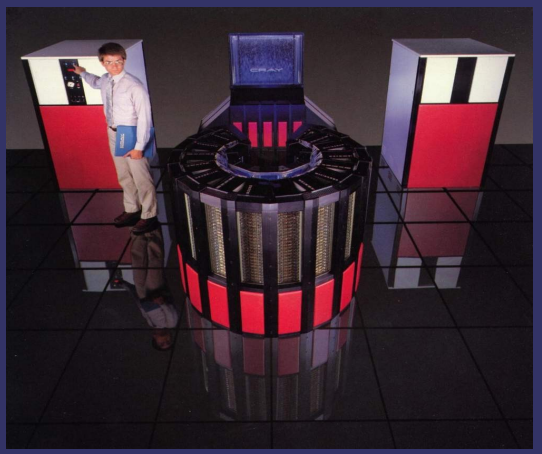
When the virtual bleeds into the physical: a music perspective for the future of human computer interaction

George Tzanetakis
Associate Professor
Computer Science (also ECE and Music)
University of Victoria

Canada Tier II Research Chair in
Computer Analysis of Audio and Music

Human Computer Interaction

Despite the breathtaking progress of computing technology (iPad tied with Cray-2, 1985 – 17 million), HCI is very limited compared to HHI



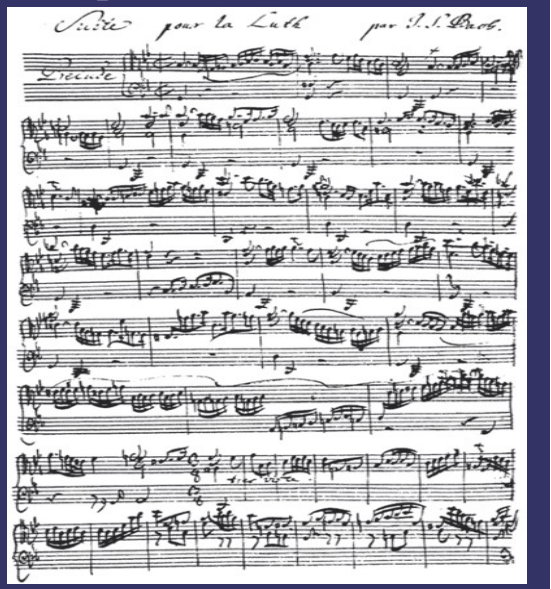
Two fundamental problems

- Lack of shared understanding
- Communication bottleneck:
  - Input: Mouse/keyboard (touch screen)
  - Output: Rectangular 2D screen, spe

Why Music ?

- The way music is created, distributed, and perceived has been and will be transformed by advances in technology
- Throughout history and throughout the world humans have been making music
- Arguably the most complex and expressive interaction with artifacts

Some examples from history



Overview

Digital Music Data

2000: Content-analysis based on Signal Processing and Machine Learning ~1000 tracks

2010: iTunes Genius, Google Music Instant Playlists, Pandora, Last.FM, Echonest, Spotify.... ~ 13 million tracks

The future: all of recorded music (48 hours uploaded to YouTube every minute)

Music Information Retrieval

Query by Humming

SoundHound
Instant Music Search and Discovery

Mobile Search -
Audio Fingerprinting

SHAZAM

Music Recommendation

Google

Personalized Radio

last·fm
the social music revolution

PANDORA
internet radio

Semantic Annotation/ Automatically Tagging Music

1) Given a music track automatically predict words that music listeners would use to describe it

2) Given a set of words describing music return a list of music tracks for which the words would be applicable

cal500.sness.net demo

Improving automatic music tag annotation using stacked generalization of probabilistic SVM output – S. Ness, A. Theocharis, L.G. Martins, G. Tzanetakis, ACM Multimedia 2009

Artist : Ac Dc

Song : Dirty Deeds Done Dirt Cheap

00:00:00.000

Track			
10cc	Emotion-Angry_/_Agressive	Emotion-Arousing_/_Awakening
2pac	Emotion-Arousing_/_Awakening	Emotion-Exciting_/_Thrilling
5th Dimension	Emotion-Exciting_/_Thrilling	Emotion-Happy
A Tribe Called Quest	Emotion-Powerful_/_Strong	Emotion-Powerful_/_Strong
Aaron Neville	Genre--_Classic_Rock	Instrument_-_Bass
Abba	Genre-Best-Rock	Instrument_-_Drum_Set
Abc	Genre-Rock	Instrument_-_Harmonica
Ac Dc	Instrument_-_Drum_Set	Instrument_-_Male_Lead_Vocals
Adam And The Ants	Instrument_-_Male_Lead_Vocals	NOT-Emotion-Calming_/_Soothing
Adam Ant	NOT-Emotion-Bizarre_/_Weird	NOT-Emotion-Laid-back_/_Mellow
Adverts	NOT-Emotion-Calming_/_Soothing	NOT-Emotion-Loving_/_Romantic
Aerobic Jonquil	NOT-Emotion-Carefree_/_Lighthearted	NOT-Emotion-Sad
Aerosmith	NOT-Emotion-Laid-back_/_Mellow	NOT-Emotion-Tender_/_Soft
Aimee Mann	NOT-Emotion-Light_/_Playful	NOT-Emotion-Touching_/_Loving
Air	NOT-Emotion-Loving_/_Romantic	Song-Catchy/Memorable
Al Green	NOT-Emotion-Positive_/_Optimistic	Song-Fast_Tempo
Alanis Morissette	NOT-Emotion-Sad	Song-Heavy_Beat
Alice Cooper	NOT-Emotion-Tender_/_Soft	Song-High_Energy
Alice In Chains	NOT-Emotion-Touching_/_Loving	Song-Like
Alicia Keys	Song-Catchy/Memorable	Song-Positive_Feelings
Allman Brothers Band	Song-Fast_Tempo	Song-Recommend
Altered Images	Song-Heavy_Beat	Song-Recorded
American Music Club	Song-High_Energy	Song-Texture_Electric
Andrews Sisters	Song-Like	Usage-At_a_party
Ani Difranco	Song-Quality	Usage-Driving
Animals	Song-Recommend	Vocals-Gravelly

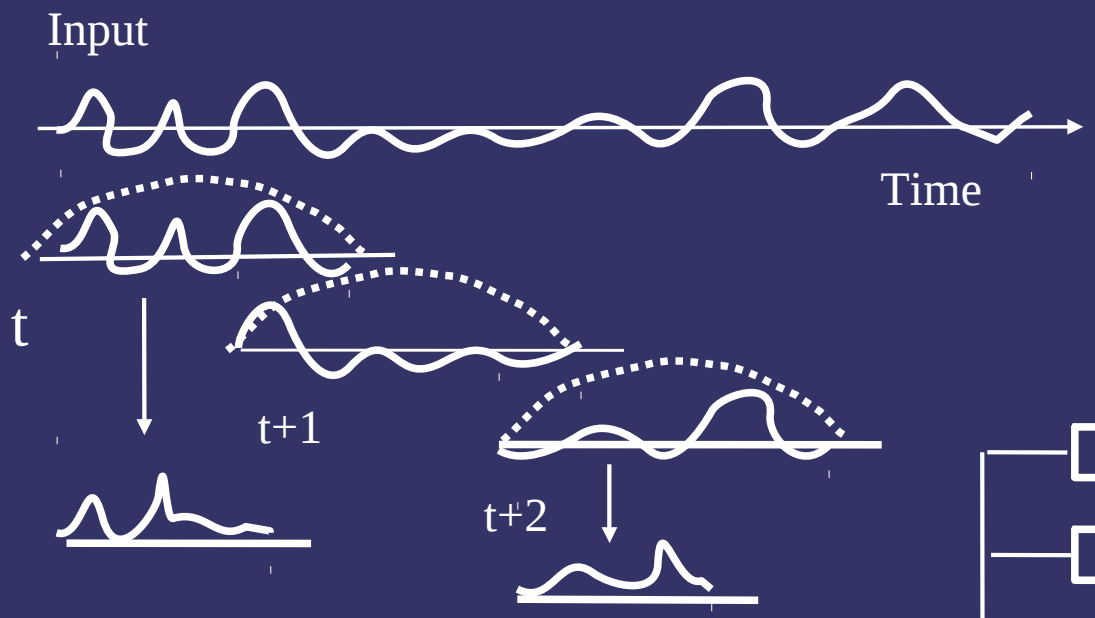
cal500.sness.net demo

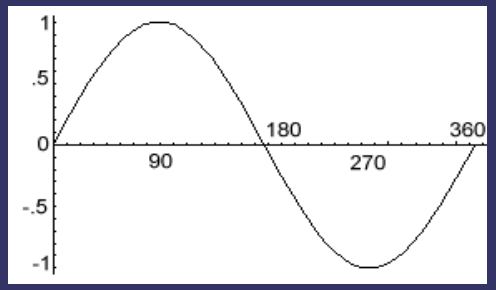
Artist : Talking Heads
Song : And She Was

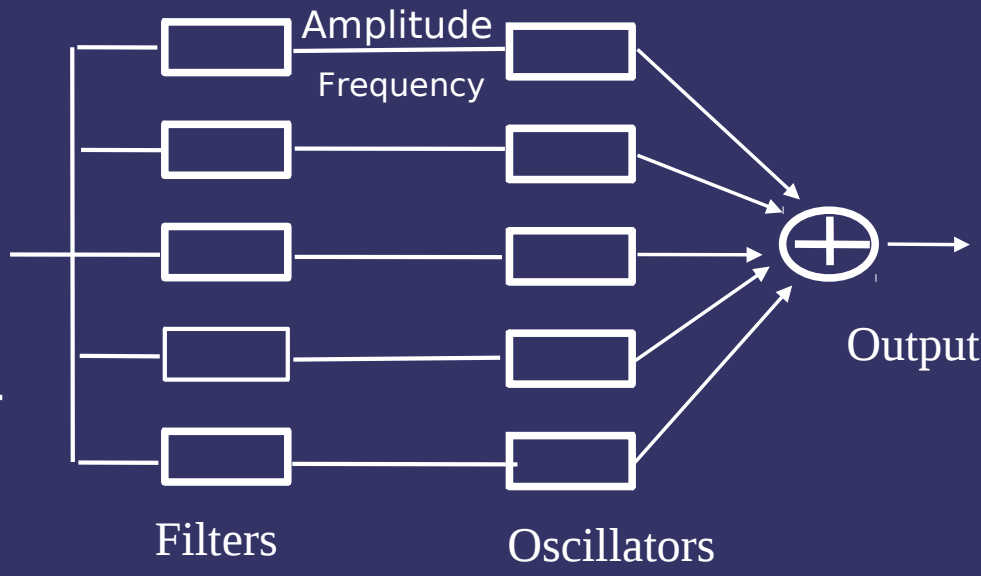
00:01:14.234

Track		
10cc	Emotion-Carefree_/_Lighthearted	Emotion-Arousing_/_Awakening
2pac	Genre-_Classic_Rock	Emotion-Carefree_/_Lighthearted
5th Dimension	Genre-Rock	Emotion-Cheerful_/_Festive
A Tribe Called Quest	Instrument_-_Backing_vocals	Emotion-Exciting_/_Thrilling
Aaron Neville	Instrument_-_Drum_Set	Emotion-Happy
Abba	Instrument_-_Male_Lead_Vocals	Emotion-Light_/_Playful
Abc	Instrument_-_Synthesizer	Emotion-Positive_/_Optimistic
Ac Dc	NOT-Emotion-Angry_/_Agressive	Genre-Rock
Adam And The Ants	NOT-Emotion-Bizarre_/_Weird	Instrument_-_Bass
Adam Ant	NOT-Emotion-Calming_/_Soothing	Instrument_-_Drum_Set
Adverts	NOT-Emotion-Emotional_/_Passionate	Instrument_-_Male_Lead_Vocals
Aerobic Jonquil	NOT-Emotion-Loving_/_Romantic	NOT-Emotion-Angry_/_Agressive
Aerosmith	NOT-Emotion-Powerful_/_Strong	NOT-Emotion-Calming_/_Soothing
Aimee Mann	NOT-Emotion-Sad	NOT-Emotion-Loving_/_Romantic
Air	NOT-Emotion-Tender_/_Soft	NOT-Emotion-Sad
Al Green	NOT-Emotion-Touching_/_Loving	NOT-Emotion-Tender_/_Soft
Alanis Morissette	NOT-Song-Changing_Energy_Level	NOT-Emotion-Touching_/_Loving
Alice Cooper	NOT-Song-Very_Danceable	Song-Catchy/Memorable
Alice In Chains	Song-Catchy/Memorable	Song-Fast_Tempo
Alicia Keys	Song-Like	Song-High_Energy
Allman Brothers Band	Song-Positive_Feelings	Song-Like
Altered Images	Song-Recorded	Song-Positive_Feelings
American Music Club	Song-Texture_Acoustic	Song-Recorded
Andrews Sisters	Song-Texture_Electric	Song-Texture_Acoustic
Ani Difranco	Song-Tonality	Song-Texture_Electric

Short Time Fourier Transform

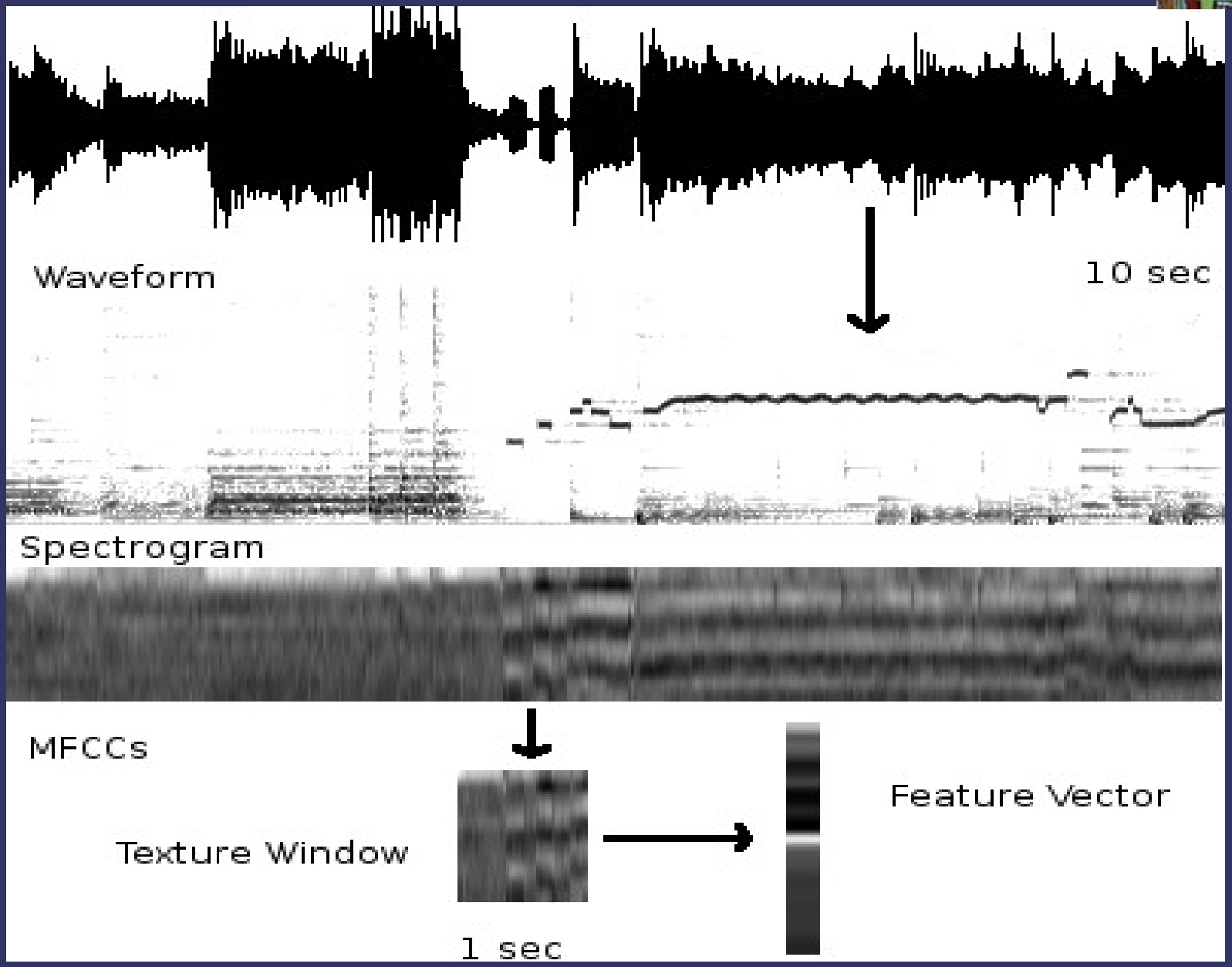




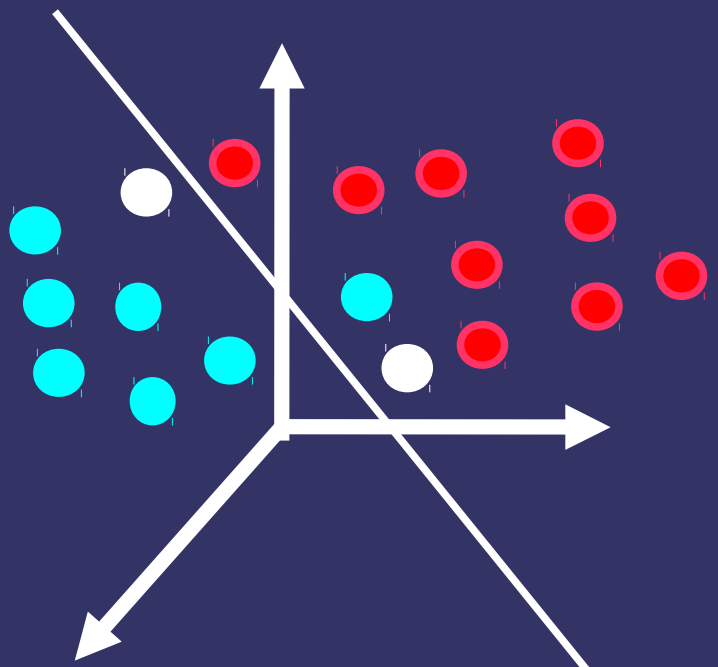


Time-varying spectra
Fast Fourier Transform FFT

Audio Feature Extraction



Statistical Supervised Learning



Decision boundary

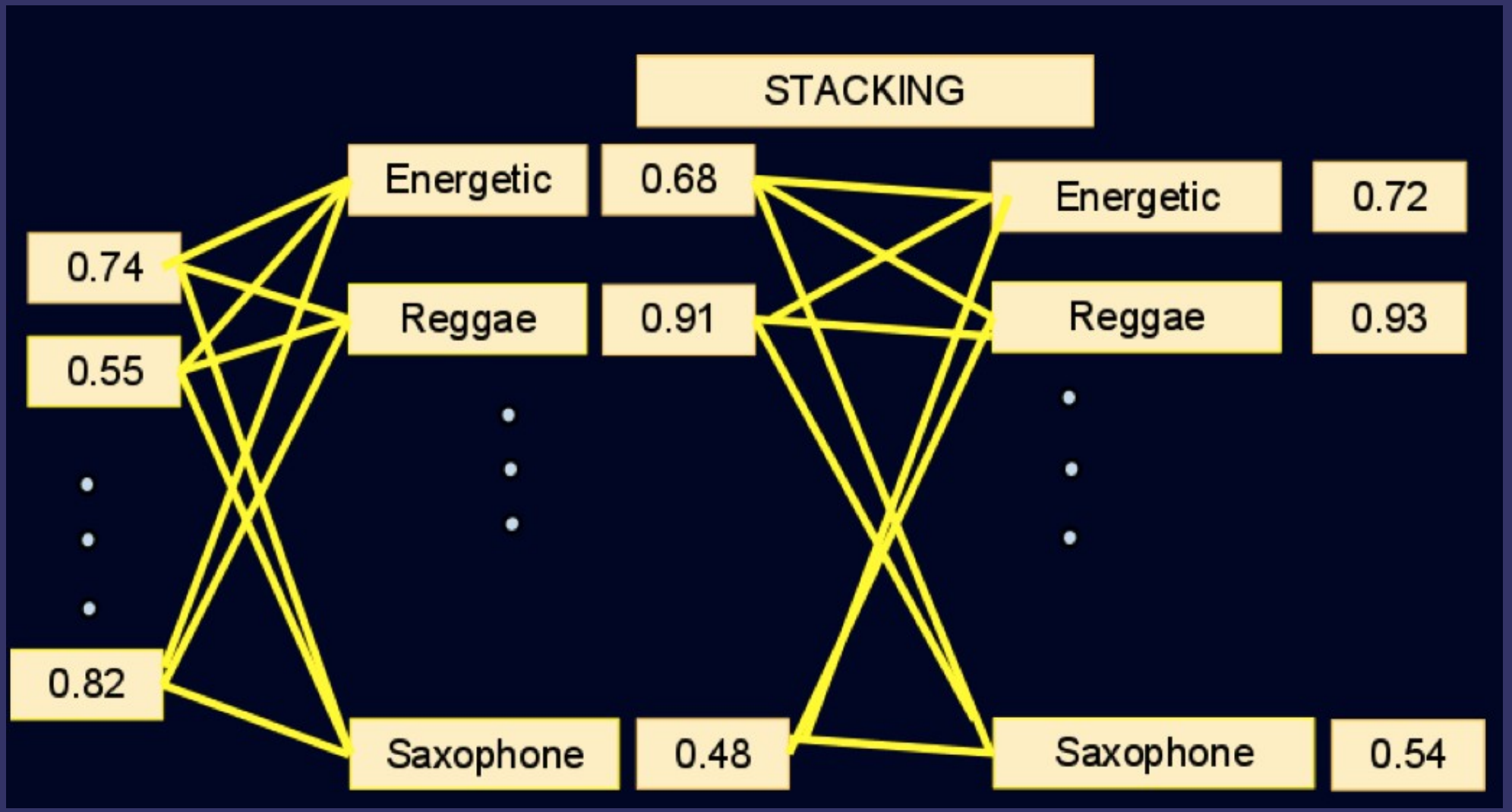
Partitioning of feature space
Generative vs discriminative models

$$P(\blacksquare|\bullet) = \frac{p(\bullet|\blacksquare) * P(\blacksquare)}{p(\bullet)}$$

■ Aggressive

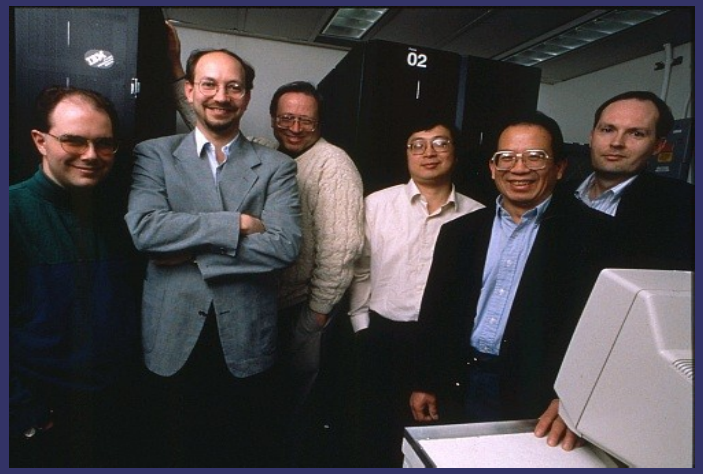
■ Not aggressive

Stacking Architecture

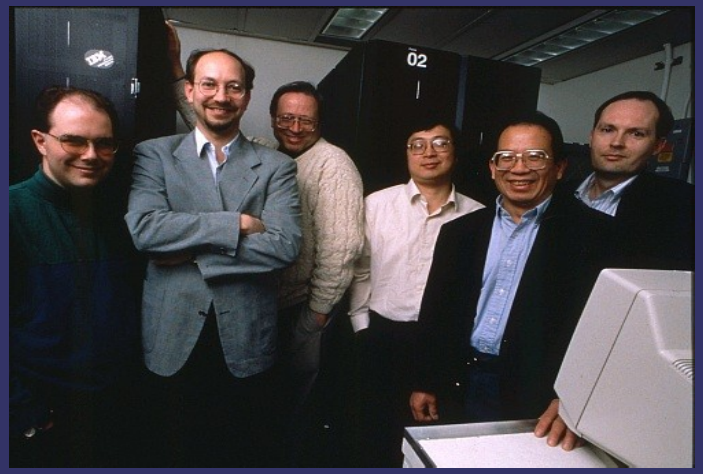


Playing/Improvising music with a computer

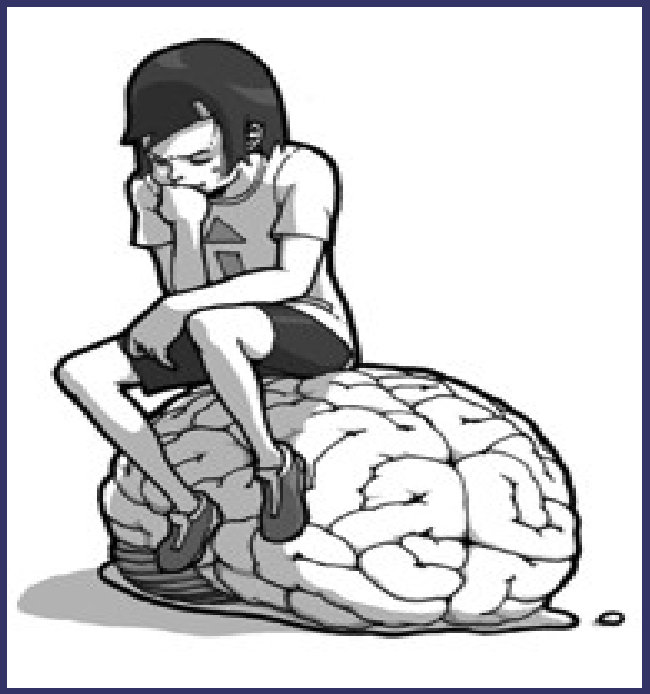
Artificial Intelligence



Artificial Intelligence



Embodied cognition

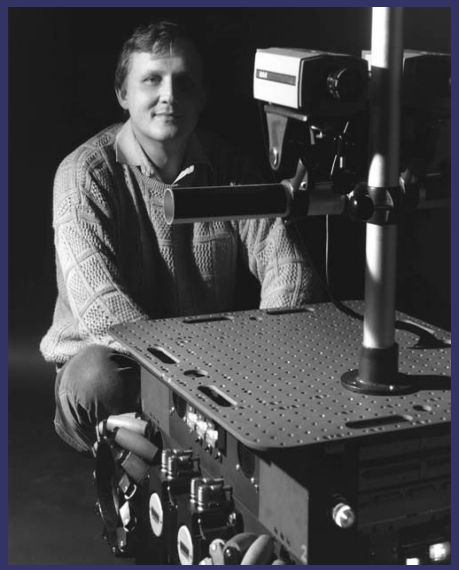


REPRESENATION and
Perception (BRAIN + ear)

ACTION and interaction
Body as mediator between
mind and environment (BODY)

Embodied Cognition in AI

Encoded in the large, highly evolved sensory and motor portions of the human brain is a billion years of experience about the nature of the world and how to survive in it. The deliberate process we call reasoning is, I believe, the thinnest veneer of human thought, effective only because it is supported by this much older and much powerful, though usually unconscious, sensorimotor knowledge. We are all prodigious olympians in perceptual and motor areas, so good that we make the difficult look easy. Abstract thought, though, is a new trick, perhaps less than 100 thousand years old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it – Moravec's paradox



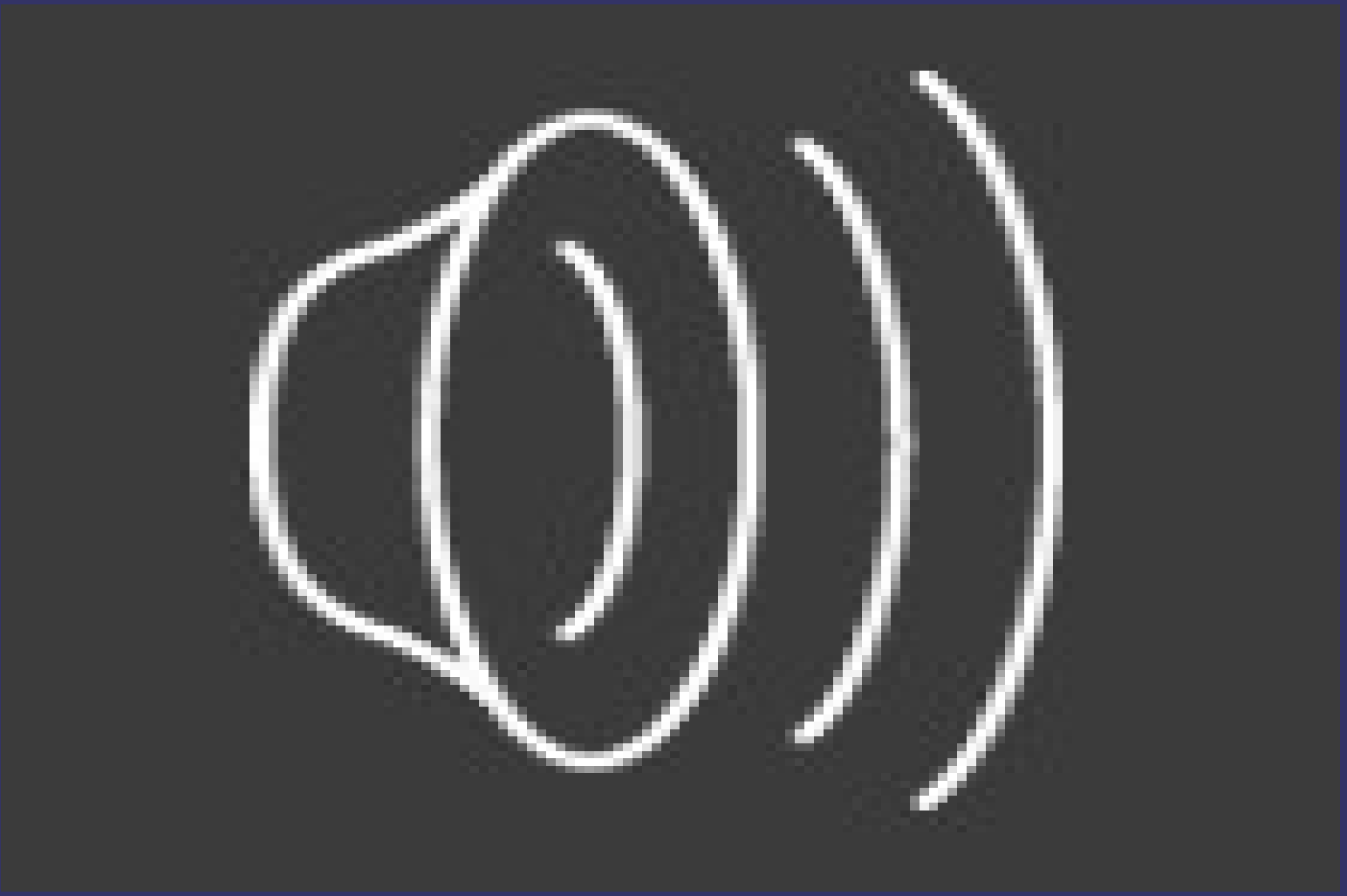
Embodied musicianship

- Focus on tool building and using the idea of embodiment to guide the design/approach rather than trying to understand cognition
- Leveraging MIR techniques for audio signals not just symbolic
- Blending the physical and the virtual
- Digital Signal Processing, Machine Learning and Human-Computer Interaction

Hyper-instruments (E-sitar)

Ajay Kapur, California Institute of the Arts

Hyper-instruments (E-sitar)

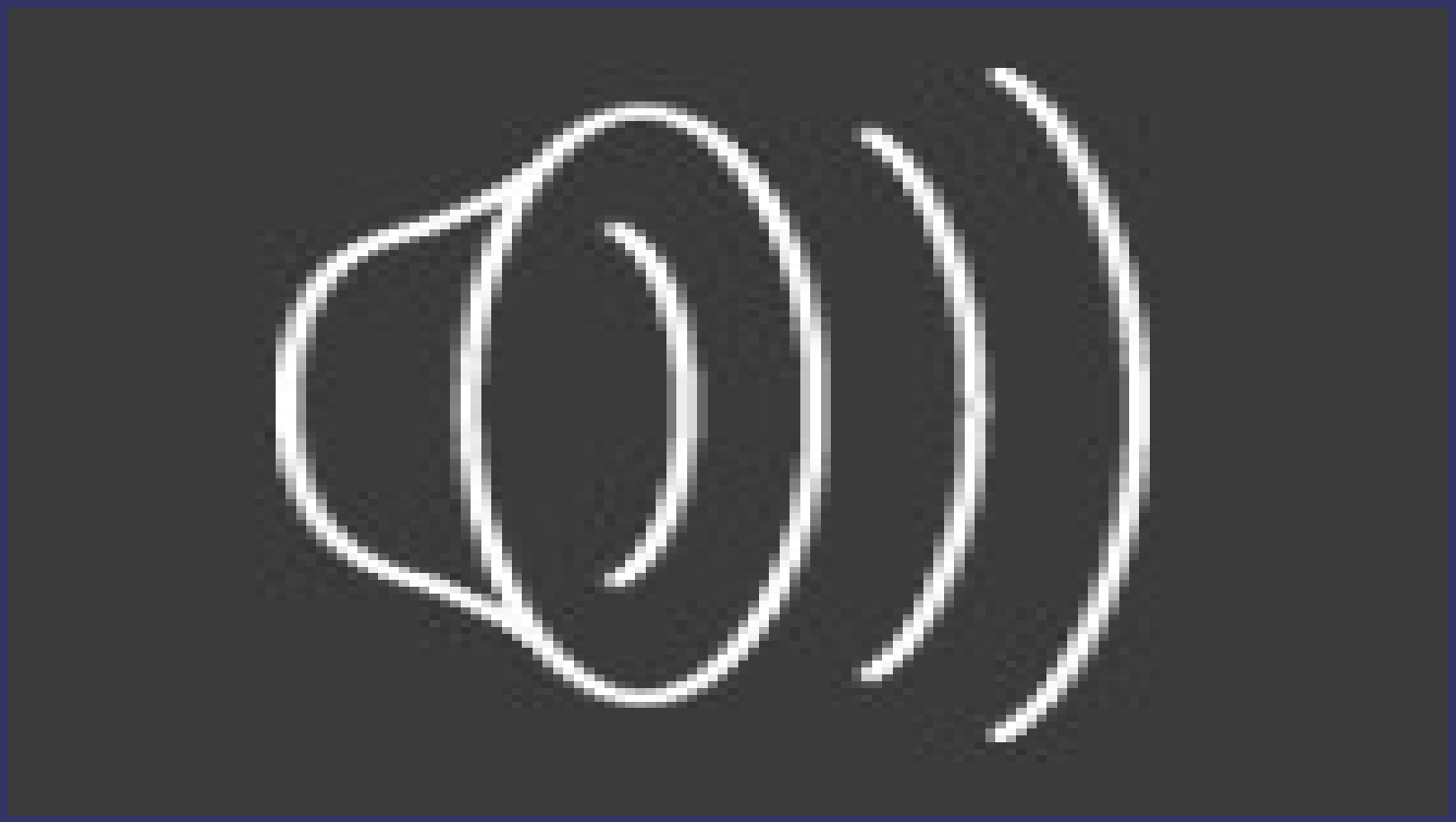


Ajay Kapur, California Institute of the Arts

Hyper-instruments
Virtual Faders

Shawn Trail
INTD PhD
UVic

Hyper-instruments
Virtual Faders



Shawn Trail
INTD PhD
UVic

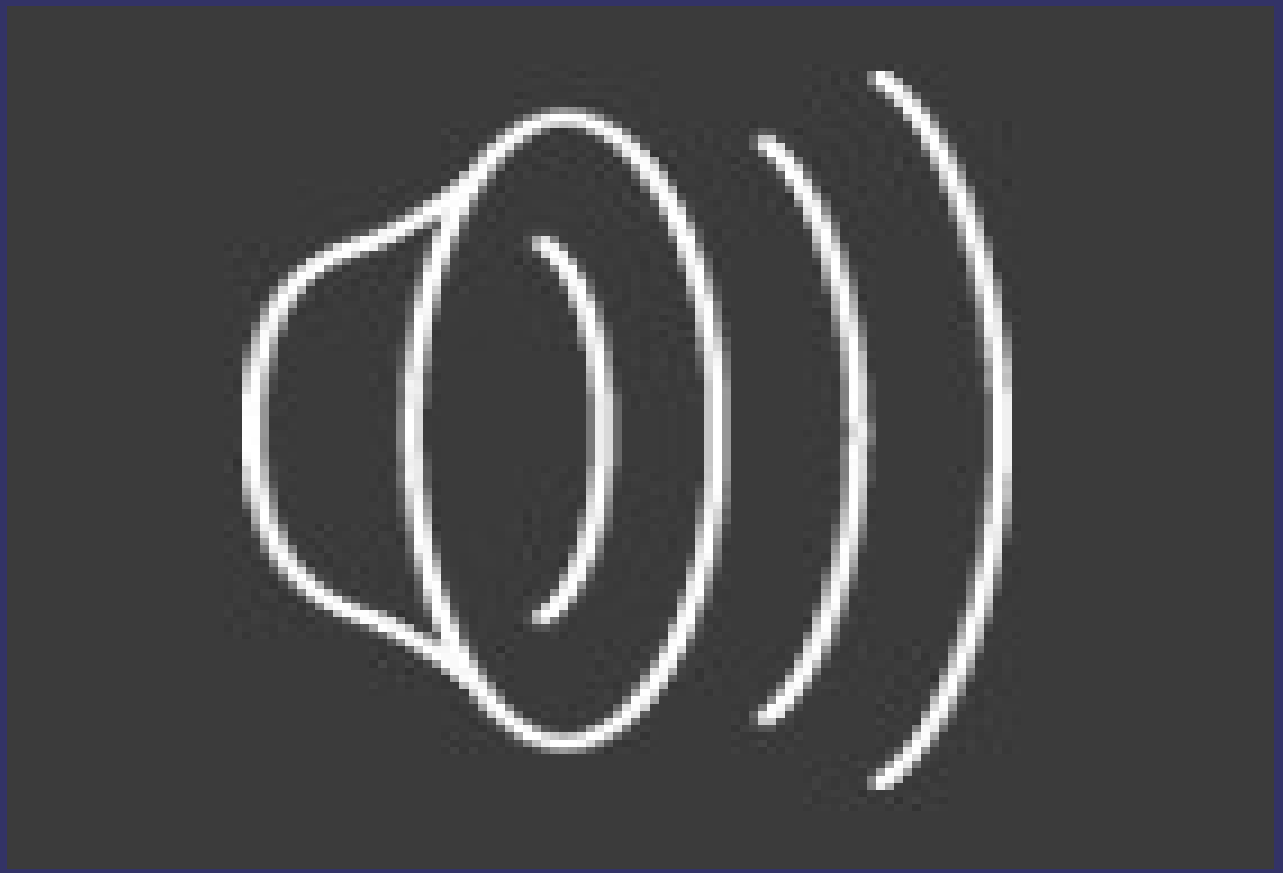
Percussion Robots

Currently essentially MIDI output devices i.e completely deaf
Idea: imbue them with self-listening capabilities

Music Robots

Dr. Andrew Schloss
School of Music
Univ. of Victoria

Music Robots



Dr. Andrew Schloss
School of Music
Univ. of Victoria

Music Information Robotics Proprioception

Music Information Robotics: coping strategies for musically challenged robots, S.. Ness, S. Trail, P. Driessen, G. Tzanetakis ISMIR 2011

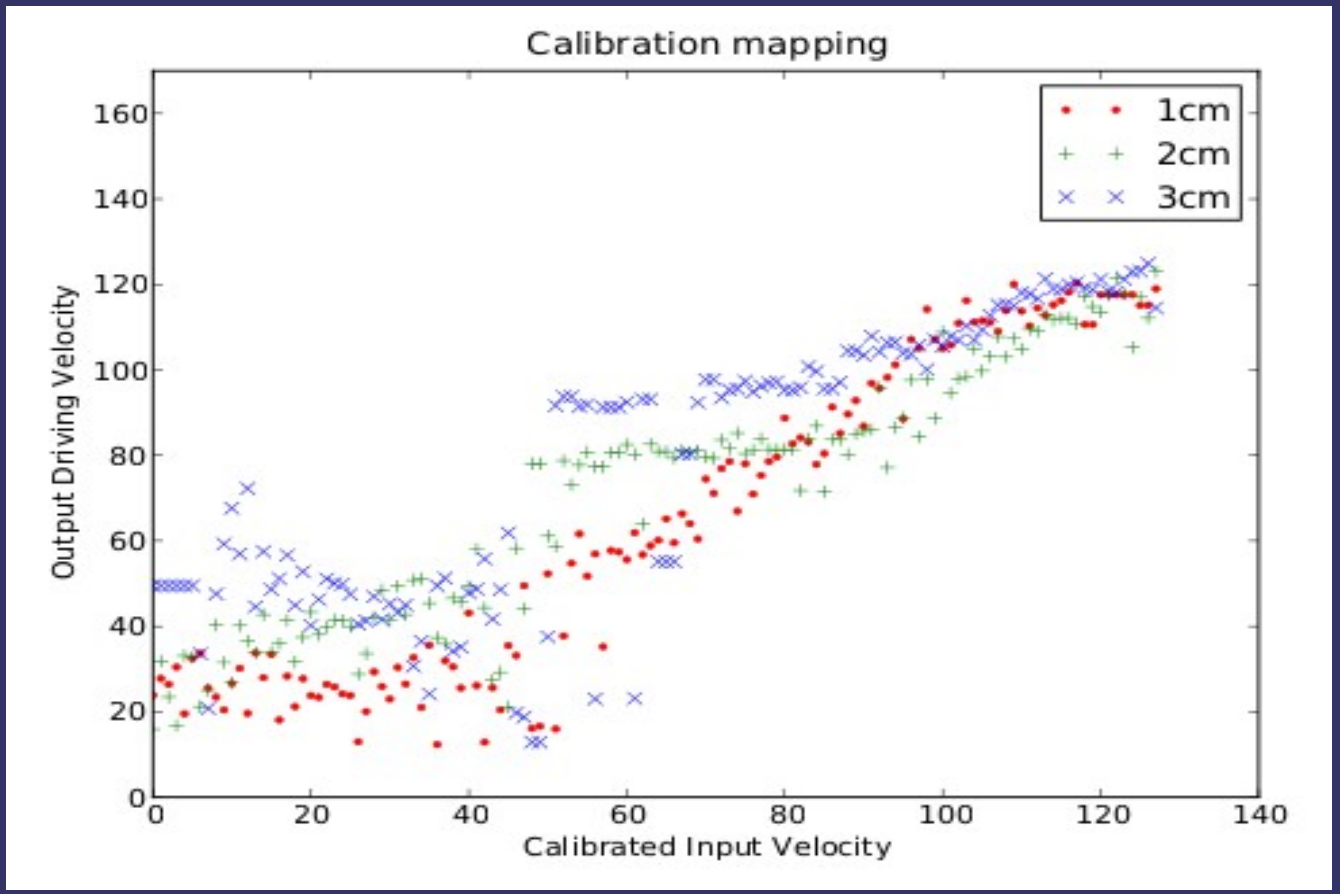
- The perception of your own body
- Automatic sound “check”
- Automatic mapping of modular percussion actuators using audio classification
- Automatic calibration/mapping to compensate for the differences in response of individual actuators to the same velocity/voltage

Modular drum classification (recognizing which drum you are banging)

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 different frame drums,
Classic audio feature extraction and SVM classification

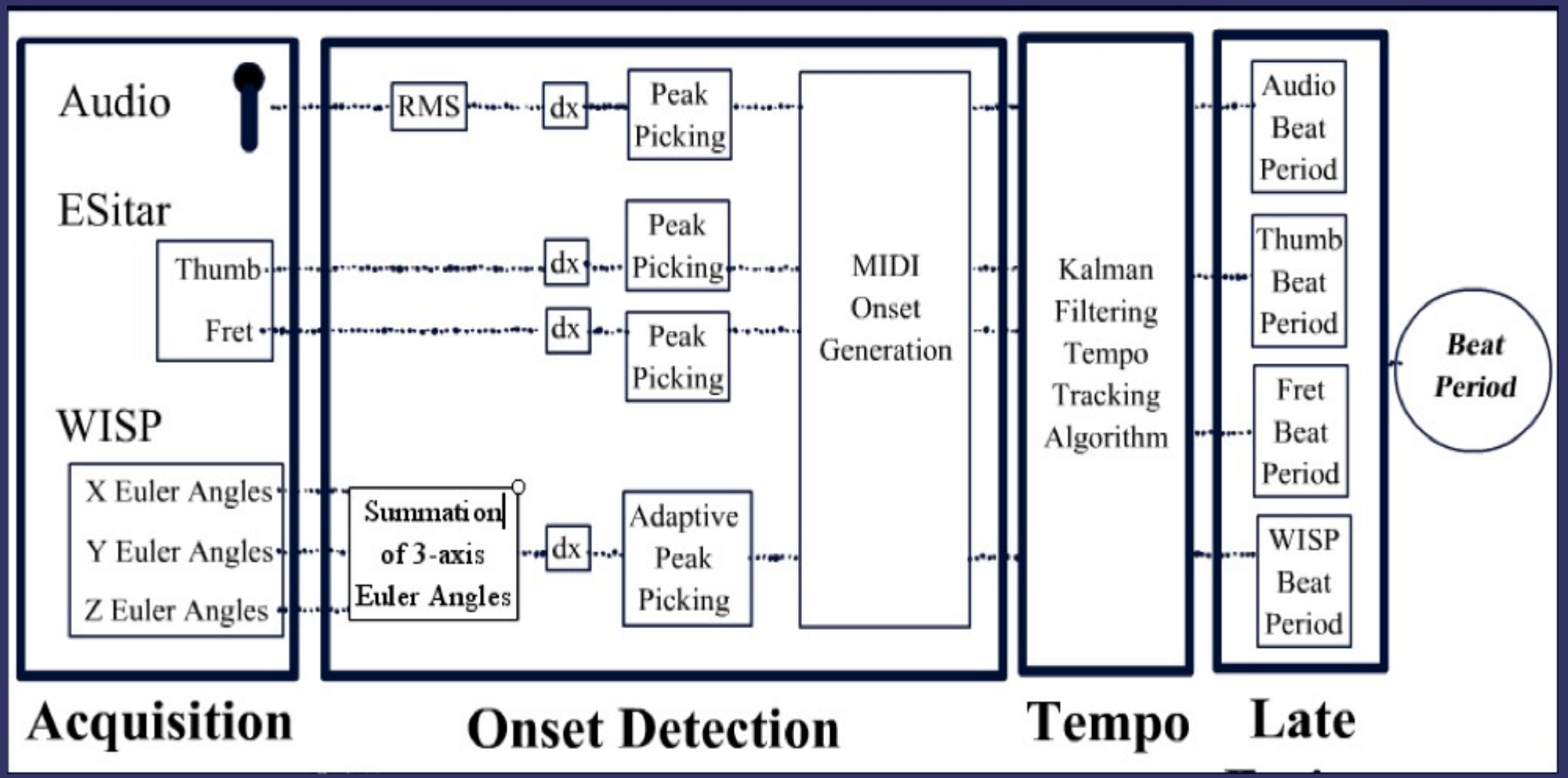
Calibration Map (adjusting how hard you play by how loud you sound)



Goal: uniform loudness, timbre for same input velocities independently of distance from drum

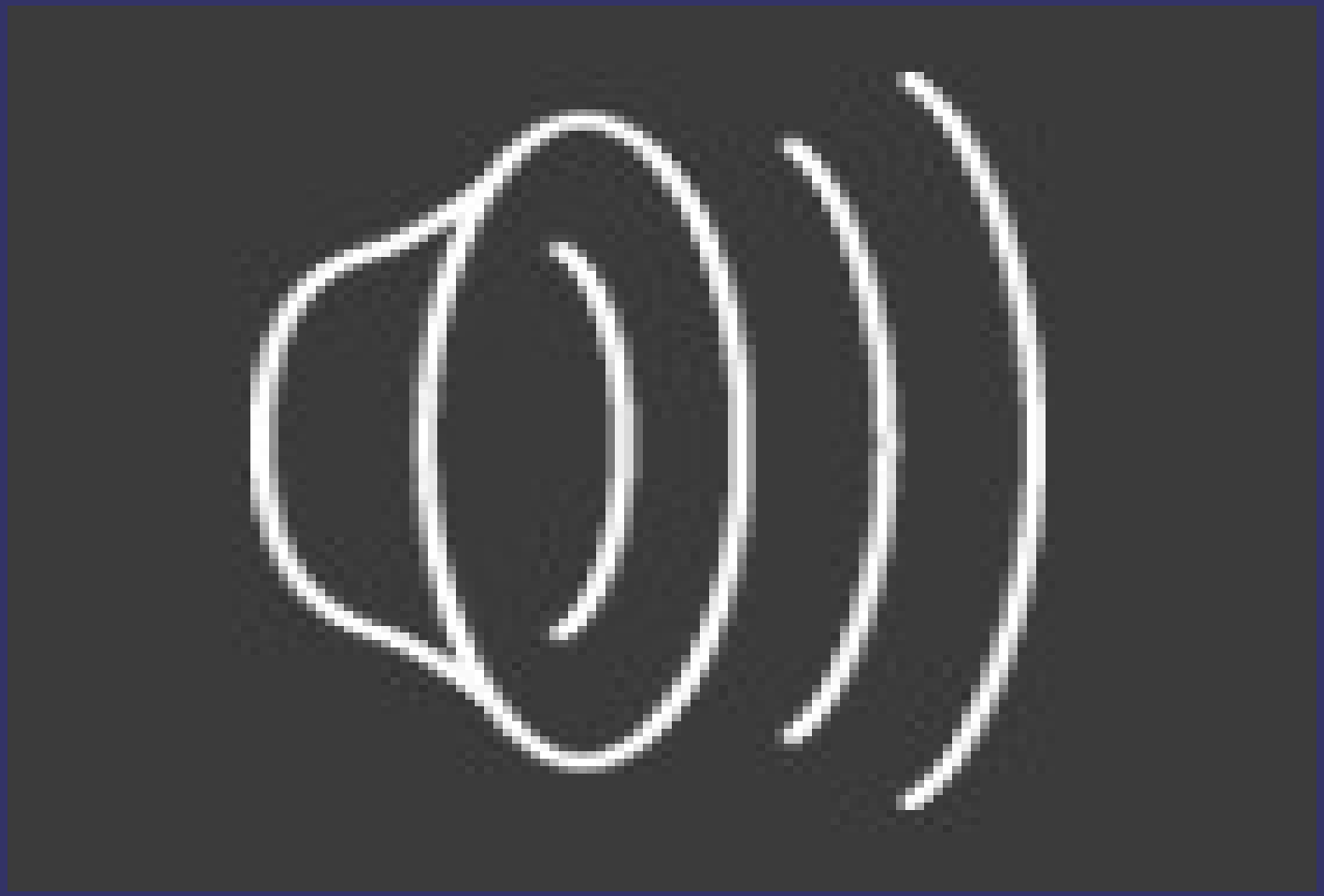
Currently done manually either in software or in the actual physical world

Real-time Multi-Modal Beat Tracking



Reacting to rhythm

Reacting to rhythm



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

G. Percival, G. Tzanetakis, N. Bailey, Sound and Music Computing 2011

- Physical modeling provides expressive control with physical parameters such as bow force and velocity
- As in a real violin continuous control is tricky and requires feedback (audio and haptic)
- Main idea: “teach” rather than “program”

Physical Modeling

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

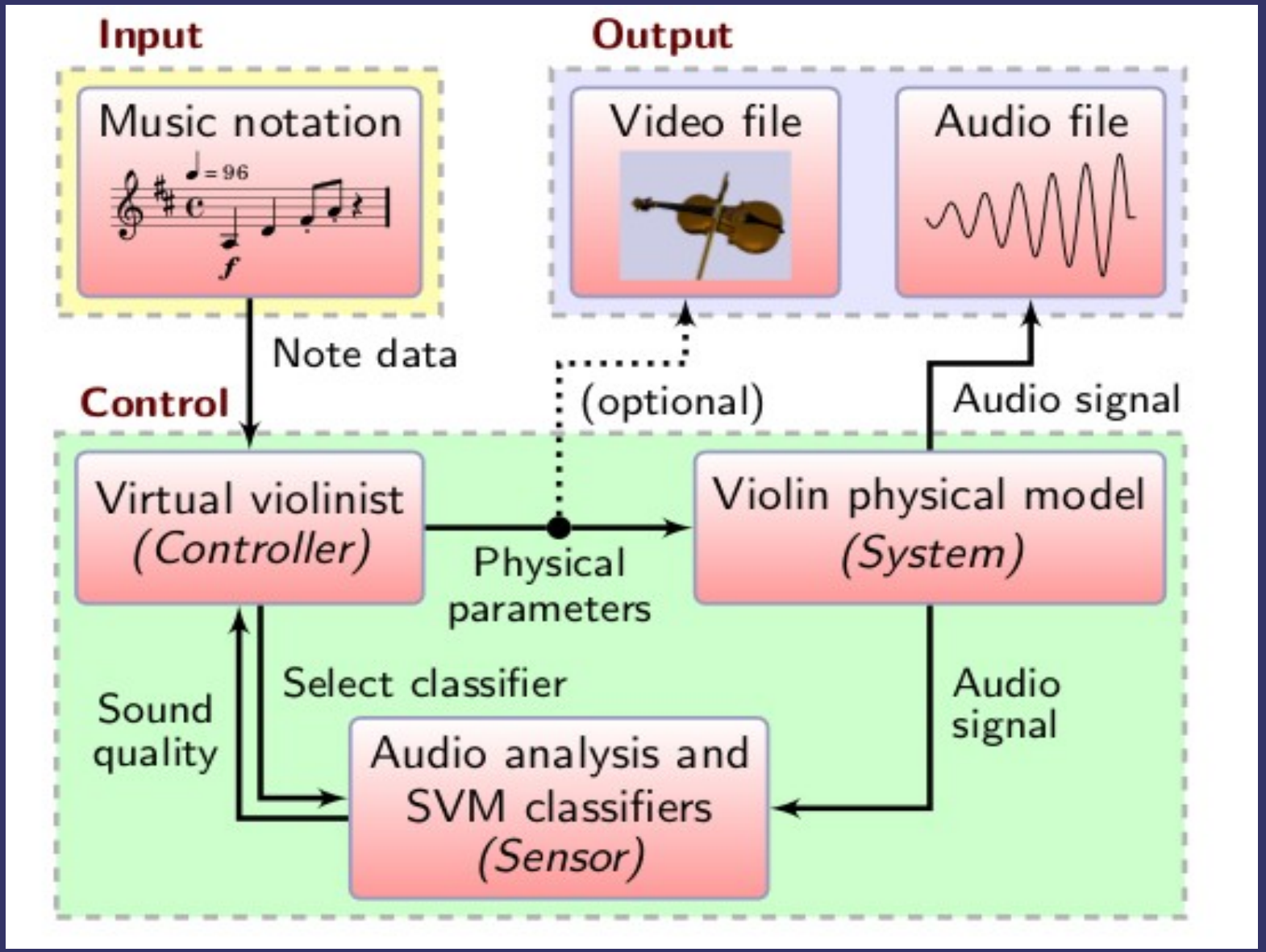
$$\rho_{\mathrm{L}} \frac{\partial^2 y(x,t)}{\partial t^2} - T \frac{\partial^2 y(x,t)}{\partial x^2} + EI \frac{\partial^4 y(x,t)}{\partial x^4} + R_L(\omega) \frac{\partial y(x,t)}{\partial t} = F(x,t)$$

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

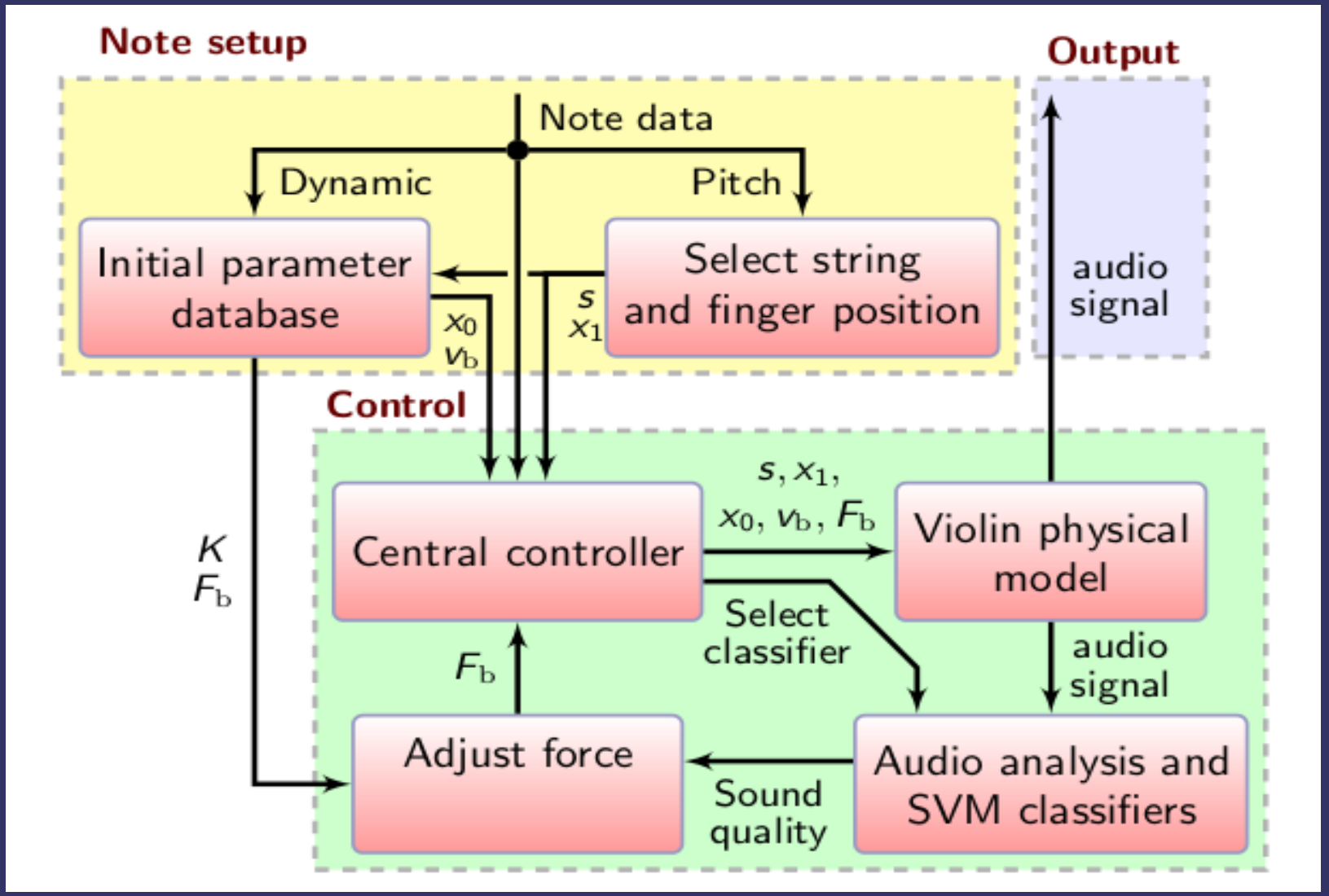
Input parameters

- Violin string number s
- Left-hand finger position x_1
- Bow-bridge distance x_0, velocity v_b, force F_{b}

System architecture



Feedback Loop



Supervised scale and exercise playing

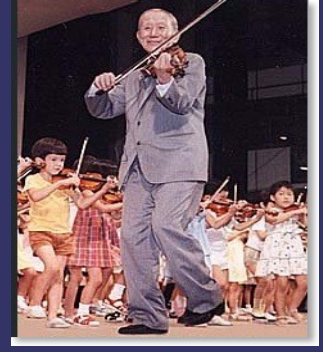
- Before training

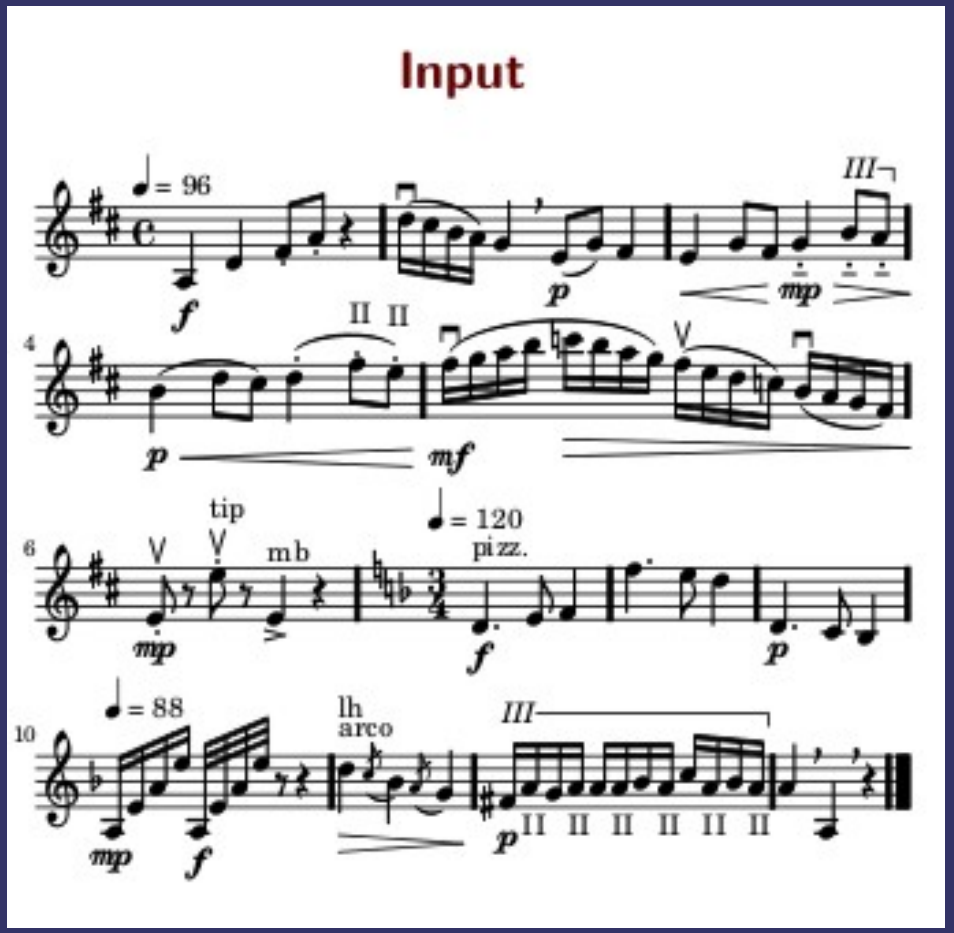
- After training

Supervised scale and exercise playing

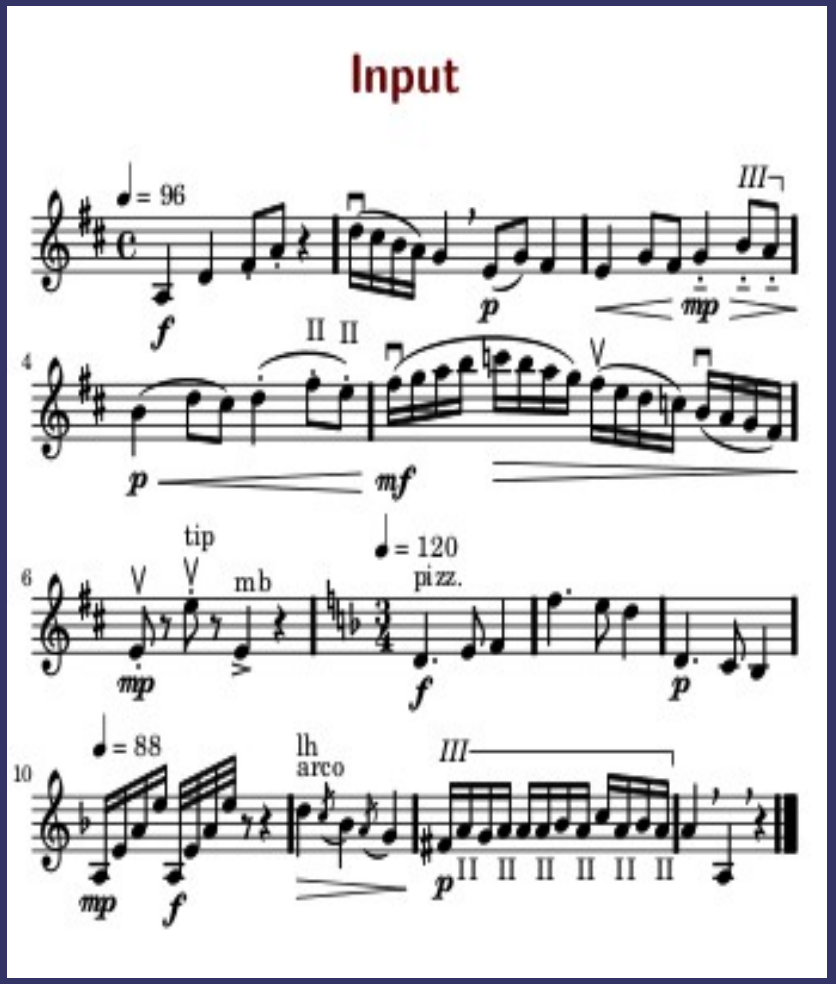
- Before training
- After training

Playing a piece





Playing a piece



Summary

- The way music is created, distributed and perceived is changing dramatically
- Music provides a beautiful testbed for exploring radical ideas in Human Computer Interaction and blending the boundaries between the physical and virtual world
- Body (physical and virtual) as mediator between action and perception

BIG THANKS TO ALL MY AMAZING STUDENTS

Mike Dean

Ajay Kapur

Steven Ness

Gabrielle Odowichuck

Graham Percival

Tiago Tavares

Adam Tindale

Shawn Trial