**Supplementary material: Acoustic Feature Characteristics**

For the description of baby cries, sound characteristics are particularly evident in three areas:

1. Brightness
2. Hardness,
3. and temporal/modulatory components
4. **Timbre features correlating positively with the perceived timbral brightness**

**Brightness** = Ratio of amplitudes below 1500 Hz to above 1500 Hz (MIRtoolbox[1] / MiningSuite[2]).

**Brightness Combination** = Ratio of amplitudes below 1500 Hz to above 1500 Hz, which, at the same time, considers the fundamental (since the fundamental also influences the perceived brightness) (MIRtoolbox[1] / MiningSuite[2] /custom-made).

**Spectral Centroid Power Spectrum** = Centroid of the frequency spectrum as the weighted average of all contained frequency components (the higher the value, the brighter the perceived timbre impression) (MIRtoolbox[1] / MiningSuite[2]).

**Highest Peak Frequency** = Frequency of the strongest amplitude value in the spectrum (the higher the frequency, the brighter the perceived timbre impression)( MIRtoolbox[1] / MiningSuite[2]).

**Spectral Skewness** = Skewness of the spectral energy distribution (the higher the value, the brighter is the perceived timbre impression)(Essentia[3]).

**Spectral Bandwidth** = Spectral bandwidth according to Klapuri and Davy (2007)[4]: Width of the distribution of the frequency content over the spectrum (the higher the value, the brighter is the perceived timbre impression)(Librosa[5]).

**Timbral Brightness** = Calculated model for perceived tonal brightness (AudioCommons Timbral Models[6]).

**Timbral Sharpness** = Calculated model for perceived timbral sharpness, based on Fastl and Zwicker (2006)[7], (AudioCommons Timbral Models[6]).

**Zero Crossing Rate** = Zero crossing rate of the signal (how often the amplitude crosses the zero line per second: the more often, the more high-frequency components are in the signal, the brighter the perceived timbre impression)( MIRtoolbox[1] / MiningSuite[2]).

1. **Timbre features correlating negatively with the perceived timbral brightness**

**Spectral Energy 150 to 800 Hz** = Spectral energy between 150 and 800 Hz (the higher the value, the duller the perceived timbre impression)(Essentia[3]).

**Spectral Energy 250 to 400 Hz** = Spectral energy between 250 and 400 Hz (the higher the value, the duller the perceived timbre impression)(Essentia[3]).

**Timbral Warmth** = Calculated model for perceived timbral warmth (the higher the value, the duller is the perceived timbre impression)(AudioCommons Timbral Models[6]).

**Timbral Booming** = Calculated model for perceived timbral booming according to Hatano and Hashimoto (2000)[8], (AudioCommons Timbral Models[6]).

**MFCC1** = First coefficient to characterize the spectral structure of a sound (MFCC = Mel Frequency Cepstrum Coefficients according to Davis and Mermelstein (1980)[9], consisting of mostly 13 coefficients for the description of a mel cepstrum of a sound. A cepstrum is a discrete cosine transformation of a mel spectrum of a sound (the higher the value of the first coefficient, the duller is the perceived timbre impression)( MIRtoolbox[1] / MiningSuite[2]).

1. **Timbre features correlating with the perceived timbral hardness**

**Timbral Hardness** = Calculated model for perceived timbral hardness according to Pearce, Brookes, and Mason (2019)[10], (AudioCommons Timbral Models[6]).

**Hardness** = Calculated model for perceived musical hardness according to Czedik-Eysenberg (2021)[11].

**Percussive Energy** = Amount of percussive components within an audio signal (corresponding to the vertical lines in a sonogram)(Librosa[5]).

**Harmonic Percussive Ratio** = Ratio between percussive and tonal components in a signal (corresponding to the horizontal and vertical lines ratio in a sonogram)(Librosa[5]).

**Tonal Energy** = Energy of the tonal parts within a spectrum (correlates negatively with the perceived hardness: the fewer tonal parts, the harder the perceived timbre)( MIRtoolbox[1] / MiningSuite[2]).

1. **Timbre features correlating with the perceived temporal/modulatory properties of a sound**

**Timbral Roughness** = Calculated model for perceived timbral roughness according to Vassilakis and Fitz (2007)[12], (AudioCommons Timbral Models[6]).

**Rhythmic Irregularity** = Variations/deviations from a continuous pulse (MIRtoolbox[1] / MiningSuite[2]).

**Attack Time Gammaton 3** = Attack time of the frequency components in the 3rd gamma filter range (corresponding to a bandpass filter at 400 Hz)( MIRtoolbox[1] / MiningSuite[2]).

**Attack Time Gammaton 10** = Attack time of the frequency components in the 10th gamma filter range (corresponding to a bandpass filter at 8000 Hz)( MIRtoolbox[1] / MiningSuite[2]).

**References**

1. Lartillot O, Toiviainen P, Eerola T. A Matlab Toolbox for Music Information Retrieval. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R, editors. Data Analysis, Machine Learning and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 261-68.

2. Lartillot O. Miningsuite: A comprehensive Matlab framework for signal, audio and music analysis, articulation audio and symbolic approaches. Proceedings of the 16th Sound and Music Computing Conference (SMC). Málaga, Spain2019. p. 489-89.

3. Bogdanov D WN, Gómez E, Gulati S, Herrera P, Mayor O. Essentia: An Audio Analysis Library for Music Information Retrieval. Proceedings of the 14th Conference of the International Society for Music Information Retrieval (ISMIR). Curitiba, Brazil2013. p. 493-98.

4. Anssi Klapuri MD. Signal Processing Methods for Music Transcription. Springer New York, NY; 2006.

5. McFee B RC, Liang D, Ellis DP, McVicar M, Battenberg E, Nieto O. Librosa: Audio and music signal analysis in python. Proceedings of the 14th python in science conference. Austin, Texas, United States of Amerika2015. p. 18-25.

6. Pearce A, Brookes T, Mason R. Timbral attributes for sound effect library searching. Audio Engineering Society; 2017. p. 2-2.

7. Eberhard Zwicker HF. Psychoacoustics - Facts and Models. 2 ed. Berlin: Springer Berlin, Heidelberg; 2006.

8. Hatano S HT. Booming Index as a Measure for Evaluating Booming Sensations. Proceedings of the 29th International Congress and Exhibition on Noise Control Engineering (Internoise). Nice, France2000.

9. Davis SB, Mermelstein P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. Ieee Transactions on Acoustics Speech and Signal Processing. 1980;28(4):357-66.

10. Pearce A, Brookes T, Mason R. Modelling Timbral Hardness. Applied Sciences-Basel. 2019 Feb 1;9(3).

11. Isabella C-E. Semantic modelling of perceptual-psychological music dimensions on the basis of acoustic signal properties. Musicological Department, Musical Acoustics and Music Psychology: University of Vienna; 2021.

12. Vassilakis PN FK. SRA: A web-based research tool for spectral and roughness analysis of sound signals. Proceedings of the 4th Sound and Music Computing (SMC. Athens, Greece2007. p. 319-25.