Stanford University CCRMA

National Institute of Advanced Industrial Science and Technology AIST

InstListener: An Expressive Parameter Estimation System Imitating Human Performances of Monophonic Musical Instruments

Kitty Zhengshan Shi
CCRMA, Stanford University
kittyshi@ccrma.stanford.edu

Tomoyasu Nakano, Masataka Goto
AIST, Japan
t.nakano@aist.go.jp m.goto@aist.go.jp

I. Introduction

InstListener is a system that takes an expressive monophonic solo instrument performance by a human performer as the input and imitates its audio recordings by using an existing MIDI synthesizer. It automatically analyzes the input and estimates, for each musical note, expressive performance parameters such as the timing, duration, discrete semitone-level pitch, amplitude, continuous pitch contour, and continuous amplitude contour.

The system uses an iterative process to estimate and update those parameters by analyzing both the input and output of the system so that the output from the MIDI synthesizer can be similar enough to the input.

Keywords: *performance imitation, expressive musical performance, iterative parameter estimation, performance synthesis by analysis, musical expression*

II. InstListener

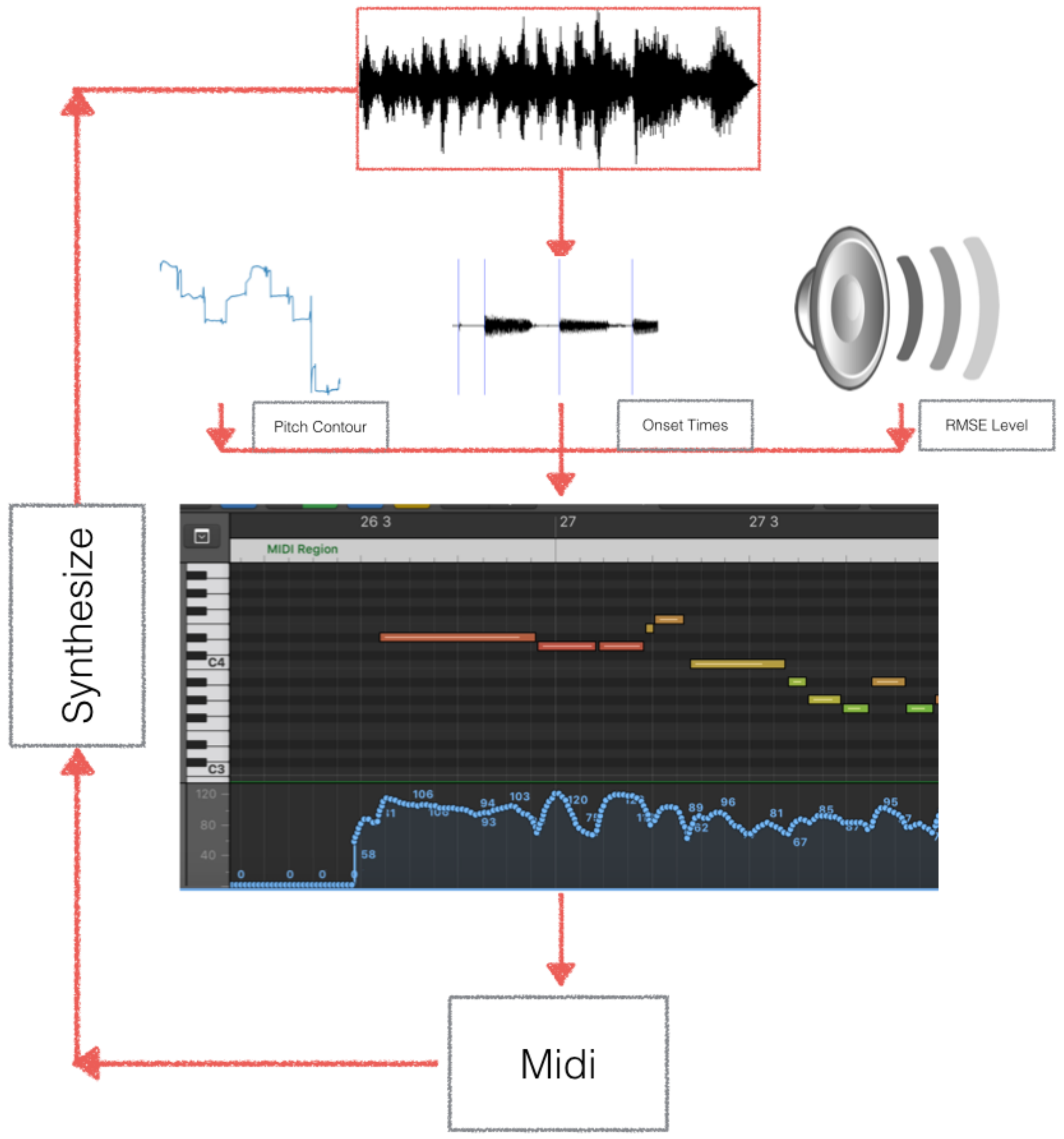


Fig 1. Workflow of InstListener

2.1 Feature Extraction
- Note onset detection
- Note pitch contour
- Root-mean-square energy

2.2 Parameter Mapping
- Pitch contour: MIDI note number and pitch bend
- RMSE: MIDI velocity level

2.3 Iterative Listening Process
- Perform dynamic time warping (DTW) between the pitch contour of the input and the pitch contour of the output.
- InstListener adjusts and updates not only pitch contours, but also onset times.

Adjustment Results of the Iterative Listening Process by InstListener

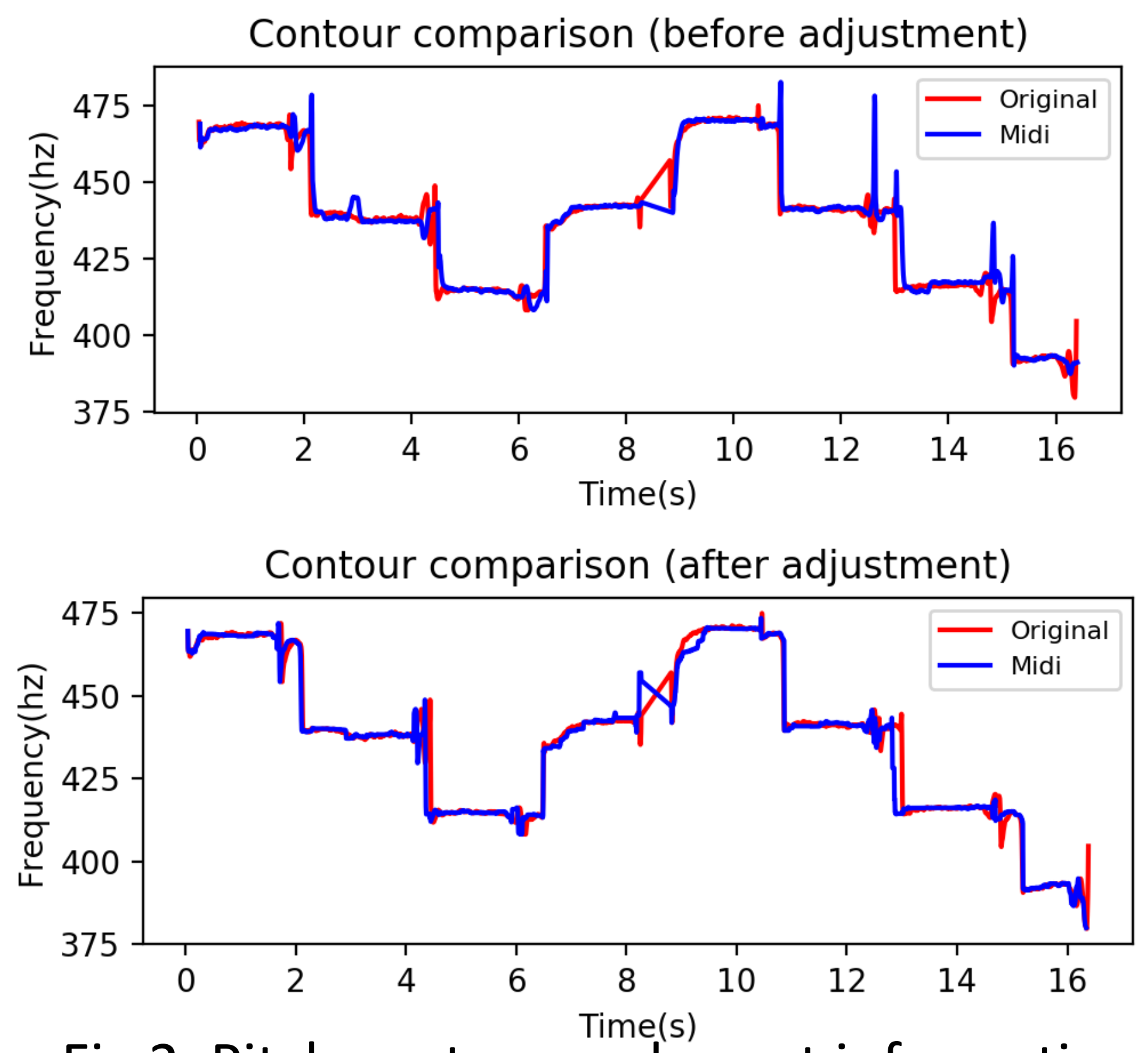


Fig 2. Pitch contour and onset information.
Top: before iterative adjustments.
Bottom: after InstListener's iterative process.

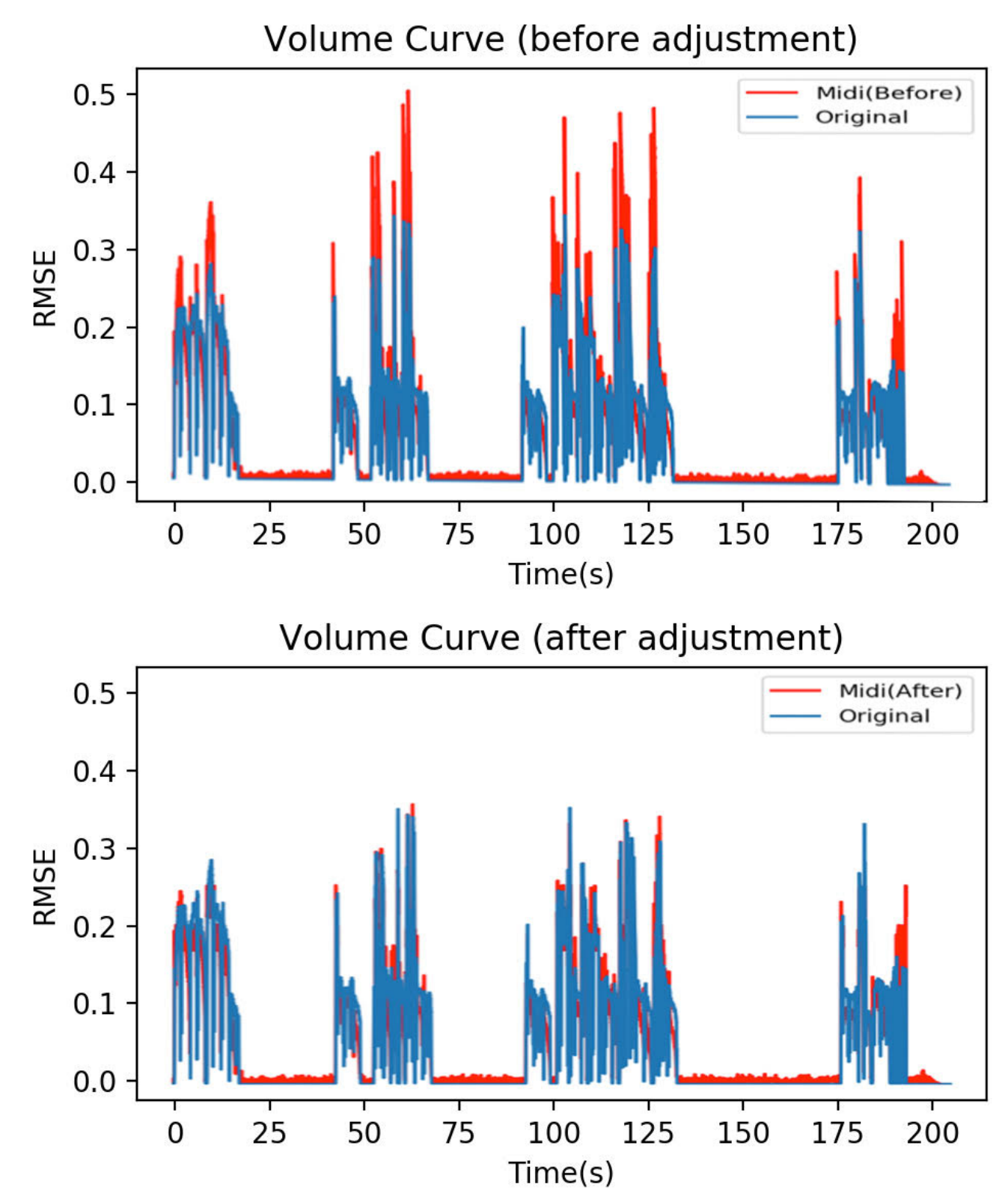
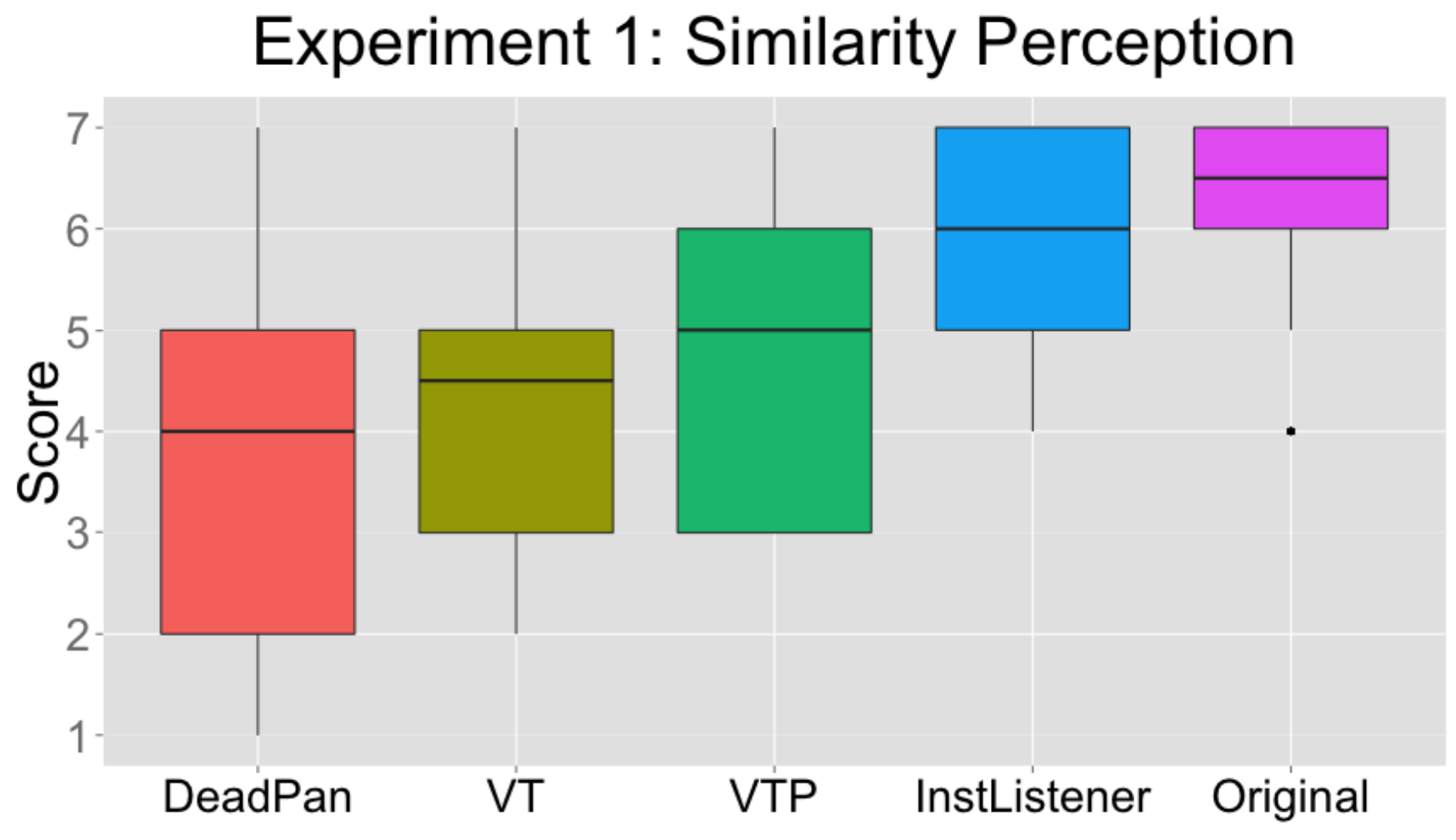
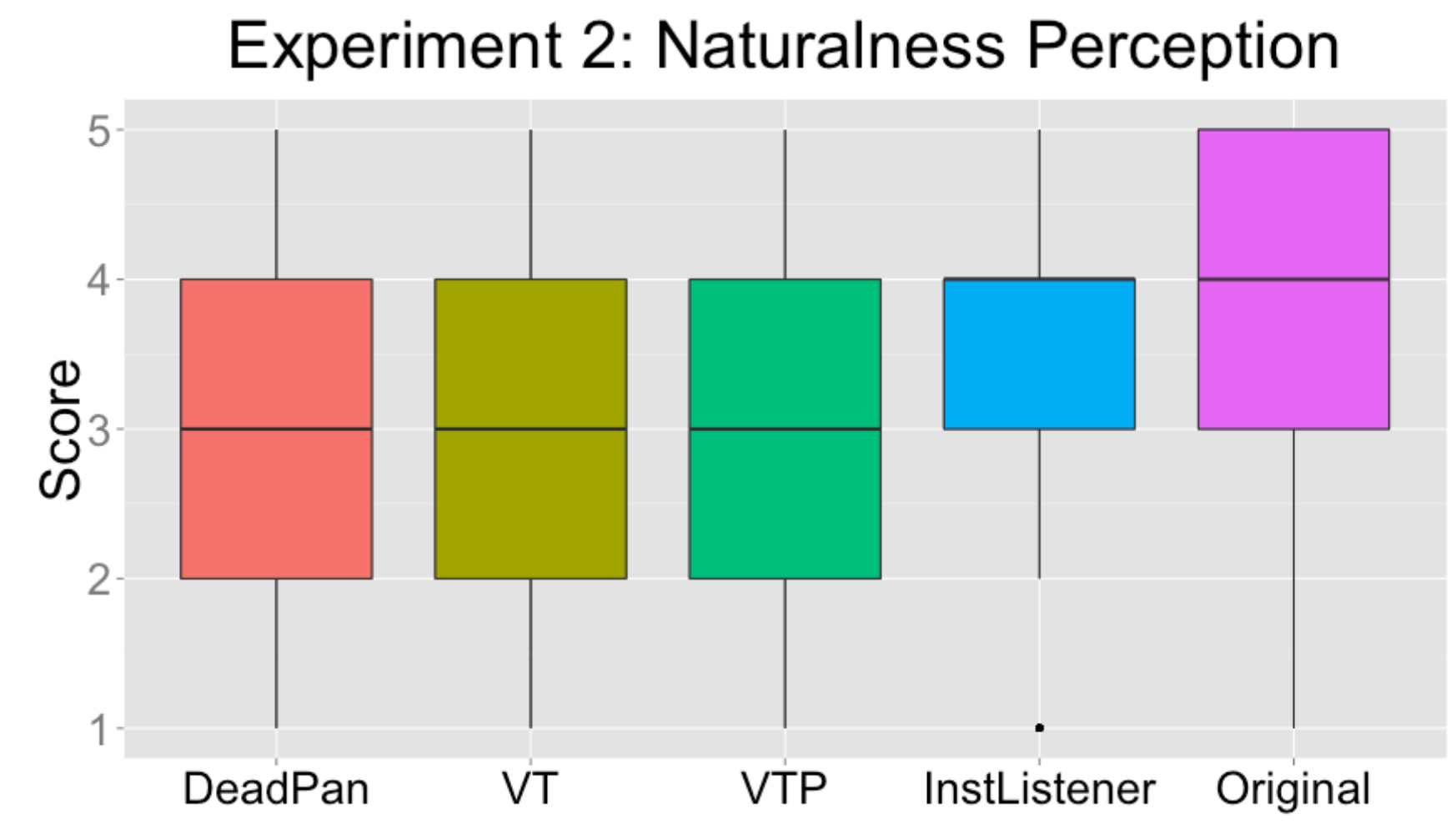


Fig 3. Volume curve
Top: before iterative adjustments.
Bottom: after InstListener's iterative process.

III. Experiments

We conducted our experiments using a crowdsourcing platform, Amazon Mechanical Turk (MTurk). We evaluate the success of imitating the original performance and the naturalness of the resulted MIDI files.





3.1 the subjects (50 in total) were asked to rate how similar the rendition is compared to the original performance on a scale of 7. InstListener was scored the highest as the most similar to the original recording in terms of musical expressions

3.2 the subjects (50 in total) were asked to rate the naturalness of a musical performance on a scale of 5.
The score for InstListener got the best score among the others except for the original performance.

DeadPan: MIDI without dynamics and quantized to 1/8 note.
VT: MIDI with velocity and timing.
VTP: MIDI with velocity and timing, plus pitch bend.
InstListener: MIDI rendition after the iterative process.
Original: recording from the original input performances by musicians.

ICASSP 2018 Calgary