# AUDIO-DRIVEN VIOLIN PERFORMANCE ANIMATION WITH CLEAR FINGERING AND BOWING

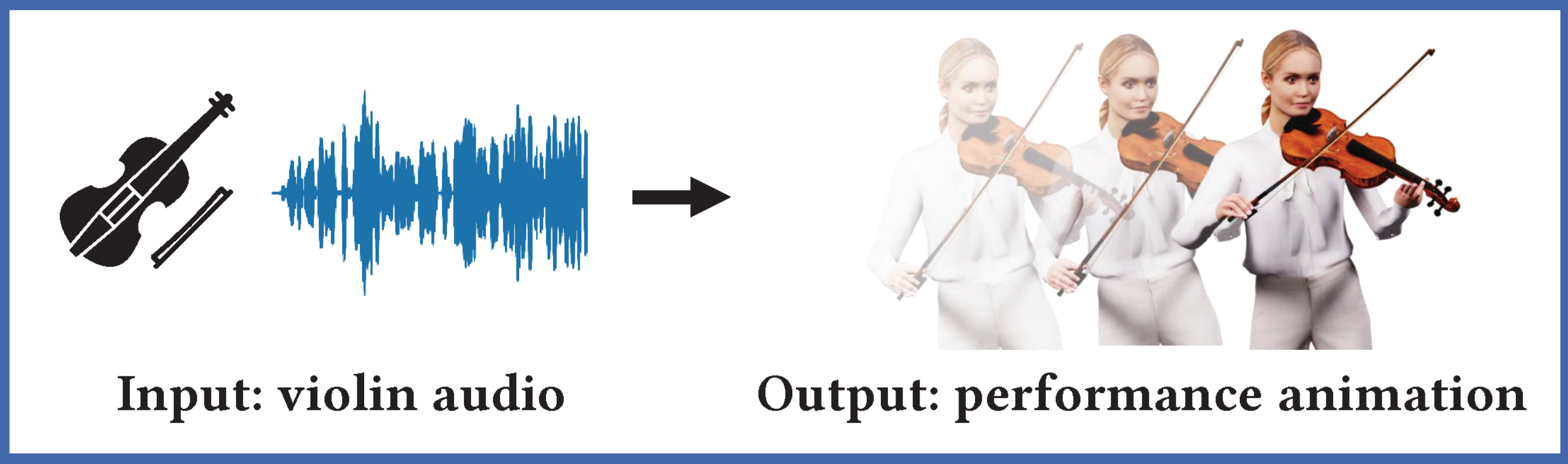
SIGGRAPH 2022
VANCOUVER+ 8-11 AUG

ASUKA HIRATA[1], KEITARO TANAKA[1], MASATOSHI HAMANAKA[2], SHIGEO MORISHIMA[3]

[1]WASEDA UNIVERSITY, [2]RIKEN, [3]WASEDA RESEARCH INSTITUTE FOR SCIENCE AND ENGINEERING

## BACKGROUND

Automatic body motion generation from audio for musical performance has attracted increasing attention recently. This technique enables anyone to easily create musical performance animations for their favorite audio without special devices.



## RELATED WORK / MOTIVATION

To generate body motion for violin performance, [Shlizerman et al. 2018] and [Kao and Su 2020] directly estimated the sequence of body joint positions from audio features.

These were pioneering works in trying to solve this task in an end-to-end manner.

However, it is still challenging to create adequate motion due to the ambiguous mapping between observed audio features and joint positions, as musical performance requires precise motion to produce the target sound.

## YOUR APPROACH / SOLUTION

In contrast to the studies above, we tackle this problem by estimating the playing procedure information for hands from audio features prior to motion synthesis in order to make the target domain simpler and more relevant to audio features.

## REFERENCES

Asuka Hirata, Keitaro Tanaka, Ryo Shimamura, and Shigeo Morishima. Bowing-Net: Motion Generation for String Instruments Based on Bowing Information. (SIGGRAPH posters 2021)

Hsuan-Kai Kao and Li Su. Temporally Guided Music-to-Body-Movement Generation. (ACM MM 2020)

Eli Shlizerman, Lucio Dery, Hayden Schoen, and Ira Kemelmacher-Shlizerman. Audio to body dynamics. (CVPR 2018)

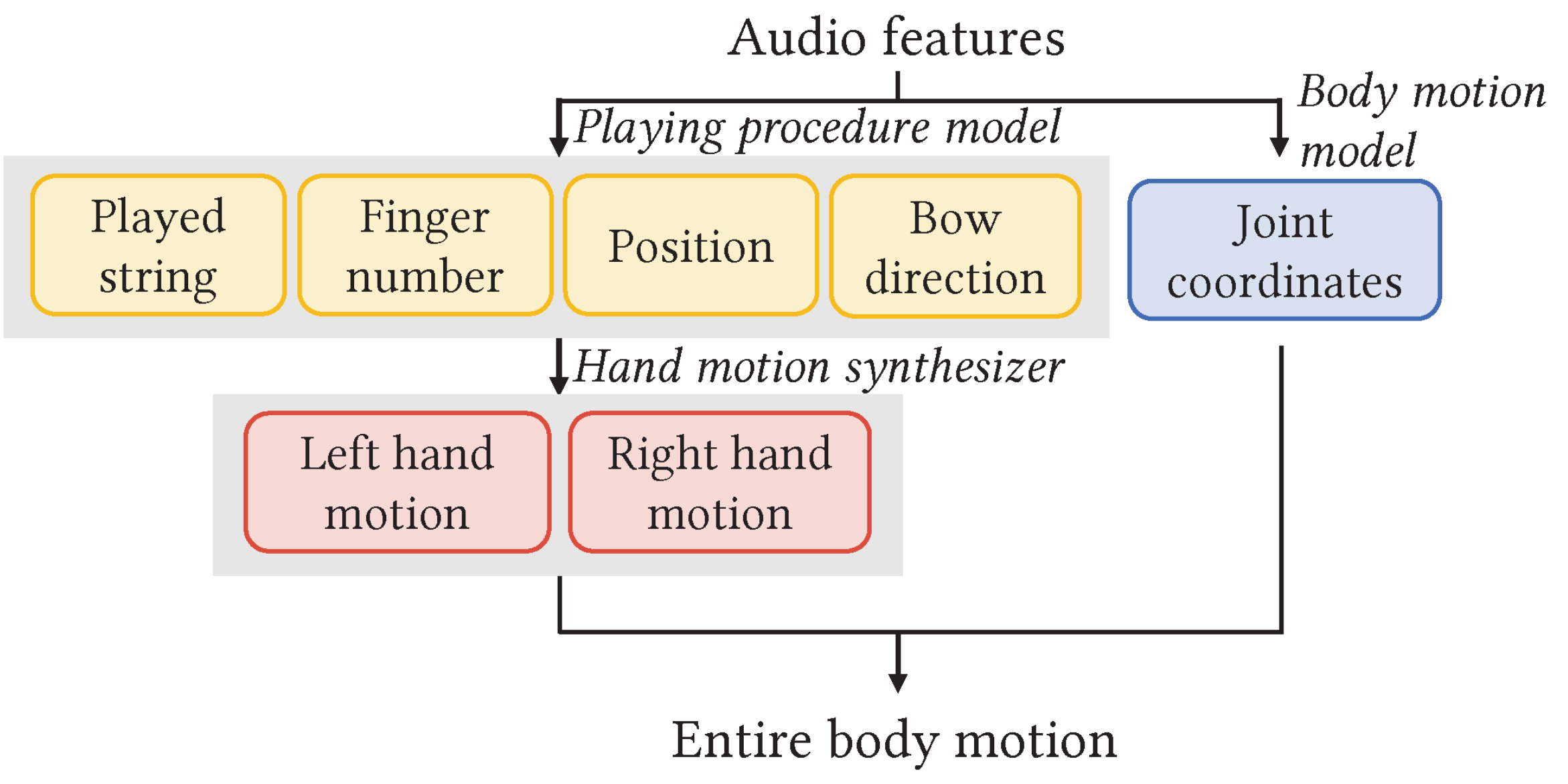
Email of Corresponding Author :
asuka112358@suou.waseda.jp

This work was supported by JSPS KAKENHI Grant Numbers JP21H05054, JP19H04137, JP21H03572.

## METHOD / PIPELINE / ALGORITHM / PROCESS

Our method consists of three components, a playing procedure model, a hand motion synthesizer, and a body motion model.

The playing procedure model estimates four hand-related features consisting of the played string, the finger number, the position, and the bow direction.



The hand motion synthesizer generates the left hand motion from the string, the finger, and the position features and the right one from the string and the bow features.

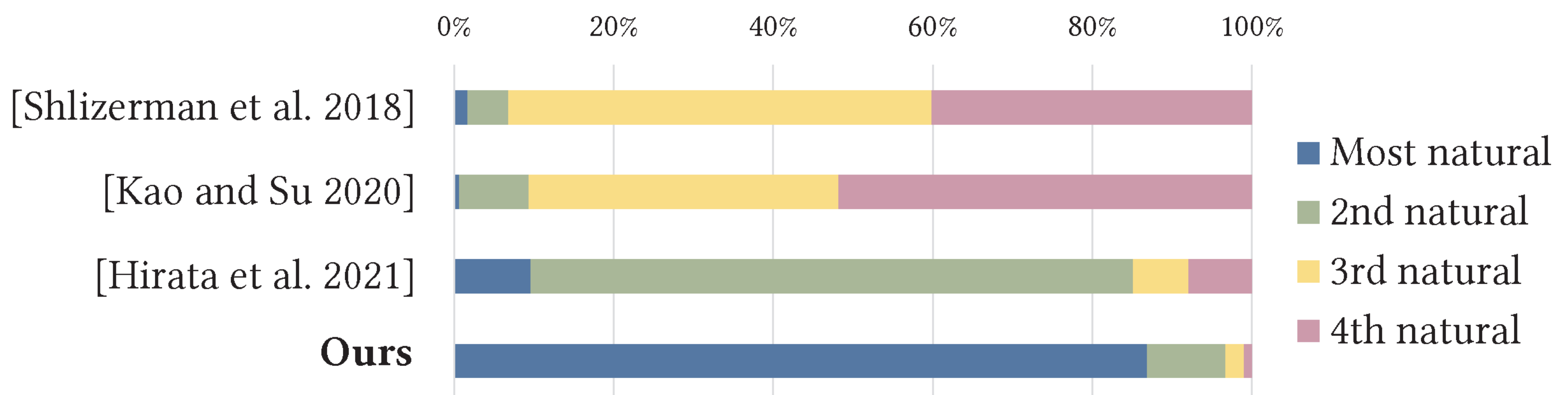
The body motion model estimates the 3D coordinates for the entire body.

## RESULTS

We evaluated our results quantitatively and qualitatively. To train our model, we constructed a new dataset consisting of one hour (30 pieces) of violin performance by a single player, which provides paired audio, joint coordinates from video, and aligned playing procedure information.

To evaluate whether the final output accurately corresponds to the input audio in terms of temporal synchronization and playing procedure accuracy, we determined the F-measure for the onset and the accuracies for the playing procedure information. The results showed that the F-measure for the onset at tolerances of 0.033, 0.067, and 0.100 seconds were 0.582, 0.867, and 0.913 and the accuracies for the played string, the finger number, the position, and the bow direction were 0.965, 0.879, 0.892, and 0.820, respectively.

We also conducted a subjective evaluation with 30 participants to compare our method with existing ones. We randomly chose ten excerpts from our dataset and synthesized the animation with each method. The participants were instructed to rank the naturalness of the synthesized animation. The results are shown in the Figure below, where we can see that our results are far more natural than the others.



Watch the result video at
bit.ly/3QpGHP7