

New Algorithms and Technology for Analyzing Gestural Data

W. A. Schloss

School of Music
University of Victoria
Victoria, BC V8W 2Y2 CANADA
+1 250 721-7931
aschloss@finearts.uvic.ca

Peter Driessen

Dept of Electrical and Computer
Engineering
University of Victoria
Victoria, BC V8W 2Y2 CANADA
+1 250 472-4234
driessen@uvic.ca

ABSTRACT

In this paper, we describe the ways in which we are analyzing gestural data as if they were an audio signal, and applying this technique to the Radio Drum, a novel controller that one of the authors uses regularly for concert performances.

Keywords

New musical instruments; gesture sensors; controllers; three-dimensional sensor

INTRODUCTION

The Radio Drum is a three-dimensional controller that has been in existence in various forms since its original development at Bell Labs in the late 1980's. [1]-[9] The Radio Drum uses capacitive sensing; a radio-frequency voltage source is conducted from the performer's mallets or sticks, and is received on the drum surface beneath. The two sticks are differentiated by using different frequencies for each one. The radio drum surface is covered with a layer of foam to provide a quiet playing surface and avoid striking the circuit board surface itself.

METHODS

We have embarked on a project to enhance this instrument, to make it more responsive, with minimum latency, and therefore more viable as a virtual percussion instrument. At the same time, it must be easily programmable, very flexible and easy to re-configure in software. Some of the ideas presented here could be applied to any input device or gesture sensor that uses analog signals that are digitized and analyzed internally. Historically, as

much as possible was hard-wired in the sensor itself—the analog signal is digitized and analyzed in the instrument, mostly due to the impossibility of doing these calculations on the host computer of the time, which also has to generate sound or MIDI, or respond in some general way to the input device. Given the considerable power of personal computers today, there is no reason not to do everything on the host, thereby affording full-screen editors, high-level analysis software like MATLAB or MSP [11], instant reconfigurability [10], and other advantages.

In our new version of the instrument, the raw analog signal is therefore not processed on the instrument, rather it is sent directly to the host computer, where a standard multi-channel audio interface is used for data acquisition. There is no need for special data-acquisition hardware anymore; with the very common and inexpensive multi-channel audio interfaces available today, offering 8, 12, 16 or 24 channels or more of A/D conversion at 16, 20, or 24 bits precision and 44.1 kHz sampling rates, we have more than enough power already available. We are aware that this rate of data conversion may be overkill for some performance devices (musical instruments) but in fact we suspect that in a lot of cases under-sampling has caused problems in the past. We no longer have to worry that we are under-sampling the input data, and if we already have an audio interface, there is no need to buy an addition data acquisition card. There is one technical problem with using audio interfaces as data acquisition systems -- they are not designed to deal with signals at very low frequencies (below 20 Hz, for example). We have fixed this problem with a chopper modulator analog circuit that amplitude-modulates the signal before it is digitized, creating sidebands that are well above 20 Hz. To demodulate, we take

advantage of the fact that almost all the audio interfaces have separate word clock in and out, which is used for the chopper waveform and enables synchronous demodulation.

Researchers from CNMAT (Wessel, Freed, et alia) have proposed similar ideas, but they are just starting to become practical. We hope to use their RIMAS box to ultimately connect to the host computer using ethernet instead of using MIDI. Finally, we are working on a version of the drum that has wireless sticks; this frees the percussionist to play the Radio Drum in the context of acoustic percussion, greatly expanding its usage in musical situations.

SIGNAL PROCESSING OF GESTURAL DATA

The gestural data from the radio drum is an analog signal band-limited to 700 Hz. There are 4 channels of this data for each of two sticks, which are reduced to 3 channels representing x,y,z displacement versus time using a transformation in [9]. This signal is processed with the objective of interpreting the gestural data to trigger and control musical events.

Typical gestures represented by this signal from each stick are: an isolated down and up motion arising from a strike or hit of the drum surface, a series of strikes in a rhythmic pattern created by the performer, and a drum roll resulting in a 'buzz' or rapid series of up and down motions of decaying amplitude as the stick bounces.

The signal processing includes two key steps. The first step is to capture all of the subtle motions associated with these gestures, so that the instrument responds in a sensitive manner to the performer's expressive technique. We seek to capture details of the entire motion represented by the x,y,z signals versus time, not just the velocity of a strike.

The second step is map the gesture signal to control musical events. This mapping can take many forms, depending on the creative ideas of the composer and/or performer.

For example, the mapping may simply trigger a sound when a strike occurs (when $dz/dt=0$ and $d^2z/dz^2 < 0$). The mapping may trigger different sounds or sequences of sounds depending on the x,y position at the strike time. The mapping may control pitch, timbre or other parameters of a sound source in response to changes in x,y,z . The most satisfying results are expected when the parameters are derived from physical models [13].

Figure 1 represents the z gesture signal for the case of a drum roll, i.e. the most rapid motions that the performer can generate.

The signal is a series of pulses with decaying amplitude. The vertical scale represents an inverted and non-linear version of the z value. The signal

peaks represent the smallest values of z where the signal is strongest, corresponding to the closest approach to the $z=0$ drum surface. We note that the signal peaks after the 4th peak are about 1700 samples or 39 msec apart, with wider spacings up to 69 msec for earlier peaks. Thus the repetition rate of the bounces in a drum roll is about 25 Hz. The nominal 3dB time width of the pulses is 15 msec.

A single isolated strike has a similar pulse shape.

FILTER DESIGN AND IMPLEMENTATION

We first design the signal processing algorithm needed to generate a trigger pulse when a strike occurs. We assume that the highest frequency in the gesture signal will be no greater than the 25 Hz rate of a drum roll, i.e. that a human cannot create successive strikes with one stick closer than 40 msec apart. Thus the algorithm which yields maximum sensitivity has two stages: a matched filter, matched to the 'pulse shape' of a strike, followed by a peak detector which finds the points where $dz/dt=0$ and $d^2z/dz^2 < 0$.

The matched filter is approximated by a FIR filter with a gaussian impulse response and 3dB time width corresponding to the nominal pulse shape of a drum roll strike. The gaussian shape is a reasonable approximation of the actual pulse shape and ensures that there are no significant frequency sidelobes above the nominal bandwidth.

The filter is implemented using MAX/MSP [11], a graphical programming environment for music and media applications, commonly used to customize the logic of building an interface to interactive media by wiring together various MAX/MSP objects. The wires can carry asynchronous control signals or synchronous signals sampled at 44100 Hz.

The FIR filter has 2048 taps, and the MSP FIR object using direct convolution does not run in real time on a 400 MHz Apple Powerbook. Thus the matched filter is implemented in the frequency domain, thus speeding up the calculation by a factor of about 20 in this case.

The filter is implemented using an overlap-add technique [12], in which the signal is segmented into overlapping windowed frames of length 2048 samples (raised cosine window), spaced at time interval of 1024 samples. These frames are zero-padded to an FFT size of 4096 and transformed. The 2048 sample gaussian impulse response is also zero-padded to the FFT size and transformed. The transforms are multiplied, and the result, a frame of length 4096, is overlap-added to the output of previous frames. A new frame is overlap-added once every 1024 samples.

Figure 2 shows the MSP implementation of the filter. The signals winpad1, ..., winpad4 are the 2048 sample window function zero-padded to 4096, delayed at intervals of 1024 samples, and have period 4096. Similarly, windg1, ..., windg4 are the filter magnitude response (FFT of the pulse shape). To simplify the figure, the real and imaginary parts of the filter response and complex multiplication with the signal transform is not illustrated.

RESULTS AND CONCLUSION

Figure 1 shows the filtered signal with circles marking the detected peaks. The sensitivity of the signal processing algorithm is outstanding. Peaks are detected at the expected 39 msec intervals even when they are no longer visible on the graph.

It remains to explore other mappings of the gesture signal to control a sound source.

ACKNOWLEDGMENTS

We would like thank Max Mathews, Bob Boie, David Wessel and Dale Stammen for inspiration, technology and vision.

REFERENCES

- [1] Mathews, M and W A Schloss The Radio Drum as a Synthesizer Controller. *ICMC Ohio State proceedings*, 1989.
- [2] Schloss, W. A. Recent Advances in the Coupling of the Language Max with the Mathews/Boie Radio Drum *ICMC Glasgow proceedings*, 1990
- [3] Schloss, W. A. Man-Machine Interaction in a Jazz Improvisation Context. *First International Workshop on Man-Machine Interaction in Live Performance* Pisa, Italy, 1991.
- [4] Schloss, W. A. and Jaffe, D. Intelligent Musical Instruments: The Future of Musical Performance or the Demise of the Performer? *Interface* 1993
- [5] Jaffe, D. and W. A. Schloss. The Computer-Extended Ensemble. *Computer Music Journal*, 1994.
- [6] Jaffe, D. and W. A. Schloss. A Virtual Piano Concerto—The coupling of the Mathews/Boie Radio Drum and Yamaha Disklavier Grand Piano in "The Seven Wonders of the Ancient World." *International Computer Music Conference*, Aarhus, 1994.
- [7] Goldstein, M and W. A. Schloss Virtual Percussion: Drumming in the Ether. *IBM Research Headquarters*, Yorktown Heights, New York, 1997
- [8] Schloss, W. A. Improvising Across Borders in Cuba *UCSD Symposium on Improvisation*. San Diego, 1999.
- [9] R. Boie, L. Ruediselli et al, AT&T Bell labs internal memo on radio drum
- [10] D. Zicarelli, Communicating with Meaningless numbers, *Computer Music Journal*, Vol 15 No. 4, 1991, pp. 74-77.
- [11] M. Puckette, D. Zicarelli, MAX/MSP reference, www.cycling74.com
- [12] J.B. Allen, L. R. Rabiner, A unified approach to short-time Fourier analysis and synthesis *Proc. IEEE*, Vol, 65, No. 1, pp. 1558-1564, Nov. 1977.
- [13] Kees van den Doel and Dinesh K. Pai. "The Sounds Of Physical Shapes." *Presence*, 7(4):382-395, 1998.

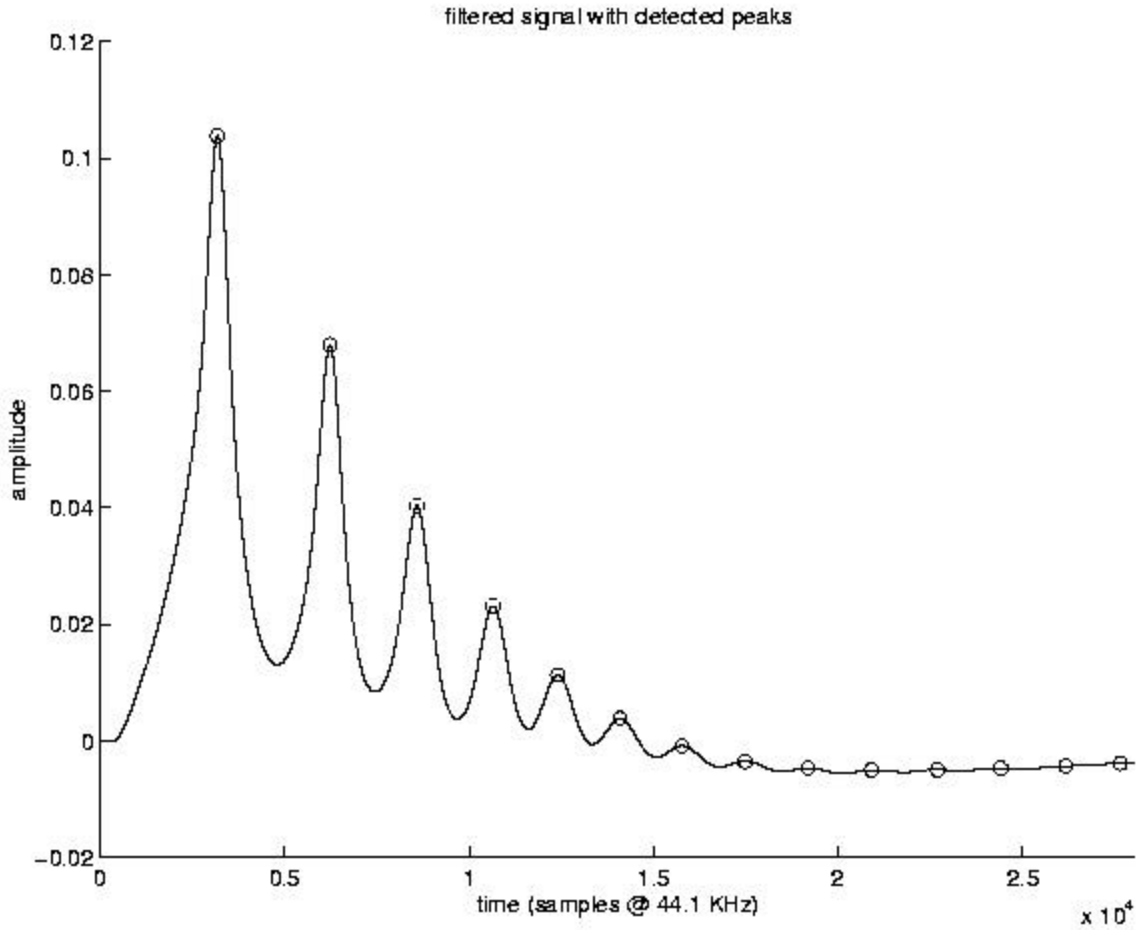


Figure 1 (above) Filtered gesture signal
 Figure 2 (below) MAX/MSP implementation

