Happy Life comes with P5
P5, ML5, Meyda and Plotly as helpful Tools in Teaching and Research

Christoph Reuter[1], Isabella Czedik-Eysenberg[1,2] Anja-Xiaoxing Cui[1]

[1] *Universität Wien, Musikwissenschaftliches Institut*
[2] *Universität Osnabrück, Institut für Musikwissenschaft und Musikpädagogik*

1 Background: Modern JavaScript Libraries

P5 is a JavaScript library for solving complex programming tasks quickly and with ease. Based on Processing [1], it was created in 2014 by Lauren McCarthy, Casey Reas and Ben Fry [2] and has been constantly expanded since then. For the field of musical acoustics, it is particularly interesting due to the additional library P5.sound. The latter enables fast and straightforward calculation of FFTs, of filtering, convolutions, sound synthesis, MIDI and many other audio-related processing operations. With the additional library ML5 introduced in 2018, which is based on TensorFlow.js, P5 has gained a machine learning interface which enables access to pre-trained models for the recognition/tracking of persons, movements, objects, faces, hands, pitches and much more.

Concurrently, Plotly.js was developed by Alex Johnson, Jack Parmer, Chris Parmer, Matthew Sundquist around 2015 [3]. Plotly.js is a JavaScript library for interactive data visualization that can be used to visualize data in more than 40 different chart types in 2D and 3D. Of particular interest for musical acoustics is the fact that the points and lines in the plots can be synchronized directly to audio and video files, which enables interactive audio-visual display of results.

Since 2015, the Meyda library has also provided an option for audio signal analysis in JavaScript. Beyond low-level audio signal processing capabilities, this library, created by Hugh Rawlinson, Nevo Segal, Jakub Fiala [4], enables real-time audio feature extraction, including not only the spectral centroid, zero crossing rate, spectral flatness etc., but also perceptual sharpness, loudness, chroma features and MFCCs.

2 Scripts and Applications

Both for the visualization of data and for data acquisition, these libraries in their combination offer completely innovative and, above all, pragmatic, robust and sustainable solutions for research and teaching in the field of musical acoustics e.g.:

2.1 Dynamic Timbre Maps

In timbre research, individual sounds of musical instruments are considered almost exclusively on their own rather than in ensemble play. For a more comprehensive view, the first two formants (F1, F2) of the respective musical instruments from reverberation-free recorded single tracks of the first movement of Beethoven's 7th Symphony [5] can be tracked using Parselmouth [6]. The obtained formant movements can then be plotted within a dynamic graph with axes corresponding to the position of F1 and F2 synchronized to the music. Each musical instrument is coded with its own color, the dynamics are symbolized by the size of the points, and the corresponding mean values (center) and standard deviations (diameter) are displayed in half-transparent ellipses of the same color. The formant centers and movements of musical instruments can thus be visualized in dynamic interplay "in the wild". This enables exploratory insight, e.g., instruments and their formants have their own timbral "territories". Similar timbres have overlapping or adjacent formant areas. Analogous representations are also possible with MFCCs or other timbral descriptors such as the Spectral Centroid.

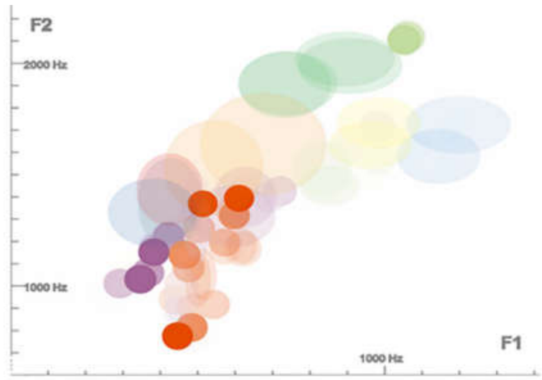

Fig. 1: Formant positions (mean values and standard deviations) of musical instruments in musical ensemble play: Horns (purple), Bassoons (orange) and Oboe (green). https://muwidb.univie.ac.at/orchestration/formants.htm

2.2 Signal Analysis Tool

The combination of the libraries P5, Plotly.js and Meyda further allows fast and uncomplicated data acquisition. For example, we created an interactive online signal analysis tool based on these libraries, which allows users to upload MP3 files and extract the audio features provided by the Meyda library as interactive line charts or - for further use - as JavaScript arrays. The extracted values can be exported in CSV format for external processing. With 20 entries per second, the accuracy of the time-synchronous output is 50 ±1 ms (see Fig. 2).

2.3 Interactive Valence Arousal Model

Next, we combined the audio feature extraction with an interactive self-report of the emotional state while listening to musical works and sounds, uploaded as MP3 files: Based on the idea of logging the respective emotional state directly via mouse movement in a valence-arousal field while listening to music ("EmuJoy")[7], the online tool we created is furthermore able to synchronize with measurements of skin conductance via Mindfield® eSense Skin Response which uses the computer's microphone jack. The emotional rating of videos or the use of other axis labels is also possible in this context. Again, all recorded values can be exported in CSV format with a temporal resolution of 50 ±1 ms (see Fig. 3).

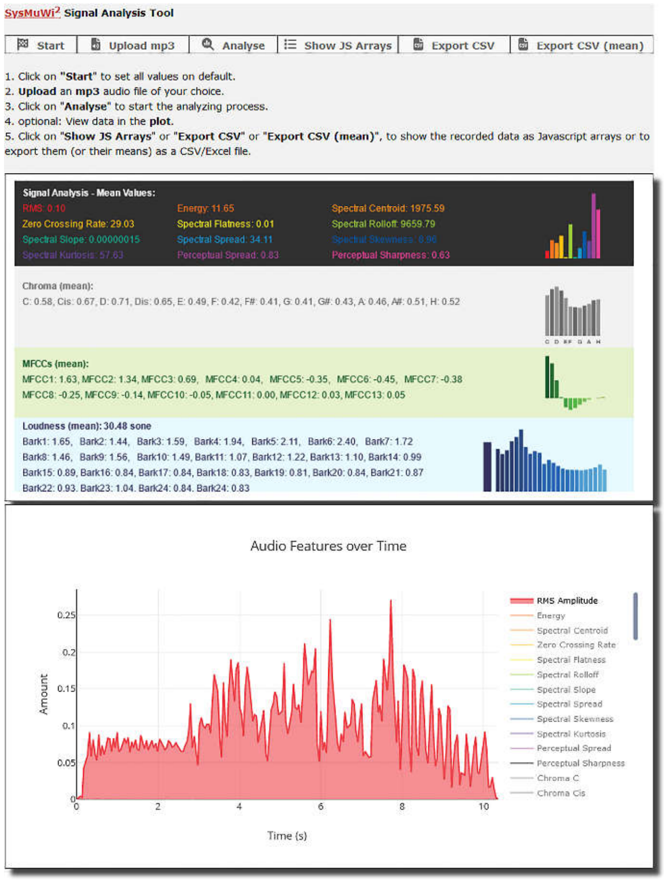

Fig. 2: Signal Analysis Tool

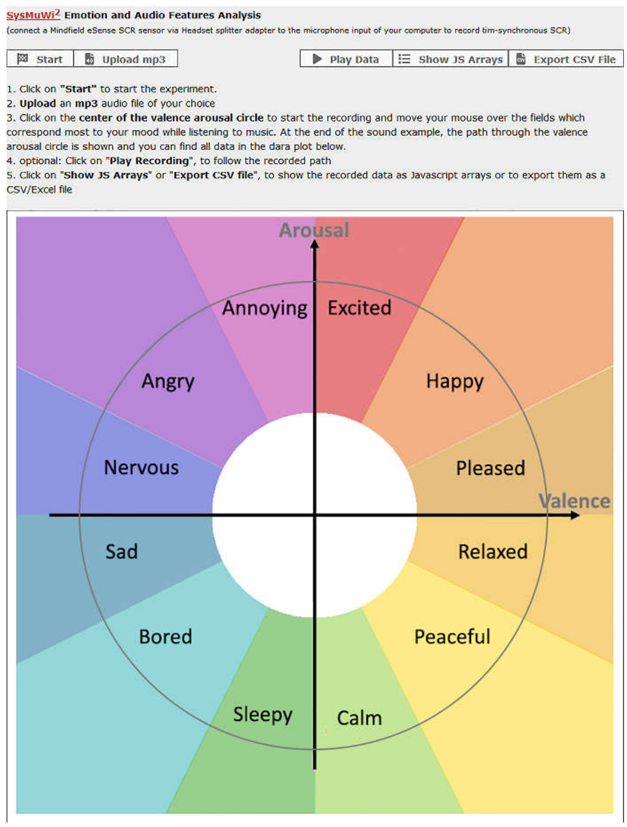

Fig. 3: Interactive Valence Arousal Model

2.4 Video Motion and Audio Feature Analysis

Using ML5 (the machine learning interface for P5), we were able to combine audio feature analysis with the detection of movements and poses of individuals in videos, uploaded as MP4 files. Thus, whole-body movements can be automatically detected from video recordings and captured via their X and Y coordinates relative to the video screen. This data can be synchronized with audio features at intervals of 100 ±2 ms and again exported in CSV format.

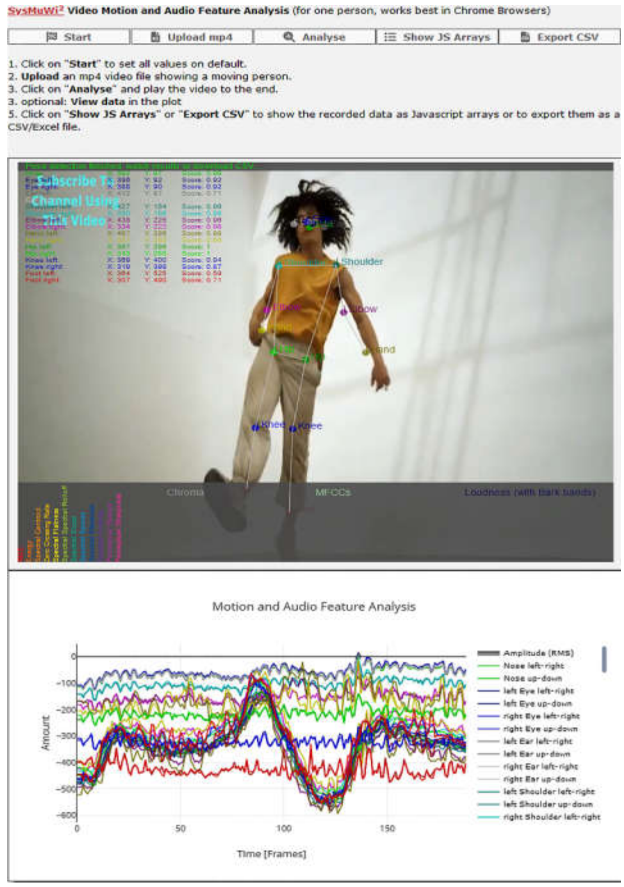

Fig. 4: Video Motion and Audio Feature Analysis (video example from https://youtu.be/watch?v= JPM3QoPyNRM)

2.5 Facial Expression and Audio Feature Analysis

Another feature that can be recognized via the ML5 library are facial expressions. We have combined pre-trained models for automatic video face expression analysis with audio feature analysis to allow for a synchronous analysis of detected facial emotions in conjunction with the audio analysis. The resulting values with an accuracy of 100 ±2 ms can be displayed as line charts as well as exported in CSV format (see Fig. 5).

2.6 Hand Tracking and Audio Feature Analysis

Besides whole-body movements and facial expressions, another promising application is making use of pre-trained models for hand-tracking via ML5 (e.g. for analysis of musicians playing their instruments). The coordinates of the individual finger positions can be displayed synchronously with the associated audio features as line charts or exported in CSV format (see Fig. 6).

Fig. 5: Face Expression and Audio Feature Analysis (video example from https://youtu.be/watch?v=QbWStbEv4LU)

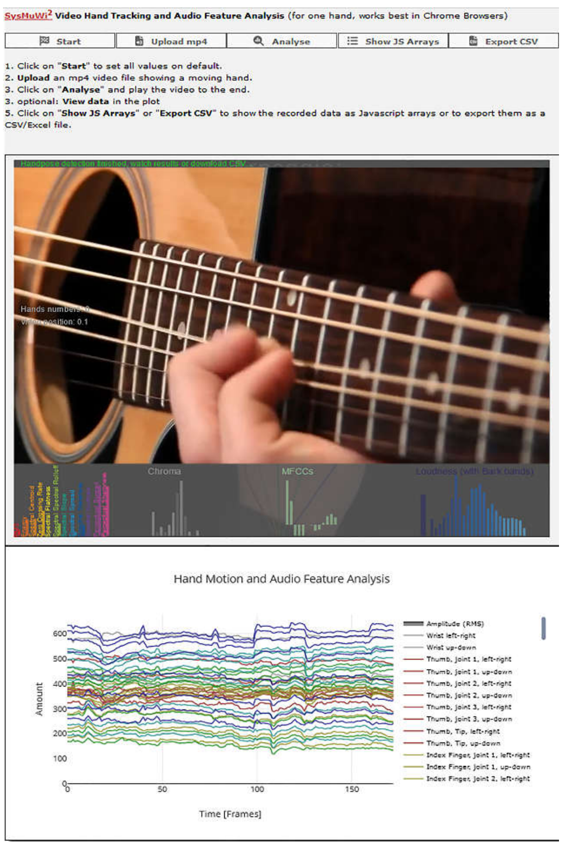

Fig. 6: Hand Tracking and Audio Feature Analysis (video example from https://youtu.be/watch?v=JPu2XQ7X5wI)

2.7 EEG Data Logger for Muse S EEG Headband®

In this application, Bluetooth-enabled input from a 4-channel EEG headband from Muse® can be recorded to establish the states of consciousness of test subjects in their delta, theta, alpha, and beta ranges at 4 ms intervals while they listen to music or any audio files. In addition, their head movements are tracked and their heartbeat is recorded. Their current fatigue, theta relaxation, alpha relaxation and beta concentration are calculated directly from the data and all values can be saved together with the audio amplitudes of their environment in CSV format. Furthermore, a synthesizer and a sampler can be controlled via the four EEG channels, which enables sonification of brain activity.

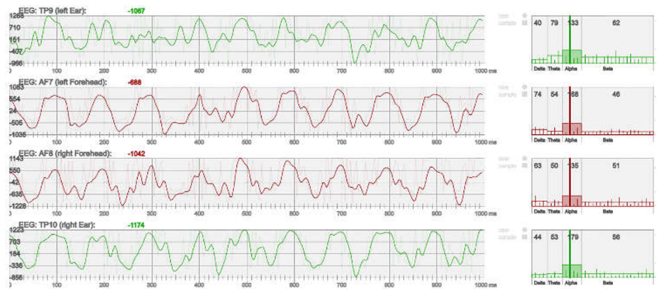

Fig. 7: EEG curves in the alpha state measured with EEG Data Logger and Muse S EEG Headband.

3 Summary and Outlook

In recent years, novel combinations of P5, ML5, Plotly.js, and Meyda have proven useful for creating online applications in research and teaching. Furthermore, the libraries provide a possible introduction to programming, especially for students. The libraries offer new ways of data acquisition and visualization and are easy and fast to learn. For the field of musical acoustics, the synchronized combinations of audio signal analysis, visual tracking of body/hand, motion and facial expressions, as well as (neuro-)physiological data are particularly valuable.

JavaScript as a proven, non-proprietary format offers extensibility, sustainability, future-proofing and promising applications in the field of musical acoustics, especially also through ML5 as an interface to machine learning models as well as strong communities for P5 and Plotly.js.

With the JavaScript libraries Essentia.js [8] and JS-Xtract [9], the audio signal analysis capabilities of the applications presented here can be greatly expanded. In the future, we also plan to incorporate methods such as motion amplification (e.g. in blood flow and pulse detection via WebCam) and additional AI models (e.g. for pitch or BPM estimation), to enable more diverse and powerful applications.

You can find all applications presented here and more at our webpage at https://arc-lab.univie.ac.at/sysmuwi2tools.
Please cite this paper if you use the applications presented here for data acquisition.

4 Acknowledgments

A big thank you to Martin Wittmann for programming the Bluetooth to JavaScript handshake for the EEG Data Logger.

Literature

[1] Reas, C., Fry, B. 2007. Processing: A Programming Handbook for Visual Designers and Artists. Cambridge, Massachusetts: The MIT Press.

[2] McCarthy, L., Reas, C., Fry, B. 2015. Getting Started with p5.js. Santa Rosa, CA: Make Community, LLC. `https://p5js.org/`

[3] Czikk, J. 2013. Ex-Silicon Valley kids to launch web startup in Montreal. Montreal Gazette July 29, 2013. `https://plotly.com/javascript/`

[4] 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. `https://meyda.js.org/`

[5] Pätynen, J., Pulkki, V., Lokki, T. 2008. Anechoic Recording System for Symphony Orchestra. Acta Acustica united with Acustica 94, p. 856-865.

[6] Jadoul, Y., Thompson, B., de Boer, B. 2018. Introducing Parselmouth: A Python interface to Praat. Journal of Phonetics, 71, 1-15.

[7] Nagel, F., Kopiez, R., Grewe, O., Altenmüller, E. 2007. EMuJoy: Software for continuous measurement of perceived emotions in music. Behavior Research Methods 39(2), p. 283-290.

[8] Correya, A., Bogdanov, D., Joglar-Ongay, L., Serra, X. 2020. Essentia.js: A JavaScript Library for Music and Audio Analysis on the Web. 21st International Society for Music Information Retrieval Conference, p. 605-612. `https://mtg.github.io/essentia.js/`

[9] Jillings, N., Bullock, J., Ryan, S. 2016. JS-Xtract: a realtime audio feature extraction library for the web. Extended abstracts for the Late-Breaking Demo Session of the 17th International Society for Music Information Retrieval Conference, 2016. `https://github.com/nickjillings/js-xtract`