All You Can Analyze: SInES Tools
An Audio/Video Analysis Lab in your Browser

Christoph Reuter[1], Isabella Czedik-Eysenberg[1], Anja-Xiaoxing Cui[1], Sarah Ambros[1], Helene Lindenbauer[1], Hao Li[2], Matthias Bertsch[3]

[1] *SInES, Musicological Department, University of Vienna, Austria*
[2] *Northwestern Polytechnical University (NWPU), XI'an, China*
[3] *Motion-Emotion Lab, University of Music and Performing Arts, Vienna, Austria*

Abstract

The Space for Interdisciplinary Experiments on Sounds (SInES) is a research lab at the University of Vienna. Its browser-based platform, SInES Tools, provides a comprehensive suite of freely available applications in the areas of:

- Audio signal analysis
- AI/Machine learning
- Sound (re)synthesis
- Video tracking tools
- Visualizers
- Helpers (statistics & Co).

SInES Tools aims to empower researchers by offering free, browser-based access to advanced research methods that would otherwise require costly software or expert programming skills. This paper presents the latest additions for 2026, showcasing significant expansions across all six core categories. Highlights include practical utilities for acoustic measurement and spectral editing, alongside advanced AI models for instrument/ensemble classification and speech emotion recognition. Furthermore, we introduce new tools for video-based motion analysis, including Eulerian video magnification, together with comprehensive 2D/3D data analyzers, visualizers and more.
.

Keywords: *audio signal analysis, music information retrieval, artificial intelligence, motion tracking, psychoacoustics*

Corresponding author: *christoph.reuter@univie.ac.at.*

1. Introduction

In recent years, the Space for Interdisciplinary Experiments on Sounds (SInES, University of Vienna), has continuously expanded SInES Tools, its suite of browser-based applications. Currently featuring more than 90 tools at https://sinestools.univie.ac.at, the platform follows a simple yet powerful core philosophy [1]:

- "It's online, it's free, it's one click away"
- "Upload file → Analyze → Download results"

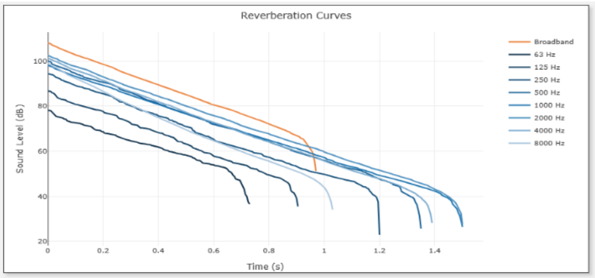
Using modern JavaScript libraries such as *Essentia.js* [2], *Meyda.js* [3], *JS-Xtract.js* [4], *TensorFlow.js* [5], *Transformers.js* [6] (a.o.) the SInES Tools bring efficient audio analysis and machine learning inference to the client side of the browser. This paper outlines novel additions (besides improvements in robustness, new batch-processing functions and Chinese translation), demonstrating how a standard web browser can function as a fully equipped audio-visual laboratory.

2. New Audio (Signal) Analysis Tools

Building on its extensive set of existing audio signal analysis applications, SInES Tools is continuously expanded to include practical tools for everyday acoustical measurement tasks, such as

- **dB Analyzer:** analyzes sound pressure and loudness from calibrated microphone recordings and calculates dB_{SPL} from dB_{FS}, alongside weighted levels (dB_A, dB_B, dB_C), LA_{eq}, LUFS, sone, Phon, sound pressure (Pa), and intensity (W/m^2).
- **Room Acoustics Tool:** calculates reverberation from recorded impulse responses. It computes standard ISO 3382 [7] metrics ($RT_{10\text{-}30}$, EDT, D_{50},

C$_{50}$, C$_{80}$, G$_{Theo}$, STI, AL$_{consTheo}$, A$_{eq}$) and features octave band filtering (63 Hz – 8 kHz), target room evaluation profiles, and an integrated real-time auralization engine (see Fig. 1).



Target: 0.8–1.2s
Too reverberant (1.33s)
Ø T30 (MEAN) 1.33 s
BASS RATIO (BR) 0.89
Hover over chart lines to hear filtered band.

On	Freq	EDT (s)	T_{10} (s)	T_{20} (s)	T_{30} (s)	T_s (ms)	G_{theo} (dB)	C_{50}	C_{80}	D_{50}	STI	%AL$_{cons}$	A_{eq} (m^2)	SNR Quality
☑	Broadband	1.29	1.32	1.34	1.37	89	18.6	-0.7	2.0	46%	0.52	10.38%	70.6	✓
☑	63 Hz	1.36	1.52	1.63	1.53	79	19.1	1.0	3.3	56%	0.49	11.41%	63.2	✓
☑	125 Hz	1.41	1.47	1.34	1.30	84	18.3	0.5	2.4	53%	0.53	9.92%	74.4	✓
☑	250 Hz	1.37	1.31	1.29	1.24	84	18.2	-0.2	2.7	49%	0.54	9.58%	77.6	✓
☑	500 Hz	1.32	1.42	1.31	1.36	87	18.6	-1.1	2.1	44%	0.52	10.32%	71.1	✓
☑	1000 Hz	1.39	1.43	1.46	1.48	93	18.9	-0.7	1.9	46%	0.50	11.13%	65.1	✓
☑	2000 Hz	1.29	1.28	1.38	1.40	90	18.7	-0.7	2.2	46%	0.51	10.56%	69.2	✓
☑	4000 Hz	1.24	1.20	1.28	1.30	86	18.4	-0.9	2.3	45%	0.53	9.94%	74.3	✓
☑	8000 Hz	0.96	1.02	1.03	1.07	68	17.5	0.8	3.8	54%	0.57	8.43%	90.4	✓

Figure 1. Room Acoustics Tool: Reverberation analysis of a lecture hall's typical acoustic properties, including highlighting of parameters that could/should be improved.

- **Tempodrom**: The application first divides the audio into 12 distinct frequency bands [according to 8] and uses novelty curves [9] to detect sudden rhythmic peaks within each band. To mimic human perception, it calculates a perceptual score that factors in half and double tempos, while heavily weighting the results toward the natural human preference of 110 to 120 BPM [10]. Finally, it determines the overall tempo by merging these results, giving the highest importance to the lower bass frequencies since they primarily drive our physical sense of rhythm.
- **Clubbernaut**: uses the *Clubber.js* library [8] to analyze harmonic tension in music. By mapping the audio spectrum to a 128-note MIDI range, it enables real-time key and chord detection to calculate the composition's harmonic tension curve.

3. New AI-powered Tools

The integration of new advanced neural networks via transformers.js [6] has significantly expanded the SInES Tools' capabilities regarding semantic audio analysis and classification, e.g. with

- **Phoneme Recognizer**: Designed for phonetic analysis (powered by the *wav2vec2-lv-60-espeak* model [13]), this tool detects and maps individual phonemes (IPA) across a temporal timeline.
- **Speech Emotion Classifier**: This tool for emotional speech analysis (based on *CLAP htsat unfused* model [12]) detects tension, nervousness, joy, and other emotional cues in speech (see Fig. 2).

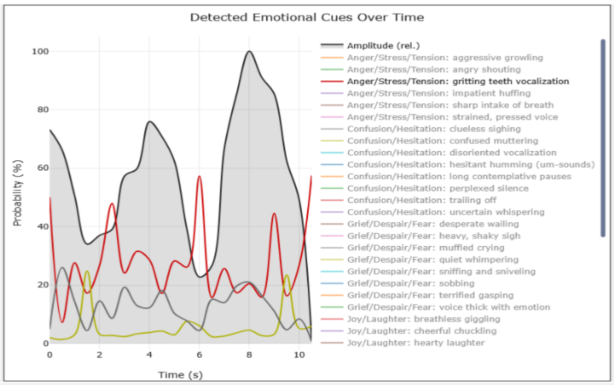


Figure 2. Speech Emotion Classifier: Automatic detection of emotional cues in talking or singing to assess the emotional state of singers and speakers. The detected emotional features can be saved as an Excel file, either synchronized with the audio amplitude or as mean values, standard deviations, and other metrics over the entire duration of the audio sample.

- **Sound Classifier**: This utility based on the *ast-finetuned-audioset* model [14] is capable of environmental and musical sound analysis, allowing users to detect and map individual sounds, noises, and timbres across a temporal timeline.
- **Instrument & Ensemble Recognizer**: Utilizing zero-shot audio classification via *Larger CLAP Music and Speech* model [12], this tool detects and maps musical textures, genres, and specific instruments across a temporal timeline.
- **Hardometer:** This tool allows to calculate an acoustic "hardness" score for uploaded music snippets based on Isabella Czedik-Eysenberg's perceptually evaluated *model of musical hardness* [15]. "Hardness" is a stylistic dimension typical for music genres such as hard rock and metal, but also some electronic music styles. The tool computes all features in exact accordance with the original implementation, ensuring full compatibility with the original model, which achieved a correlation of r = 0.91 with perceptual ratings on a test set (see Fig. 3).

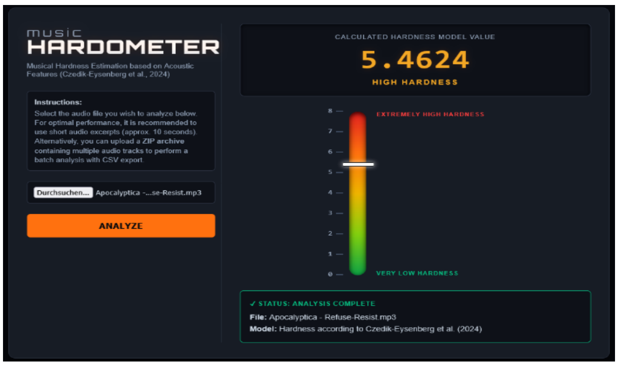


Figure 3. Hardometer: Musical "hardness" score calculation for uploaded music snippets. The tool also allows batch processing and download of the extracted values.

4. New (Re)Synthesis Tools

Between 2025 and 2026, SInES Tools was expanded to include a suite of rare spectral manipulation and synthesis tools. Key highlights include:

- **Spectral Editor:** offers professional audio restoration and forensic analysis. It features an interactive logarithmic spectrogram with versatile selection tools (Lasso, Rectangle, Horizontal Line) for precise frequency editing and artifact removal. Key capabilities include a fingerprint-based denoising engine, real-time threshold visualization, and time-domain effects like normalization and fades.
- **Modulation Transfer**: analyzes and transfers vibrato, tremolo, jitter, and shimmer between sounds (see Fig. 4). This category of audio synthesis applications also features timbral transfer tools like the **Bark Band Vocoder**, **Spectral Cross Synthesizer**, **Spectral Stamper**, and **Spectral Morpher**.

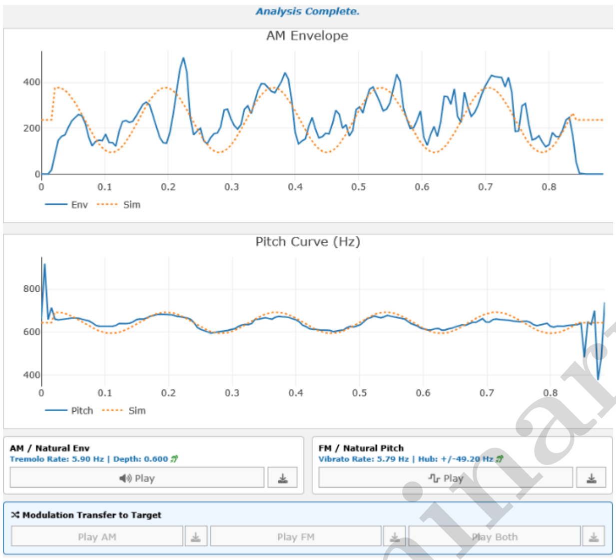


Figure 4. Modulation Transfer: Detected sound modulations such as vibrato, tremolo, jitter, and shimmer can be saved as numerical values in a .csv file and transferred to any other sounds.

- **Harmonic/Percussive Separator**: uses the FitzGerald algorithm [16] (spectral median filtering) to isolate sustained ("harmonic") and transient ("percussive") components (see Fig. 5). Related separation and filtering tools include the **Formant & Pitch Shifter**, **Spectral Gator**, **Harmonic Frequency Extractor** and **Harmonicity Analyzer**.
- **Spectral Warper**: Transforms harmonic overtone series into inharmonic structures by shifting partials (according to [17]) to explore the timbral influence of this manipulation on consonance and dissonance perception. It is complemented by extreme time-scaling tools like the phase-coherent **Time Stretcher**.

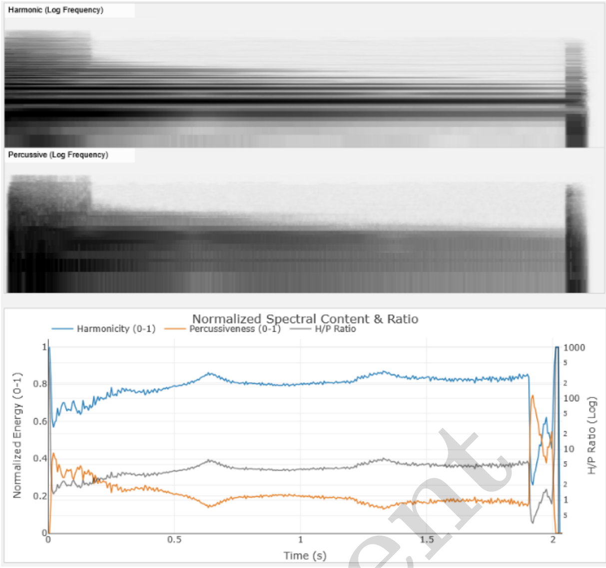


Figure 5. Harmonic/Percussive Separator: The sustained and transient components of a sound are separated and can be saved as both data and sounds.

5. New Video Analysis Tools

In addition to the already existing hand, face, expression and pose tracking SInES tools (based on *BlazePose* [18] and *MediaPipe* models [19]), the following new applications have been developed between 2025 and the 2026 conference, among others:

- **Multiple Person Pose Tracker**: Based on the *MoveNet* model [20], this tool provides frame-by-frame 2D video pose tracking for up to 6 persons.
- **Point Tracker**: Utilizing *OpenCV.js* with the Kanade-Lucas-Tomasi (KLT) feature tracker [21], this tool tracks up to 8 user-defined points of interest in videos.
- **Video Motion Analysis Tool**: Based on *OpenCV.js* and *RITMO VideoAnalysis* [22], this tool tracks the quantity of motion, motion centroid, motion periodicity, and optical flow (left – right / up – down) in a wide range of videos (see Fig. 6).

Figure 6. The Video Motion Analysis Tool displays the intensity, extent, and direction of movement of individuals, groups, and objects. The results can be saved as videos, images, and numerical values.

- **Video Motion Amplification Tool**: This tool magnifies nearly imperceptible movements as well as very subtle color differences in videos (according to [23]), which is highly useful for modal analysis, monitoring of the vibration characteristics of machines or human pulse or blood flow visualization.

6. New Motion Visualizers

To make sense of the vast amounts of tracking data generated, two powerful new visualization tools have been deployed. These visualizers for 2D and 3D coordinates bridge the gap between abstract data of tracked bodies, hands, faces and points and the visualization and statistics of human movement.

- **3D Data Analyzer:** This comprehensive visualizer allows for visualizing, analyzing, and editing 3D Qualisys MoCap data as well as body, hand, face and point tracking data from videos (as stand-alone and in combination). The motion trajectories for all body parts and points can be displayed and edited individually. It includes most of the functions of the *MATLAB MoCap Toolbox* [24], the *RITMO VideoAnalysis* [22] software and the PCA Movement Synergy Analysis [25] and it allows root centering (converting from MoCap to PoseTracking) as well as space reconstruction (converting from PoseTracking to MoCap). The recorded movements can be transferred to arbitrary 3D avatars which are equipped with the widely used *Mixamo* or *RPM* rig. The 3D avatars with baked animations can be saved as .glb files just like the videos of the moving figures or comprehensive CSV tables containing all results (see Fig. 7).

Figure 7. The 3D Data Analyzer is a powerful online toolbox for analyzing and visualizing 3D data, right at your fingertips, showing a PCA movement analysis of two dancers.

- **2D Data Analyzer:** works similarly to the 3D Data Analyzer with nearly the same functions and possibilities. In addition, the 2D movements of the body, hands, and faces can be recorded in *OpenPose ControlNet* mode to animate persons in photos and videos [26]. Facial expressions can be transferred to other images of faces using face mesh warping across 68 landmarks.

7. New Helpers

Data acquisition and pre-processing are often the most time-consuming aspects of empirical research. The SInES Helpers section has been heavily expanded to assist researchers before and after the actual analysis phase. In addition to the already existing range of statistical SInES tools (for t-tests, correlation analysis, time series analysis, inter-rater reliability, ANOVA, PCA, cluster analysis, etc.) and data visualization and processing tools (visualizers, box plotters, heatmapper, etc.), new applications for recording and editing audio and video files were added in 2026, such as

- **Audio Recorder**: This high-performance web-based tool records audio examples, includes microphone calibration, noise-print denoising, recordings of sounds from other websites and applications and downloads them as .wav files.
- **Video Recorder**: works like the Audio Recorder and allows users to record videos from a webcam or screen or upload videos to trim, crop, and download them as .webm or .mp4 files.
- **Batch Audio Loudness Normalizer**: allows users to upload audio files, align their perceptual loudness either individually, or match and maximize the loudness of an entire set of stimuli for listening experiments.
- **Sound Segmenter**: Operating as an AI sample slicer, this tool segments and slices audio samples by events and pitches (including automatic pitch detection), allowing users to download them in a batch as individual .wav files.

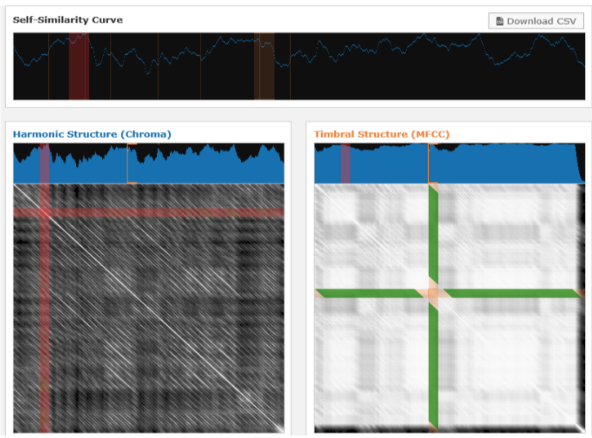


Figure 8. Audio Thumbnailer identifies the excerpt within a musical work that is the most prototypical in terms of harmony and timbre for the entire piece.

- **Audio Thumbnailer**: This tool creates audio thumbnails and timbral prototypes of longer musical works using Self-Similarity Matrices according to [27] (see Fig. 8).
- **Performance Wormhole**: generates performance worm visualizations from tempo and dynamics data (according to [28]), as well as from user-defined audio features or other time-series data, and exports them as videos, images, or Excel files (see Fig. 9).

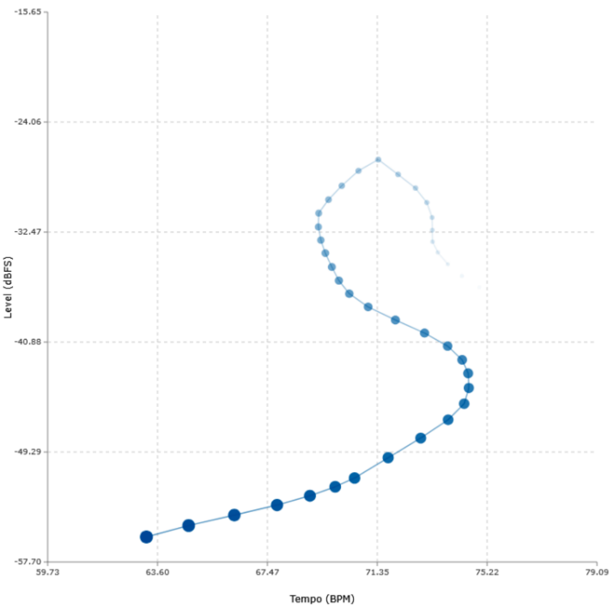


Figure 9. The Performance Wormhole can be used to visualize typical patterns in the interpretations of musical works. The method can be applied to all kinds of audio features and other data.

8. Conclusion and Outlook

With more than 90 apps, the SInES Tools democratize advanced research by turning any standard web browser into a fully equipped, client-side audio-visual laboratory, ensuring maximum data privacy. Driven by the core philosophy "It's online, it's free, it's one click away" and "Upload file → Analyze → Download results", recent expansions eliminate the need for costly software or expert programming skills. With powerful new AI audio classifiers, advanced motion tracking utilities, and comprehensive 2D/3D data visualizers, we invite researchers, students and enthusiasts to seamlessly integrate these professional-grade and experimental tools into their research workflows.
Looking ahead, the SInES group will further develop the platform as an open gateway to interdisciplinary, browser-based research. Upcoming features include:

- **Physiological Data Integration**: Browser-based capturing and analysis using *Muse2* EEG headbands [29] and *EmotiBit* sensors [30].
- **Enhanced Analytical Power**: The integration of further cutting-edge AI models, alongside innovative video and audio signal analysis tools.

9. Acknowledgments

We would like to extend our sincere thanks to the Faculty of Philosophy and Cultural Studies at the University of Vienna for its generous financial support of the SInES project's Chinese translation.

References

[1] C. Reuter, I. Czedik-Eysenberg, A.-X. Cui, S. Ambros, F. Klooss, G. Feller, I. Schachner, M. Roos, H. Lindenbauer, H. Li, and M. Bertsch, "SInES Tools, Helpers, Visualizers as your Scientific Realizers," in *Proc. of the 52nd Annual Conference of the German Acoustical Society (DAGA)*, (Dresden, Germany), pp. 1002–1005, 2026.

[2] A. Correya, J. Marcos-Fernández, L. Joglar-Ongay, P. Alonso-Jiménez, X. Serra, and D. Bogdanov, "Audio and Music Analysis on the Web using Essentia.js," in *Transactions of the International Society for Music Information Retrieval*, vol. 4 no. 1, pp. 167–181, 2021.

[3] H. Rawlinson, N. Segal, and J. Fiala, "Meyda: an audio feature extraction library for the Web Audio API," in *Proc. of the 1st Web Audio Conference* (WAC), (Paris, France), pp. 1–6, 2015.

[4] N. Jillings, J. Bullock, and R. Stables, "JS-Xtract: a realtime audio feature extraction library for the web," in *Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, (New York, USA), pp. 1–3, 2016.

[5] D. Smilkov, N. Thorat, Y. Assogba, A. Yuan, N. Kreeger, P., Yu, ... and M. Wattenberg, "TensorFlow.js: Machine Learning for the Web and Beyond," in *Proc. of the 2nd SysML Conference*, (Palo Alto, USA), pp. 1–13, 2019.

[6] J. Lochner, et al., *Transformers.js: State-of-the-art Machine Learning for the Web*. GitHub 2023. https://github.com/huggingface/transformers.js/ [Accessed: July 10, 2026].

[7] International Organization for Standardization. Acoustics, "*Measurement of room acoustic parameters Part 1: Performance spaces"* (ISO Standard No. 3382-1:2009), 2009.

[8] A. Klapuri, A. Eronen, and J. Astola, "Meter Analysis of Musical Signals," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 342–355, 2006.

[9] M. Müller, *Fundamentals of Music Processing,* Cham: Springer 2015 (pp. 303ff.).

[10] D. Moelants. "Preferred tempo reconsidered," *in Proc. of the 7th ICMPC,* (Sydney, Australia), pp. 580–583, 2002.

[11] Y. Gravezas, *Clubber.js: Application of music theory in audio reactive visualizations*. GitHub

2016. https://github.com/wizgrav/clubber [Accessed: July 10, 2026].

[12] Y. Wu, K. Chen, T. Zhang, Y. Hui, T. Berg-Kirkpatrick, et al., "Large-Scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Rhodes Island, Greece), pp. 1–5, 2023.

[13] Q. Xu, A. Baevski, and M. Auli, "Simple and effective zero-shot cross-lingual phoneme recognition," in *Interspeech 2022* (Incheon Korea), 2022.

[14] Y. Gong, Y.-A.Chung, and J. Glass, "AST: Audio Spectrogram Transformer," in *Interspeech 2021*, (Brno, Czech Republic), 2021.

[15] I. Czedik-Eysenberg, O. Wieczorek, A. Flexer, and C. Reuter, "Charting the Universe of Metal Music Lyrics and Analyzing Their Relation to Perceived Audio Hardness," in *Transactions of the International Society for Music Information Retrieval*, vol. 7, no. 1, 2024.

[16] D. FitzGerald, "Harmonic/Percussive Separation using Median Filtering," in *Proc. of the 13th International Conference on Digital Audio Effects (DAFx-10)*, (Graz, Austria), 2010.

[17] W.A. Sethares. *Tuning, Timbre, Spectrum, Scale*. London: Springer 2015.

[18] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device Real-time Body Pose tracking," in *arXiv:2006.10204*, 2020.

[19] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M.G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A Framework for Building Perception Pipelines," *arXiv:1906.08172,* 2019.

[20] Y. Zhang, and V. Bazarevsky, "Next-Generation Pose Detection with MoveNet and TensorFlow.js," in *Google TensorFlow Blog*, 2021, https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html [Accessed: July 10, 2026].

[21] C. Tomasi, and T. Kanade, *Detection and tracking of point features*, Technical Report, Carnegie Mellon University 1991.

[22] A.R. Jensenius, R.I. Godøy, and M.M. Wanderley, "Developing Tools for Studying Musical Gestures within the Max/MSP/Jitter Environment," in *Proc. of the International Computer Music Conference*, (Barcelona, Spain), pp. 282–285, 2005.

[23] H. Wu, M., Rubinstein, E., Shih, F., Durand, and W. Freeman, "Eulerian Video Magnification for Revealing Subtle Changes in the World," in *ACM Transactions on Graphics*, vol. 31, no. 4, Article 65, pp. 1–8, 2012.

[24] B. Burger, and P. Toiviainen, "MoCap Toolbox – A MATLAB toolbox for computational analysis of movement data," in *Proc. of the 10th Sound and Music Computing Conference (SMC)* (Stockholm, Sweden: KTH Royal Institute of Technology), 2013.

[25] F. Bigand, E. Prigent, B. Berret, and A. Braffort, "Decomposing spontaneous sign language into elementary movements: A principal component analysis-based approach," in *PLoS ONE* vol. 16, no. 10: e0259464, 2021.

[26] L. Zhang, A. Rao, and M. Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models," in *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, (Paris, France), pp. 3836–3846, 2023.

[27] J. Paulus, and A. Klapuri, "Music Structure Analysis Using a Probabilistic Fitness Measure and a Greedy Search Algorithm," in *IEEE Transactions on Audio, Speech, and Language Processing* vol. 17, no. 6, pp. 1159–1170, 2009.

[28] S. Dixon, W. Goebl, and G. Widmer, "The Performance Worm: Real Time Visualisation of Expression based on Langner's Tempo-Loudness Animation," in *Proc. of the International Computer Music Conference* (Gothenburg, Sweden), pp. 361–364, 2002.

[29] Muse2 *EEG headbands*: https://choosemuse.com/ [Accessed: July 10, 2026].

[30] S. Montgomery, N. Nair, H. Chen, and S. Dikker, *Validating EmotiBit, an open-source multi-modal sensor for capturing research-grade physiological signals from anywhere on the body* (October 5, 2022). Available at SSRN: https://ssrn.com/abstract=4700861 [Accessed: July 10, 2026].