I see something you can only hear
Visualising musical performance with SInES tools and sonic cam

Christoph Reuter[1], Carolin Schmid[1], Xincan Yang[2], Jian Yang[3]

[1] *SInES, Musicological Department, University of Vienna, 1090 Vienna, E-Mail: christoph.reuter@univie.ac.at*
[2] *Fritz Kreisler Department, University of Music and Performing Arts Vienna, Austria*
[3] *Department of Music Engineering, Shanghai Conservatory of Music, China*

SInES Tools

The current rapid development in the field of audio signal analysis and artificial intelligence as well as the possibilities of the sonic camera expand the possibilities for the visualisation of audio features.

Since 2023, the Space for Interdisciplinary Experiments on Sound (SInES) of the Vienna Systematic Musicology offers a range of free online tools at https://sinestools.univie.ac.at/ with enables researchers, students and the interested public to gather a variety of music- and movement-related data from audio (wav, mp3) and video (mp4) files [1][2]. Data on pitch, timbre, harmony, tempo, rhythm, dynamics and much more can be extracted from audio files, while body, hand and facial movements (including emotions) and arbitrary points can be tracked in video files. Furthermore, skin conductance and finger temperature can be recorded while listening to music using eSense sensors from Mindfield [3], as well as states of consciousness via a four-channel EEG Muse S headband [4].

Combined with videos recorded by a normal or sonic camera, a range of insightful information about musical instruments and their players can be discovered. Some examples are presented below:

Where on the E string …

Due to the high temporal resolution of the acoustic camera, it is possible, for example, to follow the path of the impulse (or the location of the strongest sound radiation) in the first milliseconds from the plucked string via the bridge to the radiation at the corpus. With the help of a sonic cam (Gfai Mikado, 96 microphones), the radiation locations for all reachable pitches on a guitar has be visualised and linked to the corresponding audio features via signal analysis, so that the characteristics of the timbres can be described depending on their point of origin. Due to the increasingly shorter wavelengths with rising pitch, this works particularly well for the notes on the high E string (see fig. 1).

Using audio signal analysis, it is also possible to classify the strings of a guitar based on their sound characteristics. The audio features spectral centroid of the percussive part, fluctuations and spectral contrast in the range between 0-200 Hz can be used to classify the strings of the guitar according to their timbral characteristics (see fig. 2).

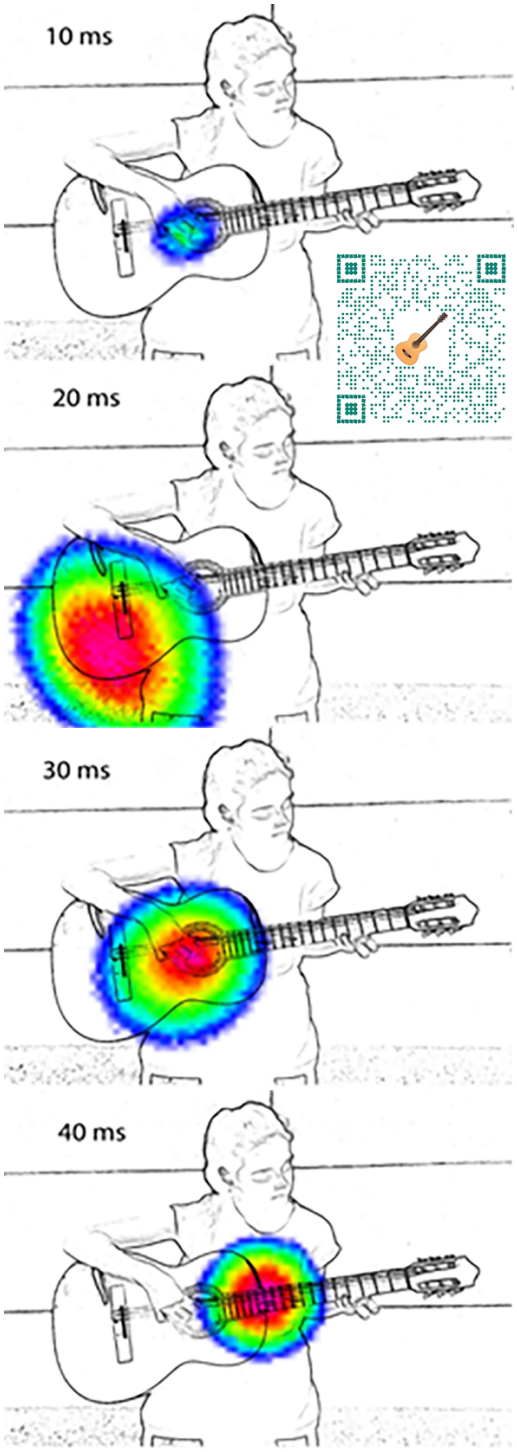


Figure 1: Sound radiation of the A5 on a guitar's high E-string played by Carolin Schmid at 10, 20, 30 and 40 milliseconds from the beginning of the sound. The radiation peak travels from the plucked string via the bridge to the body. All pitches can be played interactively at [5] https://muwiserver.univie.ac.at/sumu2024/gitarre.htm

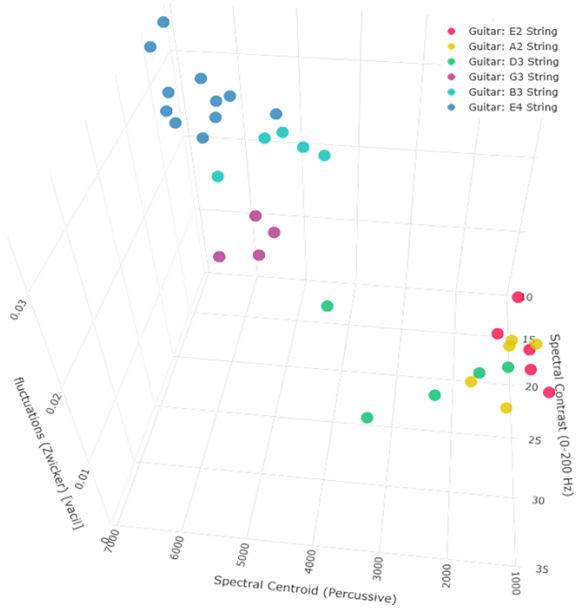


Figure 2: Classification of guitar strings according to the timbral characteristics of their notes (at https://muwiserver.univie.ac.at/sumu2024/gitarre.htm [5], at the bottom of the page).

SHAPED and HandBoX

SHAPED (**S**ound Radiation, **H**and, **A**natomy, **P**itch & **E**xpression **D**isplay): The SHAPED visualiser allows the merging of data obtained with the SInES tools for body, pitch, hand and expression tracking based on sonic camera and normal camera recordings. Thanks to the pitch tracking model used in SInES Tools [6], even fast-played notes can be determined with a high degree of accuracy. Arrow keys can be used to jump from note to note in order to intuitively visualise information about the body and hand position, sound radiation, intonation, and even players expression (shown by color in the face) as well as a range of other timbral parameters for each segment of the piece.

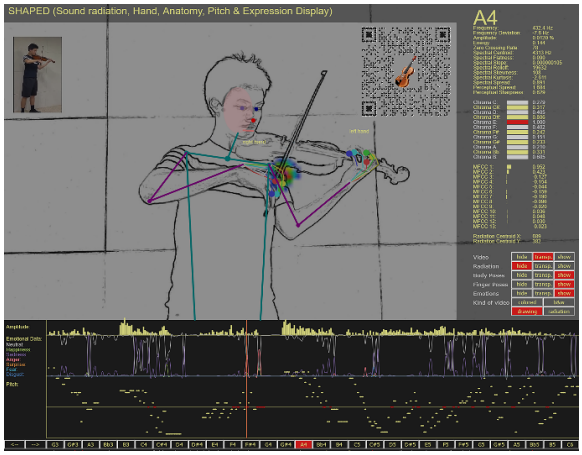


Figure 3: SHAPED: Sound Radiation, Hand, Anatomy, Pitch & Expression Display: Combination of data from SInES tools and acoustic camera with display options in the menu at the bottom right as well as additional displays for amplitudes, expression/emotion of the player and recognised selectable pitches underneath. (Violin played by Xinxan Yang https://muwiserver.univie.ac.at/sumu2024/violin_data [7]

HandBoX (**Hand Bo**dy e**X**pression Visualizer): The HandBoX visualiser has a similar structure to SHAPED and can be used to display a combination of body and hand movement (including movement trails), players' emotion and audio data collected with the SInES tools taken from any music videos (e.g. YouTube or other video collections).

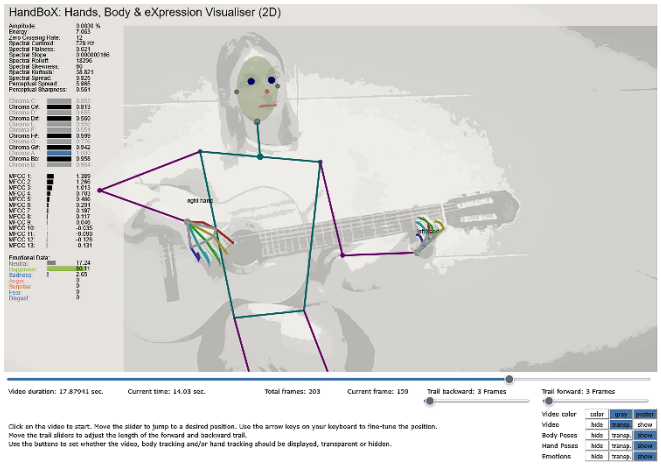


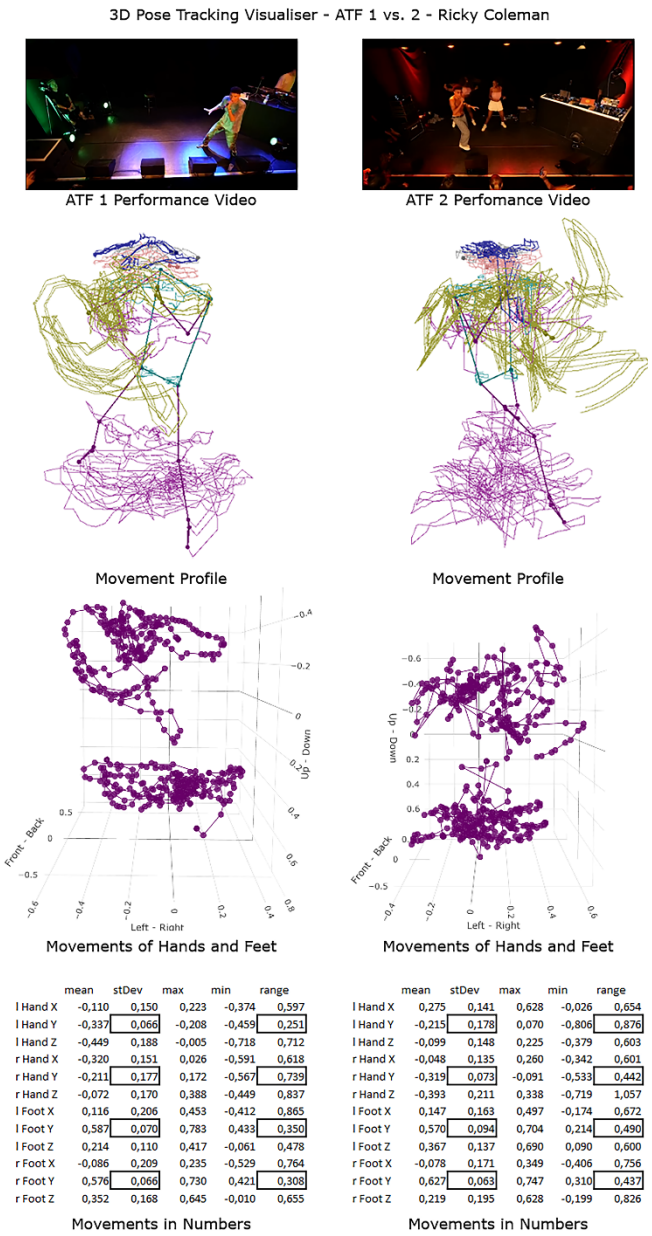
Figure 4: HandBoX (Hand Body eXpression Visualizer): displaying body, hand, expression and audio data of musical instruments and their players (Guitar played by Marie Krajci, in https://muwiserver.univie.ac.at/sumu2024/handbox) [8].

3D Hand and Pose Tracking Visualiser

With the 3D Hand Tracking Visualiser, the hand movements of musicians, speakers, conductors etc. recorded with the SInES tools from normal two-dimensional videos can be visualised as three-dimensional models.

Figure 5: Hand movements of a guitarist as 3D models with adjustable movement trails, motion trajectories, motion and audio feature plots and automatic correlation calculation (free stock video from https://youtu.be/w1u9CuPyaEE, in https://sinestools.univie.ac.at/handtracking_visualiser.htm) [9].

The same applies to the 3D Pose Tracking Visualiser, in which the movements of dancers, musicians, actors, speakers etc. can be displayed both frame by frame and summarised via movement trails:



	mean	stDev	max	min	range
l Hand X	-0,110	0,150	0,223	-0,374	0,597
l Hand Y	-0,337	0,066	-0,208	-0,459	0,251
l Hand Z	-0,449	0,188	-0,005	-0,718	0,712
r Hand X	-0,320	0,151	0,026	-0,591	0,618
r Hand Y	-0,211	0,177	0,172	-0,567	0,739
r Hand Z	-0,072	0,170	0,388	-0,449	0,837
l Foot X	0,116	0,206	0,453	-0,412	0,865
l Foot Y	0,587	0,070	0,783	0,433	0,350
l Foot Z	0,214	0,110	0,417	-0,061	0,478
r Foot X	-0,086	0,209	0,235	-0,529	0,764
r Foot Y	0,576	0,066	0,730	0,421	0,308
r Foot Z	0,352	0,168	0,645	-0,010	0,655

	mean	stDev	max	min	range
l Hand X	0,275	0,141	0,628	-0,026	0,654
l Hand Y	-0,215	0,178	0,070	-0,806	0,876
l Hand Z	-0,099	0,148	0,225	-0,379	0,603
r Hand X	-0,048	0,135	0,260	-0,342	0,601
r Hand Y	-0,319	0,073	-0,091	-0,533	0,442
r Hand Z	-0,393	0,211	0,338	-0,719	1,057
l Foot X	0,147	0,163	0,497	-0,174	0,672
l Foot Y	0,570	0,094	0,704	0,214	0,490
l Foot Z	0,367	0,137	0,690	0,090	0,600
r Foot X	-0,078	0,171	0,349	-0,406	0,756
r Foot Y	0,627	0,063	0,747	0,310	0,437
r Foot Z	0,219	0,195	0,628	-0,199	0,826

Figure 6: Visualisation of the 3D pose tracking data of the dance movements of rapper Ricky Coleman during the same song before and after a coaching session. The performance before coaching (left) shows more horizontal movements of the arms and legs, while the number and intensity of vertical arm and leg movements increase after coaching (right) (https://muwiserver.univie.ac.at/hiphop-gesten/RC.htm Video: Nikita Kudakov) [10] [11].

Conclusion

The data collected with the SInES tools can be visualised in a lot of different ways, with or without a sonic camera. The combination of the sonic camera and the SInES tools opens up new possibilities for finding connections between sound radiation, audio features and musician movements. Combined visualisers such as SHAPED and HandBoX offer the possibility of evaluating large quantities of videos from video collections such as YouTube in an easily accessible and user-friendly way, while the 3D hand and pose tracking visualisers can be used for teaching and coaching musicians, dancers, conductors, actors, speakers, etc.

Literature

[1] Reuter, C., Czedik-Eysenberg, I., Cui, A.-X. (2023). Happy Life comes with P5 - P5, ML5, Meyda and Plotly as helpful Tools in Teaching and Research. Proceedings of DAGA2023, 49. German Annual Conference on Acoustics (p. 991-994). Hamburg.

[2] Reuter, C., Czedik-Eysenberg, I., Cui, A.-X. (2023). Moves & Grooves. Analyse von Audiosignalen, Tracking von Körper, Hand und Gesicht/Ausdruck. OCG Journal 48/4, p. 20-23.

[3] Mindfield eSense Sensors: https://mindfield-shop.com/geraete/esense/

[4] Muse EEG headbands: https://choosemuse.com/

[5] Sound radiation on guitar strings: https://muwiserver.univie.ac.at/sumu2024/gitarre

[6] Kim, J.W., Salamon, J., Li, P., Bello, J.P., (2018). CREPE: A Convolutional Representation for Pitch Estimation. arXiv:1802.06182.

[7] SHAPED - Sound radiation, Hand, Anatomy, Pitch & Expression Display: https://muwiserver.univie.ac.at/sumu2024/violin_data

[8] HandBoX - Hand Body eXpression Visualizer: https://muwiserver.univie.ac.at/sumu2024/handbox

[9] 3D hand tracking visualizer: https://sinestools.univie.ac.at/handtracking_visualiser.htm.

[10] 3D pose tracking visualizer: https://muwiserver.univie.ac.at/sumu2024/body_tracking_visualiser

[11] HipHop gestures before and after coaching: https://muwiserver.univie.ac.at/hiphop-gesten/RC