

به نام خدا



دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

گزارش پروژه

مبانی بینایی کامپیوتر

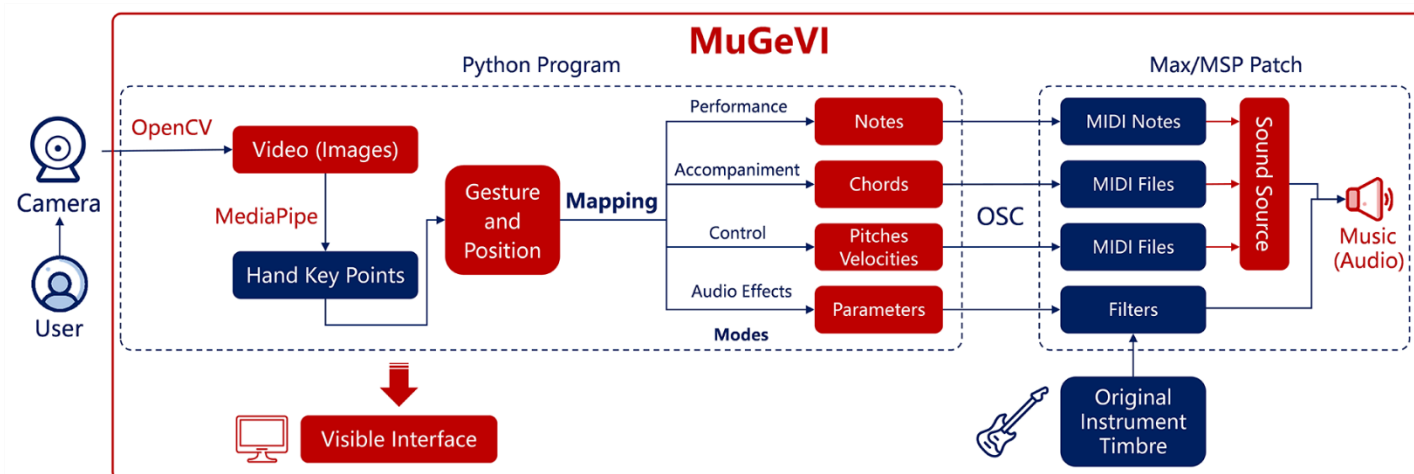
دیبا روانشید

-810199431-

نیمسال دوم 1402-03

*لطفا توجه فرمایید تمام سوالات به دقت پاسخ داده شده اند و در فایل کد، کد هر سوال مشخص شده است. سپاس

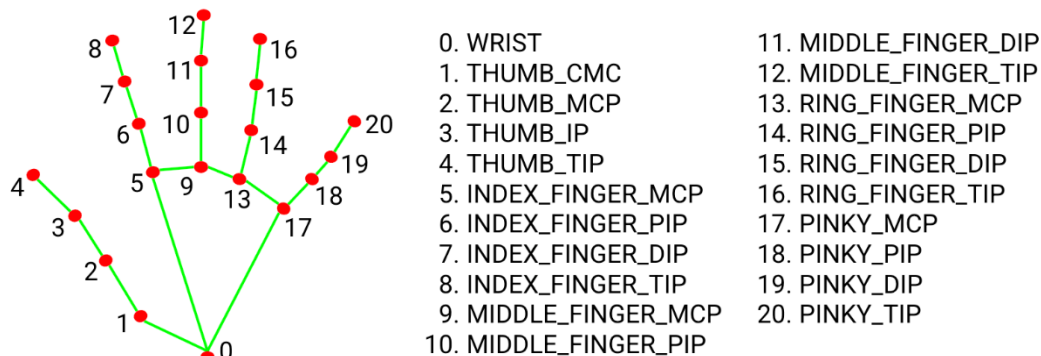
مقدمه:



:Mediapipe

مدیپایپ یک فریم ورک متن باز است که توسط گوگل توسعه داده شده است. این فریم ورک بیشتر برای **real-time object detection and tracking** استفاده می‌گردد.

در این پروژه از **hand tracking** این کتابخانه استفاده شده است که 21 نقطه کلیدی دست را شناسایی می‌کند.



در ادامه بر اساس نحوه قرار گیری بعضی از نقاط از این 21 نقطه حالت های مختلف برنامه را پیاده سازی می‌کنیم.

OSC (Open Sound Control) یک پروتکل ارتباطی است که برای تبادل اطلاعات کنترل صوتی و مولتی‌مدیا استفاده می‌شود. این پروتکل توسط مرکز تحقیقات موسیقی کامپیوتری (CNMAT) در دانشگاه کالیفرنیا، برکلی توسعه داده شده است. OSC طراحی شده است تا جایگزینی برای MIDI باشد و محدودیت‌های آن را برطرف کند. در اینجا توضیحات کاملی در مورد OSC ارائه می‌دهیم:



پروتکل Open Sound Control (OSC) برای پاسخ به نیازهای خاصی در صنعت صدا و مولتی‌مدیا توسعه داده شده است. دلایل اصلی نیاز به OSC و کاربردهای آن عبارتند از: دلایل نیاز به OSC محدودیت‌های MIDI، پهنای باند محدود: MIDI تنها می‌تواند 7 بیت (0-127) داده را برای هر پیام ارسال کند که برای بسیاری از کاربردهای مدرن کافی نیست. سرعت انتقال پایین: MIDI با سرعت 31.25 کیلوبیت در ثانیه عمل می‌کند که برای کاربردهای با حجم بالای داده مناسب نیست. ساختار پیام محدود: پیام‌های MIDI محدود به یک ساختار از پیش تعریف‌شده هستند که انعطاف‌پذیری کمی در تعریف پیام‌های سفارشی دارد. نیاز به انعطاف‌پذیری بیشتر پشتیبانی از انواع داده‌های بیشتر: برخلاف MIDI که فقط از اعداد صحیح 7 بیتی پشتیبانی می‌کند، OSC می‌تواند انواع مختلف داده‌ها شامل اعداد اعشاری، رشته‌ها، و باینری‌ها را ارسال کند. پشتیبانی از شبکه‌های مدرن: OSC می‌تواند از پروتکل‌های شبکه مانند UDP و TCP برای انتقال پیام‌ها استفاده کند که امکان انتقال داده‌ها را در شبکه‌های محلی و اینترنت فراهم می‌کند.

ویژگی‌ها و مزایا

1. قابلیت تطبیق‌پذیری بالا: OSC می‌تواند با هر نوع داده‌ای کار کند، شامل اعداد، رشته‌ها، آرایه‌ها و باینری‌ها.
2. پشتیبانی از شبکه‌های مختلف: OSC می‌تواند از طریق شبکه‌های مختلف مانند UDP، TCP، و حتی WebSockets انتقال یابد.
3. ساختار سلسله‌مراتبی: OSC از آدرس‌دهی شبیه به URL استفاده می‌کند که به شما اجازه می‌دهد تا پارامترهای کنترل را به صورت سلسله‌مراتبی سازماندهی کنید.

4. زمان‌بندی دقیق: OSC از زمان‌بندی دقیق برای پیام‌ها پشتیبانی می‌کند، که این امکان را فراهم می‌کند تا عملیات‌های زمان‌بندی شده با دقت بالا انجام شوند.
5. پشتیبانی از پیام‌های همزمان: چندین پیام می‌توانند در یک بسته OSC قرار گیرند و به صورت همزمان ارسال شوند.

ساختار پیام‌های OSC

پیام‌های OSC شامل دو بخش اصلی هستند:

1. آدرس:

- آدرس‌ها در OSC شبیه به مسیرهای فایل یا URLها هستند و از کاراکترهای `^` برای جدا کردن سطوح سلسله‌مراتبی استفاده می‌کنند. برای مثال: `^synth/frequency/`

2. آرگومان‌ها: آرگومان‌ها می‌توانند انواع مختلفی از داده‌ها باشند مانند اعداد صحیح، اعداد اعشاری، رشته‌ها و باینری‌ها.

OSC Messages

PACKET SIZE	OSC ADDRESS PATTERN	OSC TYPE TAG STRING	ARGUMENTS			
48	/sounds/sine	,ffff	401.5	0.0	1.0	0.0

نحوه‌ی ارسال و دریافت پیام‌های OSC

برای ارسال و دریافت پیام‌های OSC، نیاز به کتابخانه‌های نرم‌افزاری یا ابزارهایی دارید که از این پروتکل پشتیبانی می‌کنند. برخی از این ابزارها عبارتند از:

- Pure Data (Pd): یک محیط برنامه‌نویسی گرافیکی برای پردازش صوتی و مولتی‌مدیا.
- Max/MSP: یک نرم‌افزار برای موسیقی الکترونیک و مدیای تعاملی.
- TouchOSC: یک برنامه برای ایجاد کنترل‌های لمسی برای دستگاه‌های موبایل.
- Processing: یک زبان برنامه‌نویسی و محیط توسعه برای هنرهای الکترونیک و بصری.

کاربردهای OSC

1. موسیقی الکترونیک: کنترل سینتی‌سایزرها، افکت‌ها و نرم‌افزارهای صوتی.
2. نمایش‌های مولتی‌مدیا: هماهنگ‌سازی نورها، ویدیوها و سایر جلوه‌های بصری.
3. تعاملات زنده: کنترل اجرای زنده با استفاده از دستگاه‌های موبایل یا دیگر کنترل‌های لمسی.
4. هنرهای تعاملی: ایجاد آثار هنری که به ورودی‌های مختلف پاسخ می‌دهند.

OSC به دلیل انعطاف‌پذیری و قابلیت‌های گسترده‌اش، به طور گسترده‌ای در صنایع مختلفی که نیاز به کنترل دقیق و زمان‌بندی دارند، مورد استفاده قرار می‌گیرد.

معرفی نرم افزار Max/MSP :

برای نصب این نرم افزار باید به وبسایت Cycling به آدرس cycling74.com که شرکت سازنده Max می‌باشد. می‌توانید مراجعه کنید و در اینجا Max را دانلود کنید. به زبانه دانلود بروید و Max را برای ادامه ویدئو دانلود کنید تا بتوانید مراحل کار را دنبال کنید. مطمئن شوید نسخه مناسب برای سیستم‌عامل خود (مک یا ویندوز) را دانلود می‌کنید و دستورالعمل‌های مربوط به سیستم‌عامل خود را دنبال کنید.

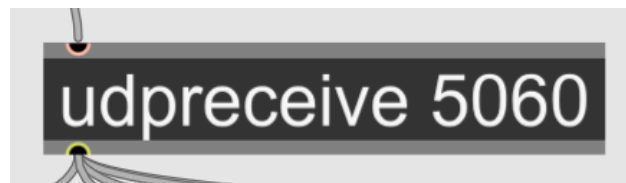
Try Max for free for 30 days, including the
RNBO Demo

VERSION 8.6.3 | JUN 18, 2024 | [RELEASE NOTES](#)

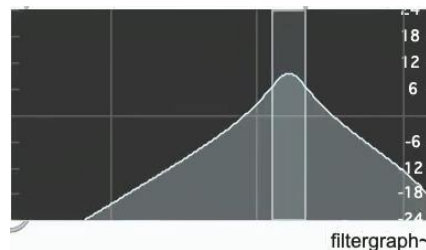
Windows 64-bit (687 MB)

Download

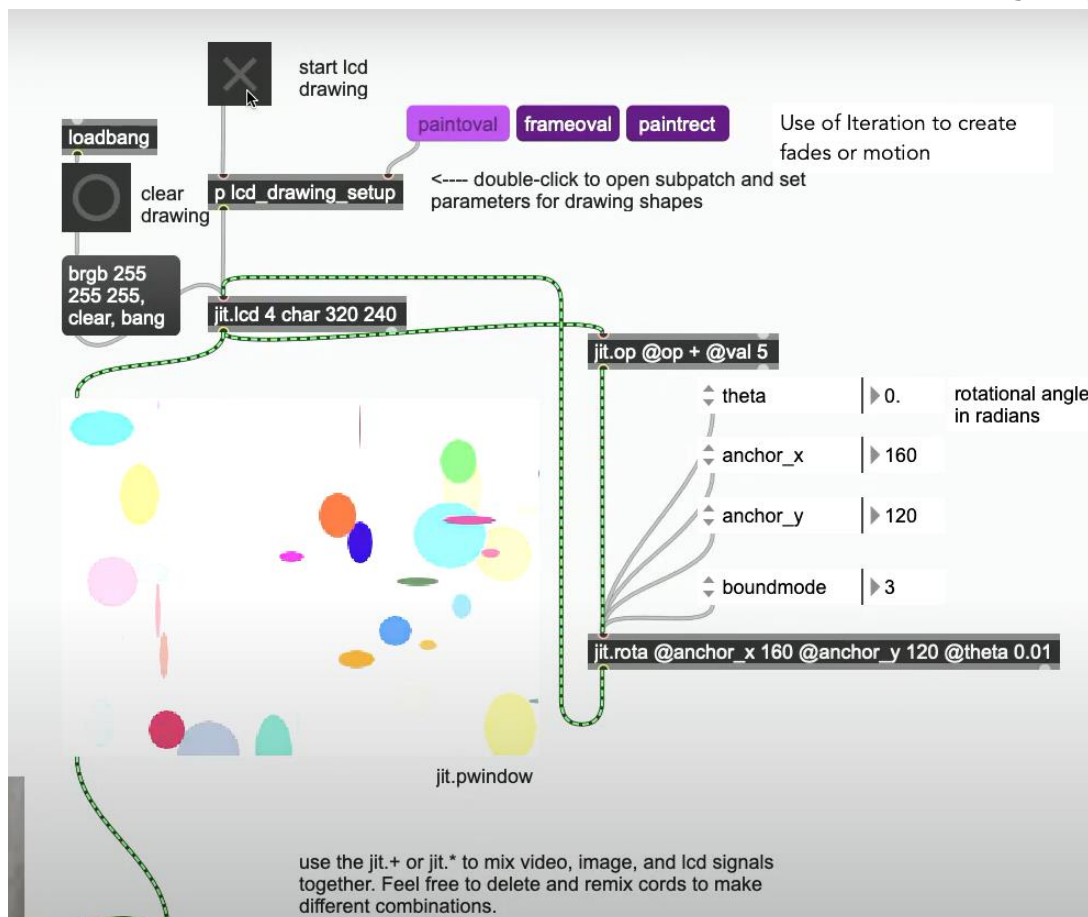
Max یک زبان برنامه‌نویسی گرافیکی است که ممکن است با آنچه قبلاً در زبان‌های برنامه‌نویسی مبتنی بر متن دیده‌اید متفاوت باشد. گرافیکی به این معناست که روی صفحه، اشیایی داریم که می‌توانیم آن‌ها را جابجا کرده و با این سیم‌های کوچک به هم وصل کنیم. این سیم‌ها را patch chords می‌نامیم زیرا گویی این زبان گرافیکی یک سینتی‌سایزر موسیقی آنالوگ با پایه پچ است، مانند اتصال کابل‌ها به ورودی‌ها و خروجی‌ها. در هر یک از این اشیاء، نیم‌دایره‌هایی در بالا و پایین وجود دارد که به ترتیب ورودی و خروجی نامیده می‌شوند. داده‌ها وارد ورودی می‌شوند، توسط شیء پردازش می‌شوند و سپس خروجی‌ها از خروجی‌ها خارج می‌شوند.



در این بخش می‌توانید ببینید که انواع مختلفی از اشیاء روی صفحه وجود دارند و ما آنها را به صورت بصری از هم متمایز می‌کنیم. شیء اصلی فقط یک شیء تخت است که ممکن است یک عنوان و شاید چند کلمه یا عدد داخل آن باشد. همچنین چیزی شبیه به دکمه‌ها وجود دارد که به آنها پیام‌ها می‌گوییم. برای مثال المان‌های مختلفی در این نرم افزار موجود است نظیر فیلترها تنظیم کننده‌ها و... که می‌توانید آنها را تغییر دهید تا به خواسته خود برسید.



اگر قصد داشته باشیم تصویری 2D بسازیم، نرم افزار Max/MSP برایش راهکار دارد بدین صورت که در کنار هم قرار دادن المان‌های متنوع می‌توانیم تصویرسازی کنیم. مثلاً در شکل زیر بلوک "start lcd drawing" را در ابتدا قرار می‌دهیم. و باقی المان‌ها نظیر رنگ‌های RGB و بلوک‌های سازنده تصویر را کنار هم قرار بدهیم تا شکل مورد نظر تشکیل شود و اگر بخواهید، می‌توانید شکل را تغییر دهید.



همچنین با این زبان گرافیکی می‌توانیم تغییراتی روی ویدیو اعمال کنیم. برای مثال در مثال پایین ویدیویی از یک مرغ را روشن می‌کنم، بنابراین شما هم باید آن را داشته باشید. می‌توانیم روی این ویدیو افکتهایی اعمال کنیم مثل محو کردن و اشباع رنگ. همچنین از عجایب این نرم افزار این است که می‌توان تغییرات روی یک ویدیو را با تغییرات یک صوت هماهنگ کرد و بلوک های آنها را با یکدیگر اقدام کرد و نتایج جالبی را مشاهده کرد.



برای مثال در دو تصویر فوق، در یکی از آنها خروجی بلوک فیلتر میانگذر و همچنین تنظیم کننده اندازه هم اتصال بدهیم، می‌توانیم بصورت متغیر با زمان تصویر خروجی را مشاهده کنیم که با زمان تغییر میکند. در صورتیکه علاقه مند هستید المان های موسیقی و کنترل کننده را بصورت دستی تعیین کنید می‌توانید از دوره یوتیوبی که در انتها گفته شده بهره ببرید.

بررسی المان‌های استفاده شده در Max/MSP

با استفاده از المان `udpreceive` داده‌ها را دریافت میکنیم. این اتصال بین پایتون و نرم افزار `max` به کمک یک `port` و `ip` خاص انجام می‌شود که به صورت زیر تنظیم شدند:

*** برای اجرای درست کدها، همزمان نرم افزار `max/msp` باید باز باشد.

```
ip = "127.0.0.1"
port = 5060
```

```
udpreceive 5060
```

هنگامی که دستوری از پایتون میفرستیم به ماژول‌های `route` فرستاده میشود و با توجه به اسم آن دستور مورد نظر اعمال میشود.

```
print receivedmess @popup 1 route /midinote /control route accompaniment_control /chord /accompaniment_play route /piano_note
```

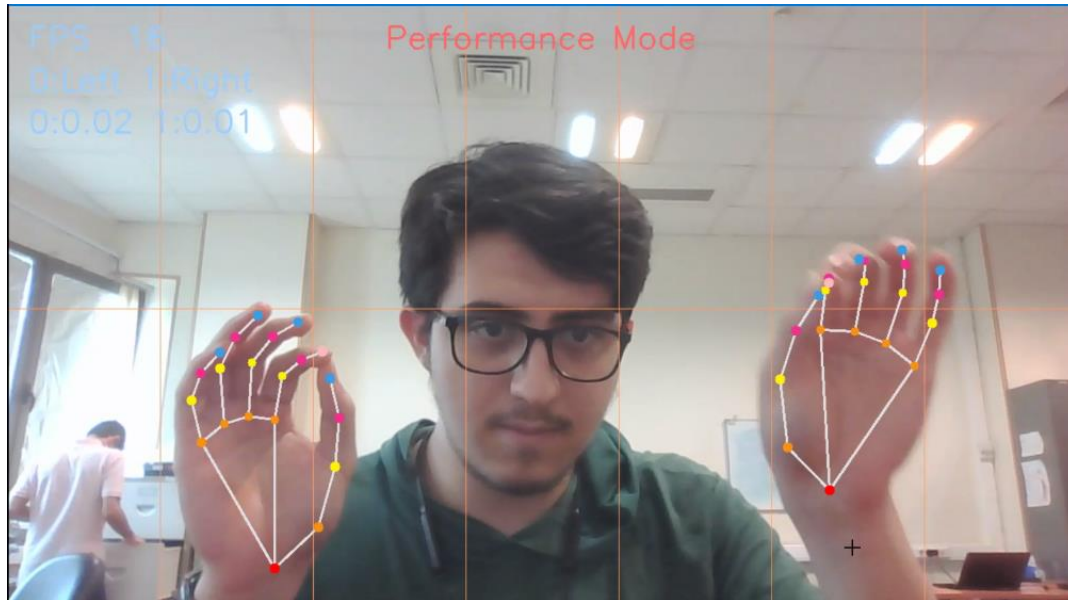
در ادامه سعی میکنیم برای هر مودی که استفاده کردیم ماژولهایش را به اختصار توضیح دهیم. برای ورود به هر مود، داخل کد دکمه‌هایی از کیبورد را اختصاص دادیم.

کد 1: Performance Mode

کد 2: Sampler Mode

کد 3: Control Mode

Performance Mode - 1



پس از دریافت نقاط کلیدی دست، اگر فاصله بین نقطه کلیدی ۴ (نوک انگشت شست) و نقطه کلیدی ۸ (نوک انگشت اشاره) پس از نرمال‌سازی کمتر از ۰.۰۳ باشد و فاصله زمانی از آخرین ارسال بیشتر از ۰.۳ ثانیه باشد، *MuGeVI* سیگنال نوت دیگری به *Max patch* ارسال می‌کند. *Max patch* با استفاده از ماژول *makenote*، پیام‌های نوت-روشن و نوت-خاموش *MIDI* را بر اساس اطلاعات گام، مدت و سرعت ایجاد می‌کند، سپس آنها را با ماژول *midiformat* به فرمت *MIDI* تبدیل کرده و در نهایت سیگنال *MIDI* را با استفاده از ماژول *midiout* به منبع صوتی ارسال می‌کند. *MuGeVI* به طور پیش‌فرض از منبع صوتی پیانو ویوتبل ویندوز استفاده می‌کند و با تغییر منبع صوتی، کاربران می‌توانند سازهای بیشتری را بنوازند. از آنجایی که *MuGeVI* بسیار قابل توسعه است، عملکرد آن می‌تواند با تغییر پارامترهایی مانند فاصله زمانی ارسال و انگشت‌گذاری غنی‌تر شود.

2- Sampler Mode:

در این بخش قصد داریم یک نسخه ساده تر از looper های مورد استفاده در ساخت موسیقی را به کمک بینایی کامپیوتر ایجاد کنیم.

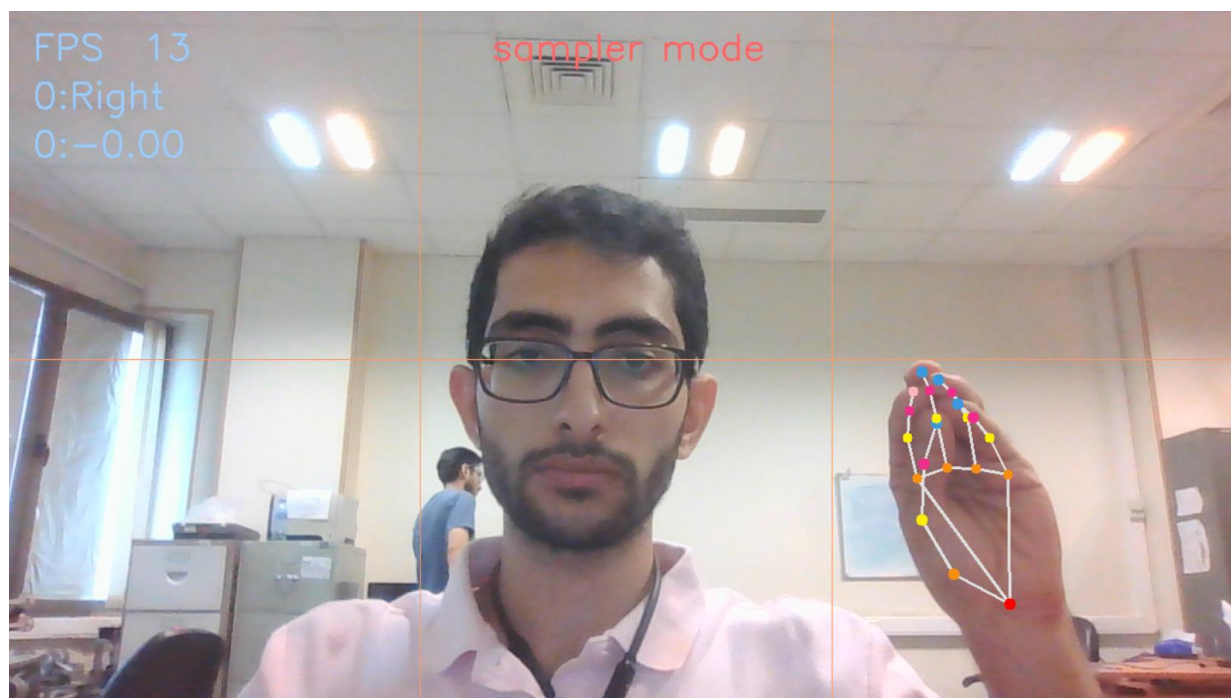
Looper ها ساختاری مثل شکل زیر دارند که برای ترکیب کردن سمپل های صوتی مختلف استفاده می شوند و این سمپل ها به صورت دائم تا انتهای ویدیو پخش می شوند. آهنگساز ها از looper برای ساختن آهنگ از ترکیب سمپل ها استفاده می کنند که به تبع در ساخت موسیقی های الکترونیک بیشتر استفاده می گردد.



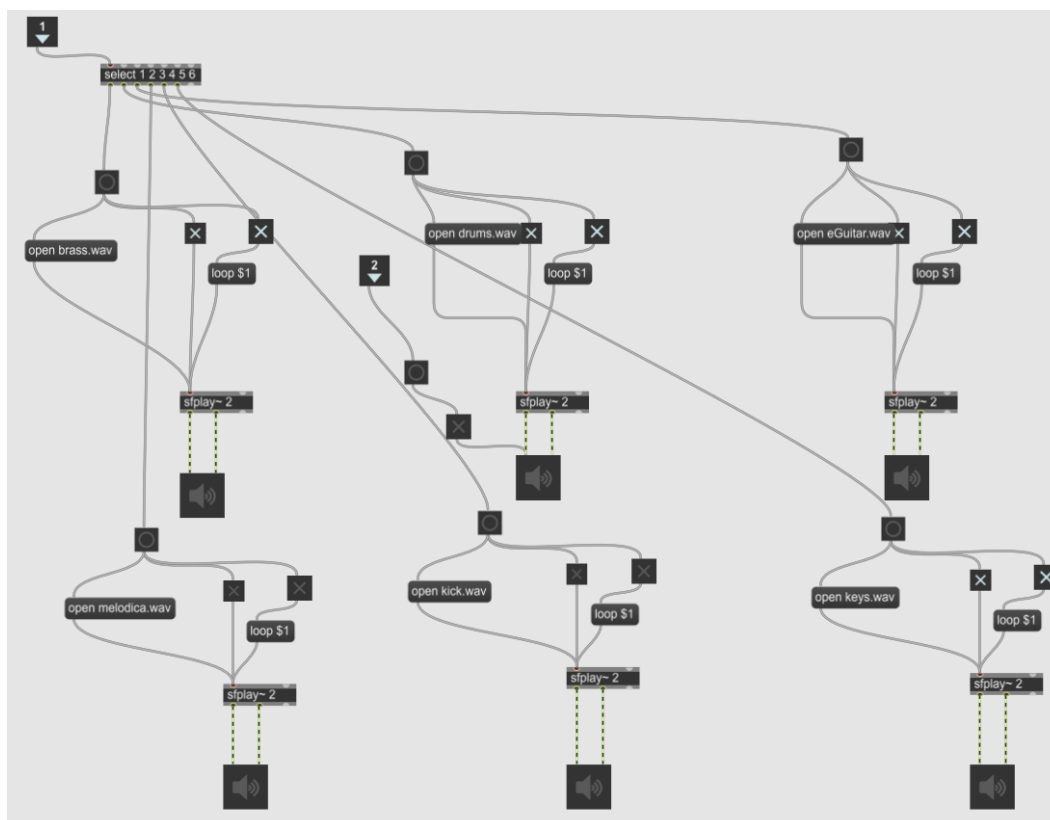
ابتدا برای شبیه سازی کلید های شکل فوق، بلوک هایی را ایجاد کردیم که هر کدام از آنها مربوط به یک سمپل هستند. توجه داشته باشید که سمپل ها از قبل باید آماده سازی شده باشند و فایل wav آنها در فولدر پروژه موجود باشد.



طبق شکل بالا، برای ساده سازی، 6 سمپل را در نظر گرفتیم. با قرار گرفتن تمام 21 نقطه کلیدی درون هر بلوک، سمپل آن به صورت loop تا انتها که کاربر برنامه را متوقف کند، پخش می شود. برای مثال در شکل زیر سمپل مربوط به بلوک پایین سمت راست، شروع به پخش شدن می کند:



سمپل ها را از قبل در نرم افزار *max* لود کردیم و آبجکت های مربوط به این مود را به شکل زیر پیاده سازی کردیم که تمام آنها درون مازول *playsoundfile* قرار گرفته اند.



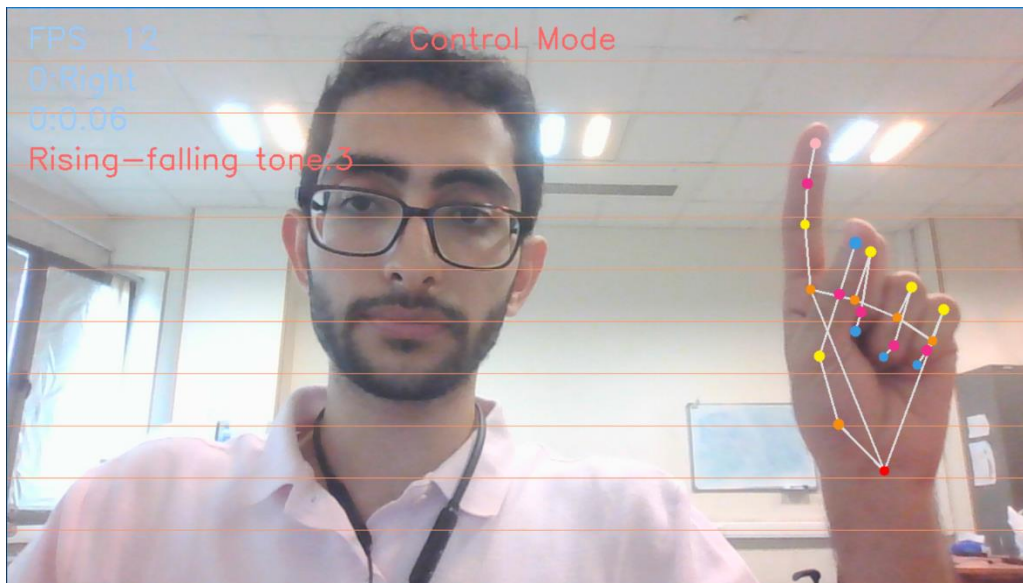
Select: با قرار گرفتن تمام نقاط کلیدی درون هر بلوک کد مربوط به آن به نرم افزار *max* ارسال می‌شود و با کمک آبجکت *select* قسمت های مربوط به آن سمپل انتخاب می‌شوند. همانطور که مشاهده می‌شود 6 بخش مشابه در شکل بالا نمایانگر، 6 سمپل مختلف هستند.

Toggle: با انتخاب سمپل مدنظر یک پیام روشن به *toggle* ارسال می‌گردد که لود شدن فایل، پخش شدن یا نشدن و همچنین تکرار آن را کنترل می‌کند.

Sfplay: این آبجکت برای *playback* فایل های صوتی استفاده می‌گردد که در اینجا دو کانال برای آن در نظر گرفته شده است.

Ezdac: بلندگوهای خروجی هستند که به *sfplay* متصل می‌شوند. در ابتدا تمام بلندگوها خاموش هستند که با دکمه *p* کیبورد فعال و با دکمه *o* غیر فعال می‌شوند.

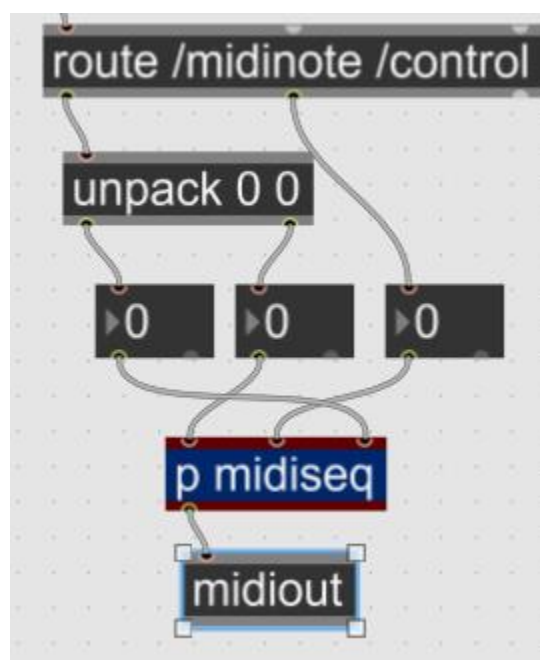
Control Mode -3



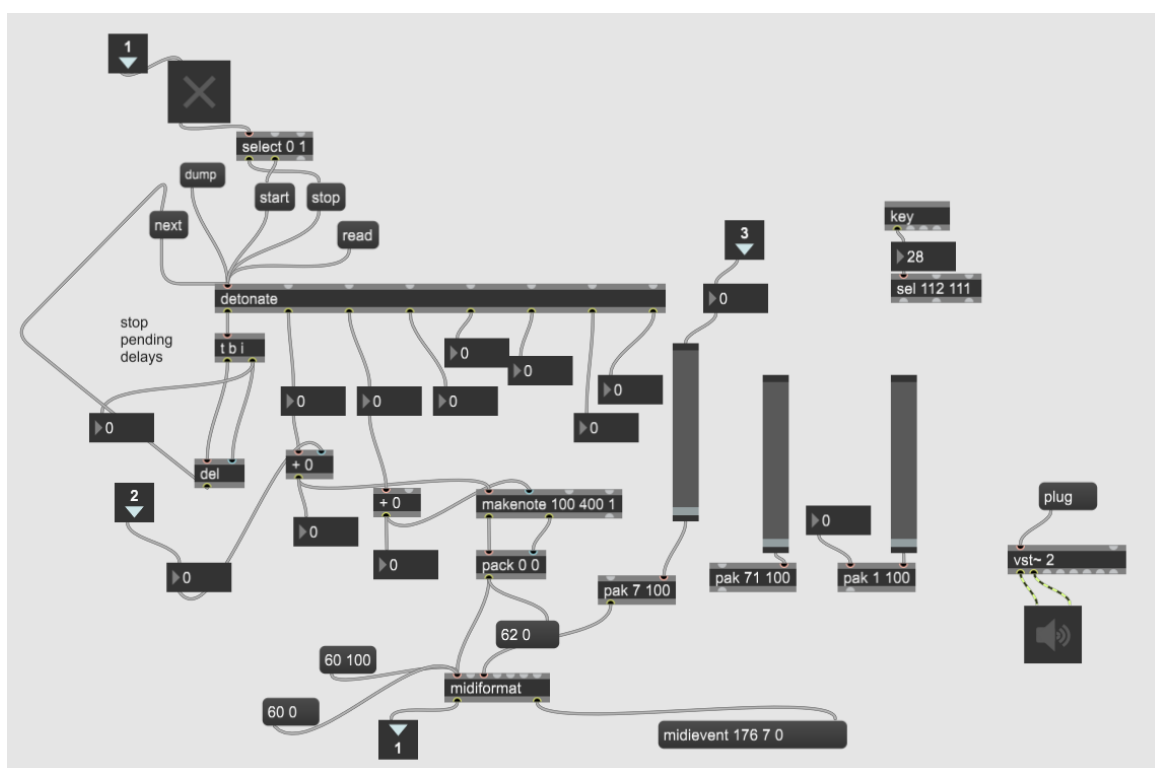
برنامه پایتون مختصات نقطه کلیدی ۸ (نوک انگشت اشاره) را تشخیص می‌دهد، تعیین می‌کند که این نقطه به کدام ناحیه عمودی تعلق دارد و موقعیت نسبی آن به صورت افقی کجاست، سپس این اطلاعات را به اطلاعات گام و سرعت تبدیل کرده و از طریق پروتکل OSC به Max patch ارسال می‌کند. سه قسمت اصلی حالت کنترل عبارتند از: (۱) پخش قابل کنترل. MuGeVI کلید p را برای پخش و o را برای توقف تنظیم می‌کند. برنامه پایتون تعیین می‌کند که آیا کلیدی فشرده شده است و متغیر کنترل سوئیچ را به Max patch ارسال می‌کند تا پخش و توقف موسیقی را کنترل کند. (۲) تغییر کلید در زمان واقعی. برنامه پایتون تعیین می‌کند که نقطه کلیدی ۸ کجاست و تعداد نیم‌پرده‌هایی که باید تغییر کنند را به Max patch ارسال می‌کند، جایی که این مقدار به گام MIDI اضافه می‌شود. (۳) کنترل حجم صدا در زمان واقعی. حرکت افقی نوک انگشت اشاره مشابه نوار کنترل حجم صدا است و Max patch اطلاعات سرعت ارسالی را دریافت کرده و سطح صدا را از طریق کنترلر MIDI شماره ۷ کنترل می‌کند.

**** برای استفاده از mode ابتدا باید درون بلوک p midiseq یک فایل با فرمت MIDI در نرم افزار لود شود.**

حال اگر به ساختار دو خواسته بالا نگاه بیاندازیم، ساختار زیر را مشاهده میکنیم:



اگر به درون *p midiseq* نگاهی بیاندازیم اجزای زیر را مشاهده میکنیم:



که اگر بطور خلاصه به المان های آن نگاهی بیندازیم متوجه میشویم که:

select 0 1 این جز برای انتخاب نت های مورد نظر بر اساس شماره نت استفاده می شوند. به عبارت دیگر، می توانید نت های مورد نظر را با استفاده از این اجزا انتخاب کنید.

Metro این جز برای تولید سیگنال های زمانی با فرکانس مشخص (مثلاً برای تولید ضرب های زمانی) استفاده می شوند. به عبارت دیگر، می توانید تعیین کنید که چه زمانی نت ها تولید شوند.

Counter این جز برای شمارش تعداد نت ها یا زمان ها استفاده می شوند. به عبارت دیگر، می توانید تعداد نت ها را در دنباله ی موسیقی شمارش کنید.

pak 7. 100 این جز برای ترکیب اطلاعات نت ها (مانند شماره نت و مدت زمان) به منظور ایجاد پیام های MIDI استفاده می شوند. به عبارت دیگر، می توانید نت ها را با استفاده از این اجزا ترکیب کنید.

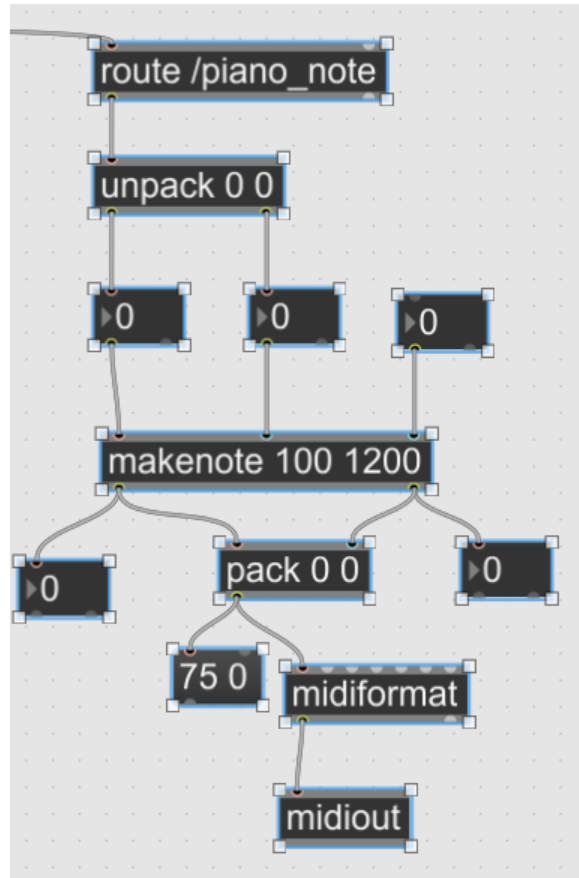
Midiformat این جز برای تبدیل اطلاعات نت ها به فرمت MIDI مناسب برای ارسال به منبع صدا (مانند سنت ها یا نرم افزارهای موسیقی) استفاده می شوند. به عبارت دیگر، می توانید اطلاعات نت ها را به فرمت MIDI تبدیل کنید.

Stop این جز برای متوقف کردن تولید پیام های MIDI به کار می روند. به عبارت دیگر، می توانید تولید نت ها را متوقف کنید.

Next این جز برای انتقال به نت های بعدی در دنباله موسیقی استفاده می شوند. به عبارت دیگر، می توانید به نت های بعدی در دنباله ی موسیقی جابجا شوید.

ساختار Piano Note برای مشاهده نت ها:

این ساختار نیز همانند دو ساختار زیر است با این تفاوت که در اینجا به تفکیک نت ها را میتوان مشاهده کرد و مشاهده کرد در هر لحظه چه نوتی مشارکت داشته است.



منابع:

https://www.youtube.com/watch?v=hJ_RaKHNo4M