A comparative latency study of hardware and software pitch-trackers

Lilit Yoo and Ichiro Fujinaga
Peabody Institute of the Johns Hopkins University
1 E. Mt. Vernon Place
Baltimore, MD 21202 U. S. A.
{lyoo|ich}@peabody.jhu.edu

ABSTRACT

In this empirical study, results show that recently developed software pitch-trackers can be as fast as some current commercial hardware pitch-trackers. Since latency was the most crucial factor in the performability of the ZETA system, comparative analyses were made of the ZETA and the AXON interfaces, as well as the fiddle~ object, a real-time audio analysis tool in the Opcode's MAX/MSP environment on a 400MHz G3.

A RetroPak-fitted ZETA violin was used to provide the audio input. The audio signal emitted by the ZETA pickup was sent independently to the ZETA VC–225 Violin MIDI Controller interface, the AXON AX–100 guitar interface, and the fiddle~ object, to be analyzed by their pitch-tracking algorithms. Both plucked and bowed (legato) articulations of the violin's open G and E strings were used as stimuli. Once the pitch was analyzed, the interfaces and the fiddle~ object triggered a MIDI xylophone sound sample on a K2000R synthesizer. Both the acoustic audio signal and the output of the synthesizer were simultaneously recorded, then compared for latency times to see how fast the pitch-tracking algorithms were able to perform the pitch-to-MIDI conversion.

Results show that the pitch-tracking of the fiddle~ object on a G3 is as fast as the AXON. The ZETA, by comparison, is slower than the fiddle~ object and the AXON. For high-pitched sounds, the average latencies were 15–35ms for the AXON, 15–45ms for the fiddle~ object, and 15–45ms for the ZETA. For low-pitched sounds, the average latencies were 20–55ms for the AXON, 25–100ms for the fiddle~ object, and 30–70ms for the ZETA. In all cases, the pitches were correctly identified.

1. Introduction

When performing on new instruments that are closely modeled on traditional ones, musicians may harbor expectations that they will behave in a similar manner given a stimulus. When situations arise where these new instruments do not behave like their traditional counterparts, musicians must understand and compensate for specific performance practice problems, such as pitch-tracking latencies. Based on previous research in the performability of the ZETA system (Yoo et al. 1998), comparative analyses were made of the ZETA VC–225 Violin MIDI Controller interface, the AXON AX–100 guitar interface, and the fiddle~ object, a real-time audio analysis tool in the Opcode's MAX/MSP environment on a 400MHz G3. Before discussing the experimental setup, features of the pitch-trackers in question will be described briefly.

The ZETA VC–225 interface has a Pitch-to-MIDI converter that is essentially a five-channel IVL PitchRider (*ZETA Owner's Manual* 1995). MIDI messages from the interface can then be sent to a synthesizer or a computer. In this study, a ZETA RetroPak-fitted instrument was used to provide audio input. It comprises of a four-channel bridge electronic pick-up that sends the acoustical signal of each string into four separate preamplifiers in a beltpack, which can be attached to the performer's waist (*ZETA Owner's Manual* 1995). There are two separate outputs from the preamp: One is the direct audio mono out and the other goes into the ZETA VC–225 interface.

For the AXON system, the composite signal of all plucked strings is detected by its six-channel electronic pickups, and is fed out of an output jack via a connector cable to an interface. The composite signal is then sent to an A/D (Analog-to-Digital) converter to convert the waveform signal into a digital signal. All attack transient information is analyzed by zero-crossings to determine the pitch. Then, the information is sent to a neural network whereby the determined pitch data is converted into MIDI information and can trigger a synthesizer or a computer (Szalay 1998a, 1998b, 1998c, 1998d). The manufacturer of the AXON system describes this as the "Transient Early Recognition System" (*AXON: Users manual AXON AX–100/AX–100SB* 1998; *AXON: AX 100 Product information* n. d.). The interface also has a MONO IN 1/4" input jack that allows any analog signal into the AXON interface.

With the advent of implementing and designing a software pitch-tracker such as the MAX/MSP fiddle~ object, users can obtain pitch extraction information in real-time as opposed to relying on pitch detection systems and their algorithms in hardware, such as the ZETA and the AXON. The fiddle~ object is based on Rabiner's and Schafer's theoretical algorithms (1978) and is described as "a monophonic or polyphonic maximum-likelihood pitch detector similar to Rabiner's, which can also be used to obtain a raw list of a signal's sinusoidal components" (Puckette et al. 1998, 109). Additional information can be obtained in the spectral envelope and developers of this object claim that "rapid changes in the spectral envelope turn out to be a much more reliable indicator of percussive attacks than are changes in the overall power reported by a classical envelope follower" (Puckette et al. 1998, 109).

2. Experimental setup

A RetroPak-fitted ZETA violin was used to provide the audio input. The audio signal emitted by the ZETA pickup was sent independently to the ZETA VC–225 Violin MIDI Controller interface, the AXON AX–100 guitar interface, and the fiddle~ object, to be analyzed by their pitch-tracking algorithms. To calculate and compare the time delays for the pitch-to-MIDI conversion, both plucked and bowed (legato) articulations of a RetroPak-fitted ZETA violin were used as the initial audio input. In this study, the violin's open G and E strings were used as stimuli.

Once the pitch of the audio input was analyzed, the interfaces and the fiddle~ object triggered a MIDI xylophone sound sample on a K2000R synthesizer. Both the acoustic audio signal and the output of the synthesizer were simultaneously recorded on two separate channels, then compared for latency times to see how fast the pitch-tracking algorithms (Kuhn 1990) were able to perform the pitch-to-MIDI conversion (Roads 1996). An example is shown below in Figure 1. The same approach to determine the latency times for the ZETA interface was used for the AXON interface. The direct audio output of the ZETA pickup was fed directly into the MONO input 1/4" jack of the AXON AX–100 Guitar interface so that it could be analyzed by its pitch detection device which was described previously. A similar approach was used for the MAX/MSP fiddle~ object whereby the direct audio output of the ZETA pickup was fed directly into a 400MHz G3.

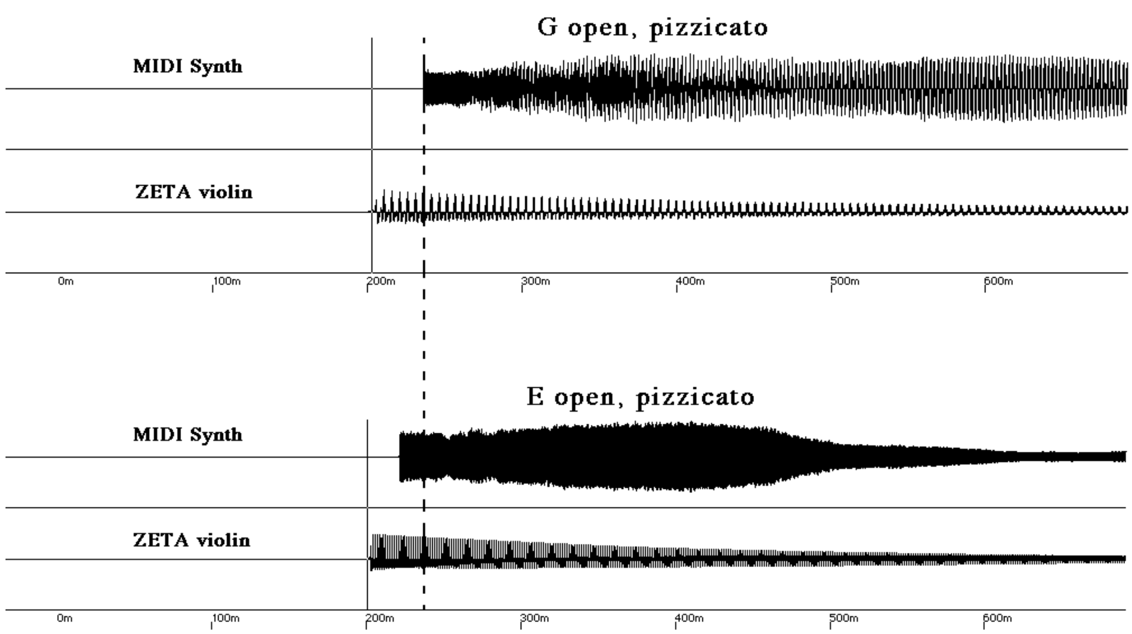

Figure 1 Time delay difference of the initial attack transients of pizzicato between two different pitches, open G string and open E string of the ZETA interface.

3. Results

Results of analyses for latency times of the ZETA, the AXON, and the fiddle~ object demonstrated that the fiddle~ object on a 400 MHz G3 can detect pitches as fast as the AXON. Average latency ranges of plucked (pizzicato) and bowed (legato) articulations were calculated for both hardware and software pitch-trackers. Latencies for various sensitivity parameters and settings of the AXON and the ZETA interfaces were compared but differences were found to be negligible. Similar findings for the fiddle~ object at 1024 samples at different I/O and signal vector sizes were obtained.

Based on the data obtained, results indicate that the fiddle~ object on a G3 can accurately detect pitches just as quickly as a hardware pitch-tracker such as the AXON interface. Average latency range results collected shown in Figure 2 below are based on 20 trials for each articulation and pitch, resulting in a combined average latency range for each pitch-tracker in question.

For plucked (pizzicato) articulations, average latency ranges were about the same for both the AXON and the fiddle~ object on a G3. Generally, bowed legato articulations were recognized slightly slower than plucked ones and pitch recognition was affected by how clean the initial attacks of the bowstroke were. Results collected demonstrate that the software pitch-tracker in question can detect pitch as quickly as that of the algorithm used by the AXON interface.

	AXON (ms)	ZETA (ms)	fiddle~ (ms)
open E pizzicato	16.0 – 19.6	17.4–24.7	16.2 – 27.4
open G pizzicato	23.9 – 29.8	31.9 – 46.4	25.9 – 48.1
open E legato	16.2 – 31.8	17.5 – 42.3	17.2 – 45.0
open G legato	25.4 – 55.0	38.7 – 67.0	26.8 – 96.9
Combined average latency range	16.0 – 55.0	17.4 – 67.0	16.2 – 96.9

Figure 2 Average latency ranges for pizzicato and legato articulations of the open E and G strings for the AXON, ZETA, and the fiddle~ object.

4. Conclusion

Comparative analyses of latencies for the AXON, the ZETA, and the MAX/MSP fiddle~ object on a 400MHz G3 were made to determine if hardware pitch-trackers were able to detect the pitch of an incoming audio input more quickly than those of software pitch-trackers. The direct audio output of a RetroPak-fitted ZETA MIDI instrument was analyzed by the above-mentioned pitch-trackers, which then triggered a MIDI xylophone sound sample of a K2000R synthesizer once the pitch was determined by their pitch detection algorithms.

By simultaneously recording the direct audio output of the RetroPak-fitted ZETA instrument and the triggered MIDI xylophone sound sample, data was obtained to determine average latency times for high-pitched and low-pitched sounds (open E and G strings). Both plucked (pizzicato) and bowed (legato) articulations were used as stimuli. Results obtained show that recently developed software pitch-tracker such as the MAX/MSP fiddle~ object on a G3 can accurately identify the pitch of an incoming signal as quickly as the AXON interface.

References

AXON: Users manual AXON AX–100/AX–100SB Version 2.0. 1998. Halsenbach, Germany: Blue Chip Music Technology.

AXON: AX 100 Product information. n. d. Halsenbach, Germany: Blue Chip Music Technology. http://www.musicindustries.com/axon/AX–100-lit.pdf

Kuhn, W. B. 1990. A real-time pitch recognition algorithm for music applications. *Computer Music Journal* 14 (3): 60–71.

Puckette, M. S., Apel, T., and D. D. Zicarelli. 1998. Real-time audio analysis tools for Pd and MSP. *Proceedings of the International Computer Music Conference,* 109–12.

Rabiner, L. R., and R. W. Schafer. 1978. *Digital processing of speech signals*. Englewood Cliffs, NJ: Prentice-Hall.

Roads, C. 1996. *The computer music tutorial.* Cambridge, MA: The MIT Press, 503–20.

Szalay, A. 1998a. Filed December, 1995. *Method for pitch recognition, in particular for musical instruments which are excited by plucking or striking*. United States Patent No. 5,780,759.

Szalay, A. 1998b. Filed January, 1996. *Method for recognition of the start of a note in the case of percussion or plucked musical instruments*. United States Patent No. 5,710,387.

Szalay, A. 1998c. Filed April, 1996. *Signal analysis device having at least one stretched string and one pickup*. United States Patent No. 5,824,937.

Szalay, A. 1998d. Filed June, 1996. *Synthesizer detecting pitch and plucking point of stringed instrument to generate tones*. United States Patent No. 5,717,155.

Yoo, L., and I. Fujinaga. 1998. ZETA violin techniques: Limitations and applications. *Journal SEAMUS* XIII(2): 12–6.

ZETA: Owner's manual. 1995a. Oakland, CA: ZETA Music.

ZETA: Suggested settings and equipment for ZETA instruments. 1995b. Oakland, CA: ZETA Music.