance *effort* in a combined visual auditory case. This presupposes a definitions of effort. Apart from a physical and physiological measure, the term is used in motion analysis, in particular in the Laban Effort dimensions [11, p. 77]. Although in this system the term is used in an extended sense, it is still relevant for our purposes, since it addresses the human perception of effort, rather than just the measurable physical one. As we will see, the transfer from subjective to objective evaluation of these dimension also forms part of this investigation (see 3.3).

An additional core concept we focus on is that of *gesturality*. As a high-level concept that encompasses direct perception, semantic content and other psychological, affective factors, it serves to frame the more detailed systematisation found principally in the literature n musical gesture. After having mapped out the use of the term 'gesture' in this field the core terminology was selected and used to structure the responses of subjective impressions from participants (see 3.2).

An interesting parallel in this specific case can be drawn by comparing these gestural categories to musical ones in Lachenmann's 'Klangtypen der neuen Musik' [18] and Smalleys spectro-morphological Sound-shapes [19].

3.2 Qualitative Methods

Segments were preselected independently by all members of the research team from the entire piece. The primary criteria for the preselection were diversity and exemplariness of the segments regarding the piece. Five segments made it into the final selection and serve as materials for the mixed method workflow with quantitative music and movement analysis as well as qualitative third-person subjective ratings and categorisations from watching videos of the segments. Notably, for the qualitative ratings and categorisation the segments were condensed to a single gesture between three and six seconds in length.

In this survey, participants (n=26) were instructed to rate the segments based on their impressions (using various judgements of previous work [20] complemented by a genuine judgement of general gesturality) and furthermore categorise them into the concepts and terminologies presented in the previous section [14, 15, 21–23]. Within the framework of the study, we looked at the ratings of the general gesturality parameter and by introducing two main category systems of:

functional distinction between communicative, sound producing, sound-facilitating and -accompanying nature of gestures [15], between 'ergotic' [1], epistemic [2] and semiotic [3] gestures [14] as well as between gestures 'helping' the production of melody, harmony/musical structure, timbre, sound level, rhythm and tempo [22].

morphological distinction between trajectory-, force- and pattern-based primitives [23] as well as between impulsive, sustained and iterative morphologies [15]. Beyond that, participants of the survey were also invited to leave comments to their respective choices of categorisation. These comments serve as a verbalised pool of data flanking the categorisation and are indicative of reasoning about features of the video segment which led to the choice of categories.

The data sets gathered for each segment, i.e., single gesture, enable the evaluation and discussion of the quantitative continuous sound and motion data from the perspective of momentary qualitative data.

3.3 Quantitative Methods

The quantitative analysis of the musical performance is based on the extraction of lower and higher level features from multi-modal recordings that consist of synchronised audio and motion capture data. From the audio data, core features such as loudness (RMS measure), centroid, brightness, ad flux are extracted [8]. From the motion capture data, position time series are extracted for three markers placed on the forehead, back of the left hand and back of the right hand of the musician and a single marker placed on the tip of the bow. Prior to any feature extraction, the position time series are smoothed using a running average with a time window of ten samples. The computation of movement features is done by a software that forms part of the Machine Learning Workbench software tool chain [24]. For each individual marker, three kinematic features and three Laban effort features are calculated. The kinematic features comprise the first three temporal derivatives of the position time series (velocity, acceleration, jerk) and their absolute scalar values (speed, scalar acceleration, scalar jerk). These lower level features directly represent physical properties of body movement. The Laban effort features comprise weight effort, flow effort and time effort [11]. These higher level features compute from kinematic input data movement properties that are more closely related to qualitative aspects of movement such as dynamics, energy and expressiveness than the kinematic data themselves. Weight effort indicates the forcefulness of movement and discriminates between powerful and gentle movement qualities. Time effort reflects a sense of urgency and differentiates between quick and sustained movements. Flow effort represents the continuity of movement and distinguishes between free and bounded movements.

The implementation of the Laban feature extraction functionality is based on a recently published review of algorithms for calculating expressive motion descriptors [25]. For each Laban effort calculation, the input kinematic data are aggregated over a window size of ten samples. Following the feature extraction calculations, all audio and motion feature time series are normalised over the duration of the entire recording and subsequently truncated to the duration of each of the three performance segments. All time series corresponding to the same segment are then merged into a single file, whose content is rendered via a simple graphical plotting routine. This routine superimposes equivalent motion features for all marker positions and stacks different motion features and audio features from top to bottom. This visualisation forms the basis for a visual interpretation and comparison between quantitative, qualitative and audiovisual performance data.

4. DATA ANALYSIS

For sake of brevity and clarity we chose to focus on three segments. [4] In the following section, a brief and non-exhaustive quantitative and qualitative analysis of each segment is presented.

4.1 Segment Two

This segment is characterised by two short alternating scratches and plucking on the strings below the bridge at the edge of the string-holder.

Quantitative Assessment The bowing movements manifest in the quantitative data as clear correlations among peaks in the kinaesthetic and effort curves of the right hand and peaks in the sound loudness. The brevity

[1] "In the first function, ergotic, there is ... only energy communication between the hand and the object." [14]

[2] "epistemic, is typically performed by our capacity of touch and muscular/ articulatory sensitivity" [14]

[3] "The third function, semiotic, is that of meaning, of communicative intent. It's the gestural function per se." [14]

[4] See video of the entire piece and the individual segments: `http://mgm.zhdk.ch/?p=2021`

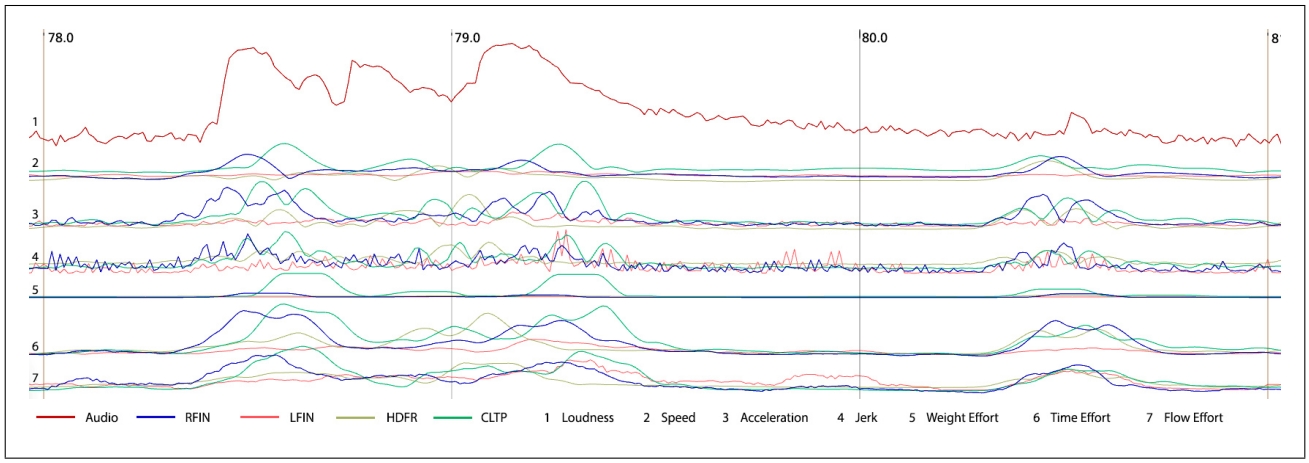

Figure 3: Continuous data plots of segment two.

of the velocity peaks in the right hand movement and their flanking by pronounced peaks in the acceleration and time effort curves are indicative of an impulsive bowing style. During bowing, peaks in the movement of the right hand and sound loudness are synchronised whereas peaks in the movement of the bow are delayed. These delayed peaks correspond to a pronounced removal of the bow after bowing. This removal exhibit a strong weight effort, time effort and flow effort, which is indicative of a forceful playing style.

Qualitative Assessment First, the survey showed a clear agreement between the participants categorising this segment as sound producing (21 of 26 participants) [15]. In the functional category system of Cadoz [14] this segment was rated ergotic in nature by the majority of participants (17/26). Looking at detailed aspects of sound production, 10 participants had the impression that the movements in this segment mainly helped the sound level whereas 7 participants chose rhythm and 5 timbre. The distribution of answers in this categorisation system [22] is therefore not showing conclusive agreement. The morphological aspects of movements [15] in this segment were clearly rated as impulsive by a large majority (23/26). Finally, concerning gestural primitives [23] participants agreed on the dominance of force-based primitives (22/26).

Regarding the rating of overall gesturality (on a scale of 1-5 with 1 being low and 5 being high in gesturality), this segment scored the lowest with a mean of 2.65.

4.2 Segment Three

In this segment, the end of the pig-sty scratching leads to the right hand slowing bowing on the bridge's face under the strings and the left hand cyclically rubbing and hitting on the fingerboard and the body of the cello.

Quantitative Assessment Percussive movements that occur when the left hand or the bow hit the instrument are characterised by a clear and strong synchronisation between peaks in movement features and peaks in sound loudness. These forceful and impulsive movements manifest as synchronisations that are visible across all movement features. The last loudness peak shows a clear correlation with the rubbing movements of the left hand but is de-correlated with the bowing movements of the right and bow tip. This deviation in feature synchronisation is a good indicator for the degree of coordination among different sound producing movements. During the rubbing movements of the left hand, the speed, acceleration, jerk and time effort curves show a repetitive pattern of pronounced peaks whose frequency and amplitude gradually increases. It is during these rubbing movements, that the movement of the forehead exhibits an interesting transition. During less effortful left hand movements, the forehead movement features are synchronised with the movement features of the right hand. As the left hand movements increase in effort, the features in the forehead movements become synchronised with the left hand movements instead. Accordingly, the amount of correlation between sound producing and non-sound producing movements can serve as an indicator for the amount of emphasis that is being put into a sound producing movement.

Qualitative Assessment As in the last segment, a clear agreement between participants was observable as this segment was categorised sound producing (21 of 26 participants). Concerning further functional categorisation, the segment was rated ergotic in nature by 13 participants whereas 8 participants chose the category 'epistemic' leaving 5 participants with a semiotic categorisation. The sound helping categorisation system again produced less agreement on a single category than other systems: 10 participants chose rhythm, 6 timbre, 4 harmony/musical structure, 3 sound level, 2 tempo and 1 'other'. Similar to that, the agreement in categorisation of morphological features was not clear-cut with half of the participants choosing impulsive (13), 7 iterative, 2 sustained and a total of 4 participants 'other'. Lastly, the absence of clear agreement continues in the categorisation of the gestural primitives with 10 participants opting for pattern-based primitives, 9 force-based, 5 trajectory-based and 2 participants 'other'.

The rating of gesturality produced an average of 4.04 with a large agreement amongst the raters rendering this segment the most gestural of all three.

4.3 Segment Four

In this segment the right hand plays a *saltando*-technique with the bow placed under the bridge. The left hand is passive and muting the strings.

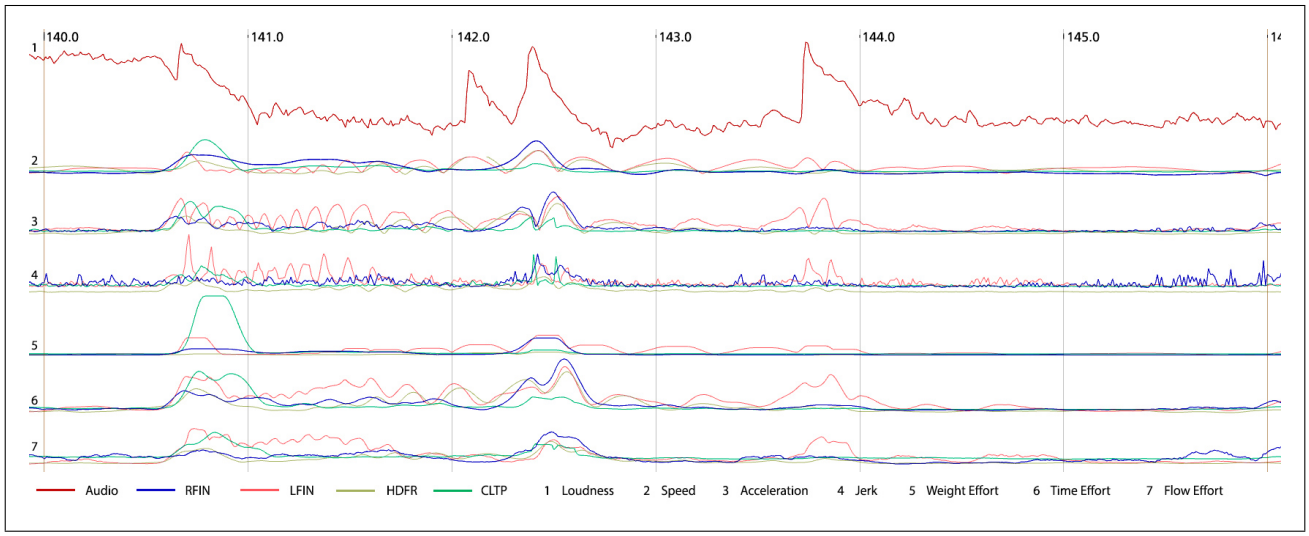

Figure 4: Continuous data plots of segment three.

Quantitative Assessment The percussive motions of the bow show repetitive peaks in jerk and elevated values in time effort and flow effort. Repetitive patterns in these motions are either actively actuated or result from a passive bouncing of the bow. These two types can be clearly distinguished based on the shape and correlation between sound loudness and the jerk and flow effort of the bow tip and right hand. For active repetitions, the features related to sound loudness and bow tip motion show constant amplitudes whereas for passive repetitions, the features decay in amplitude. Similarly, for active repetitions, the features of the right hand show high values. For passive repetitions, the feature values are low. In two cases, the head movement shows a peak in speed which precedes a percussive bow movement. These peaks operate as signifiers for the upcoming sound producing movements.

Qualitative Assessment In the last of the analysed segments, the agreement on a sound producing nature of the segment was not distinct with only 14 participants choosing this option. All other options were chosen to an equal extent with 4 participants at each of the categories of communicative, sound-facilitating and sound-accompanying. In the functional categorisation system, again 'ergotic' was favourably chosen by 12 participants whereas epistemic and semiotic scored with 6 and 8 choices respectively. Regarding the detailed aspects of sound production, 15 participants agreed on rating the movements helping the rhythm with no other category scoring more than 4. A similar conclusive agreement is observable in the morphological categorisation as the majority (17 participants) clearly rated the movements in the segment as being iterative. Finally, the overall rather distinct categorisation of this segment is also evident in the gestural primitives with 18 participants rating the segment as pattern-based.

Assessing the perceived general gesturality the segment was rated gestural with a mean of 2.92. This marks the segment as being the second most gestural of the 3 selected segments.

5. DISCUSSION

The following discussion starts with an evaluation of the similarities and differences between the quantitative and qualitative analysis results. We will argue that some of the differences result from the specifics of the experimental setup and propose means to modify and extend the method. From this, we will try to deduce general principles for employing quantitative and qualitative analysis in complementary ways.

Segment Two There exists a good correspondence between the quantitative data that indicate a impulsive and forceful playing style and the qualitative rating in the morphological categories.

On the other hand, the agreement among participants is not clear when it comes to the sound helping categorisation. For instance it is difficult to discuss why a majority (10/26) perceived the movements as primarily helping the sound level. In looking at Fig. 3, it is at least obvious that the motion and Laban derivatives accumulate in amplitude under the increased sound level. This is contrasted by respective sections of low movement and sound levels delineating an interaction of movement effort and related changes in sound level. On the other hand, comments like:

"Rhythm is the most pervasive feature of this short sequence. Of my memories of the seen, rhythm is the most important" [5] underline salient features of the alternative choices (here: helping the rhythm). The theorised factors for the choice of the sound helping category furthermore ease the understanding of the large agreement in the morphological categorisations of impulsiveness and force-based primitives. Impulsiveness is most probably perceived by the synchronised and energetic movements of the head and the rest of the body. In these moments the force applied emerges and becomes tangible which may be related to peaks in the acceleration data of these compact movements.

[5] "Der Rhythmus ist die aufdringlichste Komponente dieser kurzen Sequenz, in meiner Erinnerung an das Gesehene hat der Rhythmus die grösste Wichtigkeit"

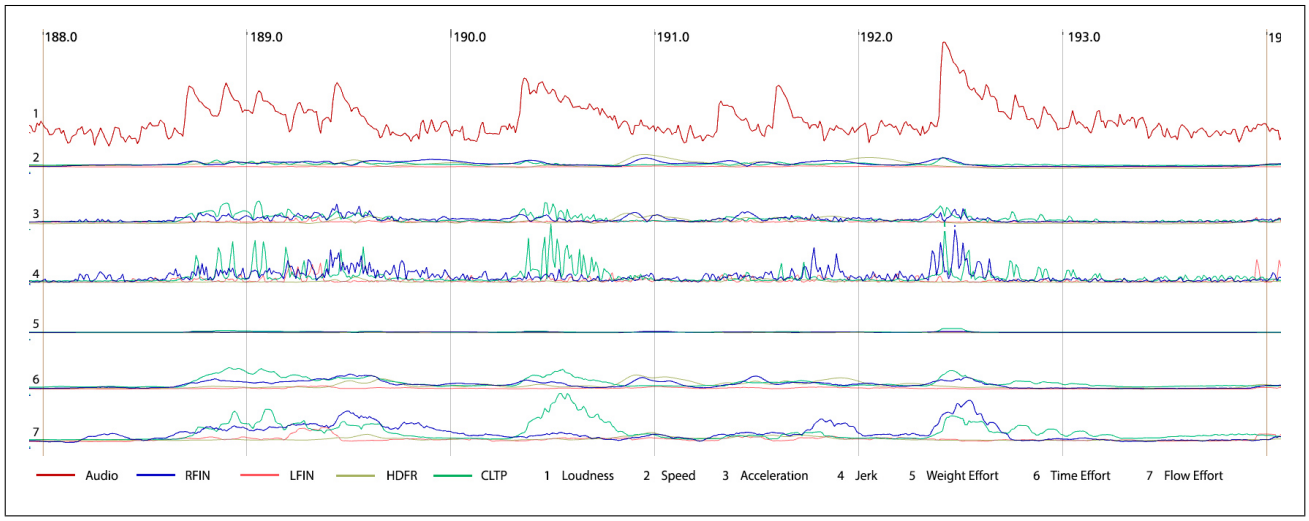

Figure 5: Continuous data plots of segment four.

Segment Three This segment is characterised by a large variety of different playing styles which causes a partial disagreement in the qualitative evaluation concerning the functional and morphological categories. Only a small majority of the participants agreed on a predominant presence of impulsiveness. With 7 choices of the category iterativity and 3 participants opting for the 4th open categorisation entitling it with "impulsive and iterative" and alike, the segment is not clearly categorised. This overall impression of broadly distributed categorisation data rather than clear-cut agreement of this segment is also impressively discernible in the equal distribution between pattern- and force-based primitives. The following comments may shed some light on the difficulties in decision: "The rhythm of the gesture and accompanying sound made me choose the pattern-based option. The sound rhythms translate well into patterns." "The beat implicates a sudden change. It is repeated but still very force-based."

A similar amount of disagreement exists in the functional categorisation. This might be caused by the gradual changes in the iterative left hand movements that convey an epistemic rather than ergotic function and the variability of the correlations between head and left hand movements which indicate a semiotic rather than ergotic categorisation. These observations may be confirmed by statements like the following annotating the epistemic choice: "The performer seemed to be experiencing the tactile structure of the different cello parts directly in her gestures. She almost seemed to be experimenting with how different sounds can be produced by different materials."

Segment Four There is a clear agreement between subjective and quantitative analysis concerning the rhythmical and pattern-based characteristics of the sound producing movements. Compared to the morphological categorisation of impulsiveness in the two other segments, participants opted for the iterative nature and pattern-based primitives for this segment.

Less agreement was observable in the functional categorisations with, for example, half of the participants choosing sound producing. Some comments on communicative aspects of head movements and mimic of the eyes with statements may explain this partial disagreement: "As I wrote earlier, the musician seems intent on communicating a mood or feeling by using her head movements and eyes." The described head and eye movements could also explain the differences in the rating of the functional categories as 8 participants chose semiotic, which is possibly related to the audience-directed communicative aspects of these movements. Furthermore, with 6 participants opting for the epistemic category, these movements can also be interpreted as personal moments of discovery by the performer in the interaction with the instrument and the material. A respective comment emphasises this observation: "I think that the epistemic function is also present here but to a lesser degree: one could explore the strings by these rhythmic movements." [6]

5.1 Gesturality

Looking at the rating of general gesturality of the segments, segment three was rated the highest followed by segment four and two. The questions which arises at this point is about the salient features in the segments which led to the respective ratings. First, we theorise that the amount and variety of movement as well as sound seems directly correlated to the gesturality scores. This observation is certainly limited to a large degree by the length of the segment and to a lesser degree by the selection of the segments, which may explain the dominance of segment three. Second, we reason that the ratings may also be influenced either by (salient) features not captured in the sound and motion data (e.g. mimic, communicative aspects of upper body/torso movements) or by more distal features like personal dispositions of the participants or features inherent to the musical material with their socio-cultural embedded-ness.

5.2 Differences

A number of reasons for the existence of differences between the subjective and quantitative analysis of the performance may exist.

[6] "Hier denke ich, dass auch die epidemische Funktion dabei ist, jedoch zu einem geringeren Anteil. Denn durch die rhythmische Bewegung könnte man auch die Saiten besser kennenlernen ."

Dimensionality differences The data acquisition process reduces the performance characteristics to a few dimensions. As a result, the quantitative analysis lacks data that is relevant in the qualitative analysis, for example data related to the timbral and melodic qualities of the music or facial expressions. The difference in the type and amount of data available for quantitative and qualitative analysis could be alleviated by including more sophisticated audio analysis and by integrating facial or gaze-tracking.

Attention differences The quantitative analysis processes all data dimensions individually and concurrently and assigns equal relevance to each of them. In the subjective analysis, by contrast, each participant focuses his or her attention on the most salient features only and the analysis results in a overall evaluation of high- level characteristics of the performance. These differences can be partially bridged by combining multiple 'MoCap' marker positions before conducting the analysis and by weighting the influence of each marker according to some salience criteria.

Contextual differences For the quantitative analysis, the sequence of observations of the performance segments is irrelevant, since the corresponding algorithms don't show any memory effects. The qualitative analysis, however, is strongly influenced by the sequence of segments, since human subjects tend to pay particular attention to differences between the segments. For this reason the sequence of presented segments was randomising in the qualitative analysis. This difference could be reduced by integrating an attenuation factor and a gradually shifting baseline into the quantitative analysis.

Correlating Sound and Movement It is difficult to analyse sound timbre effects in relation to sound loudness and movement features. In the case of the present composition with its array of noisy extended playing techniques, the task of extracting the physical motion and effort from the sonic content is challenging. Additional sound dimensions such as spectral centroid, flux, noisiness etc. could to be taken into account. The non-standard playing techniques makes using the Laban Motion Dimensions and their computable descriptors difficult to use. These descriptors seem useful to evaluate full body movements in dance (i.e. many joints) and less for evaluating movements of individual body parts in instrument playing (i.e. single joints).

Quantitative versus Qualitative Evaluation Potentially important expressive aspects of the performance are not detectable from 'MoCap' data alone because:

- facial expression changes are indicators of expressivity but not acquired in 'MoCap' data.
- small head movements have strong visual impact (carrier of semantic information) but do not appear as significant peaks in 'MoCap' data.
- upper body movements (important sound facilitating movements) show little significant features in 'MoCap' data.

Large differences in absolute values of motion features among different markers, e.g., a bow tip motion and a head movement, do not correspond to relevance in subjective interpretation. A small head movement, for example, might be considered a more important sound accompanying or facilitating gesture than a large bow motion, because it carries different signification: the former is perceived as an expressive gesture, whereas the latter as a controlled instrumental action.

6. CONCLUSIONS AND OUTLOOK

The in-depth analysis and discussion on morphological as well functional characteristics, as well as the differential observation both between qualitative and quantitative methods, and within the a single analysis category, changed the focus of this project. This meant moving from a hypothesis based on 'effort' to the wider and more versatile concept of salient features. As the interpretation shows, quantitative and qualitative data can be put into relationship and their mutual complementarity can be shown. This means that effects and statement observed through the qualitative methods may be better understood thanks to the interpretation of data carried out with the quantitative methods. The principal remaining incompatibilities have to do with the semantic, expressive content on the level of the musical performance (see 5.2). The highest level of perception is also indicated in the participant's overall ratings on 'gesturality', which in this piece inherent to the musical material (see 5.1). A stronger correspondence between the data-driven analysis and the subjective impressions by participants could be achieved by adding facial measures. Nevertheless, in order to have more encompassing interpretation the complementarity of quantitative data and qualitative statement is useful, and does not need to be brought into total alignment.

Starting from the hypothesis that expressivity in musical performance is carried by the perception of salient features in movement, we come to the following insights: Perceived gesturality is depending on salient features in movement. In a quantitative measure these are salient contrast values, whereas in the subjective, qualitative domain they reside in aspects that are not by necessity those of the instrumental actions, but rather those containing communicative information, or in an epistemic sense emphasise the exploration of the instrument through touching and action.

The results of this enquiry indicate that it would be useful to include more dimensions for the methods used to achieve a finer differentiation through: different sensors (facial tracking, physiological sensor, muscle tension sensors, skin conductance, heart rate, brain responses etc.); extending questionnaire techniques (weighting/ranking instead of forced-choice; extended setups such as quantitative continuous measurements of audience response (e.g., by means of pressure sensors) as well as live ratings.

The multi-methodological approach conducted in this research is guided by a motivation that is strongly rooted in musical practice. This motivation is based on the desire to enable computer-based interactive systems to respond to higher level qualitative and expressive cues in a musical live performance. In such a setup, the normal expressivity of a human performer would not need to be constrained

by functional necessities for controlling a computer-based instrument. Rather, this expressivity becomes as an intrinsic carrier of meaning that is readable not only by the audience but also by the accompanying computer system. The identification of correspondences between a quantitative analysis of sensorial data and qualitative and subjective evaluations of audiovisual performance media serves as a very first step towards this goal. This step helps to inform the design and implementation of feature extraction and machine learning algorithms that are able to mimic human audiences in the recognition of expressivity in a real-time, live-performance stage context. In order to have these feature-recognition processes available in an interactive system, those quantitative measures that are as close as possible to qualitative subjective interpretation could prove to have a powerful effect on expressivity for interactive compositions in music and dance.

Acknowledgments

This investigation was carried out within the 'Motion Gesture Music' Project at the Institute for Computer Music and Sound Technology of the Zurich University of the Arts, and was funded by the Swiss National Science Foundation Grant No. 100016_149345.

7. REFERENCES

[1] G. Fauconnier and M. Turner, *The way we Think: Conceptual Blending and the Mind's Hidden Complexities*, 2008th ed. New York, USA: Basic Books, 2003.

[2] E. Webb and K. E. Weick, "Unobtrusive measures in organizational theory: A reminder," *Administrative Science Quarterly*, pp. 650–659, 1979.

[3] N. K. Denzin, *The research act: A theoretical introduction to research methods*. New York, USA: McGraw-Hill, 1978.

[4] R. B. Johnson, A. J. Onwuegbuzie, and L. A. Turner, "Toward a definition of mixed methods research," *Journal of mixed methods research*, vol. 1, no. 2, pp. 112–133, 2007.

[5] J. C. Schacher, "Moving Music – Exploring Movement-to-Sound Relationships," in *Proceedings of the International Symposium on Movement and Computing (MoCo'16)*, Thessaloniki, Greece, 5.–6. July 2016.

[6] M. Fingerhut, "Music information retrieval, or how to search for (and maybe find) music and do away with incipits," in *IAML-IASA Congress, Oslo*, 2004.

[7] J. S. Downie, "Music information retrieval," *Annual review of information science and technology*, vol. 37, no. 1, pp. 295–340, 2003.

[8] O. Lartillot and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction from Audio," in *International Conference on Digital Audio Effects, DAFx*, 2007, pp. 237–244.

[9] E. Muybridge, *The human figure in motion*. Courier Corporation, 2012.

[10] B. Burger and P. Toiviainen, "MoCap Toolbox-A Matlab toolbox for computational analysis of movement data," in *Proceedings of the Sound and Music Computing Conference, SMC 2013*, Stockholm, Sweden, 2013.

[11] R. Laban, *The Mastery of Movement*, 4th ed. Alton, Hamphire, UK: Dance Books Ltd., 1950 (1980/2011).

[12] H. Knight and R. Simmons, "Layering laban effort features on robot task motions," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. ACM, 2015, pp. 135–136.

[13] E. Davies, *Beyond dance: Laban's legacy of movement analysis*. Routledge, 2007.

[14] C. Cadoz, M. M. Wanderley *et al.*, "Gesture-Music," *Trends in Gestural Control of Music*, vol. 12, pp. 71–94, 2000.

[15] R. I. Godøy and M. Leman, *Musical Gestures: Sound Movement and Meaning*. New York, USA: Routledge, 2010.

[16] R. I. Godøy, A. R. Jensenius, and K. Nymoen, "Chunking in music by coarticulation," *Acta Acustica united with Acustica*, vol. 96, no. 4, pp. 690–700, 2010.

[17] T. Orning, "Pression – a performance study," *Music Performance Research*, vol. 5, pp. 12–31, 2012.

[18] H. Lachenmann, "Klangtypen der neuen Musik," in *Musik als existentielle Erfahrung*. Wiesbaden, Germany: Breitkopf & Härtel, 1966, pp. 1–20.

[19] D. Smalley, "Spectromorphology: explaining sound-shapes," *Organised sound*, vol. 2, no. 2, pp. 107–126, 1997.

[20] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, I. M. Dalca, and D. J. Levitin, "Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance," *Cognition*, no. 118, pp. 157–170, 2011.

[21] A. Jensenius, M. Wanderley, R. Godøy, and M. Leman, "Musical Gestures, Concepts and Methods in Research," in *Musical Gestures, Sound, Movement and Meaning*, R.-I. Godøy and M. Leman, Eds. New York, USA: Routledge, 2010.

[22] I. Poggi, "Body and Mind in the Pianist's Performance," in *Proceedings of the 9th international conference on music perception and cognition (—CMPC9)*, M. Baroni, A. R. Addessi, R. Caterina, and M. Costa, Eds., Bologna, Italy, 2006, pp. 1044–1051.

[23] I. Choi, "Cognitive Engineering of Gestural Primitives for Multi-modal Interaction in a Virtual Environment," in *International Conference on Systems, Man, and Cybernetics*, vol. 2. IEEE, 1998, pp. 1101–1106.

[24] J. C. Schacher, C. Miyama, and D. Bisig, "Gestural electronic music using machine learning as generative device," in *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME'15,*, Baton Rouge, USA, May 31 – June 3 2015.

[25] C. Larboulette and S. Gibet, *A Review of Computable Expressive Descriptors of Human Motion*. Vancouver, Canada.: Proceedings of the Second Workshop on Motion and Computing MOCO'15, August 14–15 2015.