INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

AUDIO LABS

Music Processing Analysis

Music Synchronization I

Exercise

Stefan Balke, Meinard Müller

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer IIS

Session Outline
Music Synchronization I

- Homework discussion
- Reading Assignments
- Comparison of log.-freq. Representations in Python

Homework

Exercise 3.3

Exercise 3.3. Let $\mathrm{BW}(p) = F_{\mathrm{pitch}}(p+0.5) - F_{\mathrm{pitch}}(p-0.5)$ be the bandwith for a pitch p as defined in (3.5). What is the relation between the bandwidths $\mathrm{BW}(p+12)$ and $\mathrm{BW}(p)$ of two pitches that are one octave apart? Give a mathematical proof for your claim. Similarly, determine the relation between the bandwidths $\mathrm{BW}(p+1)$ and $\mathrm{BW}(p)$ of two neighboring pitches.

Homework
Solution 3.3

Solution to Exercise 3.3. Using (3.2), we obtain

$$F_{\mathrm{pitch}}(r+12) = 2^{(r+12-69)/12} \cdot 440 = 2 \cdot F_{\mathrm{pitch}}(r)$$

for any $r \in \mathbb{R}$. Applying this to $r = p - 0.5$ and $r = p + 0.5$, we obtain from (3.5):

$$\begin{aligned}\mathrm{BW}(p+12) &= F_{\mathrm{pitch}}(p+12+0.5) - F_{\mathrm{pitch}}(p+12-0.5)\\ &= 2 \cdot F_{\mathrm{pitch}}(p+0.5) - 2 \cdot F_{\mathrm{pitch}}(p+0.5)\\ &= 2 \cdot \mathrm{BW}(p).\end{aligned}$$

In other words, increasing the pitch by one octave increases the bandwith by a factor of two. Similarly, one shows that $\mathrm{BW}(p+1)/\mathrm{BW}(p) = 2^{1/12}$ (see also Exercise 1.6).

Homework
Exercise 3.4

Exercise 3.4. Given an audio signal at a sampling rate of $F_s = 22050$ Hz, we want to compute a log-frequency spectrogram as in (3.4). As a requirement, all sets $P(p)$ (as defined in (3.3)) for all pitches corresponding to the notes C2 ($p = 36$) to C3 ($p = 48$) should contain at least four Fourier coefficients. To meet this requirement, what is the minimal window length N (assuming that N is a power of two) to be used in the STFT? For this N, determine the elements of the set $P(36)$ explicitly.

Homework

Solution 3.4

Solution to Exercise 3.4. For pitch $p = 36$ (corresponding to C2), one obtains $F_{\mathrm{pitch}}(p) = 65.41$ Hz, $F_{\mathrm{pitch}}(p-0.5) = 63.54$ Hz, $F_{\mathrm{pitch}}(p+0.5) = 67.32$ Hz, and $\mathrm{BW}(p) = 3.78$. Furthermore, for $F_{\mathrm{s}} = 22050$ Hz and $N = 32768$, it follows from (2.28) that $F_{\mathrm{coef}}(96) = 63.93$ Hz, $F_{\mathrm{coef}}(101) = 67.29$ Hz, and $P(36) = \{96, \ldots, 101\}$. In other words, $P(36)$ contains six elements for $N = 32768$. This also shows, that $P(p)$ must contain at least four elements for $p > 36$. Finally, for the window size $N = 16384$, the set $P(36)$ contains only three elements. Therefore, $N = 32768$ is the minimal window length with the desired property.

Homework

Exercise 3.5

Exercise 3.5. The tuning of musical instruments is usually based on a fixed reference pitch. In Western music, one typically uses the **concert pitch** A4 having a frequency of 440 Hz (see Section 1.3.2). To estimate the deviation from this ideal reference, a musician is asked to play the note A4 on his or her instrument over the duration of four seconds. Describe a simple FFT-based procedure for estimating the tuning deviation of the instrument used. How would you choose the parameters (sampling rate, window size) to obtain an accuracy of at least 1 Hz in this estimation?

Homework
Solution 3.5

Solution to Exercise 3.5. Playing a note A4, one can expect dominant frequencies in a neighborhood of 440 Hz (corresponding to the fundamental frequency) and its integer multiples (corresponding to the harmonics). One basic procedure is to first compute a DFT of the recorded signal to obtain a spectral representation. Then, one may look for the frequency index k_0 that yields a maximal magnitude coefficient $|F_{\mathrm{coef}}(k_0)|$ in a neighborhood of 440 Hz (e.g., plus/minus a semitone). The difference $F_{\mathrm{coef}}(k_0) - 440$ Hz then yields an estimate of the tuning deviation.

Assuming a sampling rate of $F_{\mathrm{s}} = 44100$ Hz, one may compute a DFT using a window size of $N = 2^{17} = 131072$ (corresponding to 2.97 sec). This yields a spectral resolution of $F_{\mathrm{s}}/N \approx 0.34$ Hz.

To obtain a more robust estimate, one may also consider spectral peak positions in suitably defined neighborhoods of the harmonics. The resulting deviations from the ideal positions of the individual harmonics can be used to derive a single tuning estimate using a suitable fusion strategy.

Homework

Exercise 3.6

Exercise 3.6. Assume that an orchestra is tuned 20 cents upwards compared with the standard tuning. What is the center frequency of the tone A4 in this tuning? How can a chroma representation be adjusted to compensate for this tuning difference?

Homework
Solution 3.6

Solution to Exercise 3.6. The detuning of 20 cents corresponds to a fifth of a semitone. Therefore, by (3.2), the center frequency of the tone A4 in the given tuning is

$$F_{\mathrm{pitch}}(69.2) = 2^{(69.2-69)/12} \cdot 440 \approx 445.1 \text{ Hz}.$$

Using this frequency as a new reference, we define the function

$$F'_{\mathrm{pitch}}(p) = 2^{(p-69)/12} \cdot 445.1.$$

Based on this modified function, we define a modified set

$$P'(p) := \{k : F'_{\mathrm{pitch}}(p-0.5) \leq F_{\mathrm{coef}}(k) < F'_{\mathrm{pitch}}(p+0.5)\}$$

for each pitch $p \in [0:127]$ (see 3.3). From this, we obtain an adjusted log-frequency spectrogram (see (3.4)), from which we can derive an adjusted chroma representation as before (see (3.6)).