

Estimating Pianists' Hand and Finger Kinematics with Markerless Motion Capture

Ivan Pilkov
Johannes Kepler University
Linz, Austria
ivan.pilkov@jku.at

Laura Bishop
University of Oslo
Oslo, Norway
l.e.bishop@imv.uio.no

Carlos Cancino-Chacón
Johannes Kepler University
Linz, Austria
carlos.cancino_chacon@jku.at

Abstract

Highly-skilled pianists demonstrate fluent and efficient hand and finger motion. Recent developments in video-based methods for markerless motion capture that rely on deep learning algorithms offer a less intrusive and resource-intensive alternative to traditional marker-based infrared ones. In this work we present an initial look into how such a markerless motion capture system could be developed and utilized for the study of pianists' hand and finger kinematics. We collect data from different pianists performing technical exercises and present an analysis of the kinematics of hand and finger motion that arise when pianists are challenged by playing at fast and slow tempi. We investigate different kinematic measures and analyze the recorded MIDI to assess the pianists' strategies for timekeeping and hand coordination and discover certain indications for smoother motions and better control at intermediate tempi. Our results are also in line with previous work in the field and point towards the suitability of this approach for studying pianists' hand movements.

CCS Concepts

• **Applied computing** → **Sound and music computing**; **Performing arts**; • **General and reference** → *Estimation*; *Measurement*.

Keywords

Motion capture, music performance, piano, hand kinematics

ACM Reference Format:

Ivan Pilkov, Laura Bishop, and Carlos Cancino-Chacón. 2026. Estimating Pianists' Hand and Finger Kinematics with Markerless Motion Capture. In *MOCO '26: Proceedings of the 10th International Conference on Movement and Computing (MOCO '26)*, April 23–25, 2026, Montpellier, France. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3802842.3802887>

1 Introduction

Skilled pianists demonstrate high levels of control and precision in their hand and finger movements [15] and studying how they adapt these to different musical challenges can inform our understanding of musical learning and expertise, as well as aid in injury prevention and recovery.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

MOCO '26, Montpellier, France

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2500-5/26/04

<https://doi.org/10.1145/3802842.3802887>

Research on pianists' motor control commonly uses infrared (IR) motion capture (MoCap), which enables fine-grained capture of small-scale, high-velocity kinematics with the use of reflective markers placed on the body that is to be tracked. These markers may be prone to detach, the setup can also be labor-intensive and intrusive, and extensive post-processing is needed to label markers and handle missing data [22]. Video-based methods for markerless motion capture that rely on deep learning algorithms offer a less intrusive and resource-intensive workflow and could provide a more ecologically-valid alternative.

This paper presents a study of the temporal and spatial control of the hands and fingers of pianists during performances under different tempo demands. We investigate the hypothesis that pianists exhibit better control of the performance in intermediate tempi between the two extremes. We record seven pianists with various backgrounds and skill levels performing selected technical exercises and record multiview video and MIDI data of their performances. We introduce a method for markerless MoCap from these multiview video recordings to capture the fast and intricate movements of their fingers. We investigate several different aspects of the resulting motions as indicators for spatial control of the performance, namely, quantity and smoothness of motion, and characteristics of finger movement trajectories. From the MIDI data we obtain measures of temporal variability and timing accuracy, coordination between both hands, as well as variability in MIDI velocity (correlated with loudness of the produced sound).

The rest of this paper is structured as follows: Section 2 discusses related work on motion analysis of music performance and markerless motion capture. Section 3 describes our data collection in 3.1 and presents our markerless motion capture for piano performance analysis in 3.2. Section 4.1 presents hand-finger kinematics analysis. Section 4.2 presents some analysis of MIDI data. Finally, Section 5 concludes the paper.

2 Background

2.1 Motion analysis of music performance

Musicians' body motion can be classified as sound-producing, sound-facilitating (i.e., supporting sound production), sound-accompanying (i.e., made in response to the music), and/or communicative (e.g., head nods to a co-performer) [3]. In this paper, we are concerned with how pianists adapt their sound-producing motion to play at faster or slower tempi. Most prior research on pianists' sound-producing motion has focused on hand and finger kinematics, although the upper arms and torso support sound production as well [23]. Analyses of IR MoCap and MIDI data are commonly used in this literature.

Some studies have compared pianists of different skill levels. Experts have been shown to demonstrate improved independence of hands, allowing for differentiated manipulation of parameters like loudness and articulation [12]. Experts are also better able to adapt their degree of finger independence to the demands of the music, compared to non-experts [25].

A number of studies have compared pianists' performance at moderate versus fast tempi. Pianists demonstrate reduced finger control, resulting in higher temporal variability, at faster compared to slower tempi [4]. Temporal variability has been linked to joint velocities and finger independence, while variability in keystroke velocity has been linked to finger posture [21]. Pianists have been found to raise their fingers higher when they play faster, perhaps allowing for heightened kinesthetic feedback that helps with finger control [4]. Other research has shown that pianists' fingers follow characteristic trajectories, measured in terms of angular velocity, that are relatively invariant to tempo changes [5]. Goebl and Palmer [8] developed a measure of keystroke efficiency that represented the degree of coordination between finger joints. Greater efficiency was associated with higher precision in timing and keystroke velocity.

2.2 Markerless Motion Capture

In contrast to IR-based MoCap, which detects physical markers placed on the body of interest, another type of system has been adopted which relies on the detection of keypoints in video frames. Such systems do not require the placement of markers on the body, which results in less-intrusive and more ecologically-valid capture. They typically utilize machine learning-based *pose estimation* algorithms to directly obtain 3D human postures from conventional RGB video recordings. Commonly known as *markerless* MoCap, they have been used in a variety of contexts and applications, such as sports [16], clinical biomechanics [24], and expressive movement analysis in music performance [18], among others.

The work presented in this paper draws on Simon et al. [19] and the FreiHAND dataset by Zimmermann et al. [29], which both employed cost-effective alternatives to expensive IR MoCap systems. These methods used a multi-view setup with triangulation and a bootstrapping approach to generate training data for their high-quality hand pose estimation algorithms.

Several studies have examined multi-camera markerless systems using pose estimation and have reported average joint center location differences of 10–50 mm [14, 26] and temporal-spatial differences of 15 mm compared to marker-based methods [26]. In comparison, Bishop and Jensenius [2] compared the accuracy of two commercial IR MoCap systems and found mean-squared-errors between the two in the order of a few mm² when capturing musicians. Moro et al. [13] compared marker based and markerless MoCap systems for capturing violin playing performances, concluding that the error obtained when adopting a 3-view camera system is in the order of a few millimeters. Additionally, as pointed out by Wade et al. [24], some applications may not require highly accurate data and instead numerous trials could be averaged to obtain reliable average results. Our approach is similar to that of Uhlrich et al. [22], who developed an end-to-end framework for full-body movement analysis, assessing between-group differences, and prediction of injury risks in clinical settings. Their results showed an accurate

prediction of kinematics with a mean absolute joint angle error of 4.5° compared to IR MoCap and confirmed the suitability of such an approach for the task.

3 Methods

3.1 Data Collection

We collected data from seven participants (3 women and 4 men; mean age 32.1 years, std. 4.8 years), with 8.9 years of piano lessons on average (std. 6.5 years), which we categorized into three skill groups: experts ($n = 4$, mean 13.8 years of formal piano training, std. 1.5 years), intermediates ($n = 2$, mean 3 years playing piano, std. 3 years), and beginner ($n = 1$, < 1 year playing piano). In the following we will refer to the experts as E1–E4 and the intermediates as I1–I2 (the beginner was omitted from the presented results because they could only play at the slowest tempi).

Our choice of material are seven piano technique exercises (numbers 1, 2, 6, 7, 11, 14, and 20) from Charles-Louis Hanon's *The Virtuoso Pianist in Sixty Exercises for the Piano* [10] which, since its introduction in the mid-19th century, has been a staple in piano teaching. These exercises are particularly useful for analyzing finger kinematics since they focus on similar difficulties and independence for the fingers in each hand and address a variety of common technical aspects related to potentially challenging finger movements, such as trills and extensions.

The participants recorded each exercise at multiple tempi: 40, 50, 60, 70, 80, 90, 100, 110, and 120 bpm. This tempo range was selected to study its impact on finger dynamics, extending both below and above the range that Hanon recommended in his instructions for the exercises (60–108 bpm). Some participants were unable to perform at the faster tempi; in particular, the intermediates sometimes only reached 90 bpm and the beginner only reached 60–80 bpm.

Separate recordings were made for the ascending and descending parts of each exercise. We always recorded the slowest tempo first, then proceeded with increasingly faster tempi. Participants were requested to follow the fingering (which finger to use to strike the key) in the original score and play non-expressively. They heard a metronome while playing and were allowed to repeat the exercise if they made errors, used a wrong fingering, or got out of time with the metronome. Most participants recorded the exercises in 2–3 separate sessions on different days. For those who recorded all of the exercises at all tempi, the recording sessions took approximately 4 hours altogether.

We recorded synchronized multiview video from three different angles (see Fig. 1) using three GoPro cameras at 60fps and 1080p resolution in linear lens mode. MIDI data were recorded on a Yamaha C1 Enspire ST Disklavier piano.

MIDI recordings were aligned to the scores using the dynamic time warping (DTW)-based method described in [17] and we semi-automatically annotated the fingerings for each performed note.

3.2 Markerless MoCap for Pianists' Hands

We developed a system utilizing hand pose estimation from RGB videos to estimate the positions of keypoints on the hands and fingers of pianists in a world coordinate metric space for each frame of the recorded videos.



Figure 1: Data capturing setup with three cameras positioned on the left, right, and above the keyboard.

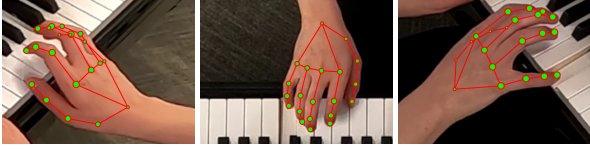


Figure 2: 2D pose estimation results on a single hand pose. The 2D poses are triangulated to obtain the 3D pose used for the analysis.

For the estimation of the 3D hand poses, we implemented a multiview triangulation (MVT) approach making use of our three-camera recording setup, as described in Section 3.1, and the Direct Linear Transform (DLT) algorithm [11] for MVT. In this way we mostly avoided the problem of hand self-occlusion as we ensured that all keypoints are visible in at least two of the viewpoints, and we obtained the 3D pose in world coordinates with accurate scaling, which allowed us to perform all steps of the subsequent analysis.

We calibrated the three cameras using Zhang’s method [28] and used two ArUco markers [6] positioned on both sides of the music stand for the estimation of the cameras’ extrinsic parameters and as origin points for our coordinate system. We used an off-the-shelf automatic hand pose estimator (HPE) [27] from Google’s repository¹ to obtain the standard 21-keypoint poses (corresponding to the 3 joints and 1 tip per finger + a wrist one) per hand for each frame in the videos (see Fig. 2). From the resulting 2D poses we obtained three sets of pair-wise MVT 3D poses.

We designated the three cameras as Left (L), Center (C), and Right (R), with C positioned above the keyboard and L and R positioned on either side of the pianist (see Fig 1). The following steps were

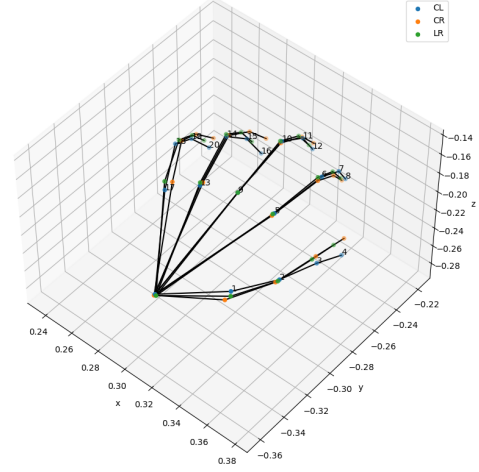


Figure 3: 3D pose estimation from multiple pairs of coordinates. Three sets of 3D world coordinates are obtained by performing MVT for each pair of 2D poses separately. The final 3D pose is computed by averaging the three sets of coordinates.

performed independently for each of the two hands, so we skip the hand index for clarity. After processing each recorded performance with the HPE model, we obtained the 2D pose $\mathbf{p}_t^c$ for each $c \in \{C, L, R\}$ per timestep t . We separately performed the DLT for each of the three pairs of poses $(\mathbf{p}_t^L, \mathbf{p}_t^C)$, $(\mathbf{p}_t^C, \mathbf{p}_t^R)$, $(\mathbf{p}_t^L, \mathbf{p}_t^R)$ to obtain the three sets of 3D world coordinates indicated as CL , CR , and LR (see Fig. 3).

We chose the middle finger’s metacarpophalangeal (MCP) joint keypoint as the root joint (due to it being the most stable and least prone to occlusion across all participants and exercises) and aligned all 3D poses to the mean position of the root joint across all three views. This resulted in an average of 3–7 mm mean-per-joint-positional-error (MPJPE) between the three sets of 3D keypoints. We then computed the mean coordinates for each of the 21 keypoints to obtain the final 3D poses used in the analysis.

We evaluated the differences between the empirically measured palm sizes of the pianists and the corresponding ones obtained from the 3D poses as a measure of the accuracy of the estimation, similar to [13], and we chose the distance between the center of the wrist and the MCP joint of the middle finger following [20]. The mean error across the randomly selected trials was 3 mm (std. 2 mm).

4 Results

4.1 Hand Motion Kinematics Analysis

We analyzed pianists’ hand-finger motion data to look into how tempo affected the quantity of motion, the smoothness, and the self-similarity of the velocity profiles (computed as the numerical derivatives of their positions) of the fingertips.

¹https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker

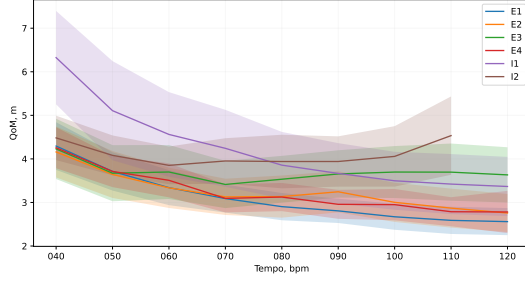


Figure 4: Mean-per-finger QoM across tempi, averaged over all exercises.

4.1.1 Quantity of Motion. We computed the quantity of motion (QoM) of the pianists’ fingertips as the sum of velocity magnitudes per performance. Since pianists performed at the same tempi, differences in performance lengths were minimal and no normalization was needed. Fig. 4 shows QoM per pianist and tempo, with data averaged across fingers and exercises. The aggregated results revealed a general decrease in QoM with increasing tempo, which suggest that pianists may have adapted to the performance demands with more efficient movements.

4.1.2 Motion Smoothness. We used the spectral arc length (SPARC) to estimate movement smoothness as proposed in the context of neurological rehabilitation [1], and applied to the analysis of music performance by González-Sánchez et al. [9]. SPARC reflects the amount of intermittency (acceleration change) in a movement and should not be computed directly for motion sequences that can be broken down into discrete movements, as we find with rhythmic motion, but rather over sequences segmented in a meaningful way [1]. We segmented our finger velocity profiles based on the timing of keystrokes and their associated fingers, which we obtained from the MIDI data aligned with the performance score. Grouping note onsets by fingers, for each finger joint, SPARC was computed for windows that started at the note onset for which this finger performed a keystroke and ended at the next such note onset. This resulted in each window capturing both “active” finger movements (i.e., during a keystroke) and “passive” finger movements that occurred while other fingers were active. The results we present here focus on SPARC computed for the fingertips. Fig. 5 shows smoothness across tempi for six of the pianists, with data averaged across exercises 1, 2, and 6. Smoothness increased for all pianists with increasing tempi, and improved at a faster rate between 40–60 bpm than between tempi above 60 bpm. Across participants, smoothness tended to be lower for the thumb and pinky fingers than for other fingers, and higher for the right (dominant) hand than the left.

4.1.3 Self-similarity in Finger Motion within Pianists. Using key-point velocities, we computed the DTW distance with the Euclidean metric for all pairs of performed tempi within pianists and exercises. This gave us a measure of how consistently pianists moved their fingers across the tempo range. Fig. 6 shows 9×9 self-similarity matrices for exercises 1–2 for the four experts. These matrices show higher self-similarity (smaller DTW distances)

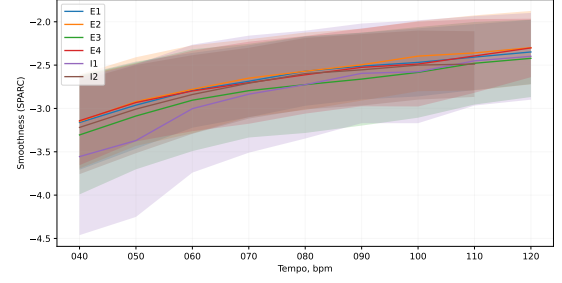


Figure 5: Average smoothness for the first three exercises for all pianists across tempi. Colored bands indicate standard deviation.

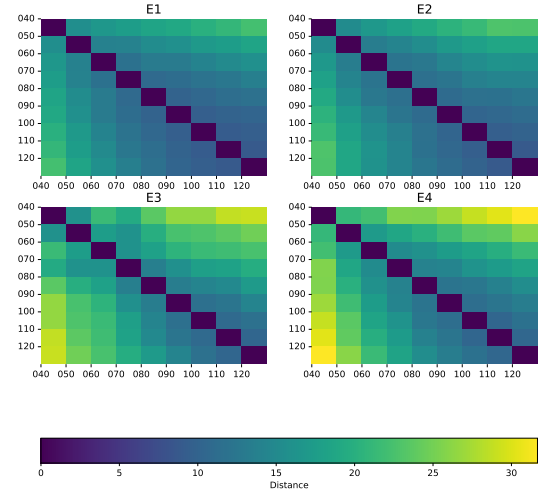


Figure 6: Averaged self-similarity matrices of the first two exercises for the four experts.

between performances with similar tempi, especially among the fastest tempi. Some pianists (E1 and E2) had more evenly distributed self-similarity across the tempo ranges than others (E3 and E4), suggesting individual differences in how pianists adapt their technique to tempo demands.

4.2 MIDI Performance analysis

We analyzed pianists’ MIDI data with the aim of showing how MIDI velocity, temporal variability, and right-left hand asynchrony differed across tempi and between expertise groups.

4.2.1 MIDI Velocity Variability. MIDI keypress velocity data were averaged per tempo and player for statistical analysis (Fig. 7 top). MIDI velocities were lower for slow tempi than intermediate, $t(38) = 2.48, p = .02$, or fast tempi, $t(34) = 4.40, p < .001$, and lower for intermediate than fast tempi, $t(35) = 2.08, p = .04$. This positive relationship with tempo is in line with prior research showing that pianists commonly raise their fingers higher and play louder at

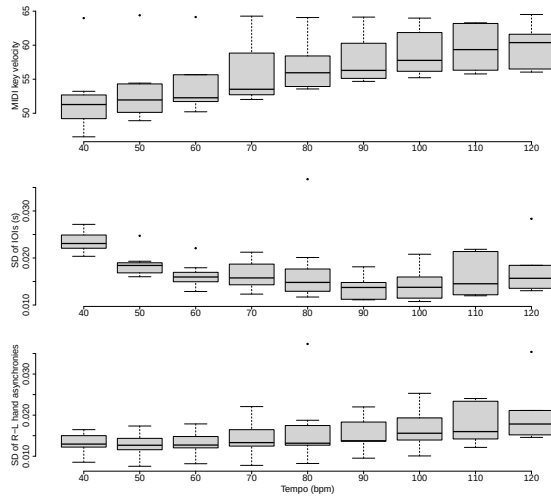


Figure 7: MIDI key velocity (top), temporal variability (middle) and right-left hand asynchronies (bottom) averaged per tempo and pianists.

faster tempi [4]. Expert pianists tended to play louder than beginning/intermediate pianists, $t(54) = 2.07, p = .04$.

4.2.2 Temporal Variability. We computed interonset intervals (IOIs) between successive notes played by each hand and took the standard deviation (std) of IOIs as our measure of temporal variability (Fig. 7 middle). Two outliers were removed before averaging stds per pianist, tempo, and hand. Left hand IOIs were more variable than right hand IOIs, $t(57) = 13.06, p < .001$, and beginning/intermediate players were more variable than expert players, $t(32) = 4.79, p < .001$. T-tests showed greater variability in the slowest tempi (40–60 bpm) than in intermediate (70–90 bpm), $t(34) = 2.19, p = .04$, or fast tempi (100–120 bpm), $t(32) = 2.45, p = .02$, and no difference between intermediate and fast tempi ($p > .05$). The negative relationship between variability and tempo suggests that time-keeping was primarily a challenge at slow tempi.

4.2.3 Right-left hand asynchrony. Right-left hand asynchronies were computed between notes that should be played simultaneously, according to the score (Fig. 7 bottom). Pianists tended to play right-hand notes slightly before corresponding left-hand notes (mean asynchrony = .004 s, std = .007 s). Previous research has shown that a “melody lead” pattern arises when pianists emphasize a melody line by playing it louder. Higher key velocities result in melody notes sounding slightly in advance of non-melody notes [7]. Indeed, our pianists played right hand notes louder than left hand notes, $t(113) = 5.52, p < .001$. Average asynchrony magnitudes did not differ between slow, intermediate, and fast tempi (t-tests all show $p > .05$). However, there was greater variability in asynchronies in fast tempi than in slow tempi, $t(22) = 3.20, p = .004$. Variability did not differ between intermediate and slow tempi or between intermediate and fast tempi (t-tests all show $p > .05$). Thus, inter-hand coordination was less stable at faster than slower tempi.

5 Conclusions

We used MIDI data and markerless MoCap to investigate pianists' kinematics when performing technical exercises and test whether they demonstrate finger motion that is more stable, smoother, and less variable at intermediate tempi compared to fast and slow tempi. Our results show some evidence for this: temporal variability was highest at slow tempi, while variability in hand asynchronies was greatest at fast tempi, suggesting that stability is maximized at tempi in between. Slow and fast tempi seem to pose distinct challenges for pianists, and our analyses of finger motion offers some clue to how they adapt. Our findings further demonstrate that markerless MoCap is a viable alternative to IR MoCap for certain types of analysis.

In addition to collecting data from more participants, future work will focus on bringing together different kinematics and MIDI measures, as well as estimating dynamics, to investigate movement efficiency and control, observe hand posture and assess injury risks, and analyze finger activity during and between keystrokes.

Acknowledgments

This work was supported by the Austrian Science Fund (FWF) under grants PAT 8820923 (“Rach3: A Computational Approach to Study Piano Rehearsal”) and PIN 1347924 (“AURA: Augmenting musical interaction via EVAs”); and by the University of Oslo and the Research Council of Norway through its Centres of Excellence scheme, project number 262762.

References

- [1] Sivakumar Balasubramanian, Alejandro Melendez-Calderon, Agnes Roby-Brami, and Etienne Burdet. 2015. On the analysis of movement smoothness. *Journal of neuroengineering and rehabilitation* 12 (2015), 1–11.
- [2] Laura Bishop and Alexander Refsum Jensenius. 2020. Reliability of two IR motion capture systems in a music performance setting. In *Proceedings of the International Conference on Sound and Music Computing (Torino)*.
- [3] Sofia Dahl, Frédéric Bevilacqua, Roberto Bresin, Martin Clayton, Laura Leante, Isabella Poggi, and Nicolas Rasamimanana. 2010. Gestures in Performance. In *Musical Gestures: Sound, Movement, and Meaning*, Rolf Inge Godøy and Marc Leman (Eds.). Taylor & Francis, Abingdon, UK, 36–68.
- [4] Simone Dalla Bella and Caroline Palmer. 2011. Rate Effects on Timing, Key Velocity, and Finger Kinematics in Piano Performance. *PLoS ONE* 6, 6 (2011), e20518. doi:10.1371/journal.pone.0020518
- [5] Shinichi Furuya and John F. Soechting. 2012. Speed Invariance of Independent Control of Finger Movements in Pianists. *Journal of Neurophysiology* 108, 7 (2012), 2060–2068. doi:10.1152/jn.00378.2012
- [6] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292. doi:10.1016/j.patcog.2014.01.005
- [7] Werner Goebel. 2001. Melody Lead in Piano Performance: Expressive Device or Artifact? *The Journal of the Acoustical Society of America* 110, 1 (2001), 563–572. doi:10.1121/1.1376133
- [8] Werner Goebel and Caroline Palmer. 2013. Temporal Control and Hand Movement Efficiency in Skilled Music Performance. *PLoS ONE* 8, 1 (2013), e50901. doi:10.1371/journal.pone.0050901
- [9] Victor Gonzalez-Sanchez, Sofia Dahl, Johannes Lunde Hatfield, and Rolf Inge Godøy. 2019. Characterizing movement fluency in musical performance: Toward a generic measure for technology enhanced learning. *Frontiers in psychology* 10 (2019), 413120.
- [10] Charles-Louis Hanon. 1928. *The Virtuoso Pianist in Sixty Exercises for the Piano*. G. Schirmer, New York, NY, USA.
- [11] Richard Hartley and Andrew Zisserman. 2003. *Multiple view geometry in computer vision*. Cambridge university press.
- [12] Sarah Kim, Jeong Mi Park, Seungyeon Rhyu, Juhan Nam, and Kyogu Lee. 2021. Quantitative Analysis of Piano Performance Proficiency Focusing on Difference between Hands. *PLOS ONE* 16, 5 (2021), e0250299. doi:10.1371/journal.pone.0250299

- [13] Matteo Moro, Maura Casadio, Leigh Ann Mrotek, Rajiv Ranganathan, Robert Scheidt, and Francesca Odone. 2021. On The Precision Of Markerless 3d Semantic Features: An Experimental Study On Violin Playing. In *2021 IEEE International Conference on Image Processing (ICIP)*. 2733–2737. doi:10.1109/ICIP42928.2021.9506356
- [14] Nobuyasu Nakano, Tetsuro Sakura, Kazuhiro Ueda, Leon Omura, Arata Kimura, Yoichi Iino, Senshi Fukushima, and Shinsuke Yoshioka. 2020. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Frontiers in Sports and Active Living* 2 (2020). doi:10.3389/fspor.2020.00050
- [15] Takanori Oku and Shinichi Furuya. 2017. Skilful Force Control in Expert Pianists. *Experimental Brain Research* 235, 5 (2017), 1603–1615. doi:10.1007/s00221-017-4926-3
- [16] David Pagnon, Mathieu Domalain, and Lionel Reveret. 2021. Pose2Sim: An End-to-End Workflow for 3D Markerless Sports Kinematics—Part 1: Robustness. *Sensors* 21, 19 (2021), 6530. Issue 19. doi:10.3390/s21196530
- [17] Silvan David Peter. 2023. Online Symbolic Music Alignment With Offline Reinforcement Learning. In *Proceedings of the 24th International Society for Music Information Retrieval Conference, ISMIR 2023, Milan, Italy, November 5-9, 2023*, Augusto Sarti, Fabio Antonacci, Mark Sandler, Paolo Bestagini, Simon Dixon, Beici Liang, Gaël Richard, and Johan Pauwels (Eds.). 634–641. doi:10.5281/ZENODO.10265367
- [18] Sujoy Roychowdhury, Preeti Rao, and Sharat Chandran. 2024. Human Pose Estimation for Expressive Movement Descriptors in Vocal Musical Performances. In *Proceedings of the 25th International Society for Music Information Retrieval Conference, ISMIR 2024, San Francisco, California, USA and Online, November 10-14, 2024*, Blair Kaneshiro, Gautham J. Mysore, Oriol Nieto, Chris Donahue, Cheng-Zhi Anna Huang, Jin Ha Lee, Brian McFee, and Matthew C. McCallum (Eds.). 834–841. doi:10.5281/ZENODO.14877458
- [19] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. 2017. Hand key-point detection in single images using multiview bootstrapping. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 1145–1153.
- [20] George Sung, Kanstantsin Sokal, Esha Uboweja, Valentin Bazarevsky, Jonathan Baccash, Eduard Gabriel Bazavan, Chuo-Ling Chang, and Matthias Grundmann. 2021. *On-Device Real-time Hand Gesture Recognition*. arXiv:2111.00038 [cs] <http://arxiv.org/abs/2111.00038>
- [21] Kenta Tominaga, André Lee, Eckart Altenmüller, Fumio Miyazaki, and Shinichi Furuya. 2016. Kinematic Origins of Motor Inconsistency in Expert Pianists. *PLOS ONE* 11, 8 (2016), e0161324. doi:10.1371/journal.pone.0161324
- [22] Scott D. Uhlich, Antoine Falisse, Lukasz Kidziński, Julie Muccini, Michael Ko, Akshay S. Chaudhari, Jennifer L. Hicks, and Scott L. Delp. 2023. OpenCap: Human Movement Dynamics from Smartphone Videos. *PLOS Computational Biology* 19, 10 (2023), e1011462. doi:10.1371/journal.pcbi.1011462
- [23] Felipe Verdugo, Justine Pelletier, Benjamin Michaud, Caroline Traube, and Mickaël Begon. 2020. Effects of Trunk Motion, Touch, and Articulation on Upper-Limb Velocities and on Joint Contribution to Endpoint Velocities During the Production of Loud Piano Tones. *Frontiers in Psychology* 11 (2020), 1159. doi:10.3389/fpsyg.2020.01159
- [24] Logan Wade, Laurie Needham, Polly McGuigan, and James Bilzon. 2022. Applications and Limitations of Current Markerless Motion Capture Methods for Clinical Gait Biomechanics. *PeerJ* 10 (2022), e12995. pubmed:35237469 doi:10.7717/peerj.12995
- [25] S.A. Winges and S. Furuya. 2015. Distinct Digit Kinematics by Professional and Amateur Pianists. *Neuroscience* 284 (2015), 643–652. doi:10.1016/j.neuroscience.2014.10.041
- [26] Matteo Zago, Matteo Luzzago, Tommaso Marangoni, Mariolino De Cecco, Marco Tarabini, and Manuela Galli. 2020. 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Frontiers in Bioengineering and Biotechnology* 8 (2020). doi:10.3389/fbioe.2020.00181
- [27] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214* (2020).
- [28] Z. Zhang. 2000. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 11 (2000), 1330–1334. doi:10.1109/34.888718
- [29] Christian Zimmermann, Duygu Ceylan, Jimei Yang, Bryan Russell, Max Argus, and Thomas Brox. 2019. Freihand: A dataset for markerless capture of hand pose and shape from single rgb images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 813–822.