

Science shines with SInES

SInES Tools for audio and video analysis – New updates and improvements

Christoph Reuter¹, Isabella Czedik-Eysenberg¹, Anja-Xiaoxing Cui¹, Marik Roos¹, Sarah Ambros¹, Matthias Eder¹, Jian Yang², Matthias Bertsch³

¹ SInES, Musicological Department, University of Vienna, 1090 Vienna, E-Mail: christoph.reuter@univie.ac.at

² Shanghai Conservatory of Music, Shanghai, China, E-Mail: yangjian@shcmusic.edu.cn

³ Motion-Emotion Lab, University of Music and Performing Arts, Vienna, Austria, E-Mail: Bertsch@mdw.ac.at

SInES Tools

Since 2023, the Space for Interdisciplinary Experiments on Sound (SInES) of the Vienna Department of Systematic Musicology has offered a range of freely available online tools at <https://sinestools.univie.ac.at/> [1], with which researchers, students and others can collect a variety of music-related data from audio and video files. The tools include applications for automatic pitch recognition, audio signal analysis, facial emotion analysis, and pose tracking of musicians, conductors, and dancers [2][3]. The SInES tools offer quick and easy access to audio signal analysis as well as motion and expression tracking, which all follow the same workflow:



Figure 1: Typical SInES tools menu.

With a click on the **start button**, all values are set to 0 so that the respective tool is ready for the analysis. Then using the **upload button** an audio (mp3 or wav) or video file (mp4) is uploaded to the browser and clicking on the **analyze button** will start the analysis. Depending on the size of the audio or video file, this can take a few seconds to a few minutes. During the wait time, a display of interactive curves for the individual sound and/or movement characteristics as well as a list of their mean values, medians, standard deviations, minima, maxima and ranges is shown. The data can be displayed as JavaScript arrays via the **Show JS Arrays** button (if you want to embed the result in other JavaScript applications) or exported as a CSV file for further processing/analysis e.g. in Excel, Matlab or Python via the **Export CSV** button. The analysis itself takes place entirely in the user's browser, i.e. the audio and video data are not uploaded to a third-party server, but remain on the user's own computer during all steps of the analysis.

Updates and improvements in Audio Signal Analysis Tools

In addition to the audio signal analysis tool [4], which is based on the updated Meyda 6.0 library [5], the integration of Essentia.js now enables a more comprehensive and in-depth analysis of audio files [6]. This includes feature extraction in the areas of pitch, chroma & beat detection [7]; spectral features [8]; analysis of frequency bands including MFCCs, ERB, Bark bands and Mel bands [9]; as well as the estimation of musical genre or mood [10]. Further new additions are new tools for formant estimation [11] based on the formantanalyzer.js library [12] as well as an Extreme Metal Vocals Analyzer [13], based on the hardness model by

Czedik-Eysenberg [14][15] and listener data on the perception of extreme metal vocals [16]. This prototype enables acoustic analyses of self-recorded vocals, which are being compared to a database of metal vocal tracks and likely semantic associations (e.g. “demonic”, “raw”, “angelic”) are being predicted.

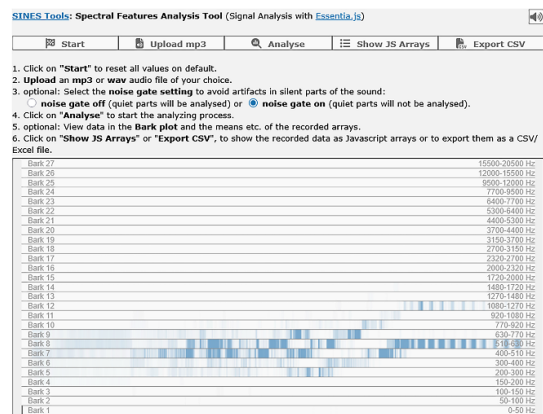


Figure 2: Spectral Feature Analysis Tool [8].

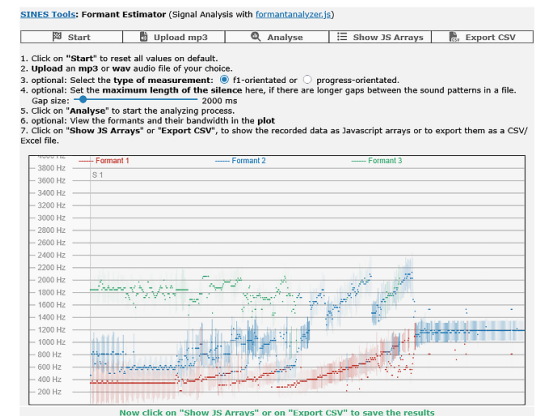


Figure 3: Formant Estimator [11].

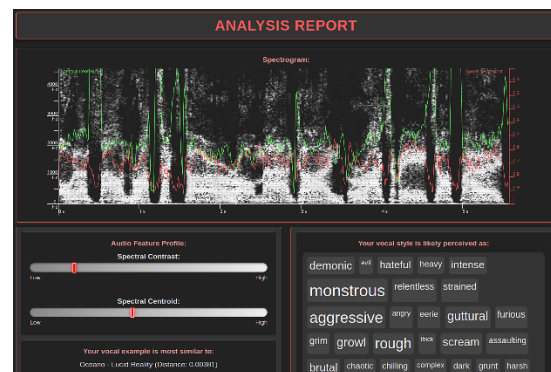


Figure 4: Extreme Metal Vocals Analyzer [13].

For pitch recognition, the new Onset, Tempo & Pitch Analysis Tool [17] based on Aubio.js [18] can be used to apply various pitch and onset analysis algorithms to monophonic audio files, while the new AI pitch estimator [19] can also be used to estimate the pitch of fast tone sequences very accurately.

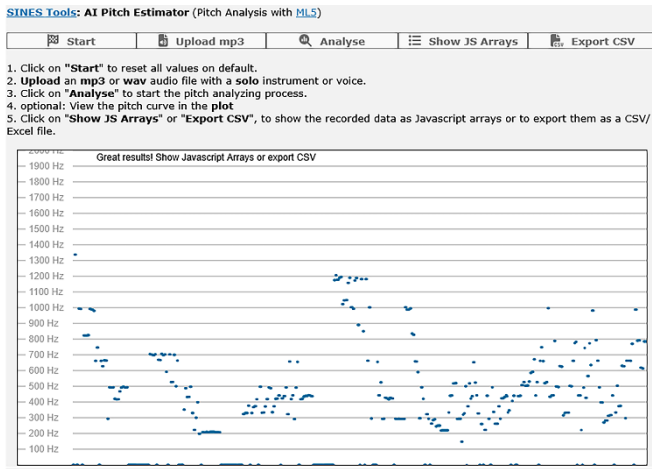


Figure 5: AI Pitch Estimator [19].

Based on the CREPE pitch detection model [20], an application for intonation control for violinists was created in 2024 together with Stefan Kölsch and Matthias Bertsch (Violin Intonation Program, VIP) [21]. The program evaluates the notes played and displays the pitch deviation in Cents, allowing the player to select the concert pitch and tuning system as a reference system and to set the error tolerance. In addition, the program has a vibrato smoothing function with an adjustable smoothing window to ensure that the pitch is correctly detected during a vibrato. While playing the violin, a high score for intonation accuracy is calculated and sound characteristics such as spectral centroid, spectral spread and MFCCs are recorded. A planned update will also enable the derivation of specific sound styles based on these features.

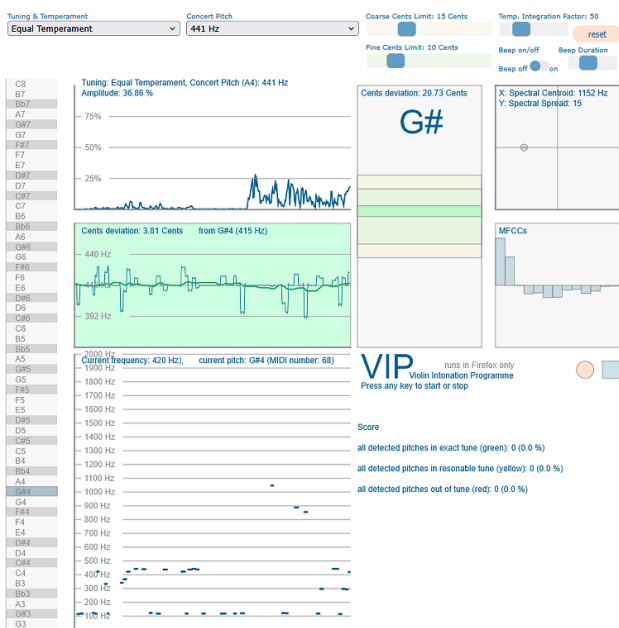


Figure 6: VIP – Violin Intonation Programme [21].

Updates and Improvements in Motion Tracking Tools

The pose tracking [22] and hand tracking [23] applications (currently supported in Chrome only) have been completely reengineered. They now enable frame-by-frame tracking of body and hand movements in three dimensions by leveraging depth information from video frames.

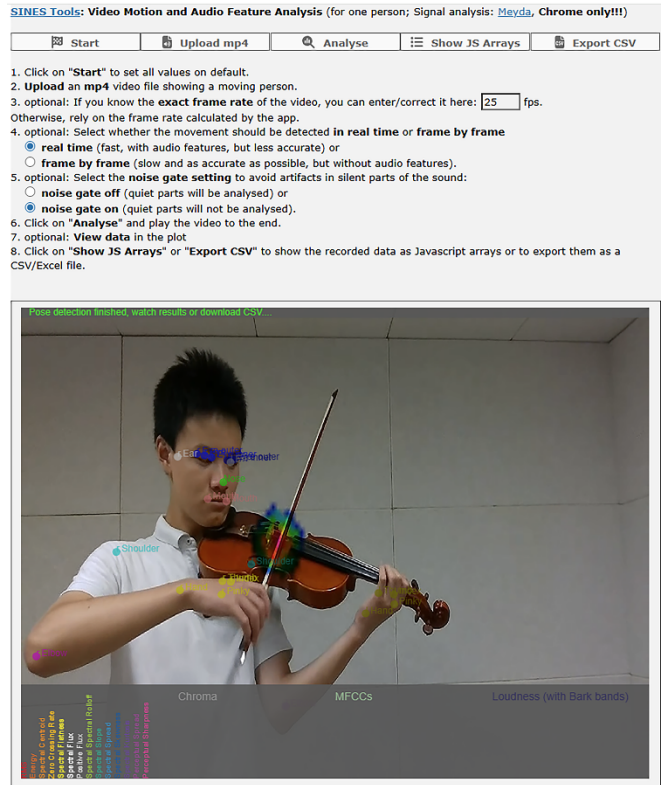


Figure 7: 3D Pose Tracking [22]

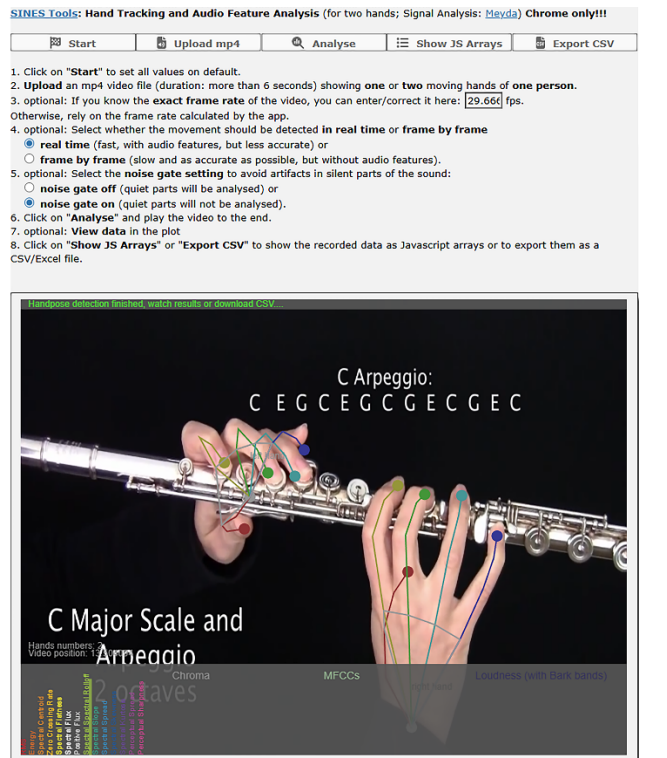


Figure 8: 3D Hand Tracking [23]

With the new Point Tracking Tool [24], eight arbitrary points in a video can be tracked in high temporal resolution. In addition to the speed of the video, the size of the point environment (template) can be freely selected as well as the radius in which the point is to be searched from frame to frame. This application is particularly suitable for tracking non-human or periodic movements of musical instruments or their strings, membranes and reeds in high-speed videos.

SINES Tools: Point Tracking (for up to eight points)

Start Upload mp4 Analyse Show JS Arrays Export CSV

- Click on "Start" to set all values on default.
- Upload an mp4 video file showing a moving object.
- Make sure that the **background** in the video is as neutral as possible and **contrasts strongly in color** with the object to be tracked.
- Click on the **object** to be tracked and mark it with **eight points**. You may name the points in the input fields as you please.

Point 1: Pixel 1 X: 426 Y: 451	Point 5: Pixel 5 X: 330 Y: 232
Point 2: Pixel 2 X: 341 Y: 366	Point 6: Pixel 6 X: 742 Y: 208
Point 3: Pixel 3 X: 511 Y: 373	Point 7: Pixel 7 X: 742 Y: 208
Point 4: Pixel 4 X: 494 Y: 98	Point 8: Pixel 8 X: 742 Y: 208

The points can be reset with [X]

- optional: Set the **template size** for the pattern to be tracked (the larger the more accurate, but also the slower): 63 (63 pixels is a good guess).
- optional: Set the **video speed** (the slower the more accurate): 0.25 (0.25 of the original speed is a good guess).
- optional: Set the **search radius** (the larger the movements, the larger the search radius should be): 20 (20 is a good guess for medium = not so fast movements).
- Click on "Analyse" and play the video to the end.
- optional: **View data** in the plot
- Click on "Show JS Arrays" or "Export CSV" to show the recorded data as Javascript arrays or to export them as a CSV/Excel file.




Figure 9: Point Tracking Application [24]

Updates and Improvements in Emotion Tracking Tools

The Facial Expression Analysis [25] has been revised so that the automatic emotion detection from a facial expression in a video is now also possible frame by frame. This means that even very brief changes in the emotions of actors, viewers, listeners, etc. can be detected in videos now.

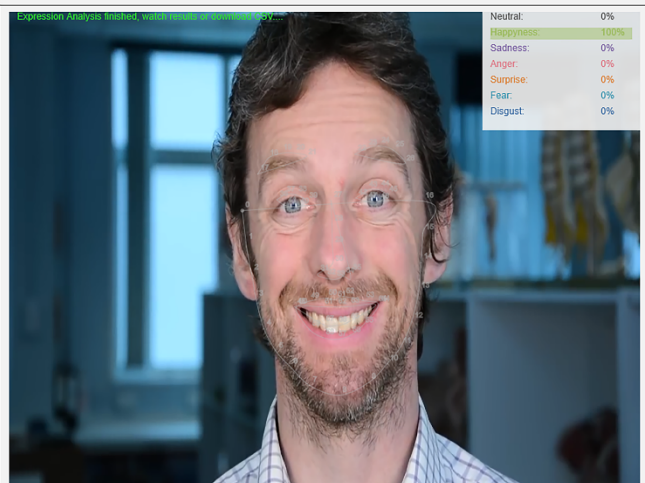
The Emotion and Audio Feature Analysis application [26] for tracking emotional responses while listening to music or sounds has also been updated. The display of the valence-arousal axes is now interchangeable, and other axes can easily be added to support annotation of custom response dimensions in the future. A Mindfield eSense sensor for skin conductance or finger temperature can now be connected via an audio-to-USB adapter. The recorded sensor data is analyzed with the original Mindfield scripts (Firefox only).

SINES Tools: Facial Expression Analysis (for one person, works best in Chrome; Signal Analysis: [Meyda](#))

Start Upload mp4 Analyse Show JS Arrays Export CSV

- Click on "Start" to set all values on default.
- Upload an mp4 video file showing the face of a person.
- optional: If you know the **exact frame rate** of the video, you can enter/correct it here: 25 fps.
- optional: Otherwise, rely on the frame rate calculated by the app.
- optional: Select whether the facial expressions should be detected in **real time** or **frame by frame**
 - real time** (fast, with audio features, but less accurate) or
 - frame by frame** (slow and as accurate as possible, but without audio features).
- optional: Select the **noise gate setting** to avoid artifacts in silent parts of the sound:
 - noise gate off** (quiet parts will be analysed) or
 - noise gate on** (quiet parts will not be analysed).
- Click on "Analyse" and play the video to the end.
- optional: **View data** in the plot.
- Click on "Show JS Arrays" or "Export CSV" to show the recorded data as Javascript arrays or to export them as a CSV/Excel file.

Expression Analysis finished, watch the results in the plot!



Neutral:	0%
Happiness:	100%
Sadness:	0%
Anger:	0%
Surprise:	0%
Fear:	0%
Disgust:	0%

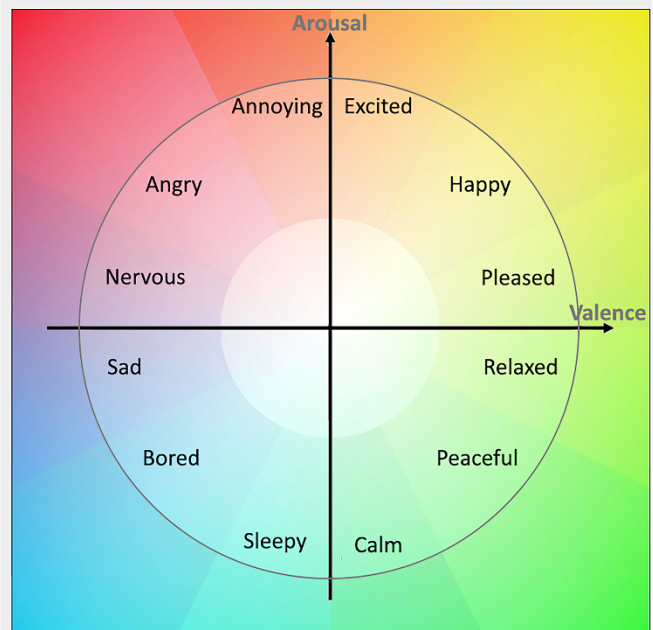
Figure 10: Facial Expression Analysis [25]

SINES Tools: Emotion and Audio Features Analysis (Signal Analysis with [Meyda](#), [Firefox only!!!!](#))

Start Upload mp3 Play Data Show JS Arrays Export CSV File

- Click on "Start" to start the experiment.
- Upload an mp3 or wav audio file of your choice
- optional: Select the **Circumplex Model of Emotions**:
 - classic (dimensions only) or labeled (dimension and labels) or colored (dimension and labels, colored)
- optional: Select the **noise gate setting** to avoid artifacts in silent parts of the sound:
 - noise gate off** (quiet parts will be analysed) or **noise gate on** (quiet parts will not be analysed).
- optional: Connect a **Mindfield eSense EDA Sensor** or a **Mindfield eSense Temperature Sensor** via an **audio to USB adapter** to your computer to record time-synchronous skin conductance or finger temperature and **attach the sensor to your fingers** (Don't forget to connect headphones to the sensor jack):
 - no sensor** or **EDA/Skin Conductance sensor** or **Finger Temperature sensor** is attached.
- Click on the **center of the valence arousal circle** to start the recording and move your mouse over the fields which correspond most to your mood while listening to music. At the end of the sound example, the path through the valence arousal circle is shown and you can find all data in the data plot below.
- Click on "Show JS Arrays" or "Export CSV file", to show the recorded data as Javascript arrays or to export them as a CSV/Excel file.
- optional: Click on "Play Data", to follow the recorded path.

Click on Start and upload a mp3 file



The diagram shows a circular plot with two axes: Arousal (vertical) and Valence (horizontal). The plot is divided into colored regions representing different emotional states:

- Top-Left (High Arousal, Low Valence):** Annoying, Angry, Nervous
- Top-Right (High Arousal, High Valence):** Excited, Happy, Pleased
- Bottom-Left (Low Arousal, Low Valence):** Sad, Bored, Sleepy
- Bottom-Right (Low Arousal, High Valence):** Relaxed, Peaceful, Calm

Figure 11: Emotion and Audio Feature Analysis [26].

It's online, it's free, it's one click away

The latest update to the SInES Tools introduces several powerful enhancements to improve functionality and user experience. The audio signal analysis uses the latest version of Meyda.js and further tools have been created based on Essentia.js, focusing on pitch, chroma, tempo, spectrum, genre and mood detection, as well as other apps for robust pitch recognition, violin intonation scoring and extreme metal vocal classification. With the new update, body and hand motion tracking of individuals in videos are now possible in three dimensions and on a frame-by-frame basis. Additionally, a new point tracking app allows for the selection and tracking of arbitrary points within a video. Facial expression tracking in videos is now also possible frame-by-frame, and skin conductance and finger temperature sensors from Mindfield can now be integrated into the Valence-Arousal application.

Acknowledgment: A big thank you to Stefan Kölsch for his support in the creation of the Violin Intonation Programme.

Literature

- [1] SInES Tools: <https://sinestools.univie.ac.at>
- [2] 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 (p. 991-994). Hamburg.
- [3] Reuter, C., Czedik-Eysenberg, I., Cui, A.-X. (2023). Moves & Grooves. OCG Journal 48/4, p. 20-23.
- [4] SInES Tools – Audio Signal Analysis Tool: https://sinestools.univie.ac.at/meyda_signalanalyse.htm
- [5] Rawlinson, H., Segal, N., & Fiala, J. (2015). Meyda: An Audio Feature Extraction Library for the Web Audio API. Proceedings of the 1st Web Audio Conference (WAC), January 2015, Paris, France.
- [6] Correya, A., Marcos-Fernández, J., Joglear-Ongay, L., Alonso-Jiménez, p., Serra, X., & Bogdanov, D. (2021). Audio and Music Analysis on the Web using Essentia.js, Transactions of the International Society for Music Information Retrieval (TISMIR). 4(1), pp. 167–181.
- [7] SInES Tools – Pitch, Chroma & Beat Analysis Tool: https://sinestools.univie.ac.at/essentia_pitch_tempo.htm
- [8] SInES Tools – Spectral Features Analysis Tool: https://sinestools.univie.ac.at/essentia_timbre.htm
- [9] SInES Tools – Frequency Bands Analysis Tool: https://sinestools.univie.ac.at/essentia_frequenzband.htm
- [10] SInES Tools – Genre and Mood Estimator: https://sinestools.univie.ac.at/genre_mood.htm
- [11] SInES Tools – Formant Estimator: <https://sinestools.univie.ac.at/formantestimator.htm>
- [12] Rehman, A., Liu, Z.T., & Xu, J.M. (2021). Syllable Level Speech Emotion Recognition Based on Formant Attention. In: Fang, L. et al.(eds) Artificial Intelligence. CICA 2021. Lecture Notes in Computer Science, vol 13070. Springer, Cham.
- [13] SInES Tools – Extreme Metal Vocals Analyser: <https://sinestools.univie.ac.at/emv/analyser>
- [14] Czedik-Eysenberg, I., Wieczorek, O., Flexer, A., & Reuter, C. (2024). Charting the Universe of Metal Music Lyrics and Analyzing their Relation to Perceived Audio Hardness. Transactions of the International Society for Music Information Retrieval, 7(1), p. 129-143.
- [15] Czedik-Eysenberg, I. (2021). Semantische Modellierung wahrnehmungspsychologischer Musikdimensionen auf Basis von akustischen Signaleigenschaften, Diss. Thesis, University of Vienna.
- [16] Czedik-Eysenberg, I., Smialek, E., & Herbst, J. P. (2024). Towards an Acoustic-Semantic Space of Extreme Metal Vocal Styles. Proceedings of DAGA2024 (pp. 979-982). Stuttgart.
- [17] SInES Tools – Onset, Tempo & Pitch Analysis: https://sinestools.univie.ac.at/pitch_rhythm_detector.htm
- [18] Aubio.js: <https://github.com/qiuxiang/aubiojs>
- [19] SInES Tools – AI Pitch Estimator: https://sinestools.univie.ac.at/pitch_detection_ml5.htm
- [20] Kim, J.W., Salamon, J., Li P., & Bello, J.P. (2018). CREPE: A Convolutional Representation for Pitch Estimation, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), April 15–20 2018, Calgary, AB, Canada.
- [21] SInES Tools – Violin Intonation Programme (VIP): <https://sinestools.univie.ac.at/vip.htm>
- [22] SInES Tools – Video Pose Tracking (3D): <https://sinestools.univie.ac.at/bewegungssanalyse3D.htm>
- [23] SInES Tools – Video Hand Tracking (3D): <https://sinestools.univie.ac.at/handbewegungssanalyse3D.htm>
- [24] SInES Tools – Point Tracking application: <https://sinestools.univie.ac.at/pointracker.htm>
- [25] SInES Tools – Facial Expression Analysis: <https://sinestools.univie.ac.at/emotiontracking.htm>
- [26] SInES Tools – Emotional Response application: https://sinestools.univie.ac.at/valence_arousal_meyda_emotion_upload_SCR.htm