Is Mayonnaise an Instrument?

Harry Fernandez, Tomas Vancura, Min Gun Kim
Team D1

Carnegie Mellon University

Problem Overview

There are two common problems with *electronic music production*

Getting started is difficult

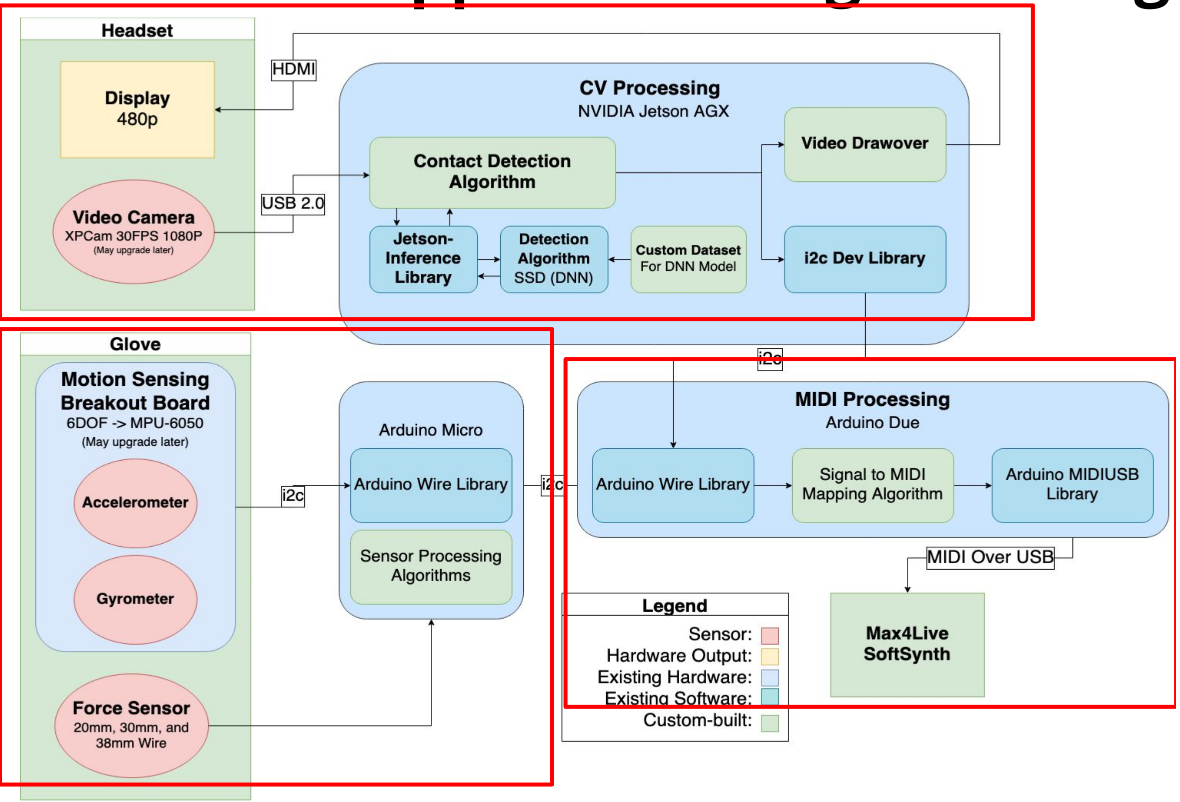
- *Complex musical concepts*
- *Digital Audio Workstation UI can be unintuitive*

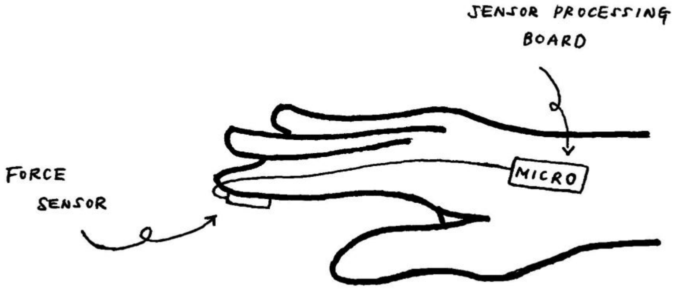
Experimentation is clunky

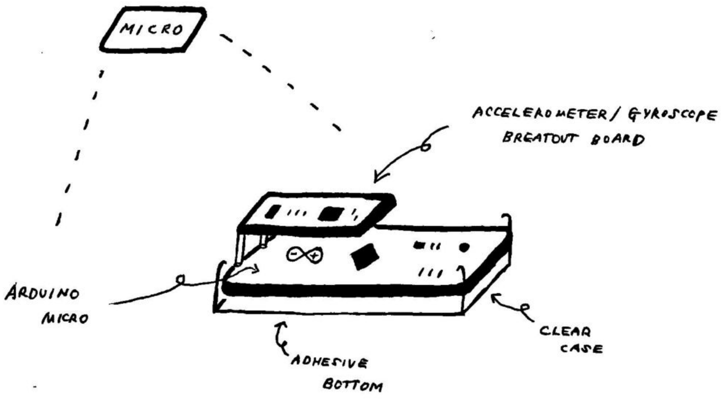
- *Difficult to set up a quick workflow*
- *Not a lot of room for creativity unless you know what you're doing*

Carnegie Mellon University

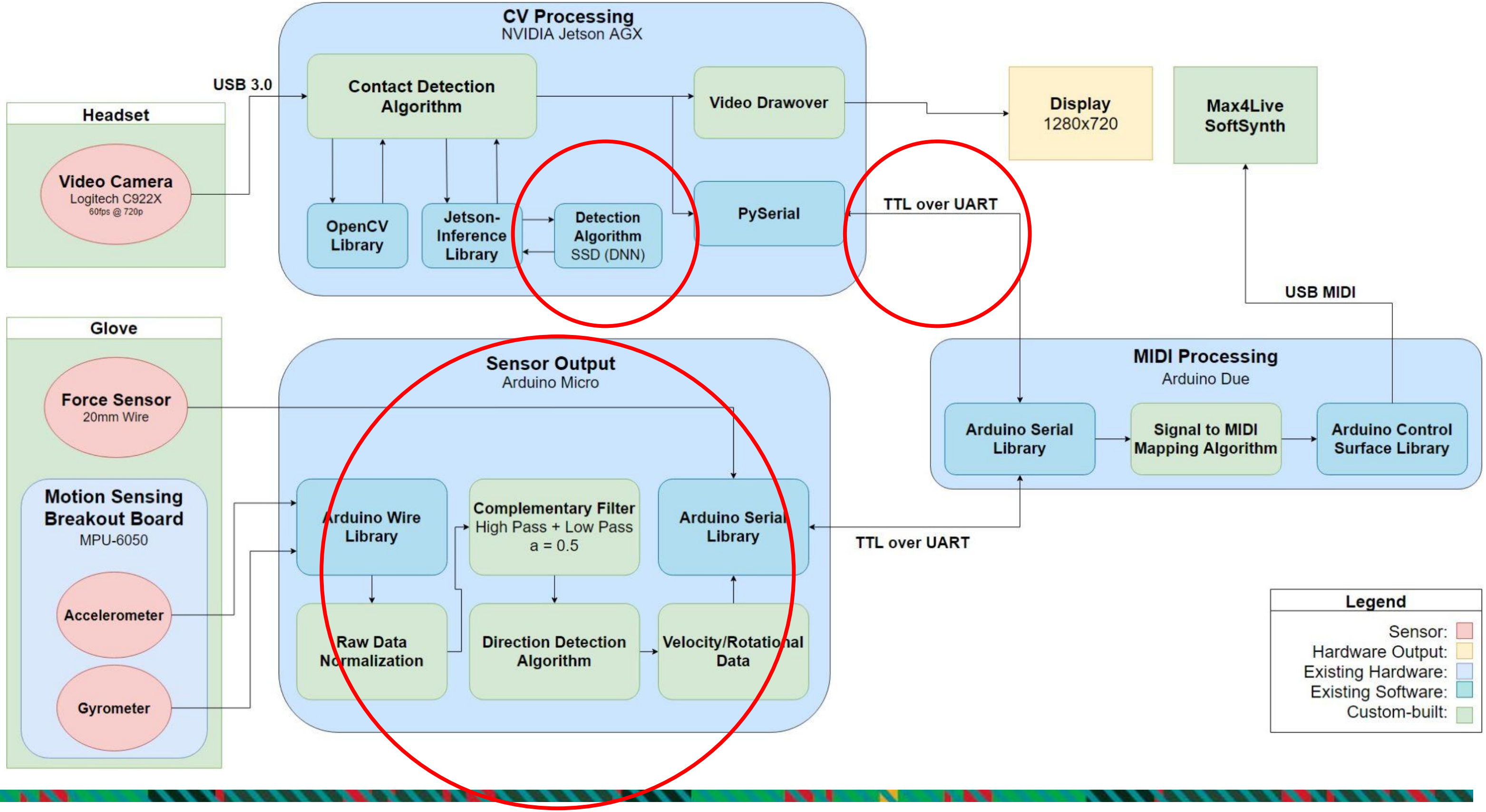
Solution Approach - Original Design







Solution Approach - Current Design

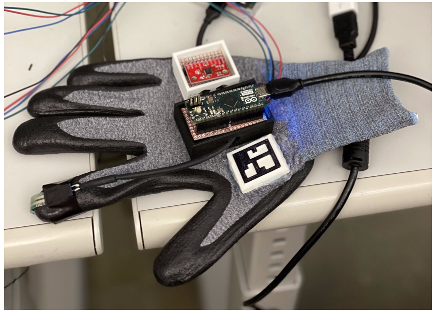


Big Design Changes

- The average user would most benefit from using this in conjunction with another controller
  - No longer a standalone product
  - Complements another device as the parameter controller
- Force sensor is no longer being used to trigger notes, solely used to determine object contact
- No helmet-mounted display
  - Proved to be somewhat clunky. It’s a possible consideration for a future design, but ultimately not worth the effort for MVP
- Supported outputs now include pre-existing software, Serum
  - Want product to be as flexible and adaptable as possible, supporting industry standard soft synths expands usability
- Presently only using rotational data as a parameter
  - Translational did not produce consistent results and was not nearly as intuitive as the rotational data as an input method

Complete Solution

CV Helmet



Motion Sensing Glove

Kevin holding mayonnaise

Carnegie Mellon University

Use-Case Requirements

Requirement	**Metric**	**Result**	**Testing Method**
Capture video	60 FPS (min. 30 FPS)	Consistently >= 30 FPS	OpenGL display API
Identify objects in video	≥ 3 distinct objects, ≤ 1m range at 90% accuracy	3 objects identified Expected case: 90-100% Worst case: 65-70%	Measure percentage of frames where objects were correctly detected, varying situation
Determine when user is touching object	≤ 30[1] ms end-to-end latency, 90% accuracy	TBD	Measure percentage of frames where held object is correctly reported
Determine position of held object	≥ 80% Movement detection accuracy	TBD	
Translate/output positional, object data to MIDI	≤ 30 ms end-to-end latency	Expected case: 0.8 ms Worst case: 2.5 ms	Timing serial delay from start to end of transaction
Visualize to a display	≤ 30 ms latency from picking up object to display	TBD	

1. Xiaoyuan Gu, M. Dick, Z. Kurtisi, U. Noyer and L. Wolf, "Network-centric music performance: practice and experiments," in *IEEE Communications Magazine*, vol. 43, no. 6, pp. 86-93, June 2005, doi: 10.1109/MCOM.2005.1452835.

Test, Verification, Validation

Testing plans

- How do we measure functionality (quantitatively)
- Verification plans (for design requirements)
- Validation plans (for use-case requirements)

Testing Results

- Tables (specification <-> performance)
- Doesn’t have to be complete

Design Tradeoffs

- What approaches did we try
- What approaches failed
- Include any quantitative results
- “Pareto tradeoff”

Accuracy Testing - CV

Two types of tests:

1. Object detection (the ability to detect and identify our chosen set of objects in the frame)
   a. Record objects in variety of situations (number of objects, on table, in hand, stationary/moving camera)
   b. 5 x 7 trials, 60 seconds each
   c. Proportion of frames with objects detected to total number of frames (all objects should always be visible in frame)
2. Potential Contact detection (the ability to predict which object is most likely being touched, if any)
   a. Record touching objects
   b. In how many of the frames when true contact occurred did the correct object get identified as potential contact
   c. Consider true positives and false negatives only

Accuracy Testing - CV

Object Detection Accuracy

Object	Stationary Camera	Moving Camera	Holding Object
Mayo Jar	99%	93%	65%
Coffee Cup	100%	72%	71%
Scissors	99%	97%	67%
Book	78%	-	-

Testing:

- 90% goal accuracy
- 60 second trials, 10 total trials
- Fixed 0.5m camera-to-object distance

$$Accuracy := \frac{N_{detected}}{N_{total}}$$

$$N = \text{number of frames}$$

Trade-offs:

- Prediction thresholds
- Chosen set of objects
- Self trained model vs. pre-trained model
- Contact robustness vs. latency

TBD: Contact Detection Accuracy

Carnegie Mellon University

Accuracy Testing - Sensors

Sensor Data	Goal	Actual
Rotational Data (Gyroscope)	STD: 15 degrees	5.74
Velocity Data (Accelerometer)	80% *error rate subject to change per velocity	TBD
Finger Force Data (Pressure sensor)	80%	98%

Rotational Data:

$$\mu = 90$$

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

Velocity and Finger Force Data:

$$Accuracy := \frac{N_{Success}}{N_{Trials}}$$

Trade-offs:

- Complementary filter coefficient:
  - alpha = 0.5 vs 0.98
- Measuring resistor: 3K vs 10K

Latency Tests

Subsystem	Goal	Actual
Accelerometer/Gyrometer	n/a	TBD
Force Sensor	18ms/frame	TBD
CV Object Detection	15ms/frame	TBD
Due <-> Micro Bus	1.5ms (with Jetson)	~0.8ms
Due <-> Jetson Bus	1.5ms (with Micro)	~1ms
Due <-> Max	n/a	<1ms
End-to-End	30ms	0.8ms - 2.5ms

Testing:

- Record time at beginning and end of transactions
- Average over ~10000 trials

Trade-offs:

- I2C vs RS-485 vs base UART
- Constant Polling vs No Polling

Carnegie Mellon University

User Validation

3 Musicians and 3 Non-musicians

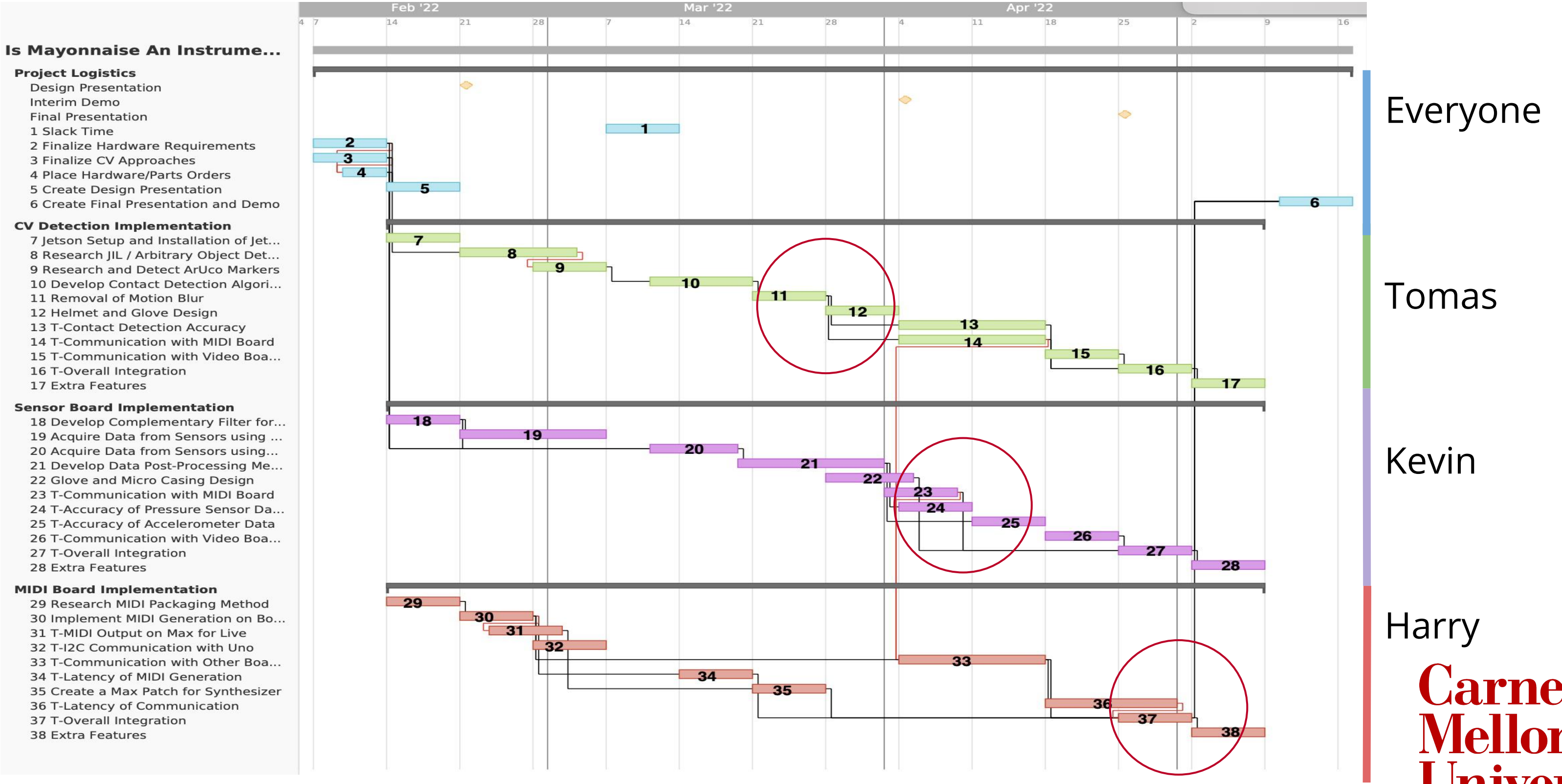
- Ask musicians to generate short music using our system
- Ask non-musicians to generate sound using our system

Procedure

1. Ask musicians to generate short music using our system
    - Ask non-musicians to generate sound using our system
2. Ask all the participants in-person about their experience
    - Take note of all the impressions they make during the experimentation
3. Ask our participants to fill out a Google form with questions regarding their experience
    - Examples of the questions
        - i. *Were you able to generate sound the way that you expected?*
        - ii. *What component of the system felt most uncomfortable?*
        - iii. *What improvements can be made to improve the user experience?*

Carnegie Mellon University

Schedule


Feb '22
Mar '22
Apr '22
Is Mayonnaise An Instrume...
Project Logistics
Design Presentation
Interim Demo
Final Presentation
1 Slack Time
2 Finalize Hardware Requirements
3 Finalize CV Approaches
4 Place Hardware/Parts Orders
5 Create Design Presentation
6 Create Final Presentation and Demo
CV Detection Implementation
7 Jetson Setup and Installation of Jet...
8 Research JIL / Arbitrary Object Det...
9 Research and Detect ArUco Markers
10 Develop Contact Detection Algori...
11 Removal of Motion Blur
12 Helmet and Glove Design
13 T-Contact Detection Accuracy
14 T-Communication with MIDI Board
15 T-Communication with Video Boa...
16 T-Overall Integration
17 Extra Features
Sensor Board Implementation
18 Develop Complementary Filter for...
19 Acquire Data from Sensors using ...
20 Acquire Data from Sensors using...
21 Develop Data Post-Processing Me...
22 Glove and Micro Casing Design
23 T-Communication with MIDI Board
24 T-Accuracy of Pressure Sensor Da...
25 T-Accuracy of Accelerometer Data
26 T-Communication with Video Boa...
27 T-Overall Integration
28 Extra Features
MIDI Board Implementation
29 Research MIDI Packaging Method
30 Implement MIDI Generation on Bo...
31 T-MIDI Output on Max for Live
32 T-I2C Communication with Uno
33 T-Communication with Other Boa...
34 T-Latency of MIDI Generation
35 Create a Max Patch for Synthesizer
36 T-Latency of Communication
37 T-Overall Integration
38 Extra Features
Everyone
Tomas
Kevin
Harry