BRNO UNIVERSITY OF TECHNOLOGY | FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023

Gitaro

Learn to play guitar
using sound detection technology

Michal Mikota

- Real-time tablature
- Change BPM on every song
- Guitar tuner
- Played note detection
- 100 + songs
- No music skills required

CREPE algorithm

This application uses the CREPE algorithm, which is based on a deep convolutional neural network operating directly on the input of the time domain. It is one of the most advanced and accurate algorithms for pitch detection, filtering out unwanted artifacts with an accuracy between 95-98\%. The algorithm takes 1024 samples from the time domain audio signal with a sampling frequency of 16 kHz as input. There are six convolutional layers whose result is a 2048-dimensional latent representation, which is connected to the output layer with a sigmoid activation corresponding to a 360-dimensional output vector.

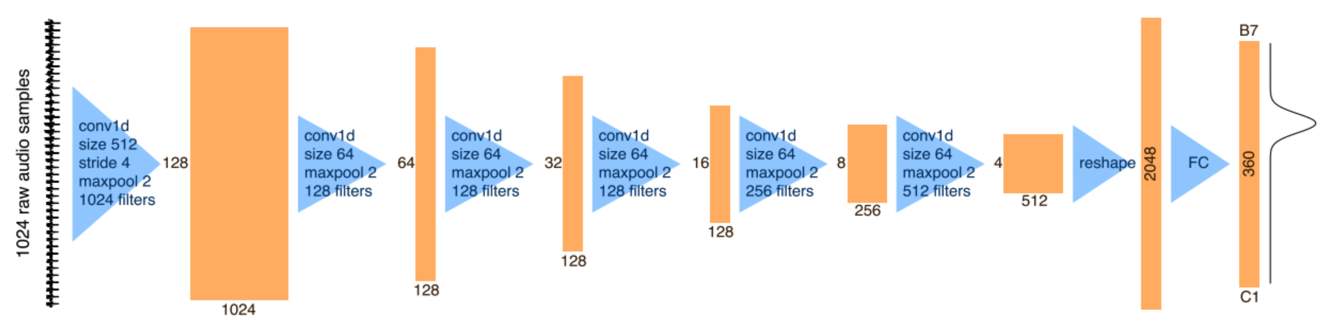


Fig. 1: The architecture of the CREPE pitch tracker. The six convolutional layers operate directly on the time-domain audio signal, producing an output vector that approximates a Gaussian curve as in Equation 3, which is then used to derive the exact pitch estimate as in Equation 2.

The pitch values used in the system are labeled as ¢1, ¢2, ..., ¢360, and are chosen to span six octaves with 20-cent intervals between the notes C1 and B7, which correspond to frequencies ranging from 32.70 Hz to 1975.5 Hz. Fref has default value 10 Hz. The estimated pitch value, ˆ¢, is determined by calculating a weighted average of the pitch values ¢i using the output yˆ, which represents the estimated frequency in Hz.

$$\hat{\text{¢}} = \frac{\sum_{i=1}^{360} \hat{y}_i \text{¢}_i}{\sum_{i=1}^{360} \hat{y}_i}, \qquad \hat{f} = f_{\text{ref}} \cdot 2^{\hat{\text{¢}}/1200}.$$

Equation.2: Calculation of the estimated pitch value from notes between C1 and B7

To train the model, we use 360-dimensional vectors as target outputs, where each dimension corresponds to a frequency bin with a width of 20 cents, which is the same as the model's output. The frequency bin associated with the actual fundamental frequency is assigned a value of one. To reduce the penalty for almost correct predictions, we follow the approach and apply a Gaussian blur to the target in the frequency domain. This blurring causes the energy around the ground truth frequency to decrease gradually with a standard deviation of 25 cents.

$$y_i = \exp\left(-\frac{(\text{¢}_i - \text{¢}_{\text{true}})^2}{2 \cdot 25^2}\right)$$

Equation.3: Application of the Gaussian blur – result causing energy around the ground truth frequency

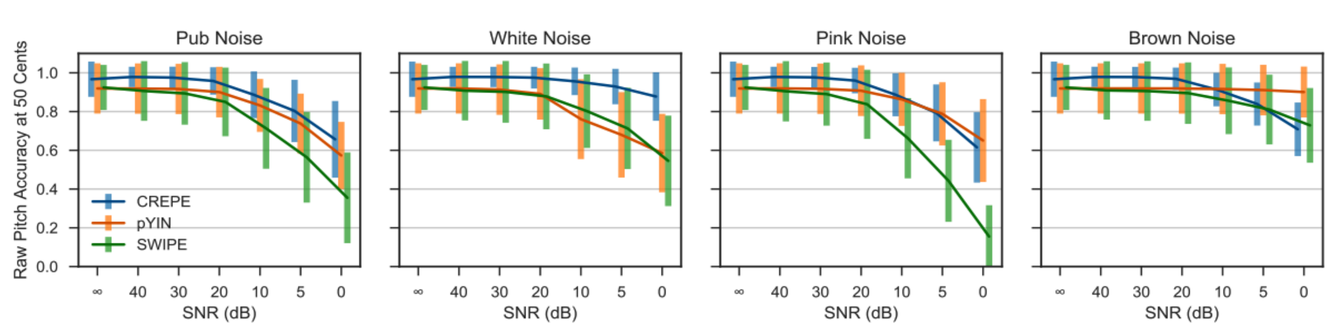


Fig. 2: Pitch tracking performance when additive noise signals are present. The error bars are centered at the average raw pitch accuracies and span the first standard deviations. With brown noise being a notable exception, CREPE shows the highest noise robustness in general.

Music tabs

The problem of aligning musical input with the corresponding score (sheet music) involves the system being able to identify the notes being played in the music and display them at the corresponding position in the score. This problem is typically addressed through multimodal machine learning, which utilizes information from the audio signal and visual information from the score.

Training such a system requires proper synchronization between the sound and the coordinates of the notes in the score images, so that the system can identify which notes are being played at a particular time and where they are located in the score. This can be challenging because sound and visual information are processed in different ways and can be affected by various factors, such as delay in data transmission.

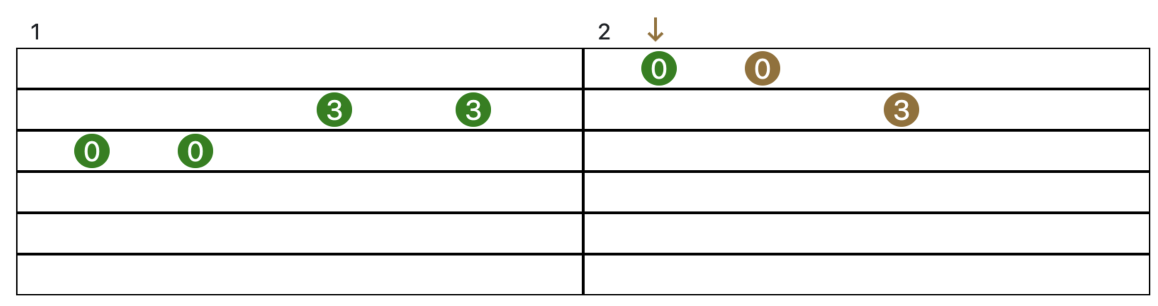
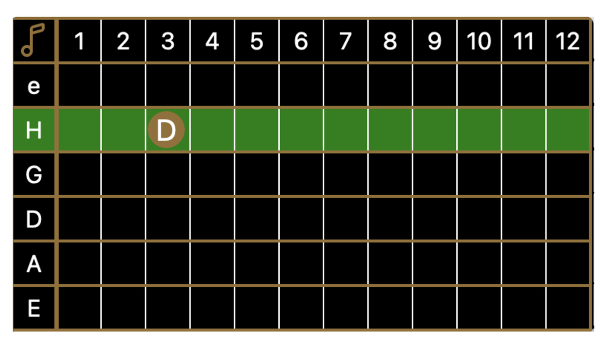
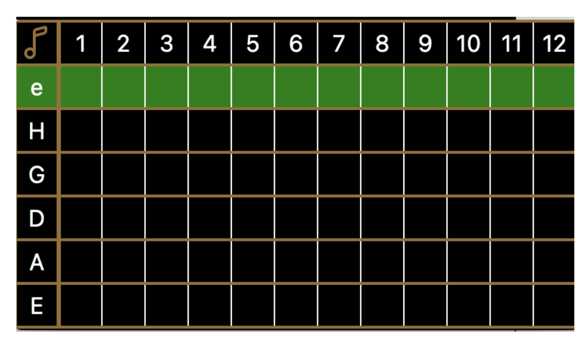


Fig.3: Generated tabs based on music note and note duration => Note ('H3', 1) which means music note is on the second string on the third fret with duration as quarter music note.

Note ('H3',2,tabData)



Note ('E0',1,tabData)



Custom guitar tuner

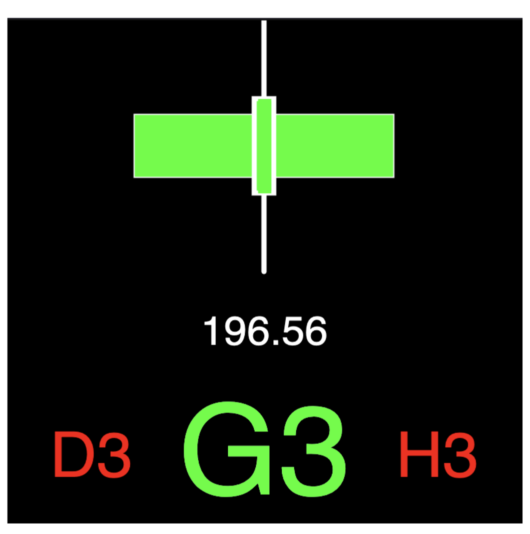


Fig.4: Preview of correct tuned string with guitar tuner

One of the most complex tools for implementing filtering and tone detection, as well as setting up a user-friendly interface, is certainly a built-in tuner.

In music or acoustics, we understand a tone as a sound that has a periodic or at least approximately periodic oscillation. It can arise in various ways, but it always involves a certain type of vibration (strings, vocal cords, etc.). The main attributes of tones in music are pitch, intensity, duration, and color. The opposite of a tone can be considered as noise (sound artifacts) that do not have a periodic waveform.

The pitch of a tone is determined by its frequency of oscillation. The higher the frequency, the higher the number of oscillations, and therefore, the higher the pitch.