Computer Vision, Speech Communication & Signal Processing Group,
Intelligent Robotics and Automation Laboratory
National Technical University of Athens, Greece (NTUA)
Robot Perception and Interaction Unit,
Athena Research and Innovation Center (Athena RIC)

Part 5
Audio-Gestural Music Synthesis
Coupling motion and sound in new musical interfaces

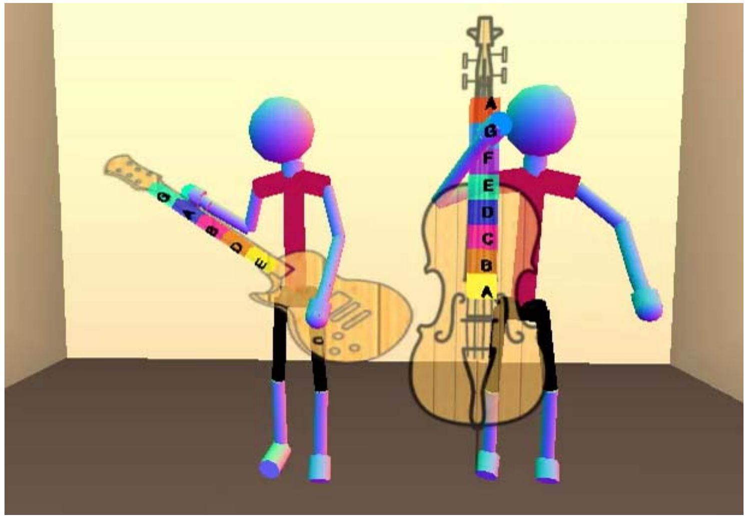
Athanasia Zlatintsi

slides: http://cvsp.cs.ntua.gr/interspeech2018

Tutorial at INTERSPEECH 2018, Hyderabad, India, 2 Sep. 2018

Overview

- iMuSciCA project
- Coupling sound with motion in new musical interfaces
- System architecture
- Modes of interaction
- Evaluation



References:

- [A. Zlatintsi, P.P. Filntisis, C. Garoufis, A. Tsiami, K. Kritsis, M.A. Kaliakatsos-Papakostas, A. Gkiokas, V. Katsouros, and P. Maragos, *A Web-based Real-Time Kinect Application for Gestural Interaction with Virtual Musical Instruments,* Audio Mostly Conf., 2018.]
- [C. Garoufis, A. Zlatintsi and P. Maragos, *A Collaborative System for Composing Music via Motion Using a Kinect Sensor and Skeletal Data,* Sound & Music Computing Conf., SMC-2018].

iMuSciCA Project: interactive Music Science Collaborative Activities

- New pedagogical methodologies and innovative educational tools to support active, discovery-based, personalized, and engaging learning
- Provide students and teachers with opportunities for **collaboration, co-creation** and **collective knowledge building**.
- Design and implement a suite of software tools and services that will deliver interactive **music** activities for teaching/learning **STEM**

STEM = **Science, Technology, Engineering** and **Mathematics** fields

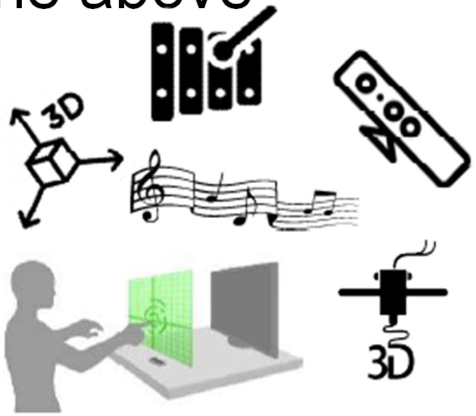
- Bring **Arts** (A) at the heart of the academic curriculum

STEM + A =S TEAM

iMuSciCA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731861.
http://www.imuscica.eu

Coupling Motion and Sound in New Musical Interfaces

- The connection between motion and sound has always been of particular interest.
- Reacting to sound via movements has been practiced since antiquity
- However, the composition of sound from human motion has only been recently explored.
- The first chronologically tangible result of the above exploration is the *theremin* ..



[R. I. Godoy and M. Leman, *Musical Gestures: Sound, Movement, and Meaning,* New York: Routledge, 2010.]
[T. Winkler, *Making motion musical: Gesture mapping strategies for interactive computer music*, Computer Music Conf., 1995]

Theremin

Theremin: early electronic musical instrument controlled without physical contact by the performer.

Right hand: changes pitch by moving it at shoulder-height back and forth between the body and the antenna. The closer the hand gets to the antenna, the higher the pitch.

Left hand: changes volume by moving it up and down over the horizontal antenna. As you lift your hand up, the volume gets louder.

- Due to the recent advances in sensors, motion tracking technology and interfacing, a lot of ground has been covered in the design of systems for the control of musical expression using gestural data!!!

Gesture and Virtual Reality Interaction for Music Synthesis and Expression

- Virtual Music Instrument: analogous to a physical musical instrument, a *gestural interface*, that could provide for much greater freedom in the mapping of movement to sound.
- Innovative interactive and collaborative application (used for STEM) with advanced multimodal interface for musical co-creation and expression
  - Musically "air control" virtual instruments without any physical contact
- Web-based application: widely accessible to everyone
- Intuitive gestural control for triggering the sound

[A. Mulder, *Virtual Musical Instruments: Accessing the sound synthesis universe as a performer*. In Proc. Brazilian Symposium on Computer Music, 1994.]

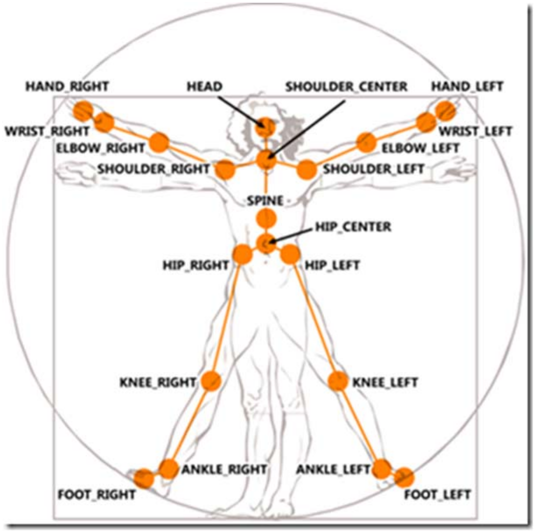
Kinect Sensor for Gesture Interaction

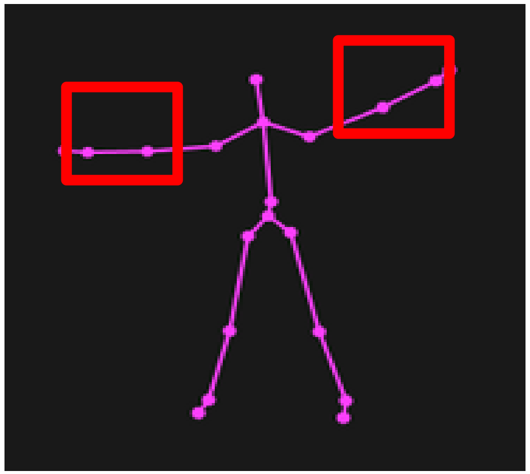
- Kinect v2 for Xbox One by Microsoft
  - inexpensive solution that minimizes intrusiveness constituting a good solution to implement high precision motion tracking,
  - gives the ability to the user to move freely in the physical space, unconstrained and without any other sensors attached to his body.
- Kinect can provide the required visual information:
  - Full HD RGB video at 30fps,
  - Depth information, recorded by the infrared camera embedded in the sensor,
  - Skeletons of up to 6 concurrent people and 25 joints, via the Kinect SDK

[https://www.microsoft.com/en-us/download/details.aspx?id=44561]
[M. Gleicher and N. Ferrier, *Evaluating video-based motion capture*. in Proc. Computer Animation, 2002.]

Skeleton Detection and Tracking

- Skeletons are inferred using depth data.
- Coordinates are provided both on the image (x,y-axis) and on the 3D world (x,y,z-axis).
  - All 25 joint positions are used to draw a full body 3D virtual avatar
  - Specific joints, such as the position of the hands, are used for recognition of specific gestures that, depending on the selected mode of interaction, generate music.





System Architecture: Server and Client

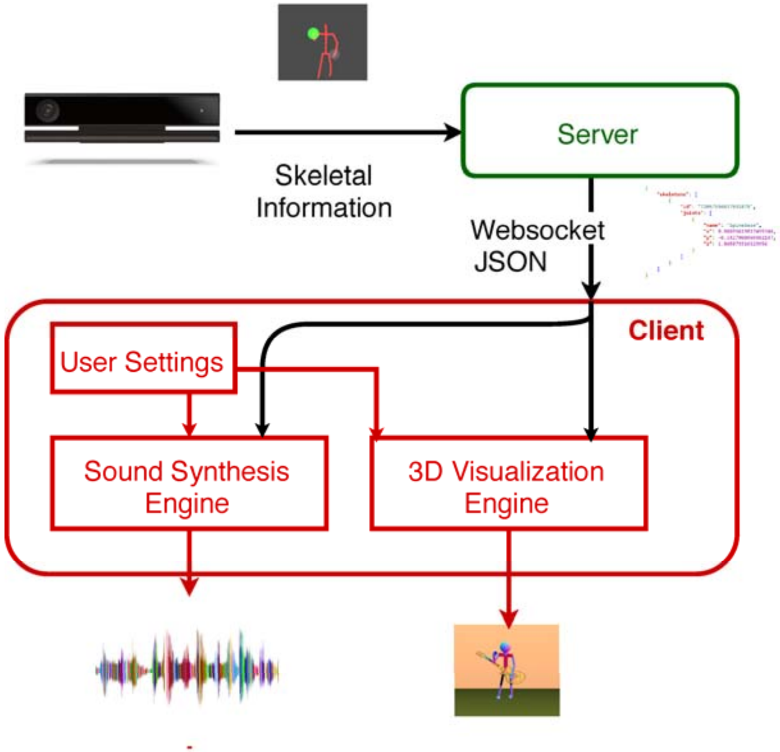
Two concrete modules

Server

- leverages the Kinect v2 API, in order to receive skeletal information from the Kinect at 30fps.
- sends the data in an appropriate format via a Websocket
- implemented in C# language

Client: runs in the user's browser and handles

- the visualization,
- the sound synthesis and
- the User Interface.



The application **has negligible memory footprint**, thus there is no bottleneck regarding the bandwidth of the user's connection.

System Architecture: 3D Visualization Engine

- Maps the world coordinates (x,y,z) that are received for each skeletal joint directly to the joints of the **3D world avatar/-s**.
- Renders semi-transparent Virtual Instruments, and overlaid colored bars with letters, denoting the generated notes.
- The 3D world that depicts the user and the instruments is built using the three.js library

https://threejs.org/

System Architecture: Sound Synthesis Engine

- Music generation is accomplished via the WebAudioFont library: a set of resources that uses sample-based synthesis to play musical instruments in browsers.
- Allows playing chords (several notes simultaneously).
- Includes an extensive catalog of instruments
- In our case:
  - a Guitar
  - a Contrabass.

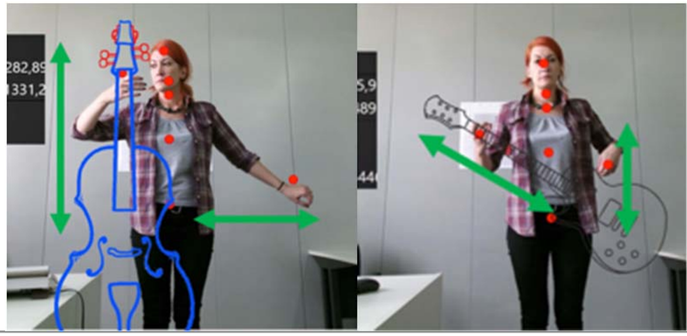
https://github.com/surikov/webaudiofont

Modes of Gestural Control and Interaction

i. The air guitar interaction
ii. The upright bass interaction (using a virtual bow)
iii. The conductor (two hands) interaction: each hand is assigned with one of the two previously named instruments

- Multiplayer interaction: for collaborative playing

- Using ``simple'' and more intuitive gestures
  - Provide the users, especially those that are not musically educated, the ability to perform various virtual instruments without constraints.



Mode 1: Air Guitar Interaction

Gesture 1 (triggering the sound): vertical movements of the right hand around the waist height.

Gesture 2 (changing the pitch): diagonal movements of the left hand from the height of the head to below the waist; enabled only when Gesture 1 is active.

Two predefined mappings:

- pentatonic scale including the notes: G4, A4, B4, D4, and E4,
- predefined chords, which are D4, F4, G4, G#4 (simulating a well-know riff).

- Visual aid: semi-transparent guitar that follows the user and color bars with note names to assist the interaction.

Mode 2: Upright Bass – Bowing interaction

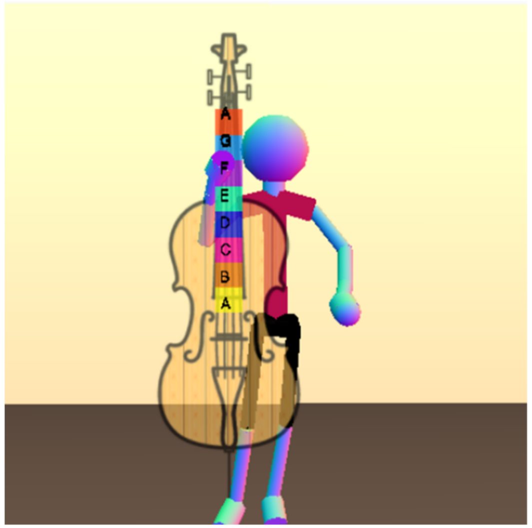
Gesture 1 (triggering the sound): horizontal movements of the right hand around the waist height.

Gesture 2 (changing the pitch): vertical movements of the left hand from the head to the waist height; enabled only when Gesture 1 is active.

Predefined mapping:

- eight notes of a scale (from top to bottom): A2, B2, C3, D3, E3, F3, G3, and A3.

- Visual aids: semi-transparent bass that follows the user and color bars with note names.



Mode 3: The Conductor (two hands) Interaction

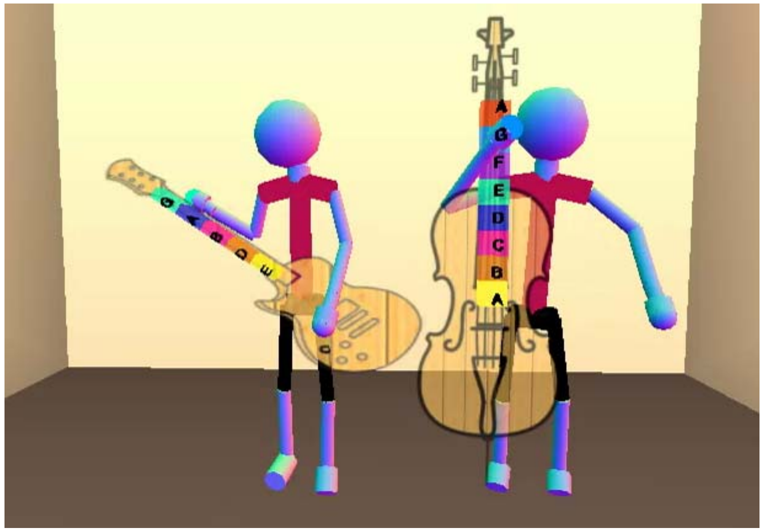
- Each hand is assigned with one of the two instruments.
- Vertical movements of the hands for triggering the notes:
  - A3, B3, C4, D4, E4, F4, G4 and A4.
- Horizontal movements of the hands, for changing the volume
  - higher volume when the two hands are further apart,
  - silencing the instruments when close to the user's spine.
- Visual aids: Color bar with note names is shown vertically, denoting which notes are played at each different height level.

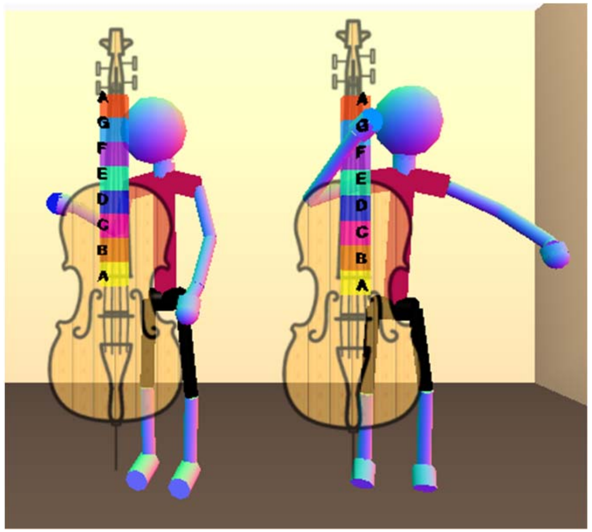
In this mode:

- The user can **"air-draw"** with the hands,
- Listen to consonant and dissonant musical intervals,
- Experiment with the virtual music performance in a more engaging, creative and fun way.

Multiplayer Interaction
Co-creation & Collaboration

- Enabling the collaboration of two or more players.
- The users can either play virtually the same instrument or choose to play the two different instruments simultaneously.





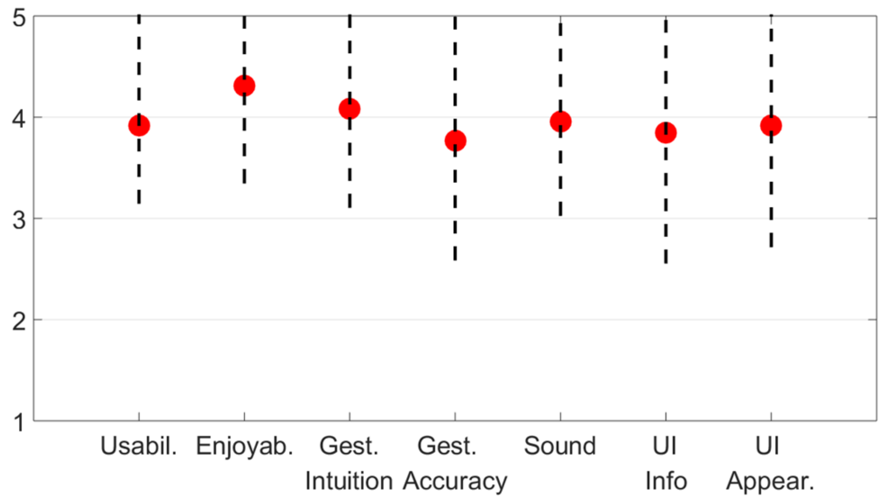
Evaluation and Usability Testing

- Subjective evaluation by 13 users.
- Questionnaires with 5-Likert point scales for: usability, gestural interaction, performance intuitiveness, UI visualization, enjoyability etc.
- The users were able to play the virtual instruments and perform collaborative interaction.

Results and Observations:

- Highly rated usability
- Enjoyable interaction
- Intuitive gestures
- Satisfactory visualization

➢ More practice would be needed to accurately play the notes.



Part 5: Conclusions

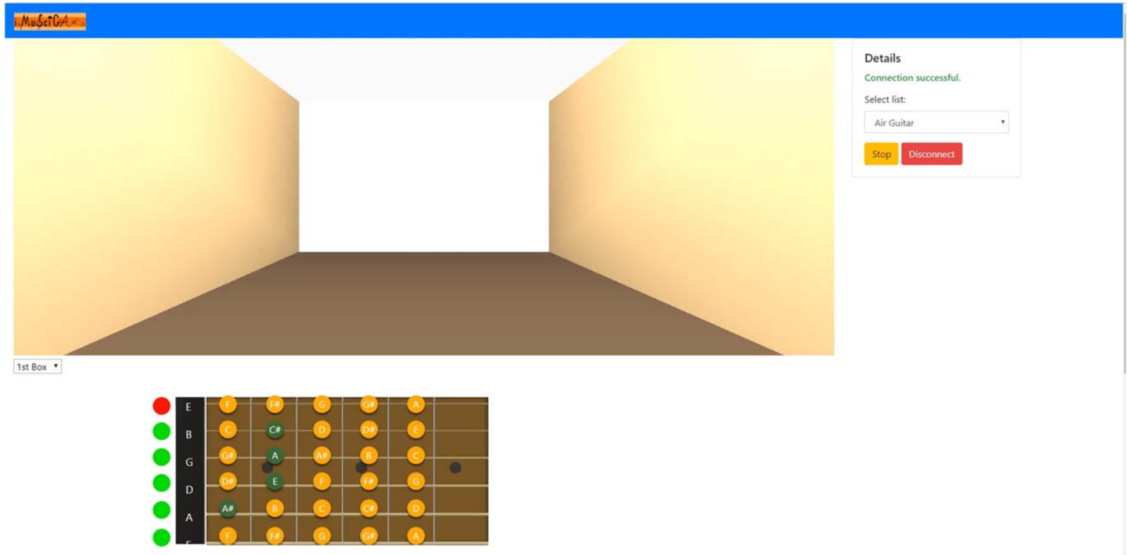
- Web-based real-time application for audio-gestural music synthesis
- Application that is easily accessible by anyone with a Kinect
- No need for prior musical education

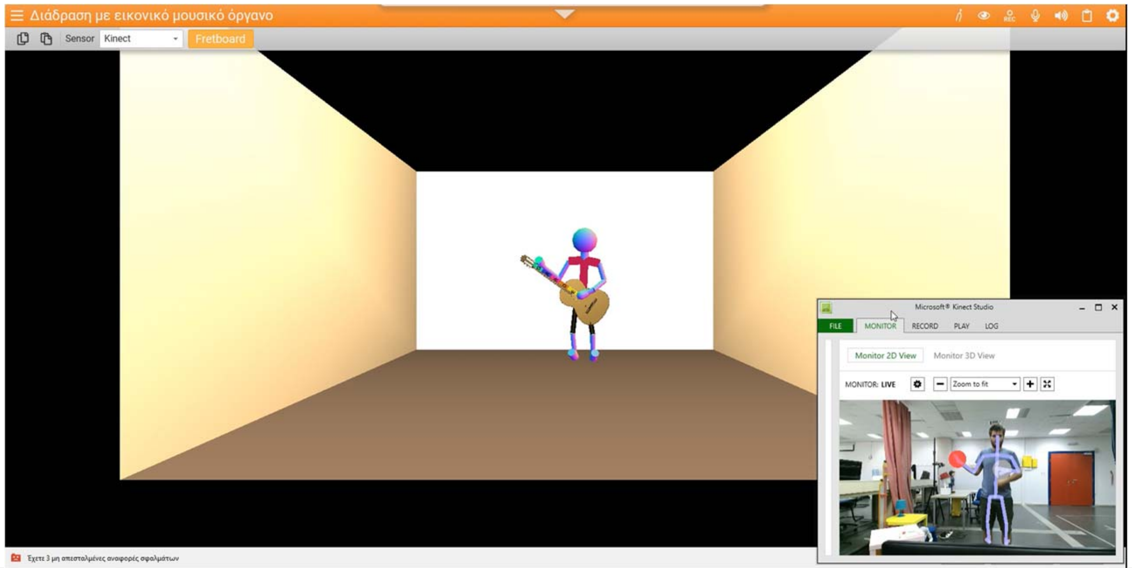
Ongoing Research

- Easily extendable: number of instruments, gestures
- Improvement of the audio-visual aids
- Increase of the educational aspects
- Further improvement of user experience and enjoyability

Tutorial slides: http://cvsp.cs.ntua.gr/interspeech2018
For more information, demos, and current results: http://cvsp.cs.ntua.gr and http://robotics.ntua.gr

Demo


Details
Connection successful.
Select list:
Air Guitar
Stop
Disconnect
1st Box
E
B
G
D
A


Διάδραση με εικονικό μουσικό όργανο
Sensor
Kinect
Fretboard
Microsoft® Kinect Studio
FILE
MONITOR
RECORD
PLAY
LOG
Monitor 2D View
Monitor 3D View
MONITOR: LIVE
Zoom to fit