

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN CÔNG NGHỆ TRI THỨC**

BÙI TRUNG HẢI – PHẠM NGỌC TUẤN

**XÂY DỰNG ỨNG DỤNG TRỢ LÝ ẢO CHO MÁY
TÍNH SỬ DỤNG GOOGLE SPEECH API**

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN CNTT

TP. HCM, 2017

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN CÔNG NGHỆ TRI THỨC**

**BÙI TRUNG HẢI – 1312165
PHẠM NGỌC TUẤN – 1312669**

**XÂY DỰNG ỨNG DỤNG TRỢ LÝ ẢO CHO MÁY
TÍNH SỬ DỤNG GOOGLE SPEECH API**

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN CNTT

**GIÁO VIÊN HƯỚNG DẪN
TS. NGÔ MINH NHỰT**

KHÓA 2013 - 2017

[illegible]

TpHCM, ngày tháng năm
Giáo viên hướng dẫn
[Ký tên và ghi rõ họ tên]

[illegible]

TpHCM, ngày tháng năm

Giáo viên phản biện

[Ký tên và ghi rõ họ tên]

LỜI CẢM ƠN

Với lòng biết ơn sâu sắc, trước hết chúng em xin chân thành cảm ơn quý Thầy Cô khoa Công Nghệ Thông Tin - trường đại học Khoa Học Tự Nhiên, những người đã ân cần giảng dạy, xây dựng cho em một nền tảng kiến thức vững chắc để chúng em có thể thực hiện khóa luận này.

Đặc biệt, chúng em xin gửi lời tri ân sâu sắc đến Thầy Ngô Minh Nhựt. Thầy đã rất tận tâm, nhiệt tình hướng dẫn và chỉ bảo chúng em trong suốt quá trình thực hiện luận văn. Nếu không có sự giúp đỡ tận tình của thầy, chúng em chắc chắn không thể hoàn thành luận văn.

Cuối cùng, chúng con xin cảm ơn ba mẹ đã sinh thành, nuôi dưỡng, và dạy dỗ để chúng con có được thành quả như ngày hôm nay. Ba mẹ luôn là nguồn động viên, nguồn sức mạnh hết sức lớn lao mỗi khi chúng con gặp khó khăn trong cuộc sống.

Để hoàn thành luận văn này là tất cả những cố gắng, nỗ lực của chúng em. Tuy nhiên, sẽ không thể tránh khỏi những thiếu sót, kính mong nhận được sự cảm thông và giúp đỡ của quý Thầy Cô và các bạn.

TP. Hồ Chí Minh, 7/2017

Bùi Trung Hải

Phạm Ngọc Tuấn

ĐỀ CƯƠNG CHI TIẾT

Tên Đề Tài: Xây dựng ứng dụng trợ lý ảo cho máy tính sử dụng Google Speech API
Giáo viên hướng dẫn: ThS. Ngô Minh Nhựt
Thời gian thực hiện: từ ngày 15/12/2016 đến ngày 15/07/2017
Sinh viên thực hiện: Bùi Trung Hải – 1312165, Phạm Ngọc Tuấn - 1312669
Loại đề tài: Phát triển hệ thống, nghiên cứu thuật toán

Nội Dung Đề Tài: Tìm hiểu các phương pháp xử lý và tương tác với tín hiệu âm thanh, tiếng nói. Nghiên cứu hệ thống Natural Language Understanding đơn giản. Ứng dụng vào xây dựng ứng dụng trợ lý ảo cho máy tính.

Nội dung chi tiết của đề tài bao gồm:

- Xây dựng ứng dụng trợ lý ảo tương tác bằng giọng nói tiếng Anh cho máy tính
- Các vấn đề quan tâm:
 - Xử lý tín hiệu số (âm thanh, tiếng nói), hệ thống chuyển đổi tiếng nói thành văn bản.
 - Tương tác audio I/O, hệ thống chuyển đổi văn bản thành tiếng nói.
 - Giao thức truyền nhận dữ liệu REST API,...
 - Hệ thống Natural Language Understanding hiệu quả để xác định ý muốn của người dùng
 - Hệ thống chạy đa nhiệm
 - Tối ưu hóa hệ thống

- Các thành phần cơ bản của hệ thống:
 - Module thu âm từ microphone
 - Module nhận dạng từ khóa wake up
 - Module chuyển đổi tiếng nói thành văn bản
 - Module chuyển đổi văn bản thành hành động.
 - Module chuyển đổi văn bản thành tiếng nói
- Ứng dụng thử nghiệm sẽ hỗ trợ các tính năng:
 - Thông báo giờ hiện tại
 - Dự báo thời tiết trong ngày
 - Phát nhạc
 - Trả lời các câu hỏi Wh-question
 - Trả lời các thông tin cơ bản của ứng dụng: tên, tuổi,...

Kế Hoạch Thực Hiện:

- 15/12/2016 – 14/01/2017: Khảo sát, tìm hiểu về các thư viện python phục vụ cho việc tương tác và xử lý tín hiệu âm thanh.
- 15/01/2017 – 14/02/2017: Thiết kế các thành phần của hệ thống.
- 15/02/2017 – 14/03/2017: Cài đặt và thử nghiệm các thành phần: thu âm, nhận dạng từ khóa wake up, chuyển giọng nói thành văn bản, chuyển văn bản thành giọng nói.
- 15/03/2017 – 14/04/2017: Tìm hiểu và xây dựng hệ thống Natural Language Understanding.
- 15/04/2017 – 14/05/2017: Cài đặt các chức năng mà ứng dụng hỗ trợ.
- 15/05/2017 – 31/05/2017: Ráp nối tất cả các thành phần, tiến hành thử nghiệm và hoàn thiện hệ thống.
- 01/06/2017 – 28/06/2017: Viết và hoàn thiện luận văn.

Xác nhận của GVHD**Ngày 08 tháng 07 năm 2013****SV Thực hiện**

MỤC LỤC

LỜI CẢM ƠN	i
ĐỀ CƯƠNG CHI TIẾT	ii
MỤC LỤC	v
DANH MỤC HÌNH ẢNH	ix
DANH MỤC BẢNG	x
TÓM TẮT KHÓA LUẬN	xi
Chương 1 Mở đầu	1
1.1 Tổng quan về đề tài	1
1.1.1 Giới thiệu về trợ lý ảo	1
1.1.2 Khảo sát thị trường trợ lý ảo	2
1.2 Mục tiêu của khóa luận	5
1.3 Nội dung luận văn	5
Chương 2 Tín hiệu âm thanh, tiếng nói. Thư viện PyAudio	7
2.1 Tổng quan	7
2.1.1 Tổng quan về âm thanh	7
2.1.2 Tổng quan về tiếng nói	9
2.2 Cách lưu trữ âm thanh trong máy tính	10
2.2.1 Các thông số của âm thanh khi lưu trữ trên máy tính	10
2.2.2 Lưu trữ không nén	10

2.2.3	Lưu trữ nén	10
2.3	Ứng dụng	11
2.4	Thư viện PyAudio	11
2.4.1	Tổng quan	11
2.4.2	Chức năng	11
2.4.3	Cài đặt	11
2.4.4	Cách sử dụng	11
2.4.5	Các ưu, khuyết điểm	11
2.4.6	Ứng dụng	12
Chương 3	Speech to text	13
3.1	Tổng quan	13
3.2	Mô hình hoạt động	13
3.3	Ứng dụng	13
3.4	Các vấn đề cần giải quyết	13
3.4.1	Dò tìm keyword	13
3.4.2	Chuyển đổi lệnh người dùng thành văn bản	13
3.5	Thư viện pocketsphinx	14
3.5.1	Tổng quan	14
3.5.2	Chức năng	14
3.5.3	Cách cài đặt	14
3.5.4	Cách sử dụng	14
3.5.5	Ưu, nhược điểm	14
3.5.6	Ứng dụng	14
3.6	Thư viện Google Speech To Text	14
3.6.1	Tổng quan	14
3.6.2	Chức năng	14
3.6.3	Cách cài đặt	14
3.6.4	Cách sử dụng	14
3.6.5	Ưu, nhược điểm	15
3.6.6	Ứng dụng	15
Chương 4	Text To Speech	16

4.1	Tổng quan	17
4.2	Mô hình hoạt động	17
4.3	Ứng dụng	17
4.4	googleTTS	17
4.4.1	Tổng quan	17
4.4.2	Chức năng	17
4.4.3	Cách cài đặt	17
4.4.4	Cách sử dụng	17
4.4.5	Ưu, nhược điểm	17
4.4.6	Ứng dụng	17
4.5	iSpeech	17
4.5.1	Tổng quan	17
4.5.2	Chức năng	17
4.5.3	Cách cài đặt	17
4.5.4	Cách sử dụng	17
4.5.5	Ưu, nhược điểm	17
4.5.6	Ứng dụng	17
Chương 5	Intent Classification	18
5.1	Tổng quan	18
5.2	Mô hình hoạt động	18
5.3	Ứng dụng	18
5.4	Thư viện Rasa NLU	18
5.4.1	Tổng quan	18
5.4.2	Chức năng	18
5.4.3	Mô hình hoạt động	18
5.4.4	Cách cài đặt	18
5.4.5	Cách sử dụng	18
5.4.6	Chuẩn bị dữ liệu	18
5.4.7	Đánh giá model	18
5.4.8	Ứng dụng	18
Chương 6	Ứng dụng Alexa	19

6.1	Tổng quan	19
6.2	Mô hình hoạt động	19
6.2.1	Các module chính	19
6.2.2	Luồng hoạt động giữa các module	20
6.3	Các chức năng chính	20
Chương 7	Kết Luận và Hướng Phát Triển	21
7.1	Kết quả đạt được	21
7.1.1	Về mặt lý thuyết	21
7.1.2	Về mặt thực nghiệm	21
7.2	Hướng phát triển	21
Phụ Lục:	Các Công Trình Đã Công Bố	22

DANH MỤC HÌNH ẢNH

1.1	Tần suất người dùng smartphone ở Mỹ sử dụng trợ lý ảo[?]	3
1.2	Những chức năng được người dùng smartphone ở Mỹ sử dụng nhiều nhất trên trợ lý ảo[?]	4
2.1	Sóng sin có biên độ 60 dB, tần số 100 Hz	8
2.2	Minh họa độ cao (Pitch) của âm	8
2.3	Minh họa âm sắc (Timbre) của âm	9

DANH MỤC BẢNG

TÓM TẮT KHÓA LUẬN

Trong xu hướng công nghệ hiện nay, vai trò của các trợ lý ảo ngày càng trở nên quan trọng. Các hãng công nghệ lớn thay nhau tung ra những trợ lý ảo của riêng mình tích hợp trên các thiết bị di động: Siri của Apple, Cortana của Microsoft, Google Assistant của Google, Alexa của Amazon,... Chức năng của các trợ lý ảo này ngày càng được mở rộng, từ những chức năng đơn giản như tra cứu, hỏi đáp, đến những chức năng cao hơn như quản lý lịch, gọi điện thoại, dẫn đường, điều khiển các thiết bị khác,... Khóa luận này có mục đích tạo ra một trợ lý ảo có khả năng chạy được trên nhiều nền tảng hệ điều hành khác nhau trên máy tính cá nhân.

Nhận diện giọng nói là một trong những thành phần quan trọng nhất của một trợ lý ảo. Nhiều công ty và nhóm nghiên cứu lớn nhỏ đã nghiên cứu và đưa ra các bộ toolkit cũng như API cho việc nhận diện giọng nói, trong đó một trong những API có chất lượng được đánh giá tốt nhất là Google Speech API của gã khổng lồ công nghệ Google. Do đó, chúng tôi muốn tận dụng chất lượng của Google Speech API để tạo nên một trợ lý ảo có độ chính xác cao về nhận diện giọng nói.

Kết quả sơ bộ mà khóa luận đạt được là tạo ra một trợ lý ảo có thể chạy trên các hệ điều hành phổ biến trên máy tính cá nhân như Windows, Linux, Mac. Trợ lý ảo có những chức năng cơ bản của một trợ lý ảo như hỏi đáp, tra cứu thông tin, trả lời các câu hỏi về thời gian, thời tiết, ngoài ra còn có thể phát nhạc theo yêu cầu và chào hỏi ở mức độ đơn giản.

Chương 1

Mở đầu

Nội dung của chương 1 giới thiệu tổng quan về đề tài, nêu ra mục tiêu của khóa luận, và cấu trúc nội dung của luận văn.

1.1 Tổng quan về đề tài

1.1.1 Giới thiệu về trợ lý ảo

Trợ lý ảo là một phần mềm trên máy tính hoặc thiết bị di động có khả năng hỗ trợ người dùng thực hiện nhiều loại công việc, nhận lệnh từ người dùng dưới dạng ngôn ngữ tự nhiên, thường là giọng nói. Nhờ khả năng nhận lệnh và phản hồi qua giọng nói, người dùng có thể ra lệnh cho trợ lý ảo mà không cần phải thao tác bằng tay trên thiết bị. Điều đó sẽ giúp tăng tính hiệu quả và tạo ra sự tự nhiên trong giao tiếp giữa người và máy, tạo ra những kênh tương tác mới khác với truyền thống, mang đến cho người dùng những trải nghiệm mới mẻ và thú vị hơn khi sử dụng những thiết bị công nghệ.

Chức năng của các trợ lý ảo rất phong phú và đa dạng, từ những chức năng bình thường như hỏi đáp, tra cứu thông tin, bật nhạc, tìm kiếm trong danh bạ, quản lý lịch, đặt báo thức,... cho đến những chức năng đặc biệt như chơi game, trò chuyện, điều khiển các thiết bị trong gia đình, thậm chí là mua sắm, đặt chỗ nhà hàng, đặt vé máy bay,...

1.1.2 Khảo sát thị trường trợ lý ảo

Tính đến hiện tại, gần như tất cả các hãng công nghệ lớn đều đã tung ra trợ lý ảo của riêng mình:

- Apple với Siri, hoạt động trên các thiết bị của Apple như iPhone, iPad, iPod Touch, Mac và Apple TV.
- Microsoft với Cortana, hoạt động trên các phiên bản mới của Windows như Windows 10, Windows 10 Mobile, Windows Phone 8.1, và các thiết bị khác như Microsoft Band, Xbox One.
- Amazon với Alexa, hoạt động trên loa thông minh Amazon Echo.
- Google với Google Assistant, hoạt động trên các thiết bị Android, loa thông minh Google Home và ứng dụng nhắn tin Allo.
- Gần đây, Samsung đã ra mắt trợ lý ảo của mình mang tên Bixby, chạy trên các dòng điện thoại Samsung Galaxy.
- Facebook cũng đã công bố trợ lý ảo của mình mang tên M, sẽ ra mắt trong năm 2017 trên các ứng dụng Facebook và Facebook Messenger.

Trong xu hướng đó, số lượng người dùng và tần suất sử dụng của các trợ lý ảo đang ngày càng gia tăng. Theo một khảo sát thực hiện vào tháng 01/2017 trên các người dùng smartphone tại Mỹ[?], có gần 27% người được hỏi nói rằng họ dùng trợ lý ảo ít nhất một lần mỗi tuần, và khoảng 22% người dùng sử dụng trợ lý ảo hàng ngày. Trong khi đó, có 28.7% số người được hỏi chưa bao giờ sử dụng trợ lý ảo.

Khi được hỏi về lý do sử dụng trợ lý ảo, 1/3 số người được hỏi nói rằng tìm kiếm bằng trợ lý ảo dễ hơn tìm kiếm bằng tay. Khoảng 1/4 số người được khảo sát nói rằng họ không thể gõ chữ trên smartphone, hoặc không thể nhìn rõ trên smartphone, hoặc vì tìm kiếm bằng trợ lý ảo nhanh hơn tìm kiếm bằng tay. Tuy nhiên, lý do lớn nhất mà nhiều người dùng sử dụng trợ lý ảo là lái xe. Có đến hơn một nửa số người được hỏi cho biết họ sử dụng trợ lý ảo trong khi lái xe.

Nhìn vào hình 1.2 có thể thấy phần đông người dùng sử dụng trợ lý ảo với mục đích bật nhạc, quản lý báo thức, hỏi thông tin dự báo thời tiết. Ngoài ra, họ còn dùng

**Frequency with Which US Smartphone Users Use
Smartphone Virtual Assistants for Search, Jan 2017**
% of respondents



Note: ages 18+

Source: HigherVisibility survey as cited in company blog, Feb 7, 2017

223269

www.eMarketer.com

Hình 1.1: Tần suất người dùng smartphone ở Mỹ sử dụng trợ lý ảo[?]

trợ lý ảo để tìm kiếm số điện thoại trong danh bạ, hỏi câu hỏi vui, bật các tin nhắn thoại hoặc xem tin tức.

Như vậy có thể nói các trợ lý ảo đang ngày càng góp một phần quan trọng trong cuộc sống của những người dùng công nghệ. Không chỉ tạo ra một trải nghiệm mới, các trợ lý ảo còn giúp tiết kiệm thời gian và công sức. Ngoài ra, các trợ lý ảo còn có thể giúp được những người khuyết tật hoặc người cao tuổi có thể tiếp cận với các thiết bị công nghệ một cách dễ dàng hơn.

Top 10 Smartphone Virtual Assistant Search Queries/Requests According to US Smartphone Users, Jan 2017

% of respondents



Note: ages 18+; among respondents who use smartphone virtual assistants at least once a day

Source: HigherVisibility survey as cited in company blog, Feb 7, 2017

223271

www.eMarketer.com

Hình 1.2: Những chức năng được người dùng smartphone ở Mỹ sử dụng nhiều nhất trên trợ lý ảo[?]

1.2 Mục tiêu của khóa luận

Nhận thấy các trợ lý ảo nêu trong phần trước đa phần chỉ hoạt động trên một số nền tảng nhất định của hãng làm ra chúng, và chủ yếu dành cho các thiết bị di động, chúng tôi thực hiện khóa luận này với mục đích chính là tạo ra một ứng dụng trợ lý ảo có khả năng chạy trên các nền tảng hệ điều hành phổ biến trên máy tính cá nhân như Windows, Linux, Mac, với các chứng năng cơ bản:

- Hỏi đáp, tra cứu thông tin
- Hỏi giờ, thời tiết
- Phát nhạc
- Trò chuyện đơn giản

Ứng dụng phải có khả năng tự kích hoạt khi người dùng gọi tên, nhận biết chính xác các câu nói của người dùng, phản hồi một cách chính xác bằng giọng nói.

Các mục tiêu nhỏ:

- Ứng dụng có khả năng tự kích hoạt khi người dùng gọi tên
- Nhận biết chính xác câu nói của người dùng
- Phản hồi một cách chính xác và tự nhiên
- Thời gian xử lý ngắn (tính từ lúc người dùng hoàn thành câu nói đến lúc ứng dụng bắt đầu phản hồi).

1.3 Nội dung luận văn

Nội dung của luận văn sẽ gồm những phần sau:

- Chương 1: *Mở đầu*: Giới thiệu tổng quan về đề tài, mục tiêu của khóa luận và cấu trúc nội dung của luận văn.

- Chương 2: *Tổng quan về tín hiệu âm thanh, tiếng nói. Thư viện PyAudio*: Giới thiệu tổng quan về tín hiệu âm thanh, tiếng nói, các thành phần của âm thanh, cách lưu trữ âm thanh trong máy tính, các thông số của file âm thanh; giới thiệu về thư viện PyAudio: chức năng, cách cài đặt, cách sử dụng.
- Chương 3: *Speech to Text*: Giới thiệu về bài toán Speech to Text, các ứng dụng, các vấn đề cần giải quyết trong Speech to Text; giới thiệu về các thư viện pocketsphinx và Google Speech to Text: chức năng, cách cài đặt, cách sử dụng, ưu và nhược điểm, vai trò của các thư viện đó trong hệ thống.
- Chương 4: *Text to Speech*: Giới thiệu về bài toán Text to Speech, các ứng dụng; giới thiệu về Google Text to Speech và iSpeech: chức năng, cách sử dụng, ưu và nhược điểm, vai trò trong hệ thống.
- Chương 5: *Intent Classification*: Giới thiệu về bài toán Intent Classification, các thuật toán để giải quyết bài toán này, các ứng dụng; giới thiệu về thư viện Rasa NLU: chức năng, cách cài đặt, cách sử dụng, vai trò trong hệ thống.
- Chương 6: *Ứng dụng Alexa*: Giới thiệu tổng quan về ứng dụng Alexa, các module trong chương trình, luồng hoạt động giữa các module, các chức năng của ứng dụng.
- Chương 7: *Kết luận và Hướng phát triển*: Nêu ra các kết quả đạt được của khóa luận, các hạn chế và hướng phát triển.

Chương 2

Tín hiệu âm thanh, tiếng nói. Thư viện PyAudio

2.1 Tổng quan

2.1.1 Tổng quan về âm thanh

Trong vật lý, âm thanh được định nghĩa là các giao động cơ học lan truyền thông qua các phương tiện truyền dẫn như: không khí, nước, chất rắn,... Các giao động cơ học đó còn được gọi là sóng âm. Sóng âm được tạo ra do có sự biến đổi về áp suất theo thời gian. Vận tốc lan truyền trong không khí của sóng âm vào khoảng 343.2 m/s.

Đối với con người, âm thanh là những giao động có thể được cảm nhận thông qua thính giác. Các giao động khi lan truyền trong không khí sẽ va đập và làm rung màng nhĩ, qua đó não bộ sẽ thu được tín hiệu âm thanh. Con người có thể nghe được âm thanh có tần số từ 16 Hz đến 20 kHz. Âm thanh có tần số cao hơn 20 kHz gọi là siêu âm, âm thanh có tần số thấp hơn 16 Hz gọi là hạ âm.

Cách đơn giản nhất để biểu diễn âm thanh là dưới dạng sóng sin với trục x là thời gian, trục y là áp suất:

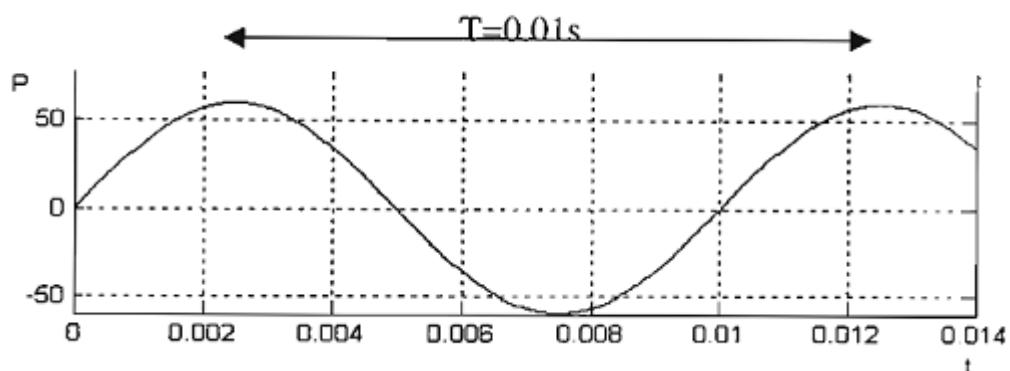
$$P = A \sin(2\pi ft)$$

Trong đó: P là áp suất, đơn vị là decibels (dB) hoặc pascals

A là biên độ của sóng, đơn vị là decibels (dB)

t là thời gian, đơn vị là giây (s)

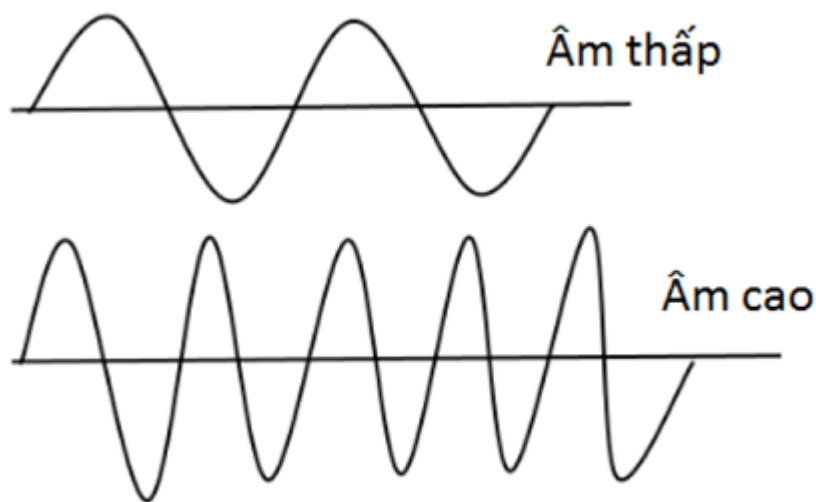
f là tần số, đơn vị là hertz (Hz)



Hình 2.1: Sóng sin có biên độ 60 dB, tần số 100 Hz

Âm thanh có một số đặc trưng cơ bản như: độ cao, độ mạnh, độ dài và âm sắc.

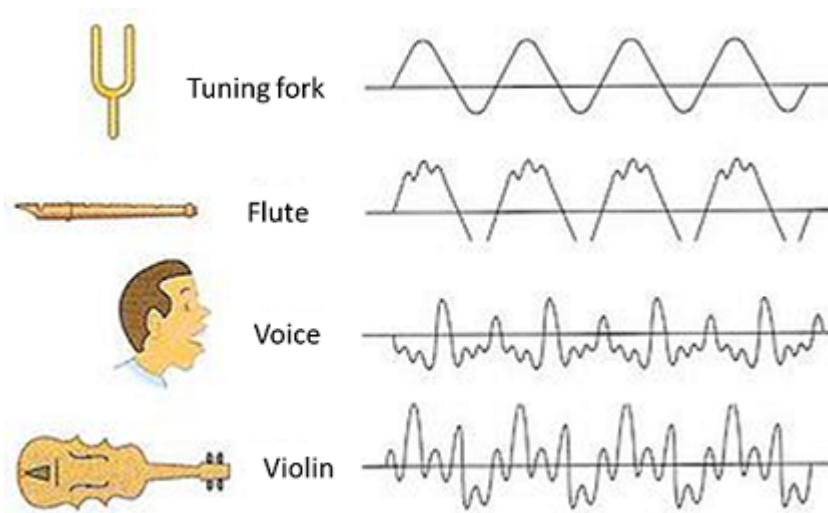
- **Độ cao (Pitch):** âm thanh luôn có một độ cao nhất định. Độ cao của âm thanh phụ thuộc vào tần số của sóng âm. Tần số càng lớn thì âm thanh càng cao, tần số càng bé thì âm thanh càng trầm.



Hình 2.2: Minh họa độ cao (Pitch) của âm

- **Độ mạnh (Intensity):** hay còn gọi là độ to của âm thanh. Độ mạnh của âm thanh phụ thuộc vào biên độ sóng âm. Biên độ càng lớn thì cường độ âm càng mạnh, biên độ càng bé thì cường độ âm càng yếu.
- **Độ dài (Duration):** là thời gian kéo dài của sóng âm.

- **Âm sắc (Timbre):** âm sắc là một đặc trưng sinh lý của âm, giúp phân biệt âm thanh do các nguồn khác nhau phát ra. Âm sắc liên quan mật thiết với đồ thị giao động âm.



Hình 2.3: Minh họa âm sắc (Timbre) của âm

Hiện nay âm thanh đã được nghiên cứu và ứng dụng rộng rãi trong nhiều lĩnh vực của cuộc sống như: truyền thông, âm nhạc, chế tạo sonar,...

2.1.2 Tổng quan về tiếng nói

Trong tự nhiên, âm thanh bao gồm nhiều loại được tạo ra từ nhiều nguồn khác nhau:

- **Âm nhạc:** âm thanh được phát ra từ các nhạc cụ.
- **Tiếng kêu:** được phát ra từ các loại động vật. Ví dụ: cá heo (1-164 kHz).
- **Tiếng động:** âm thanh phát ra từ sự va chạm giữa hai vật.
- **Tiếng ồn:** là những âm thanh không mong muốn.
- **Tiếng nói:** là những âm thanh được phát ra từ miệng con người

Ta có thể phân loại tiếng nói dựa theo thanh:

- **Âm hữu thanh:** là âm khi phát ra có sự dao động của đôi dây thanh quản.
- **Âm vô thanh:** phát ra khi đôi dây thanh quản không dao động. Thí dụ phần cuối của phát âm English, chữ sh cho ra âm sát.

Hoặc theo âm:

- **Nguyên âm:** là âm phát ra có thể kéo dài. Tất cả nguyên âm đều là âm hữu thanh, nghĩa là tuần hoàn và khá ổn định trong một đoạn thời gian vài chục ms.
- **Phụ âm:** là âm chỉ phát ra một nhát, không kéo dài được. Có phụ âm hữu thanh và phụ âm vô thanh.

Tiếng nói đóng vai trò quan trọng trong hoạt động giao tiếp giữa con người, nó là phương tiện giao tiếp nhanh, tiện lợi và phổ biến nhất.

2.2 Cách lưu trữ âm thanh trong máy tính

Âm thanh trong tự nhiên có dạng liên tục

Máy tính lưu trữ dạng rời rạc -> cần lấy mẫu âm thanh theo 1 tần số.

2.2.1 Các thông số của âm thanh khi lưu trữ trên máy tính

sample rate

bitdepth

channel ...

2.2.2 Lưu trữ không nén

file wav

2.2.3 Lưu trữ nén

file mp3

2.3 Ứng dụng

Âm thanh là công cụ tương tác giữa người sử dụng và ứng dụng.

Người sử dụng sử dụng tiếng nói để ra lệnh

Ứng dụng dùng tiếng nói để phản hồi

2.4 Thư viện PyAudio

2.4.1 Tổng quan

Là thư viện viết bằng Python hỗ trợ tất cả các Hđh

Hỗ trợ người dùng tương tác với âm thanh trên máy tính dễ dàng.

<https://people.csail.mit.edu/hubert/pyaudio/docs/>

2.4.2 Chức năng

Thu âm từ microphone của máy tính dưới dạng dữ liệu thô

Phát âm thanh ra loa của máy tính từ dữ liệu thô

2.4.3 Cài đặt

pip

build từ source

2.4.4 Cách sử dụng

2 cách sử dụng: blocking, non-blocking

các thông số của ứng dụng: samplerate, bitdepth, channel, ...

2.4.5 Các ưu, khuyết điểm

Ưu: hỗ trợ nhiều hđh, cài đặt đơn giản, nhẹ, dễ sử dụng. Nhược: ít tính năng, chưa hỗ trợ phát từ file mp3, chưa hỗ trợ thu từ nhiều micro

2.4.6 Ứng dụng

PyAudio được sử dụng trong module Microphone của ứng dụng. Giúp thu âm và chuyển cho các module khác để xử lý.

Chương 3

Speech to text

3.1 Tổng quan

3.2 Mô hình hoạt động

3.3 Ứng dụng

3.4 Các vấn đề cần giải quyết

3.4.1 Dò tìm keyword

Yêu cầu

Hoạt động liên tục -> nên hoạt động offline Xử lý nhanh từng frame của âm thanh
Độ chính xác khá cao.

Giải pháp

sử dụng thư viện pocketsphinx

3.4.2 Chuyển đổi lệnh người dùng thành văn bản

Yêu cầu

Chỉ hoạt động khi người dùng ra lệnh Độ chính xác rất cao.

Giải pháp

sử dụng thư viện google speech to text

3.5 Thư viện pocketsphinx

3.5.1 Tổng quan

3.5.2 Chức năng

3.5.3 Cách cài đặt

3.5.4 Cách sử dụng

các thông số của thư viện

3.5.5 Ưu, nhược điểm

Ưu: hỗ trợ offline, dò keyword Nhược: độ chính xác kém

3.5.6 Ứng dụng

Module wake up, giúp kích hoạt hệ thống khi người dùng gọi wakeup word

3.6 Thư viện Google Speech To Text

3.6.1 Tổng quan

3.6.2 Chức năng

3.6.3 Cách cài đặt

3.6.4 Cách sử dụng

Các thông số của thư viện

3.6.5 Ưu, nhược điểm

Ưu: độ chính xác cao Nhược: yêu cầu internet, phải gửi toàn bộ file âm thanh 1 lúc, không stream được

3.6.6 Ứng dụng

Module TTS, giúp chuyển lệnh người dùng thành văn bản.

Chương 4

Text To Speech

4.1 Tổng quan

4.2 Mô hình hoạt động

4.3 Ứng dụng

4.4 googleTTS

4.4.1 Tổng quan

4.4.2 Chức năng

4.4.3 Cách cài đặt

4.4.4 Cách sử dụng

4.4.5 Ưu, nhược điểm

4.4.6 Ứng dụng

4.5 iSpeech

4.5.1 Tổng quan

4.5.2 Chức năng

4.5.3 Cách cài đặt

4.5.4 Cách sử dụng

Chương 5

Intent Classification

5.1 Tổng quan

5.2 Mô hình hoạt động

5.3 Ứng dụng

5.4 Thư viện Rasa NLU

5.4.1 Tổng quan

5.4.2 Chức năng

5.4.3 Mô hình hoạt động

5.4.4 Cách cài đặt

5.4.5 Cách sử dụng

5.4.6 Chuẩn bị dữ liệu

5.4.7 Đánh giá model

5.4.8 Ứng dụng

Chương 6

Ứng dụng Alexa

6.1 Tổng quan

6.2 Mô hình hoạt động

6.2.1 Các module chính

Microphone

Chức năng: Các vấn đề và cách giải quyết:

Recorder

Wakeup

Text To Speech

Speech to Text

Intent Classification

Intent Processor

6.2.2 Luồng hoạt động giữa các module

6.3 Các chức năng chính

Thông báo giờ

Chức năng chi tiết: Cách thức hoạt động:

Thông báo thời tiết

Phát nhạc

Giao tiếp cơ bản

Trả lời câu hỏi Wh-question

Chương 7

Kết Luận và Hướng Phát Triển

7.1 Kết quả đạt được

7.1.1 Về mặt lý thuyết

7.1.2 Về mặt thực nghiệm

7.2 Hướng phát triển

Phụ Lục: Các Công Trình Đã Công Bố

Hội nghị quốc tế:

- **K. Tran** and B. Le, “Demystifying Sparse Rectified Auto-Encoders,” in *Proceedings of the Fourth Symposium on Information and Communication Technology*, ser. SoICT’13. New York, NY, USA: ACM, 2013, pp. 101–107. [Online]. Available: <http://doi.acm.org/10.1145/2542050.2542065>