Learning Indirect Acquisition of Instrumental Gestures using Direct Sensors

George Tzanetakis, Ajay Kapur, Adam Tindale
Department of Computer Science (also ECE, Music)
University of Victoria, Canada

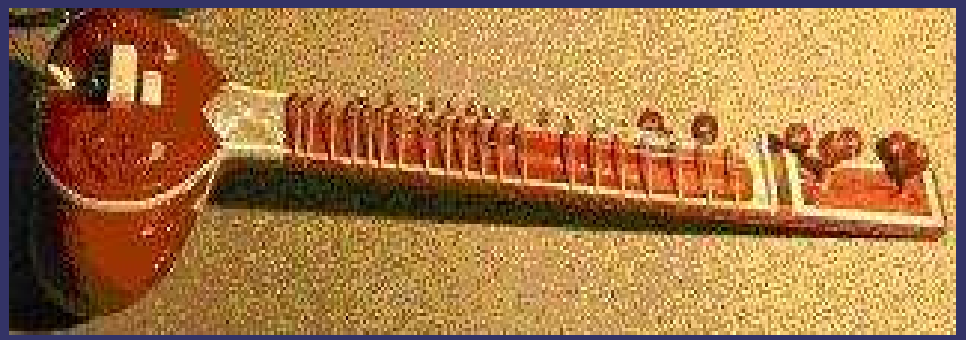
Musical Instruments

Musical Instruments vs Computer User Interfaces

- Tactile, Auditory
- Not user-friendly
- Hard to learn
- Delicate control
- Rich feedback
- Beautiful

- Visual
- User-friendly
- Easy to learn
- Clumpsy control
- Limited feedback
- Ugly

E-Sitar Direct Sensors (Ajay Kapur)



E-sitar augmented acoustic instrument
Thumb sensor – strumming
Network of resistors for frets
Communicate performance information to the computer

Indirect Acquisition

- No modification to acoustic instrument
- 1 or more microphones
- Signal processing to extract information
- Pitch detection is the classic example

Indirect Acquisition

Direct Sensors vs Indirect Acquisition

- Instrument modification
- Hard to build
- Custom-made
- Clean and mostly accurate information

- No instrument modification
- Already in place for recording purposes
- Noisy and inaccurate signal
- Require specialized DSP techniques

The main idea

- Use direct sensors to “learn” indirect acquisition
- Use augmented instrument for training
  - Record acoustic signal
  - Train model to associate direct sensor with the acoustic signal
  - Evaluate and iterate
- Use trained model in non-augmented acoustic instrument

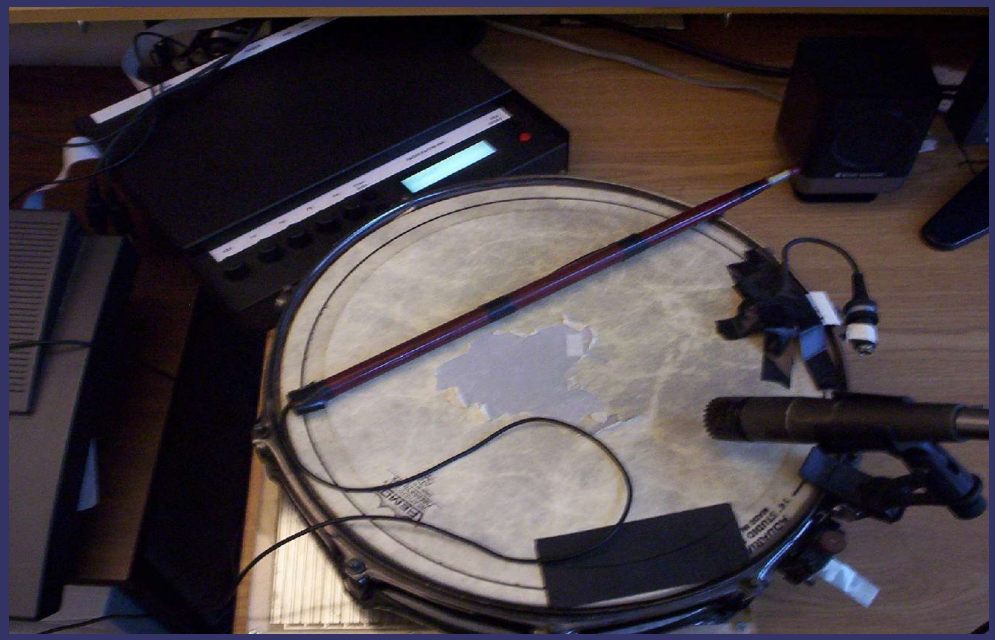
Advantages

- Hard-to-built augmented instrument is only used for training
- No modifications required
- Unlimited supply of training data for the machine learning model
- TRAIN BY PLAYING is much more fun than TRAIN BY ANNOTATING

Two case studies

- Regression

- Classification



Audio Feature Extraction

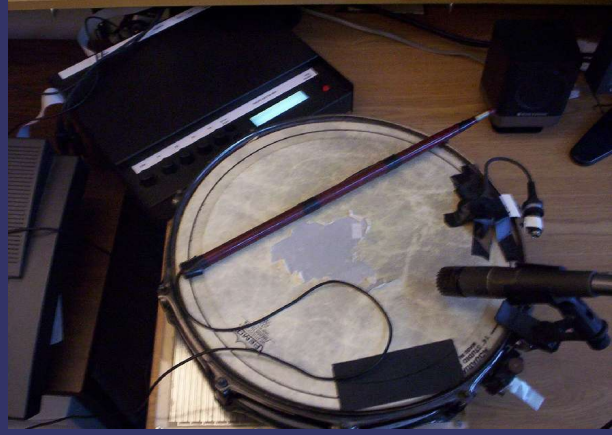
- Short-Time Fourier Transform based
- Spectral centroid, rolloff, flux
- RMS energy
- Calculated every 20 msec
- Means/variances over 1 second

Virtual thumb sensor

- Regression
  - Predict continuous value (train with direct FSR sensor) from audio features
- Upward/downward strokes + velocity

	10	20	30	40
Random Output	0.14	0.14	0.14	0.14
Linear Regression	0.28	0.33	0.28	0.27
Neural Network	0.27	0.45	0.37	0.43
M5' Regression Method	0.28	0.39	0.37	0.37

Snare drum Strike Location

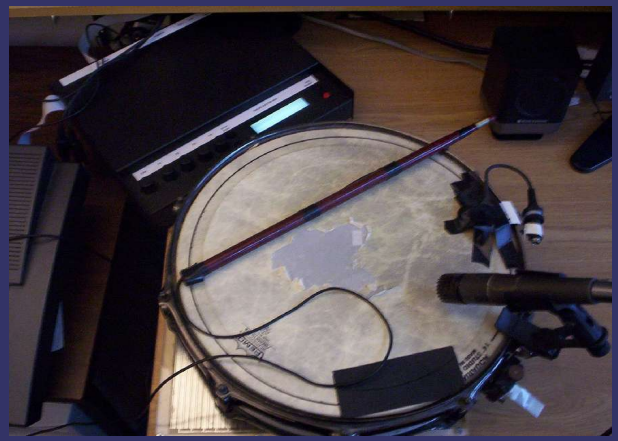


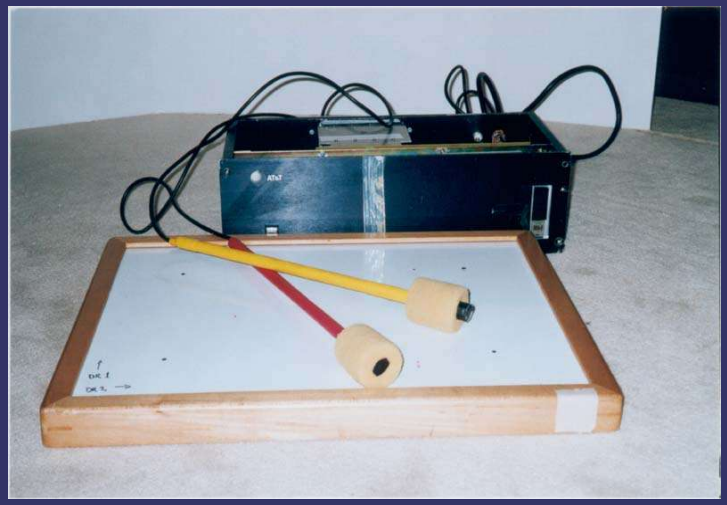
- Goal: strike location on the drum surface (edge, middle, center) (classification)
- Train using radio drum (1000 hits)
- Microphone to record the audio signal

	ZeroR	NB	MLP	MLR	SMO
Snares	53	92	91	91	92
No Snares	57	93	94	95	95
Improvisation	59	79	77	78	78

The radio drum as a direct sensor

Capacitance-based
2 x 3D controller
Each stick send the X,Y,Z position





Conclusions and future work

- Machine learning is great but requires lots of good quality training data
- Using direct sensors to train indirect acquisition allows unlimited training data to be collected for detecting music gestures
- More features: LPC, sinusoidal modeling
- More gestures
- Archiving and analysis of music performance