University of Victoria
Engineering

Past, Present and Future of AI: a music and audio perspective

George Tzanetakis,
University of Victoria, Canada

Canadian Research Software Conference
2022

University of Victoria Engineering

Outline

- Background
- Artificial intelligence in music and beyond
- Past 1998-2010 (traditional ML)
- Present 2010-2022 (deep learning)
- Future challenges and opportunities > 2022

University of Victoria Engineering

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
- Summary: uses AI for music and audio

University of Victoria Engineering

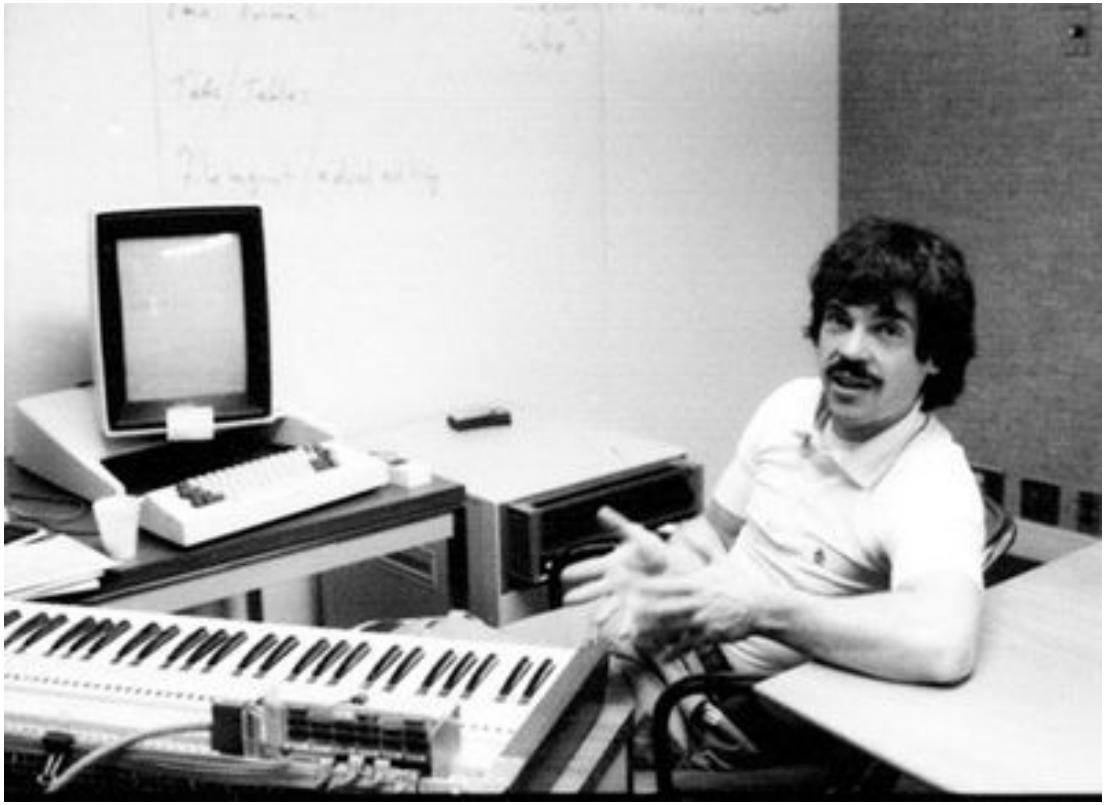
Music Background

- 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

University of Victoria
Engineering

Allan Kay

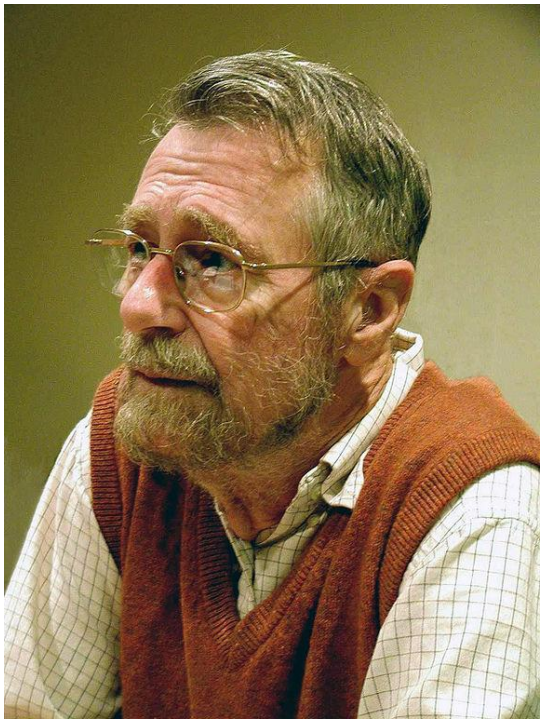
The best way to predict the future is to invent it



University of Victoria Engineering

Why ?

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



University of Victoria Engineering

Maybe it actually is interesting

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

University of Victoria Engineering

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”

Old driving vision: the great celestial jukebox
New driving vision: a virtual musician

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

George Tzanetakis, University of Victoria

University of Victoria Engineering

Past Projects (1998-2008) Traditional ML

Projects from my own body of work beyond your typical ML system (for music think Spotify):

- **Software:** Marsyas
- **Bioacoustics:** Orchive
- **Affective computing:** Laughing rats
- **Physical computing:** Morpheus guitar pedals

University of Victoria Engineering

2002-2017

Marsyas 2.0 was started after completing my PhD in 2002 while doing a PostDoc at CMU:

- Framework for audio analysis and synthesis
- Many projects in both academia and industry
- Prototyping of real-time interaction with integrated machine learning
- Declarative dataflow architecture (separation of computational structure from execution)
- >200K LOC, > 1000 citations
- Active development stopped around 2017

University of Victoria Engineering

2007-2013

Orchive

20TB archive of Orca Vocalizations recorded at Orcalab over 30 years

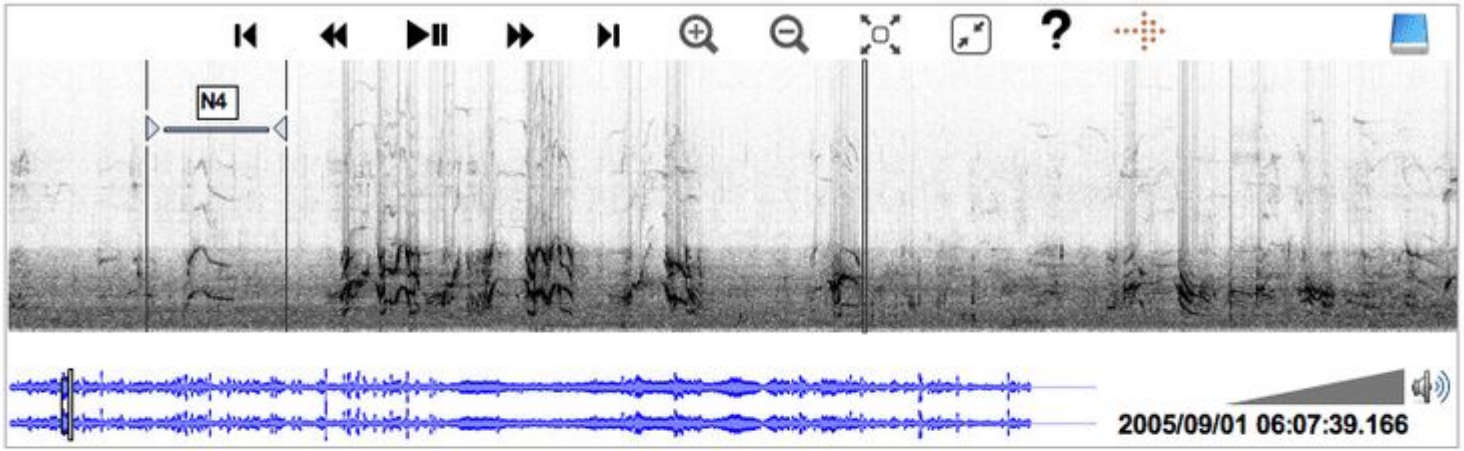
- Partially funded by Canarie/Ocean Networks
- Web interface - human in the loop annotation
- Similarity-based labeling
- Classification/temporal segmentation of presence of Orca vocalizations
- Call types (~50 for North Resident Killer Whales)
- SVM + time/frequency features

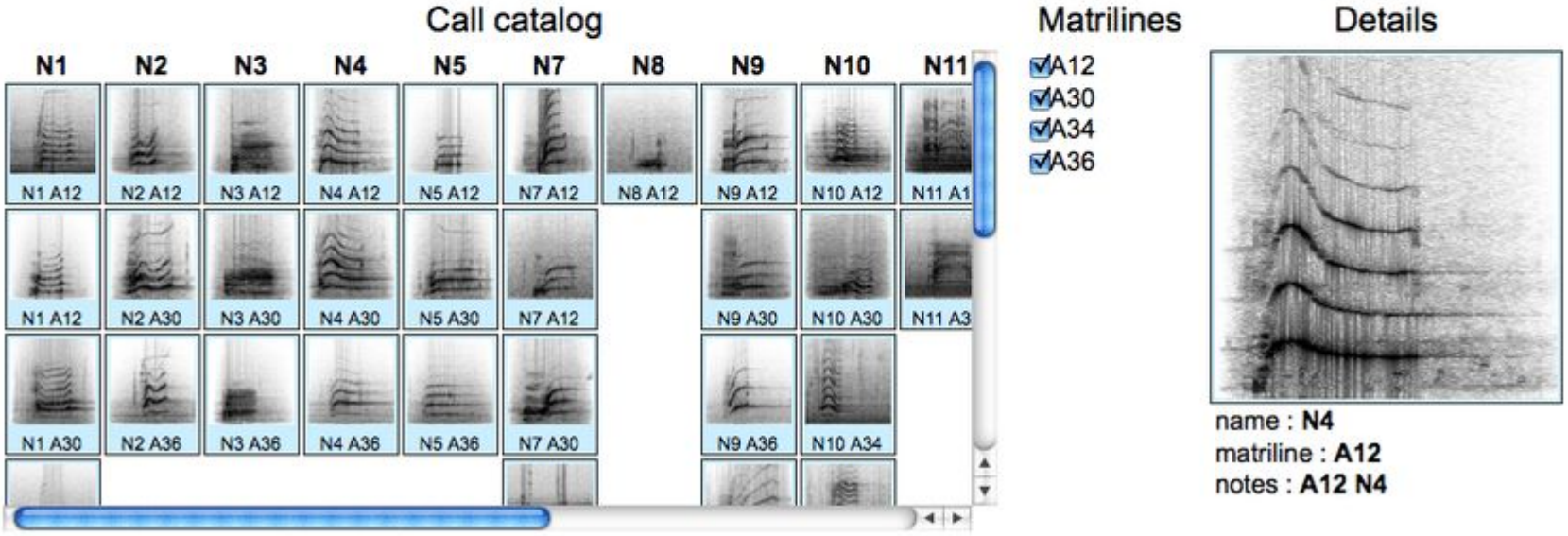
George Tzanetakis, University of Victoria

University of Victoria
Engineering

2007-2013

Orchive





George Tzanetakis, University of Victoria

University of Victoria Engineering

2009

Laughing Rats

Rats laugh (for example when tickled) - ultrasonic chirps - monitoring important for mood altering drug development

- Automatic detection and pitch shifting
- Time/Frequency features + SVM
- Phase-vocoding for lower pitch without time-stretching
- Affective computing

George Tzanetakis, University of Victoria

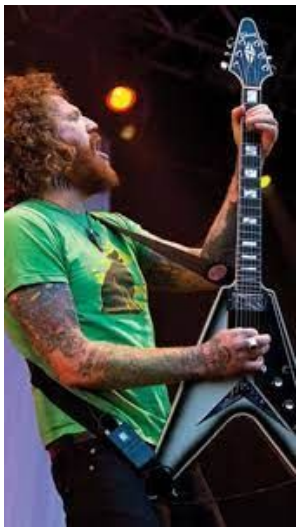
University of Victoria Engineering

2010-2011

Morpheus DropTune

Guitar pedal for lowering the pitch without time stretching (tuning strings lower) - tradeoff between low latency and good frequency resolution

- Short/Long time frequency window blending
- SVM+audio features for classification
- Literally “hardwired” into the embedded code



George Tzanetakis, University of Victoria

University of Victoria Engineering

Present Projects (2010-2022)

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

- **Perception:** Teaching a virtual violinist to bow
- **Communication:** Markov logic networks
- **Embodiment:** Music robots
- **Expressivity:** Hybrid synthesis for expressive drumming

University of Victoria Engineering

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)

University of Victoria Engineering

2011

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

George Tzanetakis, University of Victoria

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

George Tzanetakis, University of Victoria

University of Victoria
Engineering

Physical Model

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

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

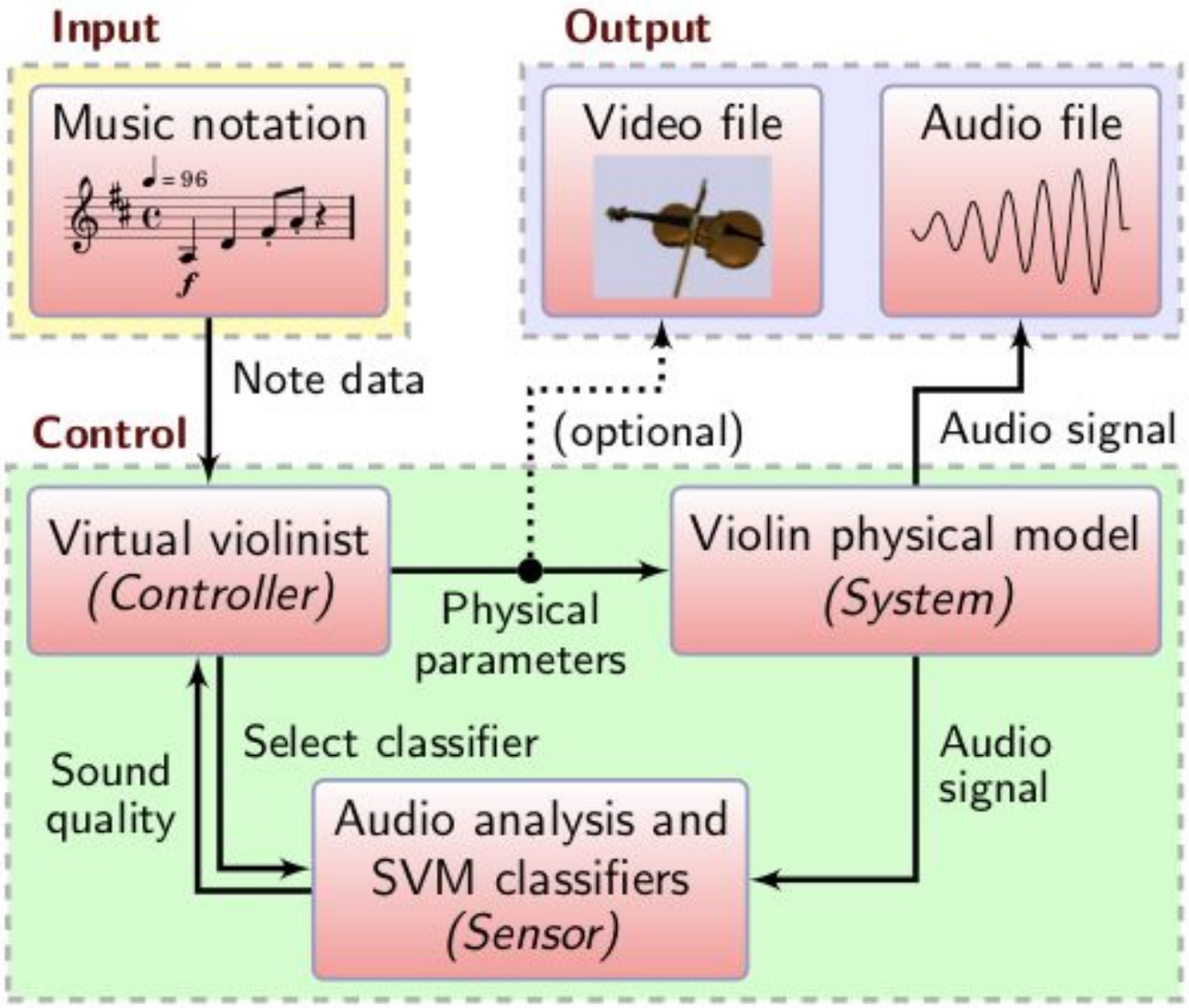
- Implemented as a C++ library, published under GNU GPLv3+

Input parameters

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

University of Victoria Engineering

System Architecture



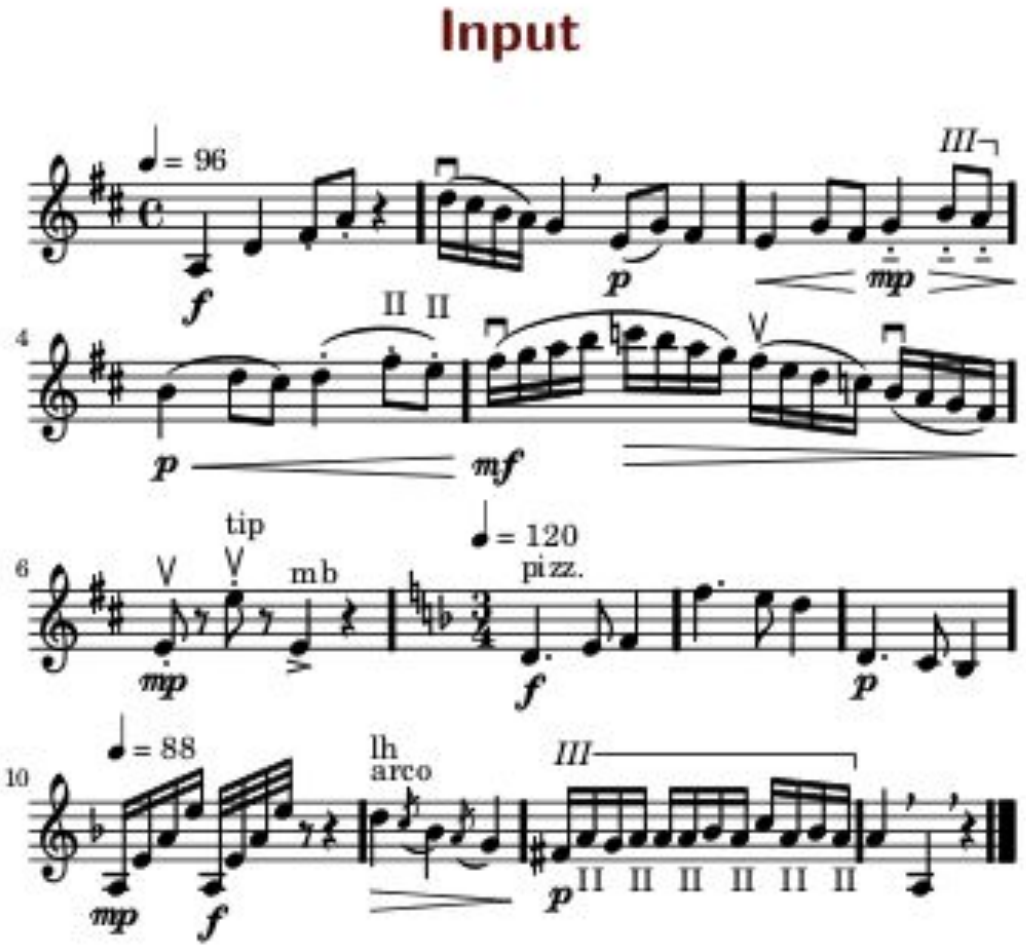
University of Victoria Engineering

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

University of Victoria
Engineering

Playing a piece



George Tzanetakis, University of Victoria

University of Victoria Engineering

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

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

University of Victoria Engineering

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)

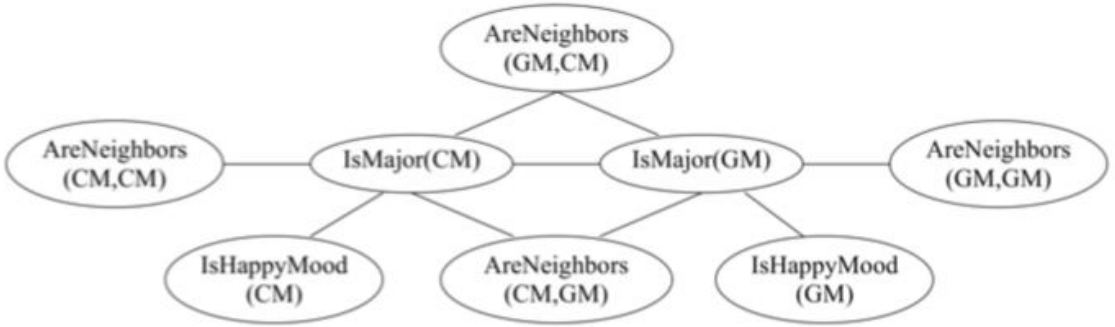
- Expressive formalism that combines probabilistic graphical models and first-order logic inference
- Highly flexible and expressive language for the harmonic analysis of audio music signals
- MLN is a set of weighted first-order logic formulas that can be viewed as a template for creating a Probabilistic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

University of Victoria Engineering

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.	$\forall$ x IsMajor(x) $\Rightarrow$ IsHappy-Mood(x)	$w_1 = 0.5$
If two chords are neighbors, either the two are major chords or neither are.	$\forall$ x $\forall$ y AreNeighbors(x, y) $\Rightarrow$ (IsMajor(x) $\Leftrightarrow$ 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

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

University of Victoria Engineering

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

University of Victoria Engineering

E-sitar and Mahadevibot (2007)

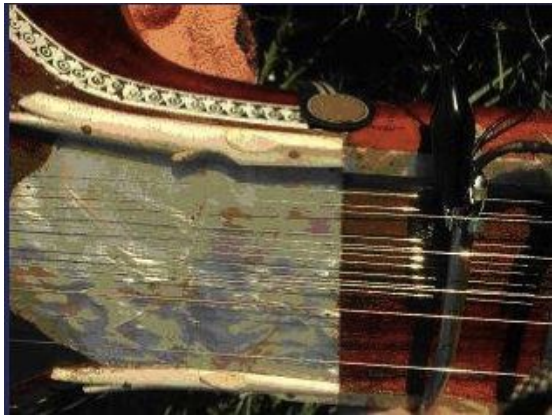
University of Victoria Engineering

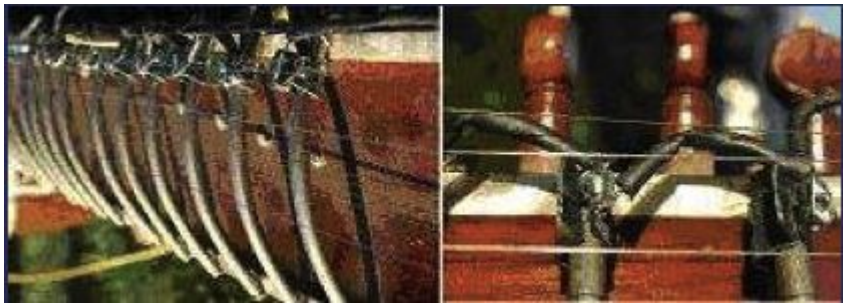
The E-sitar I

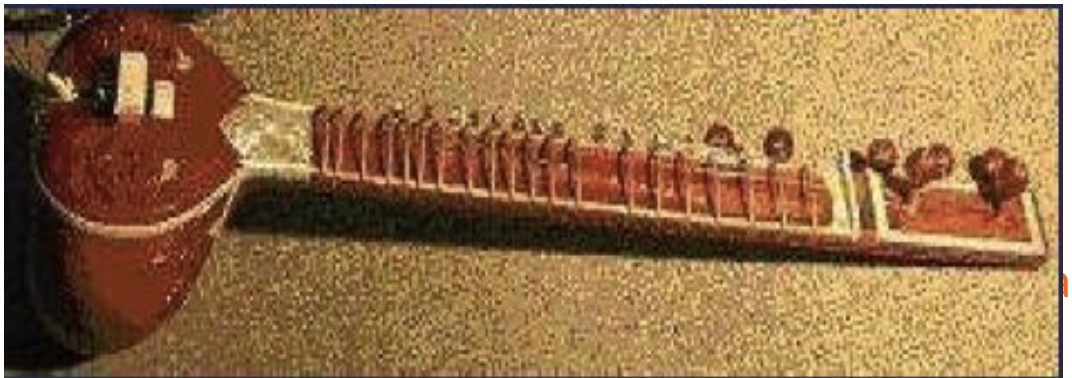
- Example of a hyper-instrument i.e an acoustic instrument that has been augmented with sensors to detect what the performer is playing
- Network of resistors for detecting what fret is being played
- Thumb pressure sensor for thumb
- Kiom (our version of the Wii-mote) for sensing elbow and head tilt

University of Victoria
Engineering

The E-sitar II

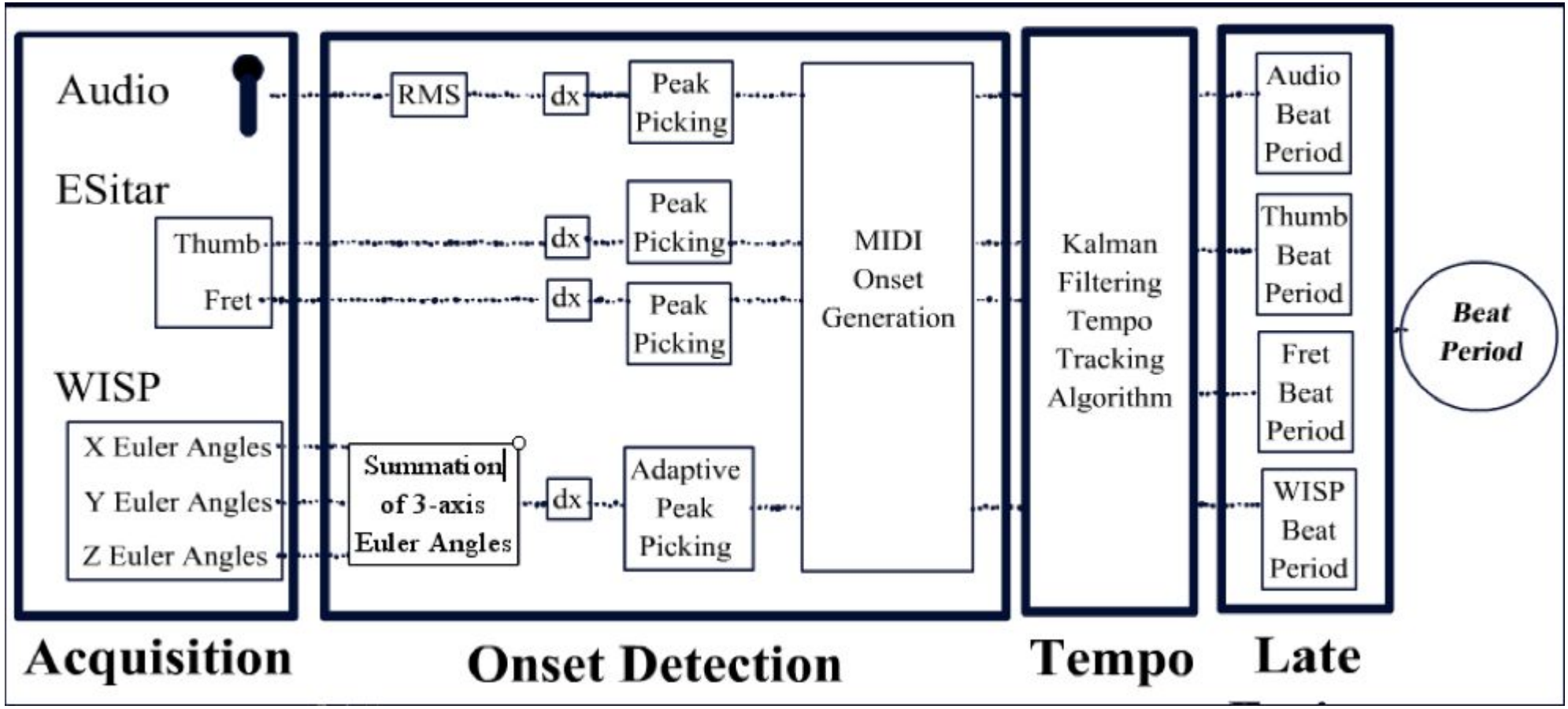






University of Victoria Engineering

Real-time multi-modal beat tracking



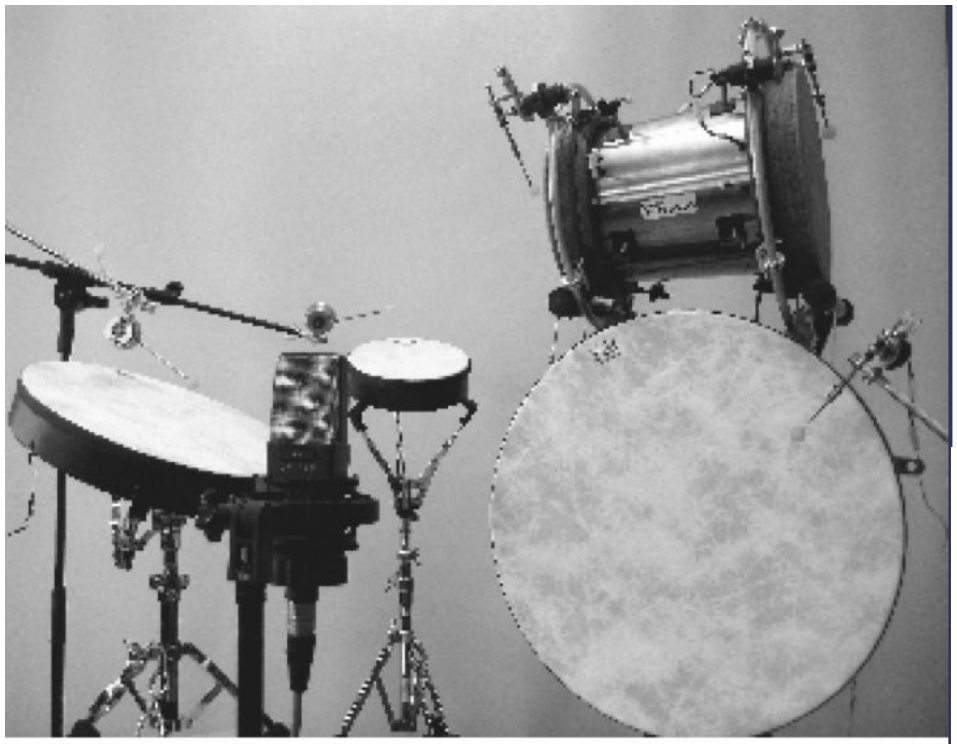
University of Victoria Engineering

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

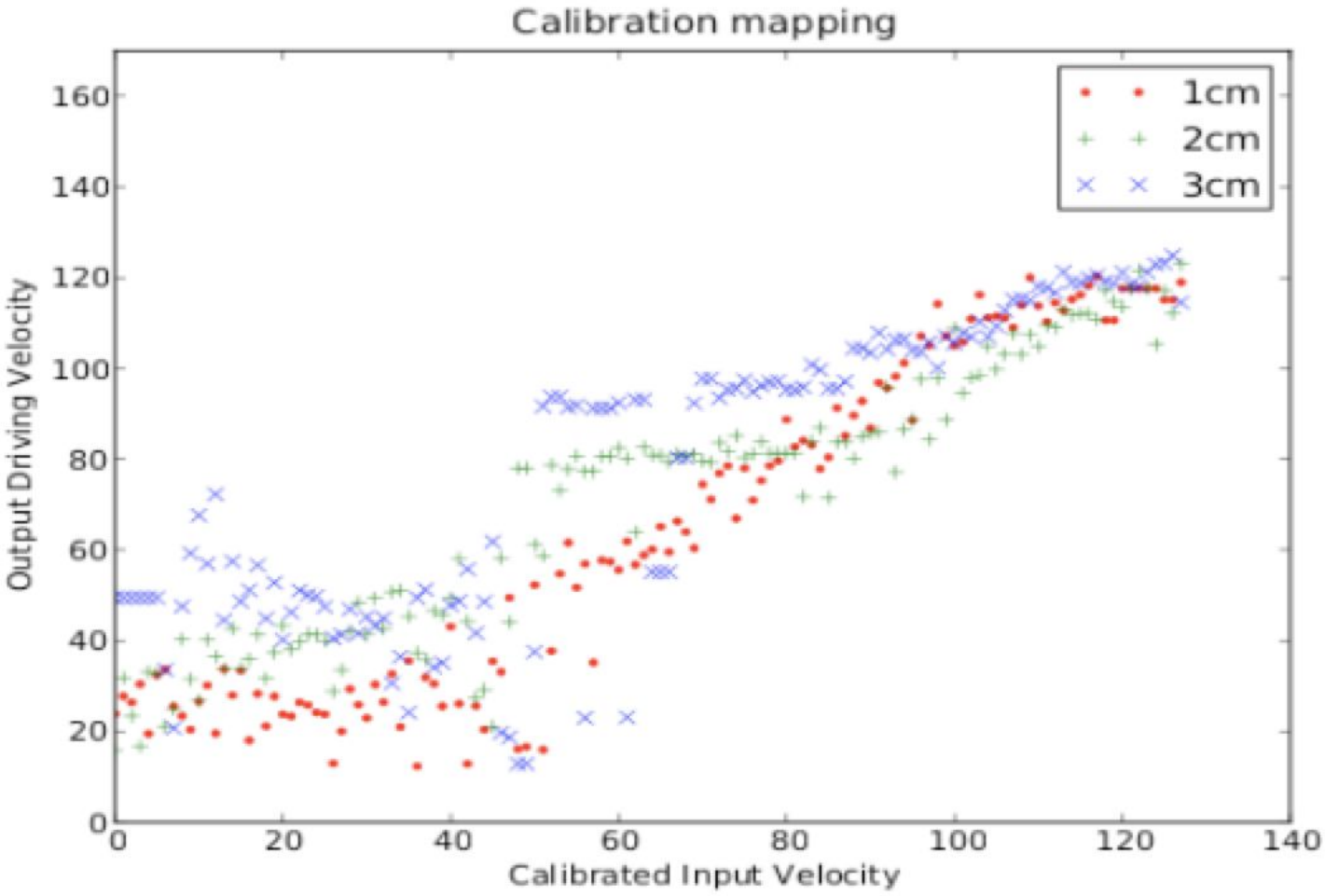


Peak offset	Percent correct	Peak offset	Percent correct
0	66.38	4	90.52
1	91.95	5	86.49
2	91.67	6	86.49
3	91.95	7	77.59

4 frame drums classification
Audio feature extraction
followed by SVM classification

University of Victoria Engineering

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.

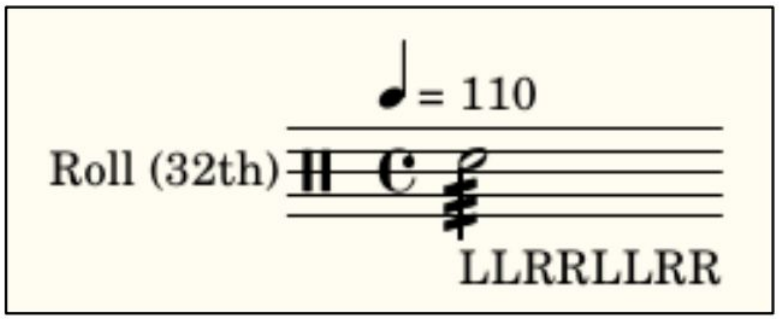
Full motion control, can be driven by data from gesture acquisition

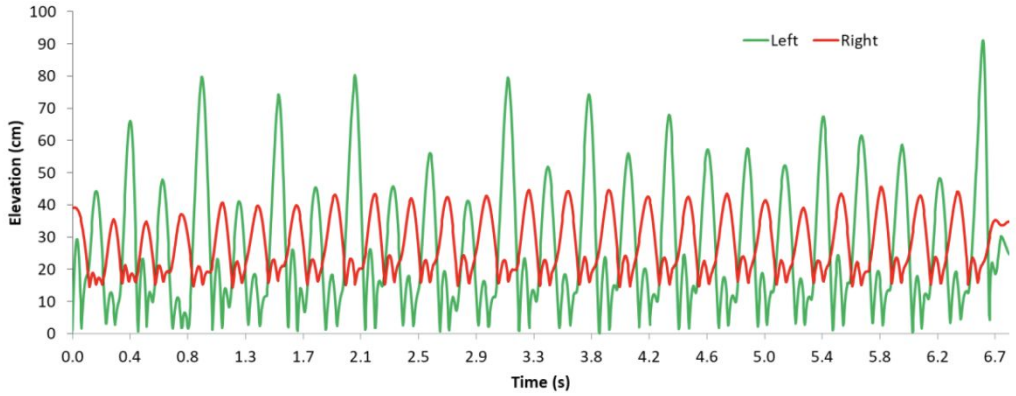
Voice coil actuators for full dynamic range and control of strike position

George Tzanetakis, University of Victoria

University of Victoria
Engineering

Gesture Acquisition

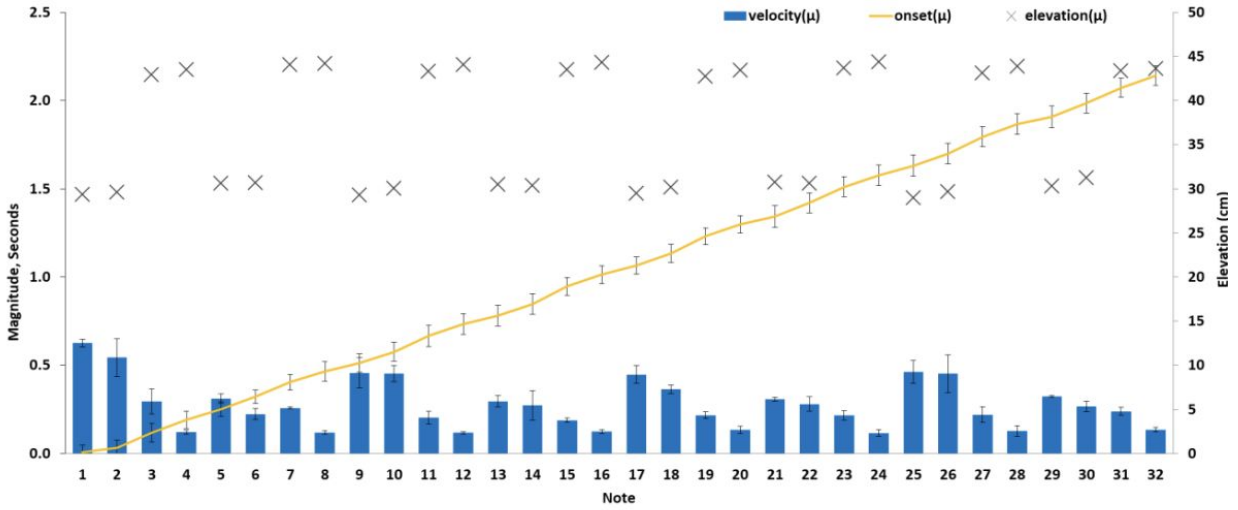


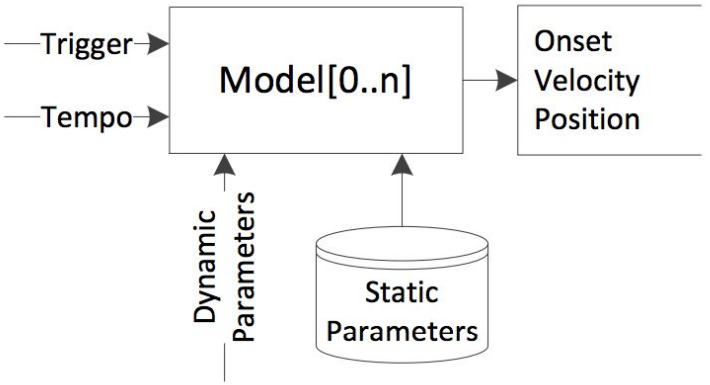


George Tzanetakis, University of Victoria

University of Victoria Engineering

Performer-specific stochastic models





Recording

Mechanical

Performer-specific stochastic

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

George Tzanetakis, University of Victoria

University of Victoria Engineering

Future work (> 2022)

Current projects:

- **ML:** Online PU-learning
- **VR/AR:**
- **HCI**: music virtual assistant - MIR combined with NLP
-

University of Victoria Engineering

Future work (> 2022)

Ideas:

- **PL:** integration of time and probability, integrative AI, pyro.AI
- **ML:** From interpretability to communication, multi-modal decomposition
- **CS:** Merging of interaction, programming, and learning (game engines)
- **VR/AR:** Virtual, physical, and human musician ensembles

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
- The challenges of perception, communication, embodiment, and expressivity also apply to general AI
- We need to work on both submarines and cyber fishes - just be honest about which is which

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