A MULTIMODAL APPROACH TO ACOUSTIC GUITAR STRUMMING ACTION TRANSCRIPTION

Sebastian Murgul[1,2] **Michael Heizmann**[1]

[1] Institute of Industrial Information Technology, Karlsruhe Institute of Technology, Germany
[2] Klangio GmbH, Alter Schlachthof 39, Karlsruhe, Germany
sebastian.murgul@klangio.com, michael.heizmann@kit.edu

ABSTRACT

Strumming is a guitar playing technique where instead of playing individual notes, the player strokes over the strings of a chord in an up- or downwards direction. Due to the very short time difference between the string hits, conventional note-based transcription methods cannot distinguish the direction of the strumming movement and hence are not suited for rhythm guitar transcription. Therefore, a multimodal approach to strumming action transcription is proposed, combining audio recordings with the motion of the hand. The audio signal is used for strumming event detection and the motion signal is used for direction classification. To measure the motion of the hand, a 6-axis gyroscope and accelerometer is mounted on the back of the hand which sends measurements to a computer. For evaluation purposes, a recording of five minutes has been labelled manually and is published along with this work. On this test dataset, the presented multimodal approach reaches an F1 score of 85 % for the up and 92 % for the down strumming transcription task.

1. INTRODUCTION

Automatic music transcription is a current task in the field of Music Information Retrieval (MIR) with the goal of retrieving a symbolic music representation from an input audio signal. For the transcription of solo instrument music, numerous new approaches and tools have been proposed over the last few years [1]. While tools for automatic guitar transcription tools like Guitar2Tabs [1] can do a decent job transcribing finger picking recordings, the input of strumming audio leads to an unsatisfactory transcription result.

In 2019, Bello et al. proposed a neural network based classification system to distinguish between up and down strokes [2]. As input feature representation, Mel-frequency Cepstral Coefficients are calculated from the manually segmented input signal. Training on these features results in a classification accuracy of 72.5 % for a CNN and 70 % for an LSTM model.

Another publication that deals with guitar strumming analysis has been published by Matsushita et al. in 2013 [3]. Here, a wristwatch-like device has been developed to analyze down strumming action musically in means of timing and strength of the notes. Guitar strumming gestures have also been analyzed more in detail using an IMU and motion capturing by Freire et al. in 2020 [4].

In this work, a multimodal approach to guitar strumming action transcription is proposed. Therefore, up and down strumming events shall be detected and classified. First, the data acquisition of the audio and the motion signal using the developed hardware is presented. Then, a method for strumming event onset detection and an approach for direction classification based on the accelerometer data is shown. Finally, the transcription results are evaluated using an annotated test dataset.

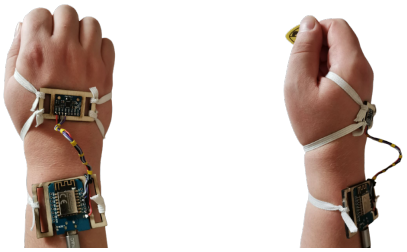

Figure 1: Prototype of the motion recording device mounted on the back of the hand.

2. DATA ACQUISITION

Since the classification of strumming direction is difficult when relying only on the audio input, a multimodal recording approach is used. Therefore, a GY-521 6-axis gyroscope and accelerometer sensor is mounted on the back of the hand of the guitar player. The sensor communicates over i2c with an ESP8266 microcontroller that sends the measurements to the computer over USB. The microcontroller is mounted around the arm wrist to not disturb while playing the guitar. The mounted device is shown in Figure 1.

[1] https://klangio.com/guitar2tabs

Late-Breaking / Demo Session Extended Abstract, ISMIR 2022 Conference

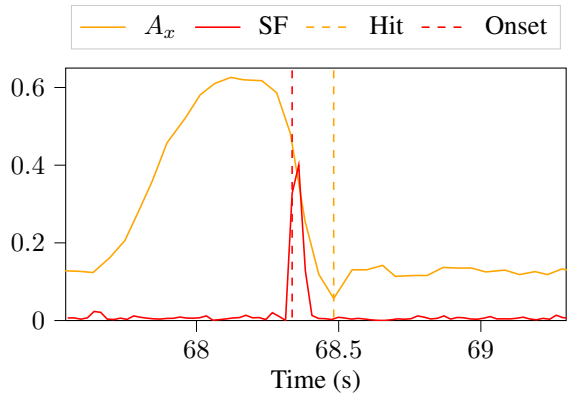

Figure 2: Latency measurement after one minute of recording.

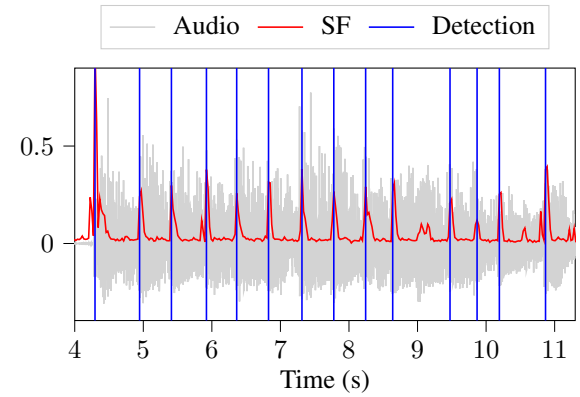

Figure 3: Strumming event onset detection using the spectral flux onset detection function.

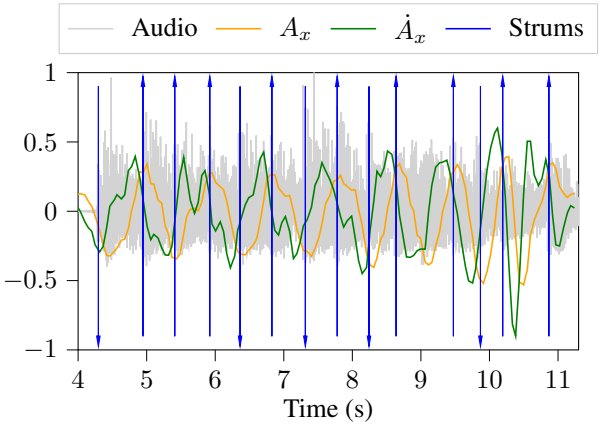

Figure 4: Classification of motion direction using the first order derivative in time direction.

The most meaningful signal provided by the sensor is the accelerometer value A_x pointing towards the fingers. To some extent, the value A_x picks up gravity and the centrifugal force caused by the strumming movement. A clearly visible peak down in A_x can be observed when the hand strums downwards, and a peak up when the hand strums upwards.

The hand motion sensor values itself do not allow detecting if the guitar strings are actually hit. Therefore, the audio signal from the guitar pickup is additionally recorded with a sample rate of 44.1 kHz. The motion signal is recorded with 20 Hz.

3. DATA PROCESSING

3.1 Synchronization

To synchronize the audio and motion signals, the latency of the hardware setup needs to be determined. Therefore, a lid has been constructed on which the motion sensor can be mounted. By closing the lid on a guitar string, the hit can be seen in the audio as well as in the motion signal. In Figure 2 a plot is shown with the spectral flux (SF) as audio onset indicator together with the motion signal A_x. By selecting the spectral flux peak and the time when the motion sensor reaches the string again, the latency can be determined as around 150 ms in this case.

3.2 Strumming Event Detection

To detect strumming event onsets, an onset detector like spectral flux [5] can be used. To remove double hit detections, a pruning with a difference of 150 ms is performed. In Figure 3, the strumming event detection results for an example recording are shown.

3.3 Strumming Event Classification

For the classification of the detected strumming events, the motion sensor signal is used. After smoothing the signal using a moving average filter, the temporal derivation $\dot{A}_x$ is calculated. When the value of $\dot{A}_x$ at an onset time is positive, an up movement is performed by the hand and the strumming event can be classified as an up strum. Vice versa for a negative value and down strumming. An example plot is shown in Figure 4.

3.4 Results

For evaluation purposes, a subset of five minutes of recording is annotated by hand. The test dataset including the audio and motion recordings as well as the labels are freely available online [2]. For evaluation, the mir_eval library is used [6]. An event is detected correctly if the onset is within 100 ms to its reference. For up strums, the multimodal approach reaches an F1 score of 85 %, for down strums 92 %.

4. FUTURE WORK

In future work, this approach will be used to create a larger scale dataset of annotated real world strumming recordings suitable for training neural networks on the task of audio based strumming tracking. By also estimating the played chords using a chord recognition method, strumming recordings could be transcribed into tablatures for e.g. rhythm guitar.

[2] https://github.com/klangio/KLANGIO-GST-MM-T

5. REFERENCES

[1] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic music transcription: An overview," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20–30, 2018.

[2] K. Bello and P. Mayol, "Classification of acoustic guitar strum using convolutional neural networks and long-short-term-memory," *Philippine e-Journal for Applied Research and Development*, vol. 9, pp. 49–57, 2019.

[3] S. Matsushita and D. Iwase, "Detecting strumming action while playing guitar," in *Proc. of the 2013 International Symposium on Wearable Computers*, 2013, pp. 145–146.

[4] S. Freire, G. Santos, A. Armondes, E. Meneses, and M. Wanderley, "Evaluation of inertial sensor data by a comparison with optical motion capture data of guitar strumming gestures," *Sensors*, vol. 20, no. 19, p. 5722, 2020.

[5] S. Böck and G. Widmer, "Maximum filter vibrato suppression for onset detection," in *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx). Maynooth, Ireland (Sept 2013)*, vol. 7, 2013, p. 4.

[6] C. Raffel, B. Mcfee, E. Humphrey, J. Salamon, O. Nieto, D. Liang, and D. Ellis, "mir_eval: A transparent implementation of common MIR metrics," in *In Proc. of the 15th International Society for Music Information Retrieval Conference, ISMIR*, 2014.