University of Victoria Engineering

Artificial perception, communication, embodiment, and expressivity in music

George Tzanetakis,
University of Victoria, Canada

Computer Science Research Week,
NUS - 2021

Outline

- Background
- Artificial intelligence in music and beyond
- My interests
  - Perception
  - Communication
  - Embodiment
  - Expressivity
- Future challenges and opportunities

Technical Background

- Main focus of research has been Music Information Retrieval (MIR)
- Involved from the early days in the field (1999-2000)
- Have published papers in almost every ISMIR conference and in most MIR topics
- Organized ISMIR 2006 in Victoria, Canada
- Tutorials on MIR in several different conferences

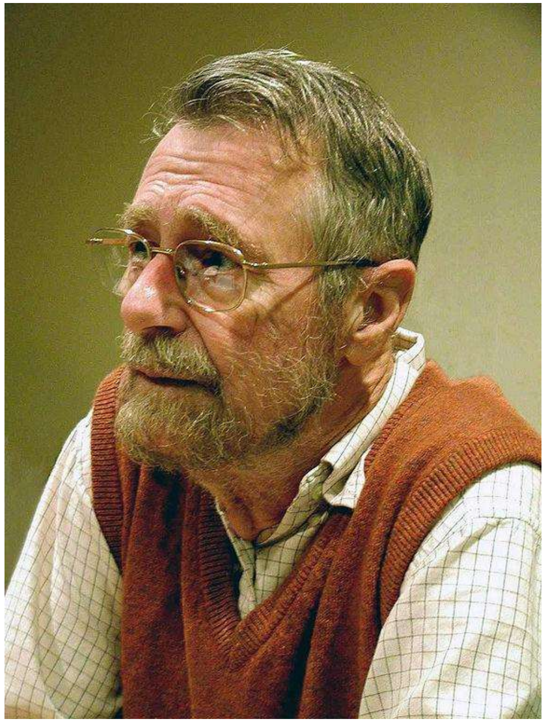
University of Victoria Engineering

Music Background

- Messing around with a piano keyboard from when I started learning piano until today
- Music theory and composition studies
- Saxophone performance (classical)
- Musical contexts and practice:
  - Rock bands in high school
  - Greek folk music in university
  - Jazz and classical music in university and graduate school
  - Today experimental music

Why ?

- The question of whether a computer can think is no more interesting than the question of whether a submarine can swim - E. Dijkstra



Maybe it actually is interesting

- Personally my main motivation is to better understand and appreciate the complexity and beauty of human music making

Artificial Intelligence (in music)

Paraphrasing my favorite quote by G. Box -
"All models are wrong some are useful"

"All artificial intelligence systems are not intelligent some are useful"

The old driving vision: the great celestial jukebox
The new driving vision: a virtual musician

Parting lesson: to build useful systems integration of all CS disciplines is needed

Deep Learning is not AI

Projects: binary CNNs, Unets for music transcription, siamese networks for singer clustering .....

Claimed no feature engineering but the reality:

ML: parameter search (blind), feature design (informed)
DL: architecture/layer/parameter search (blind) loss function (informed)

Projects

Projects from my own body of work beyond your typical ML system:

- **Perception:** teaching a virtual violinist to bow
- **Communication:** markov logic networks and a programming language for stream processing
- **Embodiment:** music robots
- **Expressivity:** hybrid synthesis for expressive drumming, soundplane, augmented reality theremin

PERCEPTION


"HIS MASTER'S VOICE"
This trademark and the trademarked word "Victrola" identify all our products. Look under the lid! Look on the label!
VICTOR TALKING MACHINE CO., Camden, N. J.

Physical Modeling Meets Machine Learning : Teaching a virtual violinist to bow

- Digital sampling can provide high-quality sounds but lacks the intimate control afforded by acoustic instruments
- Physical modeling synthesis works by directly simulating the physics of sound production rather than storing waveforms
- It has the potential to provide expressive control but like real instruments this control is not trivial and needs to be learned

University of Victoria Engineering

Main idea

- As in a real violin correct bowing requires feedback (both auditory and haptic)
- Learn the mapping of control-parameters to good sound rather than explicitly program it
- Teach rather than program
- Basically develop a virtual ear
- Graham Percival - Masters at UVic, PhD at the University of Glasgow, PostDocs at UVic and NUS

Quote: With great control comes great fragility

Physical Model

- No recordings of violin performance; we use physics [1]
  - Wave equation for a stiff string with modal dampening

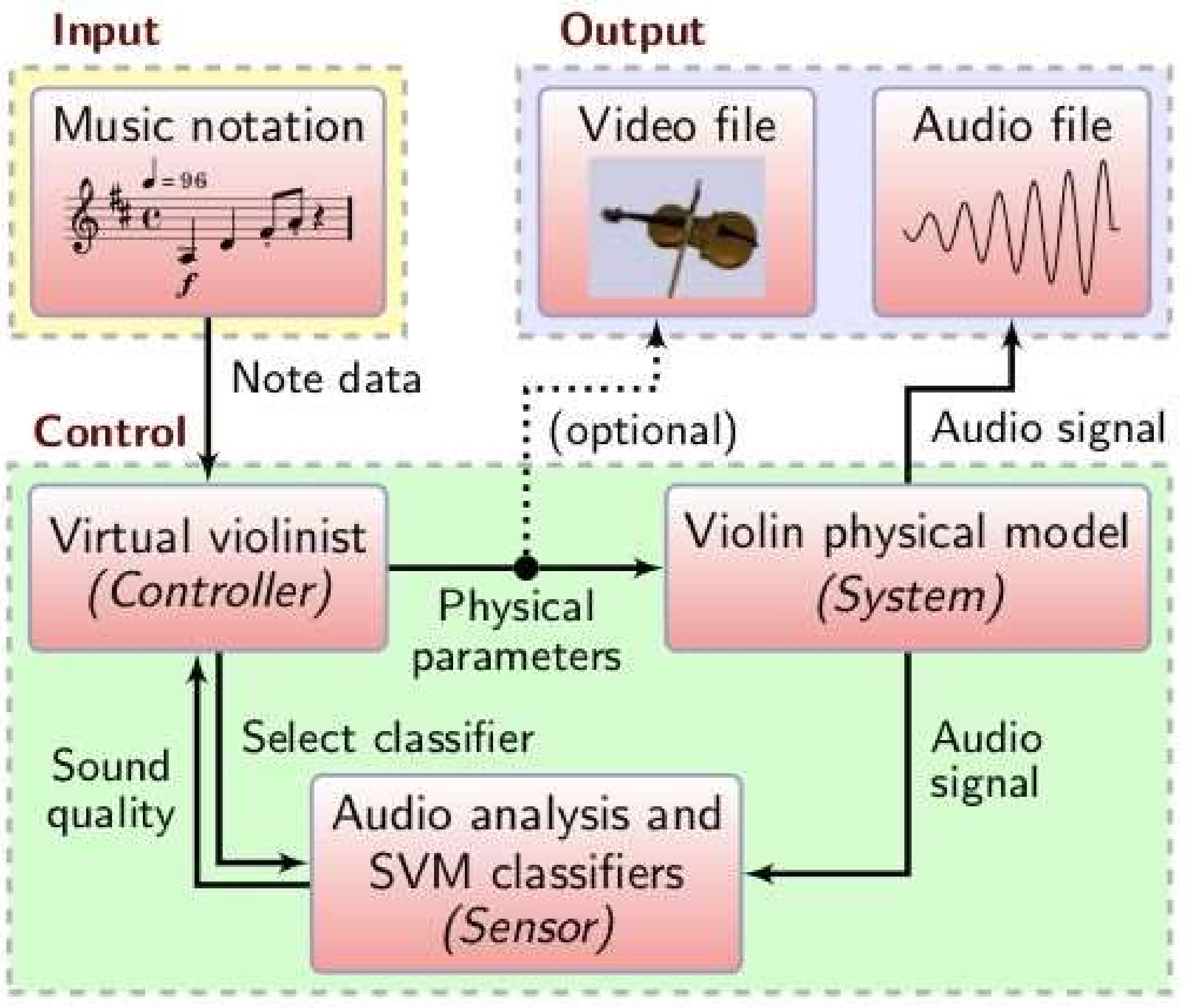
$$\rho_{\mathrm{L}} \frac{\partial^2 y(x,t)}{\partial t^2} - T \frac{\partial^2 y(x,t)}{\partial x^2} + EI \frac{\partial^4 y(x,t)}{\partial x^4} + R_L(\omega) \frac{\partial y(x,t)}{\partial t} = F(x,t)$$

- Implemented as a C++ library, published under GNU GPLv3+

Input parameters

- Violin string number s
- Left-hand finger position x_1
- Bow-bridge distance x_0, velocity v_b, force F_b

System Architecture

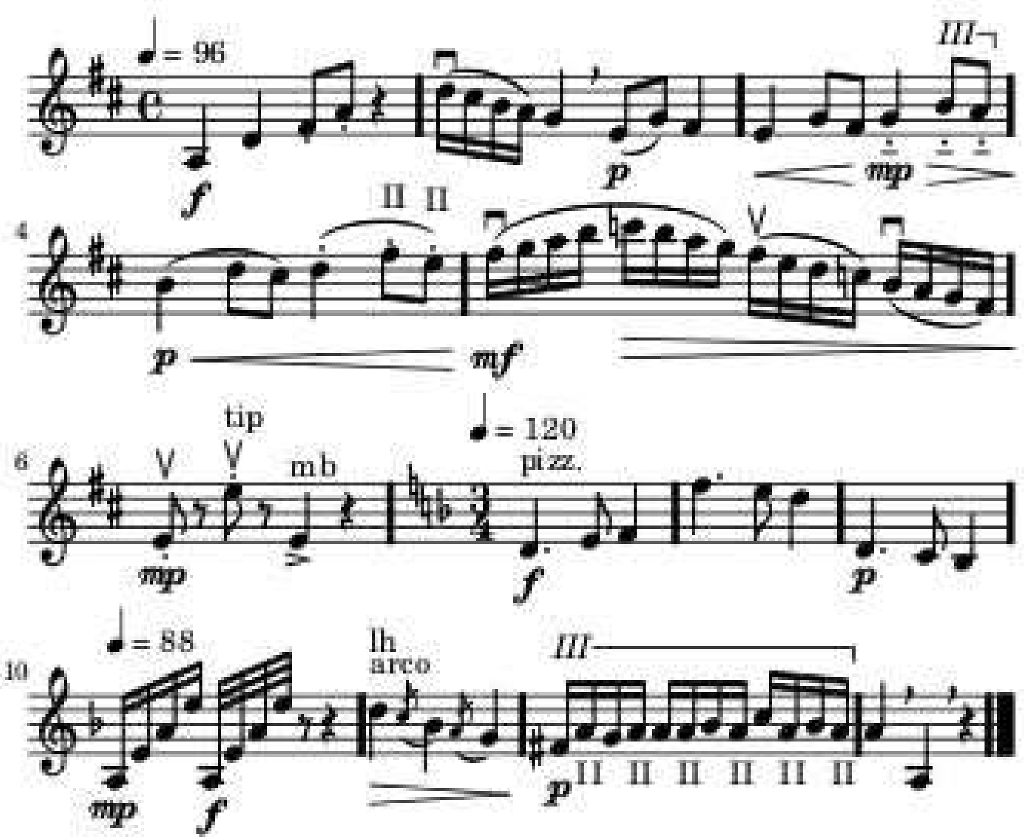


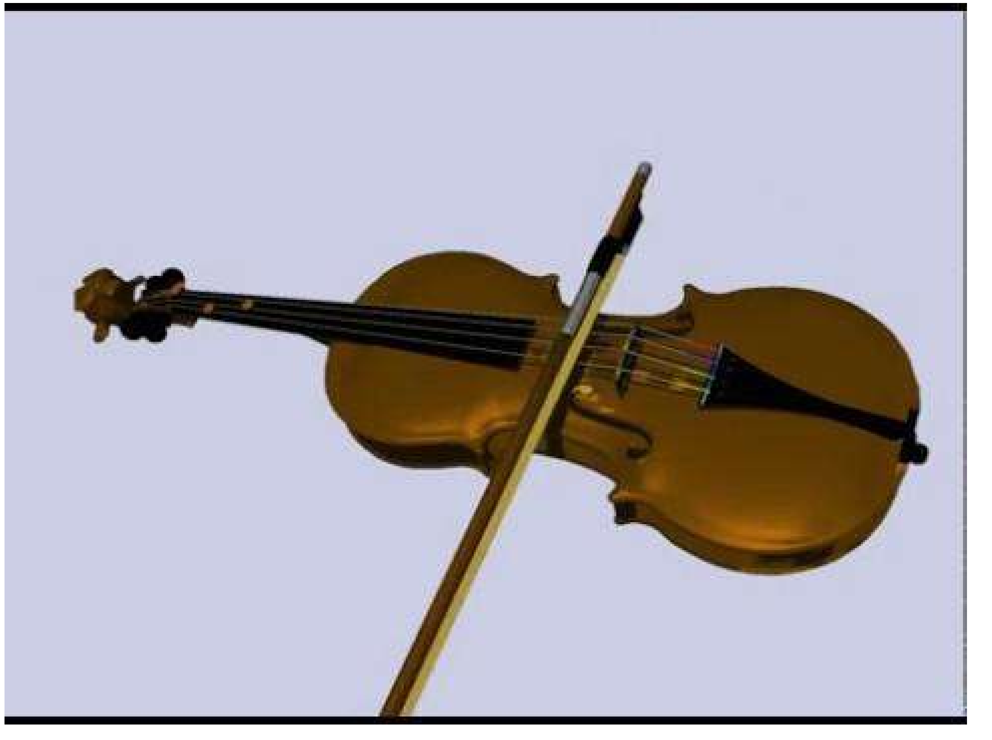
Before and after training

The virtual violinist plays scales and simple exercises. A human teacher rates each notes on a scale from 1 to 5. After several rounds of training the virtual violinist has learned the mapping of control parameters to good sound

Playing a piece

Input





COMMUNICATION

George Tzanetakis, University of Victoria

Three views of Human-Machine Communication

- Human-Computer Interaction (pressing buttons, viewing screens, listening to sounds, gloves with sensors, virtual reality)
- Programming Languages (structured textual or visual ways of creating software and hardware systems)
- Machine learning (collection of annotated data typically by humans)

University of Victoria Engineering

Arpp programming language (Jakob Leben)

- Syntax based on recurrence equations (write code like you write math)
- Supports infinite and multi-dimensional and multi-rate arrays (streams)
- Efficient compilation using polyhedral compilation
- https://arrp-lang.org/

Arpp examples I

Write code like you write the math - using the same equations

```
y[0] = 0;
y[n] = b*x[n] - a*y[n-1];
```

Work with signals at different rates

```
y[n] = x[n*hop]
```

Arpp examples II

Work with multi-dimensional streams

```
y[n,k] = x[n+k] * w[k]
```

Do math with entire signals

```
x[n] = n;
y = sin(x/100*2*pi) * 0.5;
```

Musical analysis of audio signals using ML

- Most existing recent approaches focus on a specific aspect (beat, tempo, chords, structure) and use data-driven ML models
- What is missing:
  - Human music perception understanding is holistic, hierarchical and multi-faceted
  - No easy way to communicate existing knowledge such as rules of harmony
  - No easy way to communicate partial knowledge dynamically

Musical analysis of audio signals using Logic

- A more traditional alternative is to formulate music analysis tasks as inferences using logic formulations
- What is missing:
  - Uncertainty about rules is difficult to handle
  - Low-level information extracted from the audio recording is difficult to integrate

University of Victoria Engineering

Markov Logic Networks (MLN)

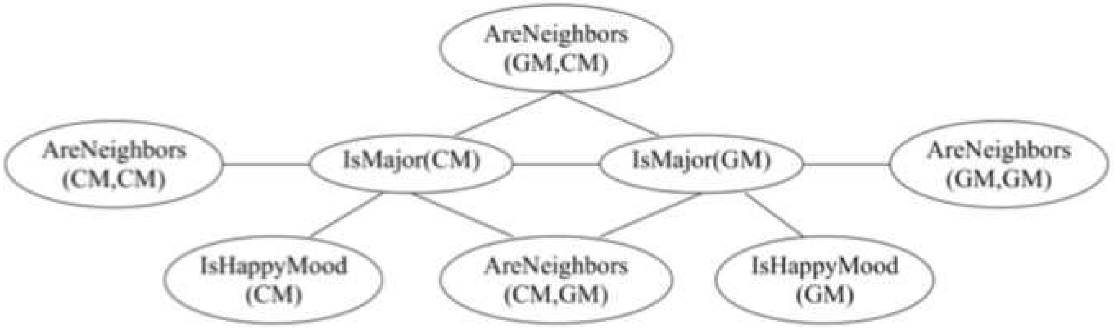
- Expressive formalism that combines probilistic graphical models and first-order logic inference
- Highly flexible and expressive language for the harmonic analysis of audio music signals
- MLN is a set of weighted first-order logic formulas that can be viewed as a template for creating a Probabilistic Graphical Model
- Softens logic rules from true/false to probabilities

MLN example

Basic idea in Markov logic: to soften these constraints to handle uncertainty. The weights reflect how strong a constraint is.

Knowledge	*Logic formula*	*Weight*
A major chord implies an happy mood.	∀ x IsMajor(x) ⇒ IsHappy-Mood(x)	w_1 = 0.5
If two chords are neighbors, either the two are major chords or neither are.	∀ x ∀ y AreNeighbors(x, y) ⇒ (IsMajor(x) ⇔ IsMajor(y))	w_2 = 1.1

Example of a first-order KB and corresponding weights in the MLN.



Ground Markov network obtained by applying the formulas to the constants CM and GM chord.

University of Victoria Engineering

MLN for Chords/Structure

Flexibility of MLN: *Prior global structural information*

Formulas added to express the constraint that *two same segment types are likely to have a similar chord progression*.

Predicate declarations	
$Observation(chroma!, time)$ $\quad$ $Succ(time, time)$	
$Chord(chord!, time)$ $\quad$ $SuccStr(time, time)$	
Weight	*Formula*
Prior observation chord probabilities:	
$log(P(CM(t=0)))$	$Chord(CM, 0)$
$\cdots$	$\cdots$
$log(P(Bm(t=0)))$	$Chord(Bm, 0)$
Probability that the observation (chroma) has been emitted by a chord:	
$log(P(o_0\|CM))$	$Observation(o_0, t) \wedge Chord(CM, t)$
$\cdots$	$\cdots$
$log(P(o_{N-1}\|Bm))$	$Observation(o_{N-1}, t) \wedge Chord(Bm, t)$
Transition probability between two successive chords:	
$log(P(CM\|CM))$	$Chord(CM, t_1) \wedge Succ(t_2, t_1) \wedge Chord(CM, t_2)$
$\cdots$	$\cdots$
$log(P(Bm\|Bm))$	$Chord(Bm, t_1) \wedge Succ(t_2, t_1) \wedge Chord(Bm, t_2)$
Probability that similar segments have the same chord progression:	
w_{struct}	$Chord(CM, t_1) \wedge SuccStr(t_2, t_1) \wedge Chord(CM, t_2)$
w_{struct}	$Chord(C\#M, t_1) \wedge SuccStr(t_2, t_1) \wedge Chord(C\#M, t_2)$
$\cdots$	$\cdots$
w_{struct}	$Chord(Bm, t_1) \wedge SuccStr(t_2, t_1) \wedge Chord(Bm, t_2)$

The predicate *SuccStr* allows considering wider windows, as opposed to consecutive frames via the *Succ* predicate.

A chroma feature is observed at each time frame:
$Observation(o_0, 0) \cdots$
$Observation(o_{N-1}, N-1)$
The temporal order of the frames is known:
$Succ(1, 0) \cdots$
$Succ(N-1, N-2)$
Prior information about position of same segment type in the structure is given:
$SuccStr(1, 10)$
$SuccStr(2, 11) \cdots$

Results for chord/structure

- Test-set: 143 hand-labeled Beatles songs $\rightarrow$ Removing songs for which the structure was ambiguous.
- Evaluation measure: chord label accuracy.

	Chord LA results	*Stat. Sig.*
MLN_chord	72.57 ± 13.51	}*yes*
MLN_struct	74.03 ± 13.90	}*no*
[36]	73.90 ± 13.79	

Figure: *MLN_struct: MLN incorporating prior structural information, MLN_chord: baseline HMM, [36]: chromagram averaged over same segment types as in [Mauch et al. 2009].*

University of Victoria Engineering

Results for chord/key

Improving chord estimation using provided key information. Joint estimation provides key estimation for free.

	Chord LA	*Stat. Sig.*
HMM	72.49 ± 14.68	**no**
Chord MLN	72.33 ± 14.78	
Prior key MLN, WMCR	$\mathbf{73.00 \pm 13.91}$	**yes**
Prior key MLN, CB	72.22 ± 14.48	no
Joint chord/key MLN	72.42 ± 14.46	no

	EE	*EE*	*E+N*	*Stat. Sig.*
Joint chord/key MLN	82.27	88.09	94.32	
DTBM-chord	48.59	67.39	89.44	yes
DTBM-chroma	75.35	85.14	95.77	yes

EMBODIMENT

Human-Machine Improvisation

- In 2004 I joined the University of Victoria as an assistant professor
- Ajay Kapur was my first PhD student
- Ajay: “I want to make a percussion robot that is able to improvise rhythmically North Indian music with me playing the Sitar”
- Me: “That’s too ambitious - focus on something more specific”
- Fortunately he ignored me

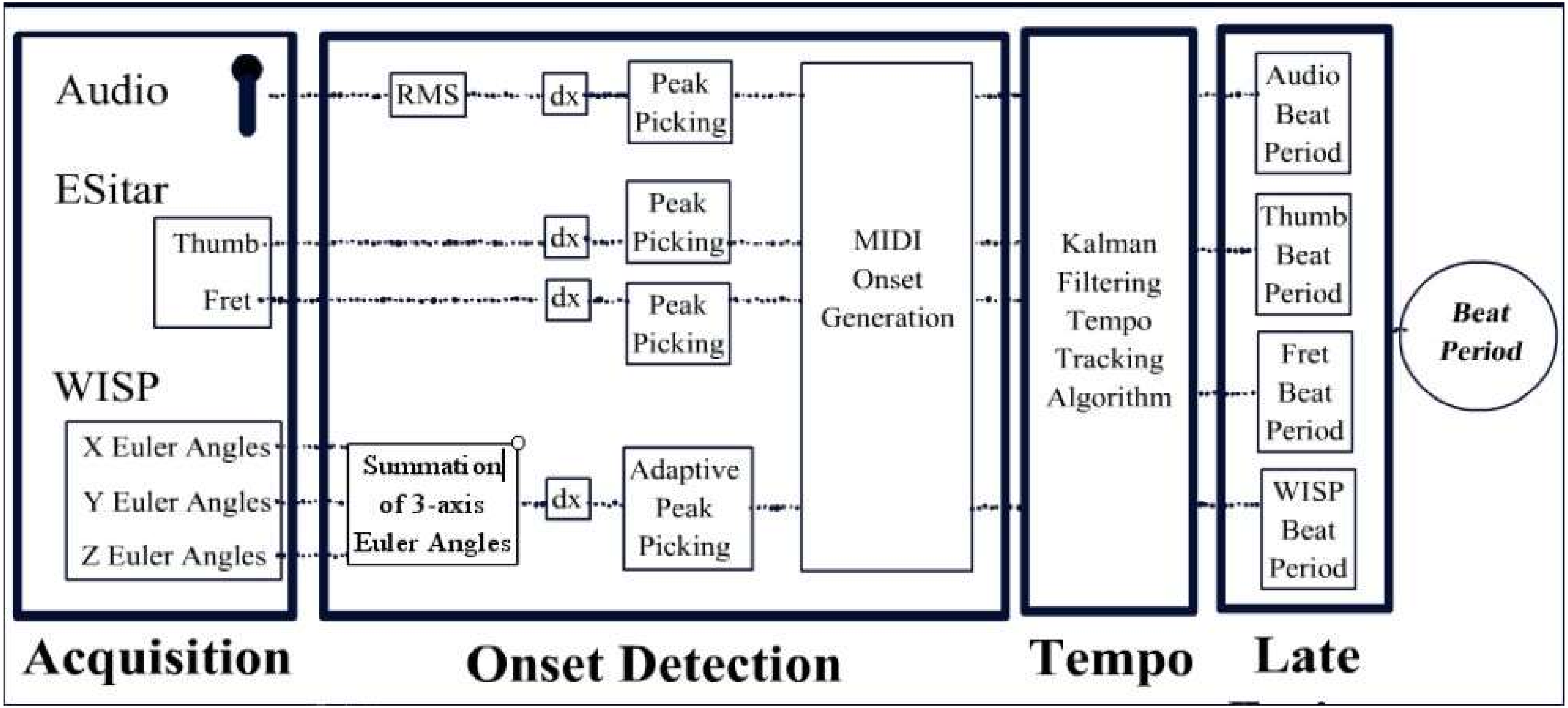
E-sitar and Mahadevibot (2007)

The E-sitar I

- Example of a hyper-instrument i.e an acoustic instrument that has been augmented with sensors to detect what the performer is playing
- Network of resistors for detecting what fret is being played
- Thumb pressure sensor for thumb
- Kiom (our version of the Wii-mote) for sensing elbow and head tilt

The E-sitar II

Real-time multi-modal beat tracking



Mahadevibot

Solenoid-based robot percussion instrument. Bobbing head visually conveys tempo information

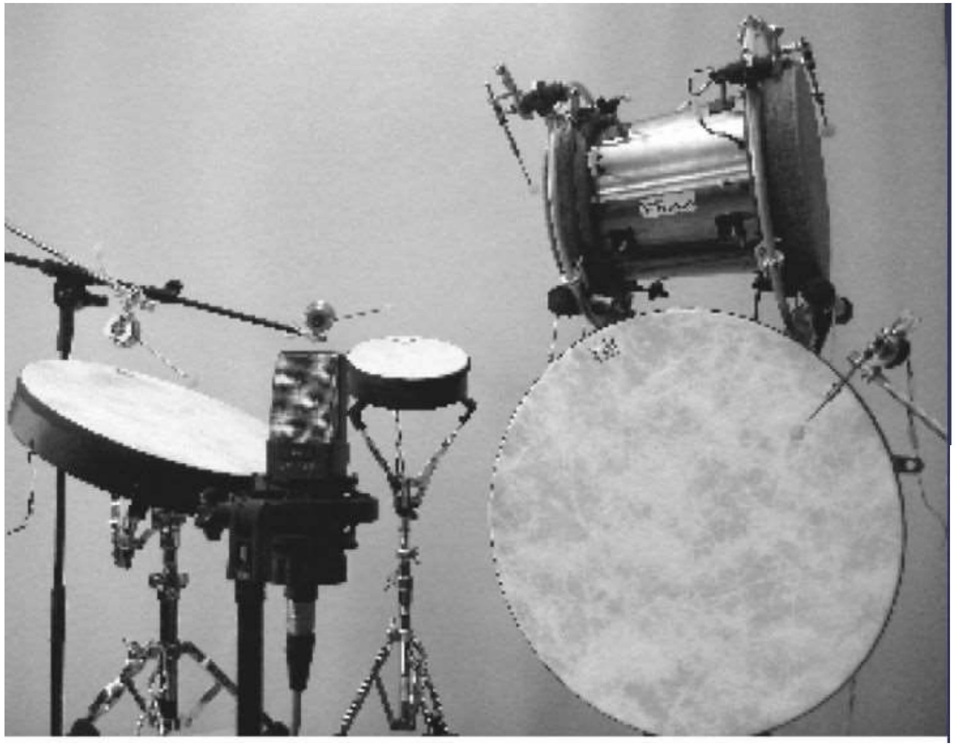
George Tzanetakis, University of Victoria

Proprioception in music robotics

- The majority of existing music robots are literally deaf i.e they only receive commands and react to them
- The ability to listen to the acoustic output has concrete practical applications
- Intelligent mapping of control messages to actuators (play hi-hat rather than solenoid #3)
- Volume calibration - play softly rather than reduce voltage

University of Victoria
Engineering

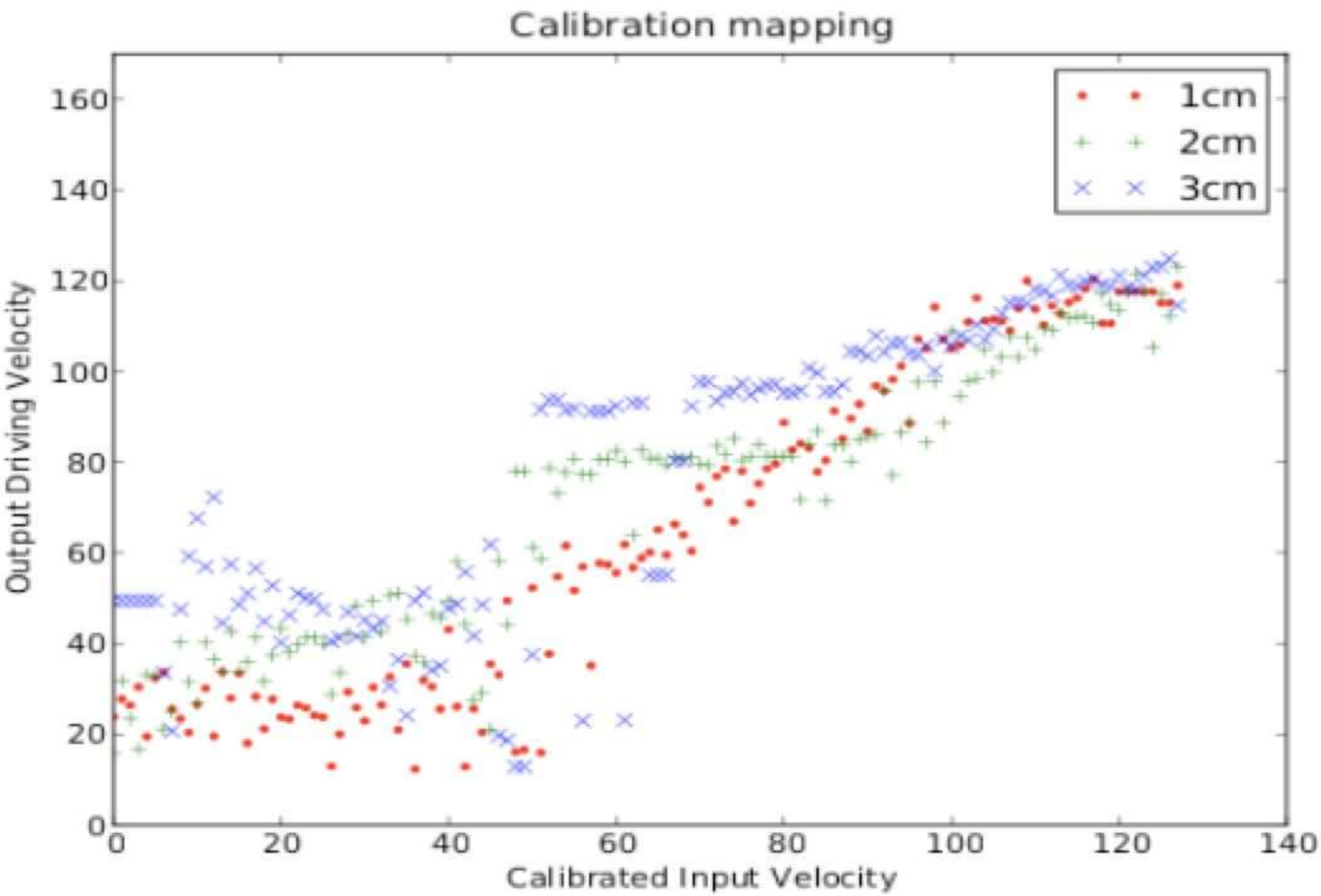
Drum classification for modular mapping



Peak offset	Percent correct	Peak offset	Percent correct
0	66.38	4	90.52
1	91.95	5	86.49
2	91.67	6	86.49
3	91.95	7	77.59

4 frame drums classification
Audio feature extraction
followed by SVM classification

Calibration map



Adjusting how hard you drive the solenoid by how loud the sound is - learning a non-linear mapping

University of Victoria
Engineering

Mechatronic Drummer
Robert van Rooyen

The most advanced percussion robot today in terms of expressiveness and dynamic range.

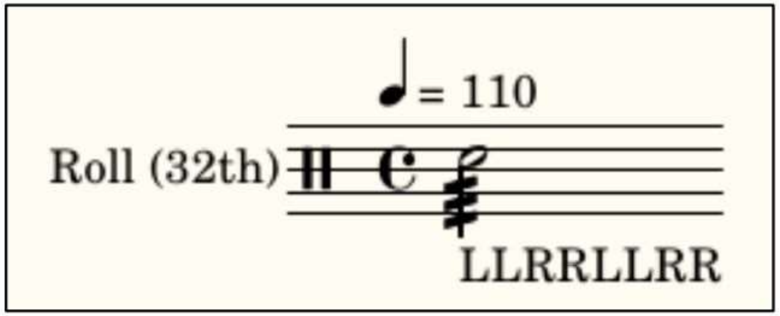
Full motion control, can be driven by data from gesture acquisition

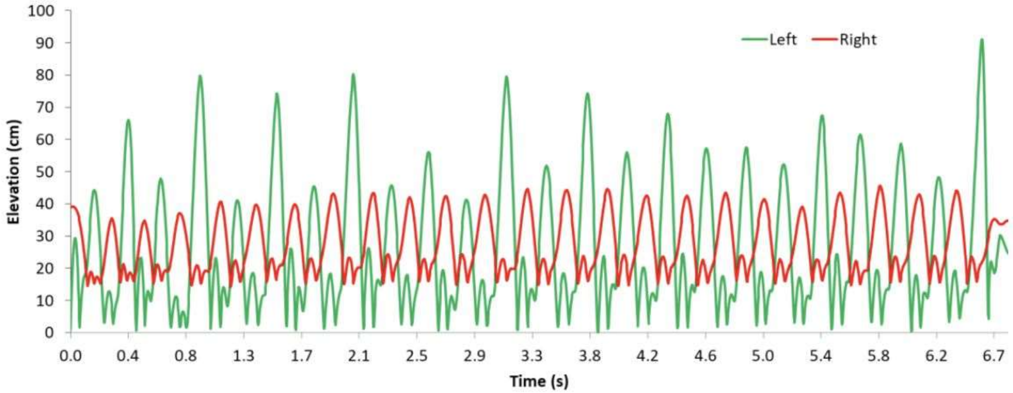
Voice coil actuators for full dynamic range and control of strike position

Mechatronic Drummer
Robert van Rooyen

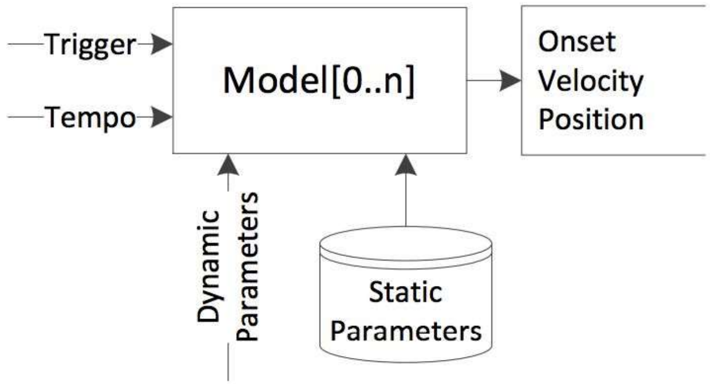
Guthman new instrument music competition - technical achievement award 2018

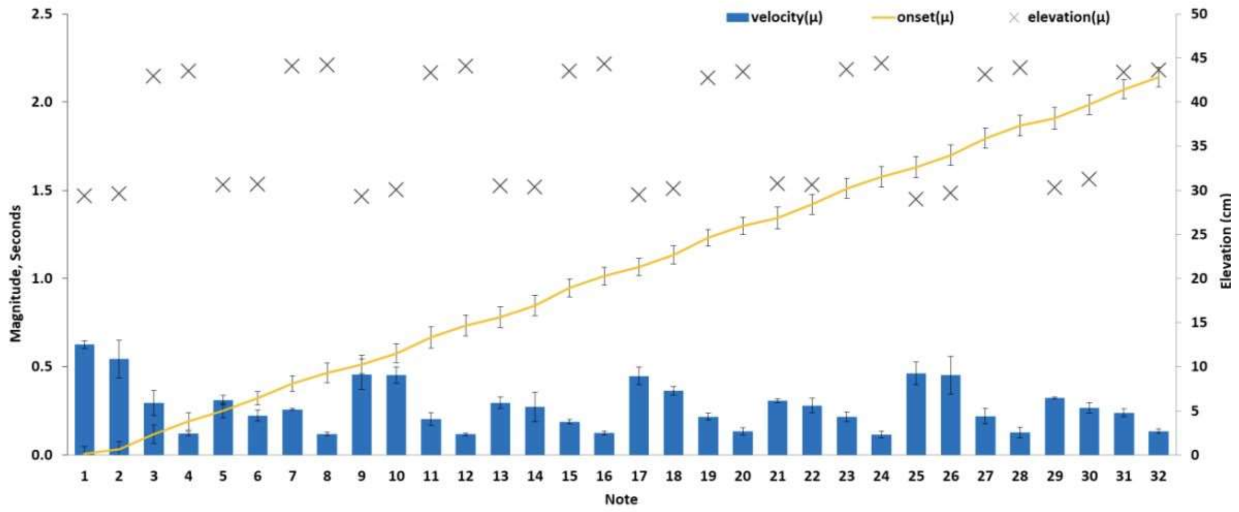
Gesture Acquisition

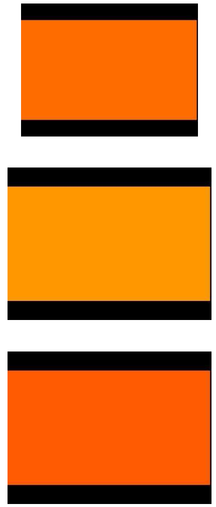




Performer-specific stochastic models





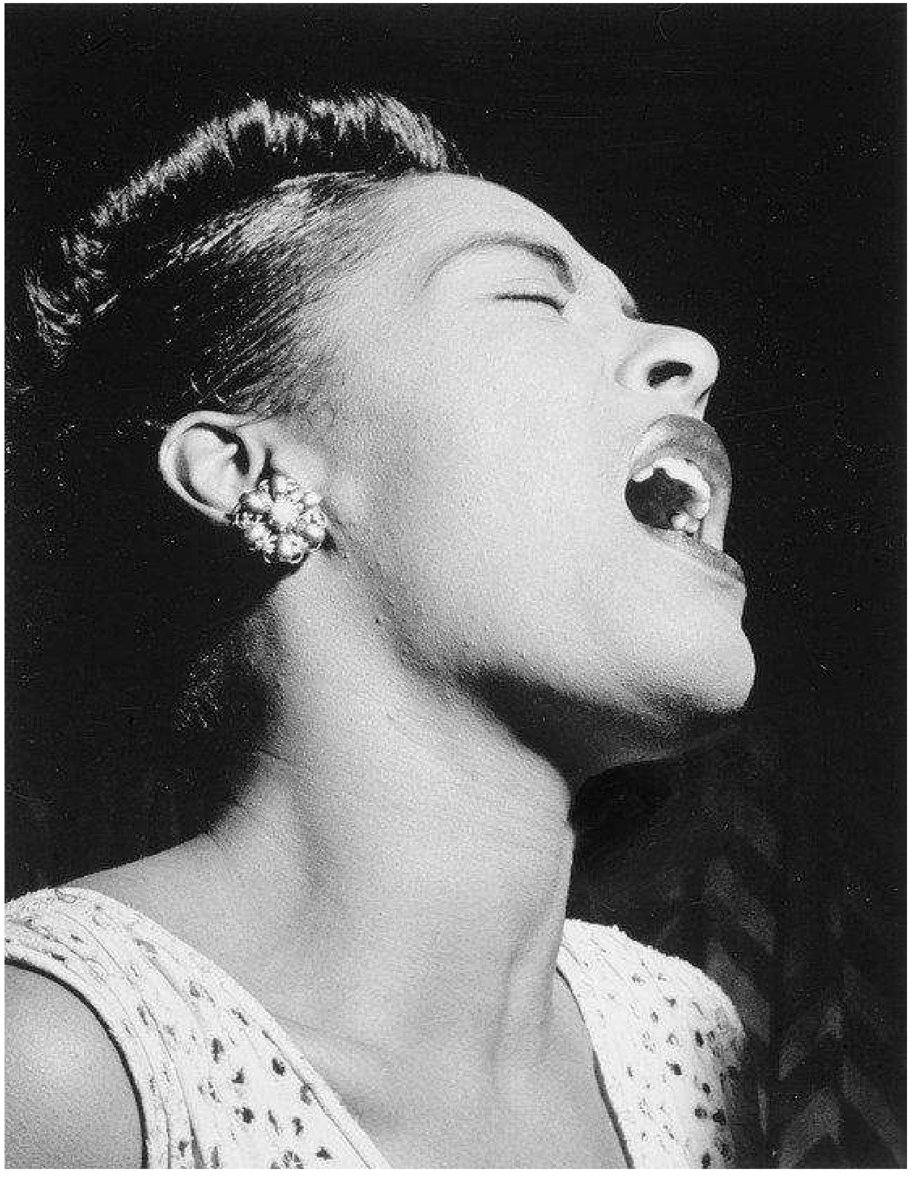


Recording

Mechanical

Performer-specific stochastic

EXPRESSIVITY



George Tzanetakis, University of Victoria

University of Victoria Engineering

Expressive drumming

- Electronic drums are simple triggers sending MIDI messages
- Not sufficient to convey the expressive nuance and physicality of percussion performance
- Adam Tindale was my second PhD student and classically trained percussionist
- Hybrid-synthesis uses a physical membrane (practice pad) to excite a synthesis model

University of Victoria Engineering

Hybrid-synthesis for expressive drumming - Adam Tindale

Intimate control with Soundplane - Randy Jones

University of Victoria
Engineering

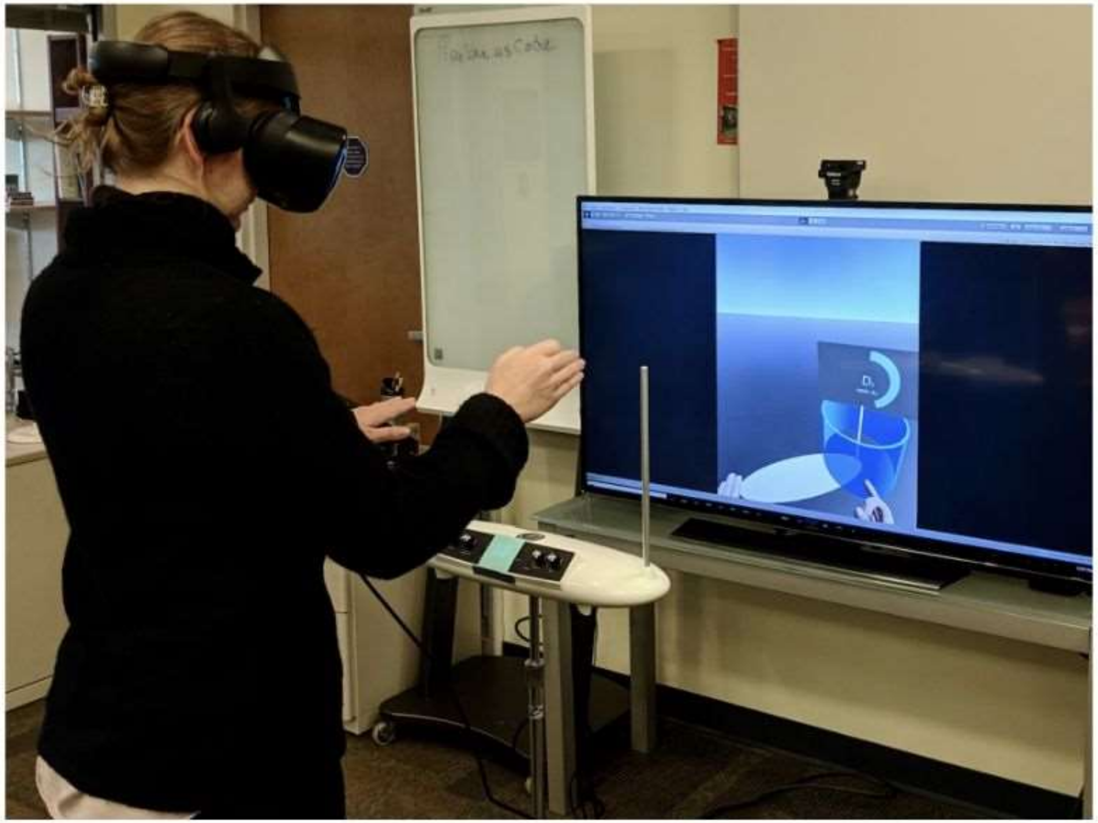
Theremin

- The Theremin is an electronic instrument invented by Leon Theremin in 1928
- It is controlled without physical contact by the performers hands
- Well-known from sci-fi movies it can be a very expressive instruments in the hands of skilled performers
- Learning to play notes is challenging because of the lack of haptic and visual feedback

Theremin

Carolina Eyck

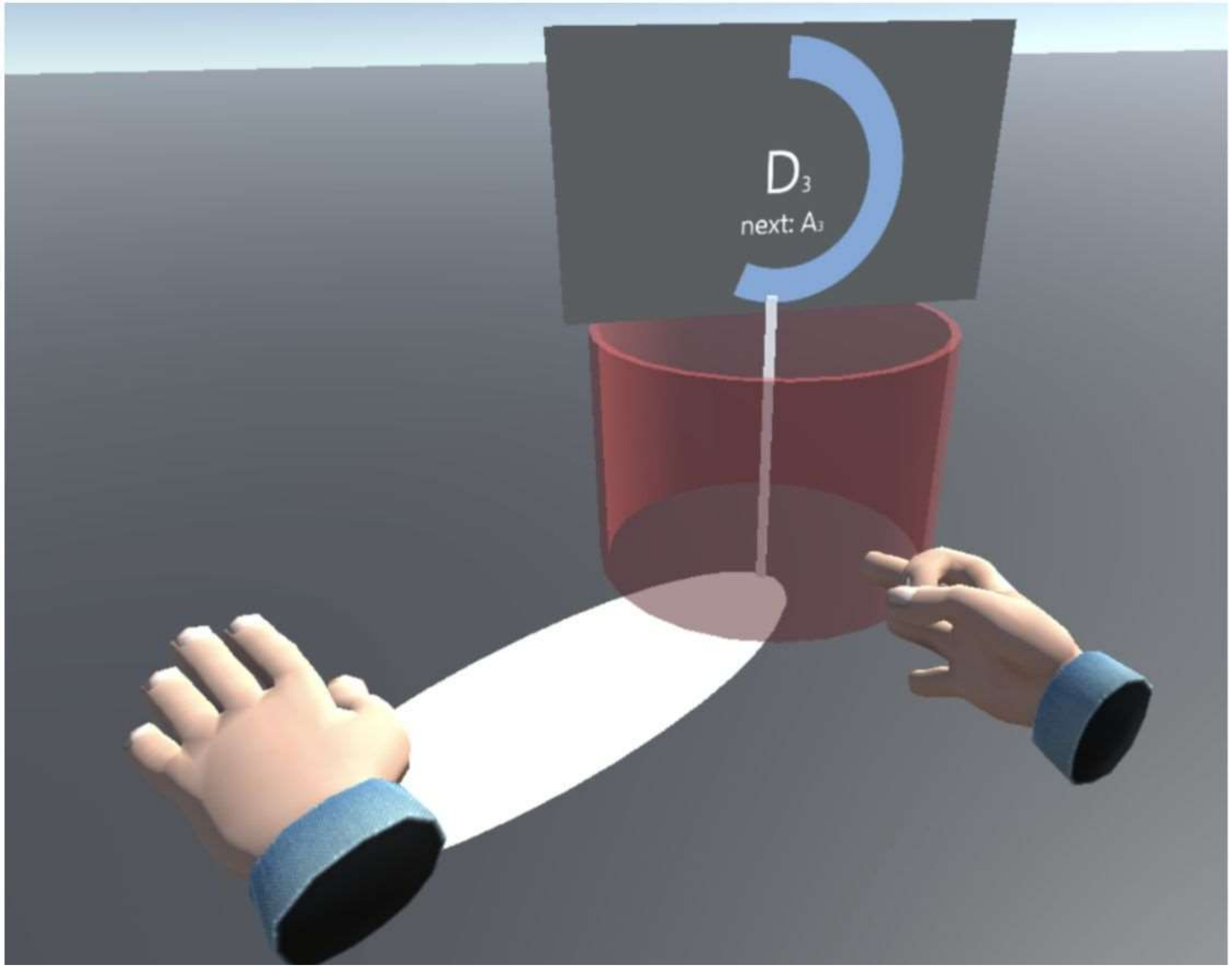
Mixed Reality Theremin - David Johnson



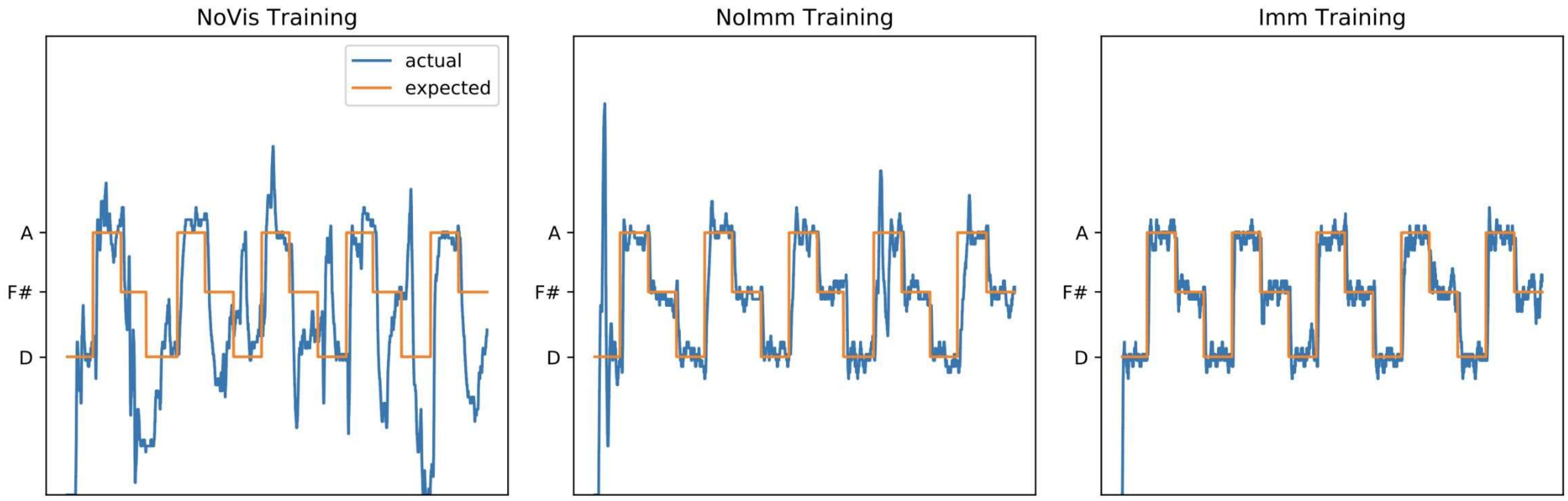
Use an actual physical Theremin for playing and sensing the hand position

In VR place a virtual representation in the right place and provide visual feedback

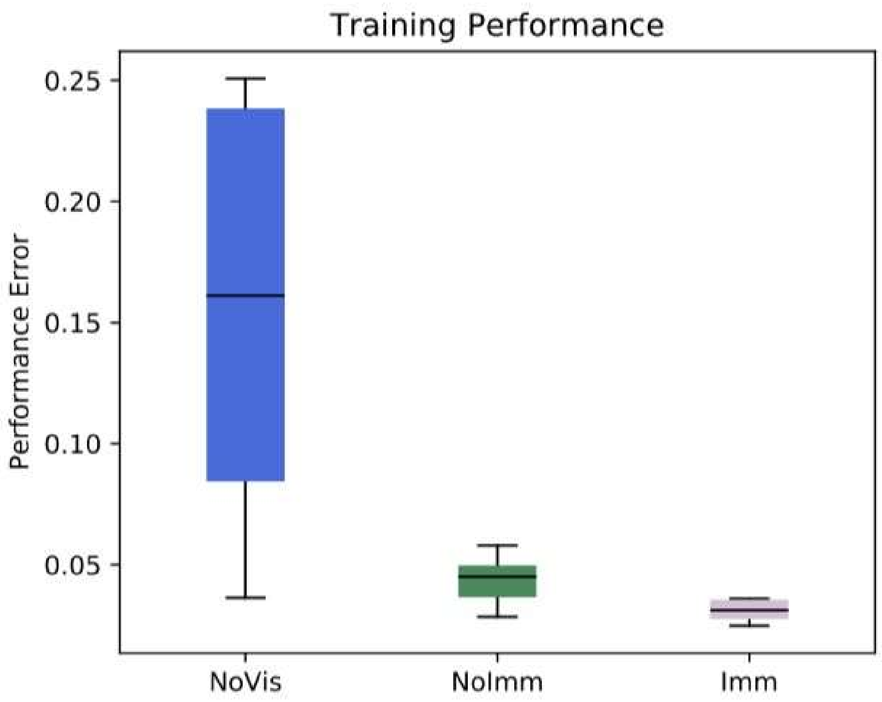
Visual Feedback

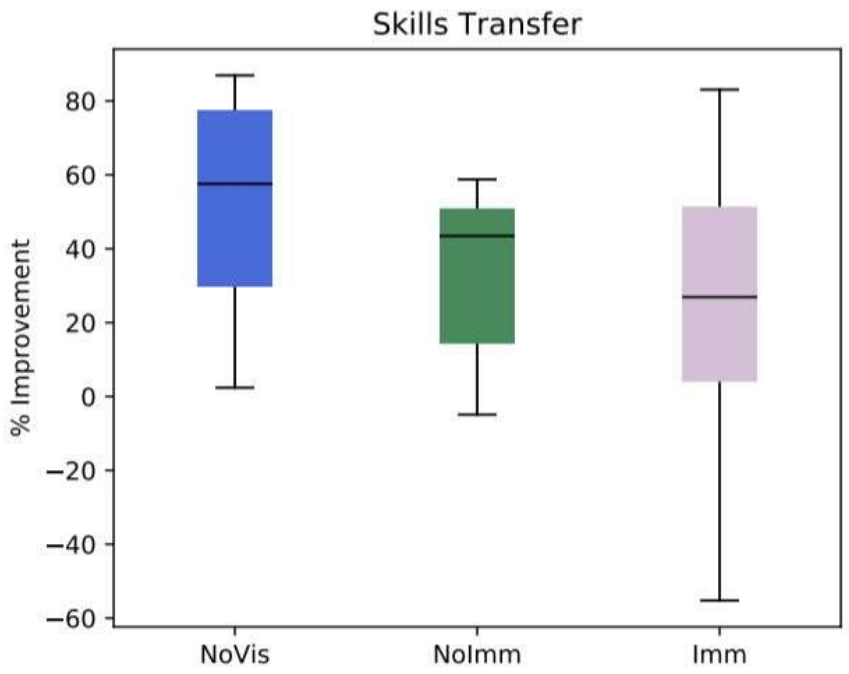


Performance data for different training environments



User study


Training Performance
Performance Error
0.25
0.20
0.15
0.10
0.05
NoVis
NoImm
Imm


Skills Transfer
% Improvement
80
60
40
20
0
−20
−40
−60
NoVis
NoImm
Imm

University of Victoria Engineering

Concluding thoughts

- Having a body (sensors and actuators) introduces layers of possible failure that provide opportunity for the sublime to occur
- Collaboration and communication between different entities - deeply personal and communal at the same time
- Music is not just pitches in time but has much richer and nuanced layers of information
- The challenges of perception, communication, embodiment, and expressivity also apply to general AI

University of Victoria Engineering

Kadenze MIR program

- Three courses:
  - Extracting information from audio signals
  - Machine learning for music information retrieval
  - Music Retrieval Systems

- https://www.kadenze.com/programs/music-information-retrieval

Dedicated to David Wessel (1942-2014)