TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

ĐẠI HỌC QUỐC GIA HÀ NỘI

BÁO CÁO
MÔ HÌNH HÓA VÀ MÔ PHỎNG

ĐỀ TÀI:

NHẬN DIỆN NỐT NHẠC SỬ DỤNG XỬ LÝ TÍN HIỆU SỐ

GV hướng dẫn: PGS.TS.Hoàng Văn Xiêm

Nhóm sinh viên thực hiện:

Họ và tên	*Mã sinh viên*
Bùi Đức Anh	**21021553**
Nguyễn Đình Hiếu	**21020685**

Hà Nội, 2023

Mục lục

Tóm tắt nội dung:

Âm nhạc không chỉ là một phần của cuộc sống hàng ngày của mỗi con người chúng ta mà nó còn có thể ảnh hướng trực tiếp đến tâm trạng, tinh thần và cảm xúc của chúng ta. Một bản nhạc về cơ bản bao gồm giọng hát và nhạc nền. Trong đó thì giọng hát phụ thuộc vào bẩm sinh và tố chất của mỗi người, còn nhạc nền thì đó là sự kết hợp giữa một hoặc nhiều loại nhạc cụ khác nhau như piano, guitar, sáo,.v.v. Và thật tuyệt vời nếu chúng ta có thể sau khi vừa nghe một bản nhạc mà có thể tự mình chơi lại bản nhạc đó.

Dự án này của chúng có thể giúp mọi người thực hiện việc chơi lại một bản nhạc mà mình yêu thích một cách dễ dàng hơn. Bây giờ, húng ta lấy bản nhạc gốc làm đầu vào, trích xuất đặc điểm của bản nhạc, phát hiện và xác định các nốt nhạc một cách chính xác nhất với mỗi nốt có một khoảng thời gian xuất hiện khác nhau. Trước tiên bản nhạc sẽ được thu âm và sử dụng thuật toán áp dụng xử lí tín hiệu số để nhận dạng đặc điểm. Thí nghiệm được thực hiện với một số bản nhạc piano đơn giản và các nốt nhạc của bản nhạc sau khi được xác định thì được so sánh với các nốt đã được xác định sẵn từ đầu cho đến khi tỷ lệ đúng cao nhất. Cuối cùng thì thử chơi bản nhạc bằng kết quả thu được từ thực nghiêm trên.

Từ khóa–Phân tích âm thanh, xử lý tín hiệu số, nốt nhạc, tần số[4]

1 GIỚI THIỆU CHUNG

Cảm âm tuyệt đối là khả năng nhớ, nhận biết và phát hiện với độ chính xác gần như là tuyệt đối về cao độ âm thanh mà không cần điểm tham chiếu. Những người có sở hữu khả năng này có thể ngay lập tức nhận diện được nốt nhạc nào đang được chơi từ những nhạc cụ phát ra âm thanh khác nhau. Việc sở hữu khả năng này đem lại giá trị vô cùng lớn trong lĩnh vực âm nhạc, giúp chủ sở hữu đáng kể trong quá trình học tập, biên soạn, sáng tác cũng như các công việc liên quan. Trên thực tế, rất nhiều nhạc sĩ thành công như Charlie Puth hay Lara Poe cũng sở hữu tài năng này. Tuy nhiên, cảm âm tuyệt đối yêu cầu một vài yếu tố di truyền cũng như cần phải tiếp xúc với âm nhạc từ khi còn rất nhỏ. Bởi vậy, trung bình chỉ có một trên mười nghìn người có khả năng này. [9]

Dự án này thảo luận về việc thiết kế một hộ thống nhận diện nốt nhạc sẽ tái tạo lại một phần khả năng cảm âm tuyệt đối. Hệ thống sẽ nhận một đoạn nhạc làm tín hiệu đầu vào. Sau đó xác định thời điểm một nốt nhạc mới được vang lên, từ đó chia tín hiệu thành những khoảng nhỏ và phân tích chúng trong miền tần số để xác định được nốt nhạc đang chơi.

Mục đích của viêc tạo ra hệ thống này để hỗ trợ cho việc chơi nhạc cơ bản của học sinh, sinh viên,.v.v. cũng như hỗ trợ các nhạc sĩ, nhà sản xuất sáng tác nhiều tác phẩm mới dễ dàng hơn. Dự án có thể được coi như một chiếc gương, khi ta đặt một vật gì đó trước gương thì ta có thể thấy nó phản chiếu tất cả các đặc điểm của vật đó.

2 CƠ SỞ LÝ THUYẾT

2.1 Giải thích từ khóa

Kí hiệu	*Giải thích*
C	**nốt Đô**
D	**nốt Rê**
E	**nốt Mi**
F	**nốt Pha**
G	**nốt Son**
A	**nốt La**
B	**nốt Si**
#	**dấu thăng**
b	**dấu giáng**
D5	**nốt Rê ở quãng 5**

2.2 Âm thanh và âm nhạc

2.2.1 Âm thanh

Âm thanh gồm 2 đặc tính chính, đó là cao độ và độ lớn[2]:

Âm thanh bản chất là sự dao động của các phần tử trong không khí. Tốc độ dao động sẽ ảnh hưởng đến cao độ của âm thanh. Tần số dao động thấp sẽ tạo ra âm thanh trầm và ngược lại. Khi tần số dao động càng cao thì con người sẽ càng khó phân biệt được giữa hai âm thanh cái nào có cao độ lớn hơn.

Độ lớn âm thanh phụ thuộc vào biên độ của tín hiệu âm thanh. Biên độ càng lớn thì âm thanh nghe được sẽ càng to.

2.2.2 Nốt nhạc

Con người có thể thể nghe được âm thanh trong khoảng 20Hz đến 20000Hz. Trong khoảng tần số này, âm thanh được chia thành nhiều nốt nhạc khác nhau và chọn ra các nốt nhạc đại diện cho các tần số khác nhau.

Hệ thống âm nhạc bao gồm 12 nốt nhạc, lặp lại trong suốt khoảng tần số trên. Nốt nhạc C sau có tần số lớn gấp đôi nốt nhạc C phía trước, và khoảng cách giữa hai nốt nhạc này được gọi là một quãng tám.[3]

2.2.3 Công thức tính tần số

Tần số tại mỗi nốt nhạc được xác định bởi công thức sau:

$$f(n) = (\sqrt[12]{2})^{n-49}.440Hz = 2^{\frac{n-49}{12}}.440Hz \quad (1)$$

Trong đó, n là số thứ tự của nốt [3].

2.3 Biến đổi Fourier

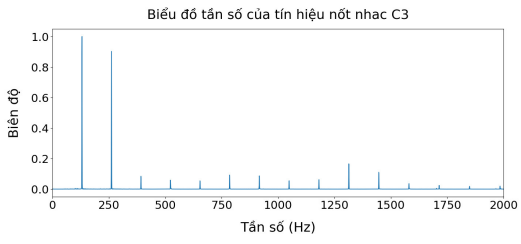
Từ mối liên hệ giữa tần số và cao độ nốt nhạc trên, có thể thấy rằng nếu như xác định được tần số của tín hiệu(âm thanh), chúng ta sẽ từ đó xác định được nốt nhạc đang được chơi là nốt nào.

Việc này có thể được thực hiện bằng cách sử dụng biến đổi Fourier lên tín hiệu đã được lấy mẫu.

Biến đổi Fourier là phương pháp toán học có thể phân tách các tín hiệu phức tạp như âm nhạc thành tổ hợp các tín hiệu sin đơn giản với biên độ và tần số khác nhau. Qua đó giúp chúng ta xác định được những tần số và cường độ của chúng trong đoạn nhạc.

Vì tín hiệu ở đây là âm thanh được lấy mẫu, chúng ta sẽ sử dụng biến đổi fourier rời rạc và được xác định theo công thức sau đây [6]:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] \cdot e^{-j\omega n} \quad (2)$$



Hình 1: Hình ảnh tín hiệu nốt C3

Hình 1 trên là tín hiệu của nốt nhạc C3 khi được biểu diễn dưới miền tần số. Từ hình ảnh chúng ta có thể thấy rõ những tần số xuất hiện trong tín hiệu tương ứng với các đỉnh nhô lên.

3 TRIỂN KHAI MÔ HÌNH PHÂN TÍCH ÂM NHẠC

3.1 Lấy mẫu tín hiệu

Âm thanh chúng ta nghe được trong đời sống là sự rung động liên tục của không khí, tuy nhiên để có thể lưu trữ lại âm thanh này vào máy tính, chúng ta cần phải liên tục lấy mẫu các giá trị tại các thời diểm liền sát với nhau. Qua đó tại tạo được âm thanh ban đầu.

Hệ thống này sẽ lấy mẫu tín hiệu âm thanh với tần số tiêu chuẩn là 44100Hz, nói cách khác, cứ mỗi 0.00002267573 giây tín hiệu này sẽ được lưu lại giá trị một lần. Con số này thường được các hãng thu âm sử dụng bởi con người chỉ có thể nghe được tần số cao nhất lên đến 20000hz, theo định lý lấy mẫu Nyquist, tín hiệu phải được lấy mẫu tại tốc độ ít nhất lớn gấp 2 lần tần số lớn nhất để tránh hiện tượng chồng phổ. Do đó, con số khoảng 40000hz là hợp lý.

Sau khi tiến hành lấy mẫu, tín hiệu âm thanh lúc này sẽ được biểu diễn dưới dạng mảng một chiều chứa các giá trị.

3.2 Biến đổi tín hiệu sang miền tần số

Tín hiệu sau khi được lấy mẫu sẽ được áp dụng biến đổi Fourier rời rạc để chuyển đổi sang miền tần số. Bằng cách này, chúng ta sẽ thu được toàn bộ tần số xuất hiện trong tín hiệu đó.

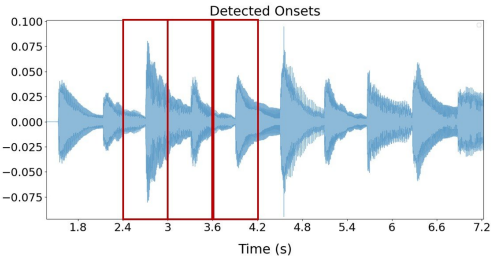
Tuy nhiên việc áp dụng trực tiếp biến đổi fourier như vậy lại không phù hợp với việc phân tích tín hiệu âm nhạc. Âm nhạc thường là một chuỗi các nốt nhạc được chơi nối tiếp nhau, nói cách khác, tại thời điểm t1, nốt đô đang vang lên nhưng khi sang t2, giai điệu chuyển sang nốt Re. Nếu áp dụng biến đổi Fourier lên toàn bộ tín hiệu, thứ chúng ta thu được là tất cả những nốt nhạc được trong toàn bộ đoạn nhạc thay vì những nốt nhạc được vang lên tại từng thời điểm nhất định. Điều này khiến cho thông tin thu được trở nên vô nghĩa.

Để giải quyết vấn đề này, chúng ta sẽ chia nhỏ tín hiệu và áp dụng biến đổi fourier lên từng phần của chúng. Kĩ thuật này được gọi là biến đổi fourier thời gian ngắn(Short Time Fourier Transform).[5]

3.2.1 Xác định thời điểm nốt nhạc bắt đầu

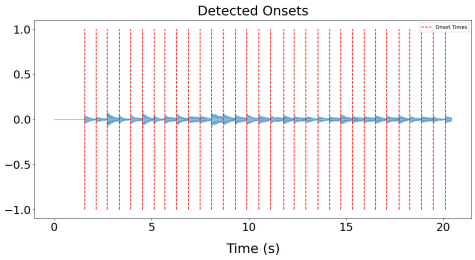
Thông thường, trong STFT, họ thường sử dụng khoảng nhảy và kích thước cửa sổ cố định. Tuy nhiên qua thực nghiệm, chúng em nhận thấy

khi áp dụng phương pháp này, phần tín hiệu tại thời điểm giữa lúc nốt nhạc trước kết thúc và nốt nhạc sau bắt đầu sẽ tồn tại tần số của cả hai nốt nhạc. Việc này khiến đánh giá nốt nhạc tại thời điểm này thiếu chính xác như hình 2.



Hình 2: Hình ảnh khi sử dụng STFT với khoảng nhảy cố định

Để khắc phục vấn đề này, chúng em quyết định chọn khoảng nhảy và kích thước linh hoạt, cụ thể là kích thước bằng thời gian từ khi một nốt nhạc bắt đầu đến khi có một nốt nhạc khác được chơi. Khi biểu diễn âm thanh trong miền thời gian, sự kiện một nốt nhạc vang lên được xác định là khi tín hiệu thay đổi đột ngột từ bé trở thành lớn. Điều này là do khi nhấn phím đàn thì búa sẽ tác động lực mạnh nhất lên dây đàn rồi rung động này sẽ yếu dần theo thời gian. Nói cách khác, việc xác định thời điểm nốt nhạc bắt đầu chính là xác định vị trí của các đỉnh trong tín hiệu âm thanh.



Hình 3: Hình ảnh phát hiện thời điểm bắt đầu của các nốt nhạc

Một mẫu tín hiệu được coi là một đỉnh khi chúng thoả mãn 3 điều kiện sau[1]:

- Giá trị của mẫu là lớn nhất trong khoảng xung quanh

$$x[n] == max(x[n - pre\_max : n + post\_max]) \tag{3}$$

- Giá trị của mẫu lớn hơn đáng kể độ lớn trung bình của các tín hiệu xung quanh

$$x[n] >= mean(x[n-pre\_avg : n+post\_avg])+delta \tag{4}$$

- Khoảng cách giữa đỉnh hiện tại và đỉnh trước là đủ lớn

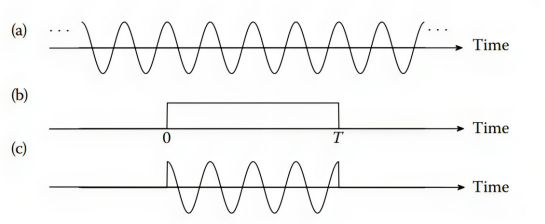
$$n - previous\_n > wait \tag{5}$$

Thông qua phương pháp chọn đỉnh trên, chúng ta sẽ thu được một mảng các giá trị đỉnh và có thể sử dụng chúng để tính toán khoảng nhảy trong tín hiệu âm thanh

3.2.2 Phương pháp của sổ

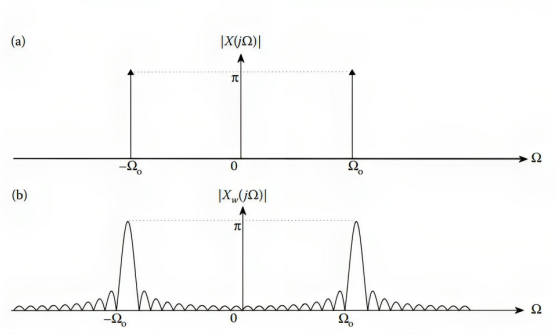
Khi tính toán phép biến đổi Fourier rời rạc, chúng ta sẽ buộc phải lấy một số lượng mẫu nhất định trong khoảng thời gian T và bỏ qua những gì thông tin tồn tại trước và sau khoảng tín hiệu này. Điều này tương đương với việc nhân tín hiệu ban đầu với một hàm cửa sổ hình chữ nhật

$$w(t) = 1, 0 \leq t \leq T \tag{6}$$



Hình 4: (a) tín hiệu, (b) hàm chữ nhật, (c) tín hiệu hữu hạn thu được

Khi áp dụng biến đổi Fourier lên phép tín hiệu mới thu được, có thể thấy rằng phổ tần số thay vì tập trung tại một đỉnh như tín hiệu ban đầu nay lại trải dài ra khắp dải tần số. Hiện tượng "rò rỉ tần số" xảy ra do quá trình lấy mẫu trong khoảng thời gian hữu hạn đã phá hủy tính chất tuần hoàn vốn có của tín hiệu do sự thiếu đồng bộ giữa điểm bắt đầu và kết thúc của tín hiệu được lấy mẫu, từ đó vô tình tạo ra thêm những 'tần số mới'[7].



Hình 5: (a) Tín hiệu gốc, (b) tín hiệu khi được lấy mẫu với sự xuất hiện rò rỉ tần số

Sự tồn tại của 'rò rỉ tần số' gây ảnh hưởng trực tiếp tới độ chính xác của việc phân tích nốt nhạc. Để giảm thiểu hiện tượng này, chúng ta có thể sử dụng các hàm cửa sổ với hình dạng khác với giá trị biên thấp, qua đó giảm thiểu rò rỉ tần số và tập trung được năng lượng vào các đỉnh tín hiệu cần phân tích.

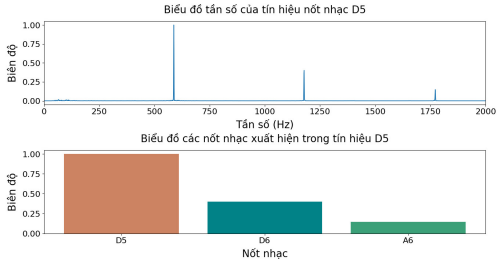
Cửa sổ blackman sử dụng trong mô hình được định nghĩa bởi công thức sau[8]:

$$w(n) = 0.42 - 0.5 \cdot \cos\left(\frac{2\pi n}{N-1}\right) + 0.08 \cdot \cos\left(\frac{4\pi n}{N-1}\right) \quad (7)$$

3.3 Chọn đỉnh và đánh giá kết quả

3.3.1 Chọn đỉnh

Sau khi biến đổi tín hiệu sang miền tần số, chúng ta sẽ thu được một đồ thị như hình 6 :



Hình 6: Hình ảnh tín hiệu nốt nhạc D5

Tại đây, có thể thấy những đỉnh trội hẳn lên chính là tần số của các nốt nhạc tồn tại trong đoạn tín hiệu. Để có thể dễ dàng hơn trong việc kết luận, tất cả biên độ của tín hiệu sẽ được chuẩn hóa bằng cách chia chúng cho biên độ của tần số có biên độ lớn nhất. Như vậy, biên độ của tất cả tần số lúc này sẽ nằm trong khoảng từ 0 tới 1.

Tiếp theo, chúng ta sẽ tiến hành lọc đỉnh. Một tần số sẽ có thể được coi là đỉnh khi thỏa mãn 2 điều kiện sau:
- Biên độ chúng lớn hơn một ngưỡng nhất định
- Khoảng cách giữa các đỉnh phải lớn hơn một khoảng nhất định. Tuy nhiên, khi càng lên cao thì khoảng cách giữa hai nốt nhạc càng lớn và ngược lại, bởi vậy khoảng cách này sẽ biến đổi tùy thuộc vào tần số đang được xem xét.

Các đỉnh đã thu được, thông qua bảng quy đổi tần số sang nốt nhạc sẽ biến đổi thành nốt nhạc. Kết quả thu được là một dãy gồm các nốt nhạc xuất hiện trong tín hiệu và tỉ lệ của chúng.

Nốt nhạc	***Tần số***
A4	**440.0000**
A#4/Bb4	**466.1638**
B4	**493.8833**
C5	**523.2511**
C#5/Db5	**554.3653**
D5	**587.3295**
D#5/Eb5	**622.2540**
E5	**659.2551**
F5	**698.4565**
F#5/Gb5	**739.9888**
G5	**783.9909**
G#5/Ab5	**830.6094**
A5	**880.0000**

3.3.2 Đánh giá kết quả

Từ bước lọc đỉnh, có thể xác định được những nốt nhạc xuất hiện trong một tín hiệu. Tuy nhiên khi thử phân tích tín hiệu D5, có thể thấy không chỉ mình tần số 587Hz của nốt nhạc này xuất hiện mà còn kèm theo một vài tần số khác với biên độ nhỏ hơn.

Lý giải cho hiện tượng này, trên thực tế, các nốt nhạc được cấu tạo nên từ hai thành phần là tần số gốc và hòa âm.

Tần số gốc có thể được xem xét như tần số cơ bản của âm thanh đoạn. Ví dụ hình 6, nếu nốt La vang lên trên piano, tần số cơ bản của nốt La đó sẽ là tần số gốc. Tần số gốc thường sẽ có biên độ lớn nhất và có thể được người nghe cảm nhận rõ nhất. Tuy nhiên trong một số trường hợp, nốt nhạc sẽ có tần số gốc bé hơn hòa âm.

Hòa âm là các tín hiệu con bổ trợ cho chất lượng âm thanh. Trên thực tế, nếu âm thanh chỉ là một sóng hình sin với duy nhất một tần số, âm thanh sẽ chẳng khác nào âm thanh của trò chơi pacman chứ chẳng thể nào màu sắc như âm thanh được tạo ra từ các loại nhạc cụ. Các hòa âm này không hề xuất hiện ngẫu nhiên, khoảng

cách giữa chúng và tần số gốc dưới góc độ toán học, sẽ tạo nên một bội số nguyên, con dưới góc độ âm nhạc, những âm này kết hợp với nhau sẽ tạo nên âm thanh êm tai, hay còn gọi là hợp âm. Bởi vậy các hòa âm này thường sẽ là các nốt nhạc cách tần số gốc 3 , 5 hoặc 8 bậc.

Hệ thống này kết luận một nốt nhạc dựa trên việc so sánh sự khác biệt giữa tín hiệu được phân tích với một set dữ liệu các nốt nhạc đã được lấy mẫu sẵn từ trước. Phương pháp này cho chúng ta kết quả khả thi khi có thể phán đoán chính xác gần như toàn bộ các nốt nhạc được từ phím đàn.

Tuy nhiên âm nhạc trên thực tế, bao gồm một tổ hợp các nhạc cụ chơi các nốt nhạc khác nhau để tạo nên những âm thanh chứa đầy màu sắc mà chúng ta thường nghe. Việc phân tích nhiều nốt nhạc cùng một thời điểm hiện tại gặp khá nhiều vấn đề do sự tồn tại của hòa âm. Khi nhiều nốt nhạc được chơi cùng một lúc, sẽ tồn tại nhiều tần số gốc khác nhau. Bên cạnh đó các hòa âm giữa những nốt nhạc này có thể chồng lên nhau, gây khó khăn cho việc đánh giá của hệ thống. Ngoài ra, việc nhấn nhá những nốt nhạc với lực không đều nhau để thể hiện cảm xúc cũng là một trở ngại không hề nhỏ. Đối mặt với vấn đề này, chúng em quyết định thay vì đánh giá luôn những nốt nhạc nào đang được chơi, sẽ coi kết quả thu được như một công cụ hỗ trợ và kết hợp cùng khả năng cảm âm cũng như nhạc lý để sắp xếp lại bản nhạc.

4 MÔ PHỎNG VÀ ĐÁNH GIÁ

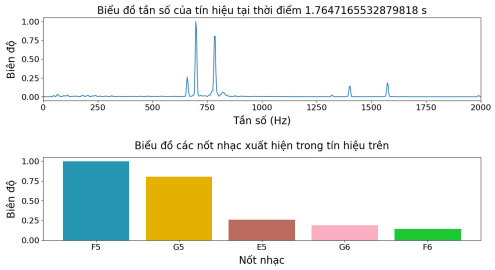
Việc mô phỏng được thực hiện bằng cách chơi một bản nhạc bằng piano, sau đó ghi âm lại và đưa lên hệ thống xử lý.

Hình 7: Hình ảnh quá trình mô phỏng

Sau khi chạy chương trình, tín hiệu sẽ được chia nhỏ và biến đổi sang miền tần số rồi phân tích những nốt nhạc tồn tại trong tín hiệu.

Nhận xét kết quả như sau: Tín hiệu thu được bao gồm tất cả những nốt nhạc được vang lên tại thời điểm đó, tuy nhiên ngoài các nốt nhạc gốc ra còn tồn tại 2 loại tần số khác:

- Các hòa âm của tần số gốc. Như đã trình bày trong phần đánh giá kết quả, tần số của những nốt nhạc cách tần số gốc 3, 5 hoặc 88 bậc cũng xuất hiện. Đặc biệt, khi chơi hợp âm(nhiều nốt nhạc cùng một lúc) thì số lượng hòa âm này sẽ lớn hơn nhiều, thậm chí còn chồng chéo lên nhau. Đây chính là nguyên nhân chính dẫn đến khó khăn trong việc xác định nhiều nốt nhạc được chơi trong cùng một thời điểm
- Tần số của các nốt nhạc trước đó. Piano có một chức năng gọi là pedal, cho phép âm thanh của một nốt nhạc được kéo dài ngay cả khi nó không còn nhận lực tác dụng từ tay nữa. Việc sử dụng pedal giúp âm thanh trở nên liền mạch, hòa hợp hơn. Tuy nhiên, dưới góc nhìn của hệ thống phân tích tín hiệu, tần số của nốt nhạc trước bị trộn lẫn vào tần số nốt nhạc sau. Ví dụ hình 8 tại nốt nhạc Sol, có thể thấy thanh tần số của nốt Fa rất lớn, tuy nhiên trên thực tế đây chỉ là giai điệu của nốt Fa được kéo dài từ trước đó và lúc này tay chỉ đang nhấn nốt Sol .



Hình 8: Các thành phần tần số xuất hiện khi mô phỏng

Tóm lại, hệ thống có thể xác định chính xác các nốt nhạc đang được chơi, tuy nhiên do sự tồn tại của hòa âm nên chỉ có thể hoạt động như một công cụ hỗ trợ khả năng cảm âm của người chơi

Tài liệu

[1] Librosa.util.peak_pick. *Read and Docs*, December 4, 2022.

[2] Pitch and loudness. *Waves and Thermodynamics*, Feb 23, 2009.

[3] Piano key frequencies. *Wikipedia*, May 24, 2023.

[4] E.S. Gopi Jay K. Patel. Musical notes identification using digital signal processing. *Journals and Books*, 2015.

[5] Nasser Kehtarnavaz. *Short-Time Fourier Transform*. Digital Signal Processing System Design, 2008.

[6] K. M. M. Prabhu. Fourier analysis techniques for signal processing. *Window Functions and Their Applications in Signal Processing*, page 18, 2014.

[7] K. M. M. Prabhu. Pitfalls in the computation of dft. *Window Functions and Their Applications in Signal Processing*, pages 67–71, 2014.

[8] K. M. M. Prabhu. Review of window functions. *Window Functions and Their Applications in Signal Processing*, page 107, 2014.

[9] Max Witynski. What is perfect pitch. *Uchicago news*, Mar 31, 1997.

Video demo sản phẩm: Tại đây