

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



Dương Đình Long

**NGHIÊN CỨU ĐIỀU KHIỂN MOBILE ROBOT
BẰNG GIỌNG NÓI DỰA TRÊN KỸ THUẬT XỬ
LÝ NGÔN NGỮ TỰ NHIÊN**

**KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY
Ngành: Kỹ Thuật Robot**

HÀ NỘI - 2023

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

Dương Đình Long

**NGHIÊN CỨU ĐIỀU KHIỂN MOBILE ROBOT
BẰNG GIỌNG NÓI DỰA TRÊN KỸ THUẬT XỬ
LÝ NGÔN NGỮ TỰ NHIÊN**

**KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY
Ngành: Kỹ Thuật Robot**

Cán bộ hướng dẫn: TS. Dương Xuân Biên

Cán bộ đồng hướng dẫn: PGS.TS. Hoàng Văn Xiêm

HÀ NỘI – 2023

LỜI CẢM ƠN

Lời đầu tiên, tôi xin kính gửi lời cảm ơn sâu sắc tới thầy TS.Dương Xuân Biên và thầy PGS.TS. Hoàng Văn Xiêm, hai thầy đã trực tiếp gợi ý đưa ra đề tài, định hướng nghiên cứu và tận tình hướng dẫn tôi trong suốt quá trình thực hiện khóa luận này. Các thầy đã tổ chức các buổi trao đổi hàng tuần, tạo điều kiện giúp giải đáp các thắc mắc, cung cấp kiến thức và định hướng nghiên cứu điều khiển mobile robot bằng giọng nói ứng dụng xử lý ngôn ngữ tự nhiên.

Tôi cũng xin cảm ơn thầy TS.Dương Xuân Biên và thầy PGS.TS. Hoàng Văn Xiêm đã luôn động viên, và giúp đỡ cũng như đưa ra những lời khuyên hữu ích cho tôi trong quá trình thực hiện và hoàn thành khóa luận đồ án tốt nghiệp nay.

Tôi xin chân thành cảm ơn các thầy, cô giáo, cán bộ trong trường Đại học Công Nghệ - Đại Học Quốc gia Hà Nội đã truyền dạy những kiến thức quan trọng xuyên suốt quá trình học ở môi trường đại học, đặc biệt là các thầy cô trong khoa Kỹ thuật Robot đã tạo điều kiện thuận lợi nhất cho tôi thực hiện, nghiên cứu và hoàn thành khóa luận.

Lời cuối, tôi xin kính chúc tất cả các thầy, cô giáo, các bạn sinh viên sức khỏe, hạnh phúc và gặt hái được nhiều thành công trong sự nghiệp cũng như trong cuộc sống!

TÓM TẮT

Tóm tắt: Trong thời gian gần đây, sự phát triển của Robot đã thu hút sự quan tâm đặc biệt từ cộng đồng nghiên cứu do tiềm năng ứng dụng rộng lớn của chúng trong nhiều lĩnh vực, đặc biệt trong công nghiệp và cuộc sống hàng ngày. Trong các robot được quan tâm rất nhiều phải kể đến Mobile Robot (Robot di động), Robot di động được phát triển về nhiều khía cạnh vì tính thực dụng, dễ sử dụng của nó. Một trong số đó là vấn đề điều khiển robot luôn được quan tâm từ các nhà nghiên cứu. Trong phạm vi của khóa luận đồ án này, tôi tập trung vào việc phát triển xây dựng hệ thống điều khiển cho mobile robot thông qua việc sử dụng giọng nói và ứng dụng xử lý ngôn ngữ tự nhiên. Mobile robot được thiết kế để di chuyển linh hoạt dựa trên các câu lệnh bằng giọng nói.

Hệ thống xử lý ngôn ngữ tự nhiên (NLP) được tích hợp để hiểu và đáp ứng đúng ý muốn từ câu nói của người dùng, Mục tiêu là tạo ra một trải nghiệm tương tác mượt mà, nâng cao khả năng giao tiếp và điều khiển robot bằng ngôn ngữ tự nhiên.

Dự án không chỉ hướng đến sự tiện lợi trong tương tác con người-robot mà còn nhằm đưa ra những ứng dụng tiềm năng trong các lĩnh vực như dịch vụ khách hàng, giáo dục và quản lý thông tin. Chúng tôi kỳ vọng rằng dự án sẽ đóng góp vào sự phát triển của công nghệ điều khiển robot và tạo ra một hệ thống thông minh và linh hoạt.

Từ khóa: Mobile Robot, NLP, Control Robot,...

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài “Nghiên cứu điều khiển Mobile Robot bằng giọng nói ứng dụng xử lý ngôn ngữ tự nhiên” do TS. Dương Xuân Biên và PGS.TS. hướng dẫn là công trình nghiên cứu của tôi. Các nội dung nghiên cứu và kết quả trong đề án này đều là trung thực và không sao chép từ công trình của người khác.

Tất cả các tài liệu tham khảo sử dụng trong đề án đều được ghi rõ nguồn gốc và tên tác giả. Nếu có sai sót, tôi xin hoàn toàn chịu trách nhiệm.

Hà Nội, ngày tháng năm 2023

Sinh viên

Dương Đình Long

DANH SÁCH HÌNH ẢNH

Hình 1.1. Robot công nghiệp lắp ráp ô tô.....	10
Hình 1.2. Robot phục vụ khách sạn.....	10
Hình 1.3. Robot di động chở hàng.....	10
Hình 1.4. Hệ thống điều khiển mực nước bể.....	12
Hình 1.5. Học máy và trí tuệ nhân tạo.....	15
Hình 1.6. Robot lau sàn điều khiển bằng giọng nói.....	17
Hình 2.1. Cấu tạo miệng người.....	21
Hình 2.2. Cơ chế hình thành âm thanh.....	22
Hình 2.3. Một số nguyên âm.....	24
Hình 2.4. Mức cường độ âm của một câu.....	24
Hình 2.5. Tần số cơ bản tai có thể nghe.....	25
Hình 2.6. Biến đổi Fourier.....	26
Hình 2.7. Đoạn âm thanh trong miền thời gian.....	26
Hình 2.8 Mức năng lượng âm của một âm.....	28
Hình 2.9. Biến đổi Fourier từ mức năng lượng âm.....	29
Hình 2.10. Phổ âm thanh.....	29
Hình 2.11. Lọc phổ âm thanh.....	30
Hình 2.12. Cepstrum.....	30
Hình 2.13. Chuyển tiếp trạng thái mô hình ẩn Markov.....	33
Hình 2.14. Biểu đồ Markov.....	34
Hình 3.1. Robot di động bánh xe.....	37
Hình 3.2. Các biến đầu ra $y(x)$	39
Hình 3.3. Âm thanh được ghi bằng Audacity.....	40
Hình 3.4. Gán nhãn âm thanh.....	41
Hình 3.5. Quá trình trích xuất đặc trưng MFCCs.....	41
Hình 3.6. Class chuyển đổi các nhãn dán dự đoán.....	47
Hình 4.1. Mạch Arduino Uno R3.....	49
Hình 4.2. Động cơ DC 5V.....	50
Hình 4.3. Mạch điều khiển động cơ L298N.....	51
Hình 4.4. Module bluetooth HC-05.....	52
Hình 4.5. Cài đặt bluetooth.....	53

Hình 4.6. Kết nối với bluetooth HC-05.	53
Hình 4.7. Kiểm tra cổng ảo.....	54
Hình 4.8. System Preferences.	54
Hình 4.9. Kết nối với bluetooth	55
Hình 4.10. Mật khẩu bluetooth là 1234.	55
Hình 4.11. Kết nối thành công.....	55
Hình 4.12. Kiểm tra cổng ảo.....	56
Hình 4.8. Mô hình điều khiển thực tế.....	56
Hình 4.9. Code tách thu âm thanh từ micro.....	57
Hình 4.10. Biểu đồ cường độ âm thanh một câu.	57
Hình 4.11. Biểu đồ cường độ âm thanh của một từ sau khi tách.....	58
Hình 4.12. Mô hình thuật toán thực tế.....	58
Hình 4.13. Mobile Robot.	60
Hình 4.14. Nối mạch robot.	61
Hình 4.15. Kết quả huấn luyện mô hình trên tập dữ liệu.....	62
Hình 4.16. Máy tính nhận âm thanh tín hiệu điều khiển.	62
Hình 4.17. Vị trí ban đầu robot.....	63
Hình 4.18. Vị trí sau khi đi thẳng.	63
Hình 4.19. Máy tính nhận giọng nói chuyển tín hiệu điều khiển.	64
Hình 4.20. Vị trí ban đầu.	65
Hình 4.21. Vị trí sau khi rẽ trái.....	65

MỤC LỤC

LỜI CẢM ƠN	2
TÓM TẮT.....	3
LỜI CAM ĐOAN	4
DANH SÁCH HÌNH ẢNH.....	5
CHƯƠNG 1. TỔNG QUAN VỀ ĐIỀU KHIỂN ROBOT	9
1.1. Giới thiệu chung về điều khiển robot	9
1.2. Một số phương pháp điều khiển robot.....	11
1.3. Xác định bài toán điều khiển của đồ án.....	18
Kết luận Chương 1	20
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT VỀ ĐIỀU KHIỂN BẰNG GIỌNG NÓI.....	21
2.1. Kỹ thuật xử lý ngôn ngữ tự nhiên.....	21
2.1.1. Nguyên lý hình thành tiếng nói.	21
2.1.2. Cơ chế hoạt động của tai.....	25
2.1.3. Biến đổi Fourier.....	26
2.1.4. Đặc trưng MFCCs.....	28
2.1.5. Mô hình ẩn Markow.	31
2.2. Mô hình điều khiển robot bằng giọng nói.	34
Kết luận chương 2	36
CHƯƠNG 3. MÔ HÌNH HÓA VÀ ĐIỀU KHIỂN ROBOT BẰNG GIỌNG NÓI	37
3.1. Mô hình toán học mobile robot.	37
3.1.1. Phân tích mô hình toán học.	37
3.1.2. Hệ phương trình động lực học.	37
3.2. Xây dựng mô hình điều khiển bằng giọng nói.	40
3.2.1. Xử lý dữ liệu giọng nói.....	40
3.2.2. Chuyển đổi tín hiệu giọng nói thành tín hiệu điều khiển.....	46

Kết luận chương 3	48
CHƯƠNG 4. THỰC NGHIỆM ĐIỀU KHIỂN ROBOT BẰNG GIỌNG NÓI.....	49
4.1. Mô hình thực nghiệm.	49
4.2. Kết quả thực nghiệm điều khiển.....	60
Kết luận chương 4.	66

CHƯƠNG 1. TỔNG QUAN VỀ ĐIỀU KHIỂN ROBOT

1.1. Giới thiệu chung về điều khiển robot

Robot và hệ thống điều khiển ngày càng chiếm vị trí quan trọng trong cả cơ bản và phức tạp của cuộc sống hiện đại. Định nghĩa đơn giản nhất về robot là một thiết bị tự động có khả năng thực hiện nhiều tác vụ một cách đáng tin cậy và hiệu quả. Từ công nghiệp đến y tế, từ nghiên cứu khoa học đến quân sự, robot đã và đang định hình mọi khía cạnh của xã hội và kinh tế.

Đa dạng về hình dạng và kích thước, robot không chỉ xuất hiện trong các nhà máy công nghiệp với những chiếc cánh tay cơ khí lớn lẫm, mà còn trong những dịch vụ y tế với những bộ phận nhỏ gọn và nhạy bén. Khả năng tự động hoá công việc không chỉ mang lại sự hiệu quả mà còn giảm bớt nguy cơ cho con người, làm tăng tính an toàn và đồng thời mở ra những khả năng mới cho sự sáng tạo và phát triển. Ứng dụng của robot và hệ thống điều khiển không ngừng mở rộng, từ việc giảm nhào lộn trong các công việc lặp đi lặp lại như phân loại sản phẩm, lắp ráp và vận chuyển, đến những nhiệm vụ đòi hỏi độ chính xác cao như chế tạo vi mạch hay lắp ghép vi mạch. Chúng không chỉ xuất hiện trong môi trường công nghiệp mà còn chinh phục những thách thức ở những môi trường nguy hiểm như không gian vũ trụ, núi lửa, và cả trong các công việc liên quan đến phóng xạ hạt nhân. Ngoài ra, robot cũng đang trở thành một phần không thể thiếu trong các lĩnh vực dịch vụ và sinh hoạt hàng ngày. Từ các robot phục vụ trong nhà hàng và khách sạn đến những thiết bị hỗ trợ gia đình, chúng đều đóng vai trò quan trọng trong việc cải thiện chất lượng cuộc sống. Nhìn chung, vai trò của robot và hệ thống điều khiển không chỉ giới hạn trong việc thay thế con người trong những công việc đơn điệu và lặp lại mà còn mở ra những khả năng mới cho sự phát triển và sáng tạo trong mọi lĩnh vực của đời sống xã hội. Chúng đang là những đối tác đáng tin cậy, làm tăng cường hiệu suất và sự an toàn trong công việc, cũng như mang lại những trải nghiệm mới và thuận lợi cho cuộc sống hàng ngày của chúng ta.

Dưới đây là một số ví dụ về robot đã giúp ích con người trong việc phát triển trong công việc cũng như cải thiện cuộc sống chất lượng cao hơn.



Hình 1.1. Robot công nghiệp lắp ráp ô tô



Hình 1.2. Robot phục vụ khách sạn.



Hình 1.3. Robot di động chở hàng.

Điều khiển robot là một lĩnh vực kỹ thuật quan trọng, đảm bảo cho robot có thể thực hiện các nhiệm vụ được giao một cách chính xác, hiệu quả và an toàn.

Mục tiêu của điều khiển robot là đạt được sự cân bằng giữa độ chính xác và tốc độ trong việc thực hiện các tác vụ cụ thể. Điều này có nghĩa là robot phải có thể hoàn thành các nhiệm vụ với độ chính xác cao, nhưng vẫn đảm bảo tốc độ xử lý đủ nhanh để đáp ứng các yêu cầu của công việc.

Điều khiển robot đóng vai trò quan trọng trong việc định hình và cung cấp sự linh hoạt cho hệ thống robot. Việc nắm vững các nguyên tắc điều khiển sẽ giúp các nhà thiết kế robot có thể tạo ra những hệ thống robot có khả năng thực hiện nhiều loại nhiệm vụ khác nhau. Lĩnh vực điều khiển robot đã trải qua một sự phát triển kỹ thuật đáng kể trong những năm gần đây. Từ những cơ cấu cơ học đơn giản ban đầu, robot ngày nay đã có khả năng thực hiện các tác vụ phức tạp nhờ sự kết hợp của công nghệ điện tử và trí tuệ nhân tạo. Việc sử dụng hệ thống cảm biến, vi mạch tích hợp và phân tích dữ liệu thông minh đã mở ra nhiều khả năng mới cho lĩnh vực điều khiển robot. Các robot hiện đại có thể sử dụng các cảm biến để thu thập thông tin về môi trường xung quanh và sử dụng các thuật toán học máy để điều chỉnh hành vi của mình cho phù hợp. Tiến bộ trong trí tuệ nhân tạo, học sâu và công nghệ cảm biến đang tạo ra những cơ hội mới cho lĩnh vực điều khiển robot. Các robot tự học hỏi, có khả năng tương tác với con người và có khả năng thích ứng với môi trường thay đổi đang dần trở thành hiện thực.

Tương lai của lĩnh vực điều khiển robot là rất tươi sáng. Với sự phát triển của công nghệ, robot sẽ ngày càng trở nên thông minh và linh hoạt hơn, đáp ứng được nhiều nhu cầu khác nhau của con người.

1.2. Một số phương pháp điều khiển robot

Điều khiển hồi tiếp

Nguyên lý hoạt động

Điều khiển hồi tiếp là phương pháp điều khiển sử dụng các thông tin cảm biến từ môi trường để điều chỉnh hành vi của robot. Phương pháp này dựa trên nguyên tắc so sánh trạng thái thực tế của robot với trạng thái mong muốn, sau đó sử dụng các tín hiệu điều khiển

để điều chỉnh hành vi của robot sao cho trạng thái thực tế của robot tiệm cận với trạng thái mong muốn.

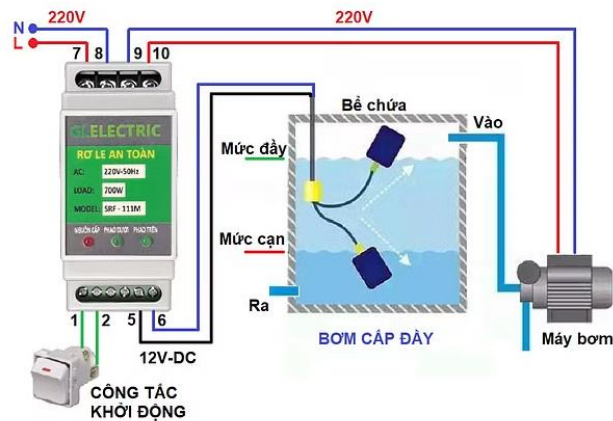
Các bước thực hiện

Đo lường trạng thái thực tế của robot: Đầu tiên, bộ điều khiển cần đo lường trạng thái thực tế của robot, chẳng hạn như vị trí, tốc độ, hoặc hướng.

So sánh trạng thái thực tế với trạng thái mong muốn: Sau đó, bộ điều khiển sẽ so sánh trạng thái thực tế của robot với trạng thái mong muốn. Nếu trạng thái thực tế khác với trạng thái mong muốn, thì bộ điều khiển sẽ cần thực hiện điều chỉnh.

Tính toán tín hiệu điều khiển: Bộ điều khiển sẽ sử dụng các thuật toán điều khiển để tính toán tín hiệu điều khiển cần thiết để đưa robot về trạng thái mong muốn.

Áp dụng tín hiệu điều khiển: Cuối cùng, bộ điều khiển sẽ áp dụng tín hiệu điều khiển lên các cơ cấu của robot để điều chỉnh hành vi của robot.



Hình 1.4. Hệ thống điều khiển mực nước bể.

Ưu điểm

Điều khiển hồi tiếp là phương pháp điều khiển đơn giản và dễ hiểu. Phương pháp này cũng có thể được áp dụng cho nhiều loại robot và nhiệm vụ khác nhau.

Nhược điểm

Điều khiển hồi tiếp có thể không hiệu quả trong môi trường phức tạp, chẳng hạn như môi trường có nhiều vật cản hoặc môi trường có nhiều yếu tố gây nhiễu.

Ứng dụng

Điều khiển hồi tiếp thường được sử dụng để điều khiển các robot di chuyển trong môi trường đơn giản, chẳng hạn như robot di chuyển trên đường thẳng hoặc robot di chuyển trong một phòng kín.

Điều khiển dự đoán

Nguyên lý hoạt động

Điều khiển dự đoán là phương pháp điều khiển sử dụng các mô hình dự đoán để dự đoán trạng thái của robot trong tương lai, từ đó điều chỉnh hành vi của robot để tránh các va chạm hoặc các tình huống không mong muốn. Phương pháp này dựa trên nguyên tắc sử dụng các mô hình dự đoán để dự đoán trạng thái của robot trong tương lai, sau đó sử dụng các tín hiệu điều khiển để điều chỉnh hành vi của robot sao cho tránh được các va chạm hoặc các tình huống không mong muốn.

Các bước thực hiện

Đo lường trạng thái thực tế của robot: Đầu tiên, bộ điều khiển cần đo lường trạng thái thực tế của robot, chẳng hạn như vị trí, tốc độ, hoặc hướng.

Dự đoán trạng thái của robot trong tương lai: Sau đó, bộ điều khiển sẽ sử dụng các mô hình dự đoán để dự đoán trạng thái của robot trong tương lai.

So sánh trạng thái dự đoán với trạng thái mong muốn: Bộ điều khiển sẽ so sánh trạng thái dự đoán của robot với trạng thái mong muốn. Nếu trạng thái dự đoán khác với trạng thái mong muốn, thì bộ điều khiển sẽ cần thực hiện điều chỉnh.

Tính toán tín hiệu điều khiển: Bộ điều khiển sẽ sử dụng các thuật toán điều khiển để tính toán tín hiệu điều khiển cần thiết để đưa robot về trạng thái mong muốn.

Áp dụng tín hiệu điều khiển: Cuối cùng, bộ điều khiển sẽ áp dụng tín hiệu điều khiển lên các cơ cấu của robot để điều chỉnh hành vi của robot.

Ưu điểm

Điều khiển dự đoán có thể hiệu quả hơn điều khiển hồi tiếp trong môi trường phức tạp, chẳng hạn như môi trường có nhiều vật cản hoặc môi trường có nhiều yếu tố gây nhiễu.

Nhược điểm

Điều khiển dự đoán có thể phức tạp và khó thực hiện hơn điều khiển hồi tiếp.

Ứng dụng

Điều khiển dự đoán thường được sử dụng để điều khiển các robot di chuyển trong môi trường phức tạp, chẳng hạn như robot di chuyển trong một thành phố đông đúc hoặc robot di chuyển trong môi trường có nhiều vật cản.

Điều khiển học máy

Nguyên lý hoạt động

Điều khiển học máy là phương pháp điều khiển sử dụng các thuật toán học máy để tự động học hỏi từ dữ liệu thực tế, từ đó điều khiển robot một cách hiệu quả hơn. Phương pháp này dựa trên nguyên tắc sử dụng các thuật toán học máy để học hỏi từ dữ liệu được tạo.

Ưu điểm

Phương pháp điều khiển học máy có thể học hỏi và thích ứng với môi trường thay đổi.

Phương pháp này có thể được sử dụng để điều khiển robot trong môi trường phức tạp.

Nhược điểm

Phương pháp điều khiển học máy có thể phức tạp và khó thực hiện.

Phương pháp này cần có lượng dữ liệu đào tạo lớn.



Hình 1.5. Học máy và trí tuệ nhân tạo.

Phương pháp điều khiển học máy thường bao gồm các bước sau:

Thu thập dữ liệu: Đầu tiên, hệ thống điều khiển sẽ thu thập dữ liệu từ môi trường. Dữ liệu này có thể bao gồm thông tin về trạng thái của robot, trạng thái của môi trường, và kết quả của các hành động của robot.

Huấn luyện mô hình: Sau đó, hệ thống điều khiển sẽ sử dụng các thuật toán học máy để huấn luyện mô hình. Mô hình học máy sẽ học cách điều khiển robot dựa trên dữ liệu đã thu thập.

Điều khiển robot: Cuối cùng, hệ thống điều khiển sẽ sử dụng mô hình học máy để điều khiển robot.

Ứng dụng

Phương pháp điều khiển học máy thường được sử dụng để điều khiển các robot tự hành, chẳng hạn như robot giao hàng hoặc robot thăm dò.

Phương pháp điều khiển bằng giọng nói

Phương pháp điều khiển bằng giọng nói là một phương pháp điều khiển robot sử dụng giọng nói của con người để điều khiển hành vi của robot. Phương pháp này dựa trên nguyên tắc sử dụng các thuật toán nhận dạng giọng nói để nhận dạng các lệnh thoại của con người, sau đó sử dụng các thuật toán điều khiển để điều chỉnh hành vi của robot theo các lệnh thoại đó.

Ưu điểm

Phương pháp điều khiển bằng giọng nói cho phép con người điều khiển robot một cách tự nhiên và trực quan. Phương pháp này có thể được áp dụng cho nhiều loại robot và nhiệm vụ khác nhau.

Nhược điểm

Phương pháp điều khiển bằng giọng nói có thể không chính xác trong môi trường có nhiều tiếng ồn. Phương pháp này có thể khó thực hiện trong môi trường phức tạp.

Ứng dụng

Phương pháp điều khiển bằng giọng nói thường được sử dụng để điều khiển các robot trong các môi trường gia đình hoặc văn phòng, chẳng hạn như robot hút bụi, robot lau nhà, hoặc robot trợ lý.



Hình 1.6. Robot lau sàn điều khiển bằng giọng nói.

Phương pháp điều khiển bằng giọng nói thường bao gồm các bước sau:

Thu thập dữ liệu giọng nói: Đầu tiên, hệ thống điều khiển sẽ thu thập dữ liệu giọng nói của con người. Dữ liệu giọng nói này có thể được thu thập bằng cách sử dụng các microphone.

Phân tích dữ liệu giọng nói: Sau đó, hệ thống điều khiển sẽ sử dụng các thuật toán nhận dạng giọng nói để phân tích dữ liệu giọng nói. Các thuật toán này sẽ xác định các từ và cụm từ trong dữ liệu giọng nói.

Tạo ra các lệnh điều khiển: Từ các từ và cụm từ được xác định, hệ thống điều khiển sẽ tạo ra các lệnh điều khiển. Các lệnh điều khiển này sẽ được sử dụng để điều chỉnh hành vi của robot.

Chuyển đổi các lệnh điều khiển thành các tín hiệu điều khiển: Cuối cùng, hệ thống điều khiển sẽ chuyển đổi các lệnh điều khiển thành các tín hiệu điều khiển. Các tín hiệu điều khiển này sẽ được sử dụng để điều khiển các cơ cấu của robot.

Các thuật toán nhận dạng giọng nói

Có nhiều thuật toán nhận dạng giọng nói khác nhau. Một số thuật toán phổ biến bao gồm:

Thuật toán phân tích phổ: Thuật toán này phân tích phổ của tín hiệu giọng nói để xác định các từ và cụm từ.

Thuật toán dựa trên Markov: Thuật toán này sử dụng các mô hình Markov để xác định các từ và cụm từ.

Thuật toán dựa trên CNN: Thuật toán này sử dụng các mạng nơ-ron tích chập để xác định các từ và cụm từ.

Các thách thức

Nhận dạng giọng nói không chính xác: Các thuật toán nhận dạng giọng nói có thể không chính xác trong môi trường có nhiều tiếng ồn.

Điều khiển robot trong môi trường phức tạp: Điều khiển robot trong môi trường phức tạp có thể khó khăn, chẳng hạn như môi trường có nhiều vật cản hoặc môi trường có nhiều yếu tố gây nhiễu.

Tương lai

Phương pháp điều khiển bằng giọng nói có tiềm năng trở thành một phương pháp điều khiển robot phổ biến trong tương lai. Với sự phát triển của công nghệ, các thuật toán nhận dạng giọng nói sẽ trở nên chính xác hơn và dễ dàng sử dụng hơn. Điều này sẽ giúp cho phương pháp điều khiển bằng giọng nói trở nên phù hợp với nhiều loại robot và nhiệm vụ khác nhau.

1.3. Xác định bài toán điều khiển của đồ án.

Bài toán điều khiển của đồ án "Điều khiển mobile robot bằng giọng nói" là thiết kế và xây dựng hệ thống điều khiển robot có khả năng nhận dạng và xử lý các lệnh điều khiển từ giọng nói của người dùng. Hệ thống này phải đảm bảo các yêu cầu sau:

Nhận dạng và xử lý các lệnh điều khiển từ giọng nói của người dùng.

Có khả năng điều khiển robot thực hiện các lệnh điều khiển.

Để giải quyết bài toán này, hệ thống điều khiển sẽ bao gồm các thành phần sau:

Thiết bị nhận dạng giọng nói: Thiết bị này sẽ có nhiệm vụ thu thập và xử lý tín hiệu âm thanh từ người dùng.

Bộ xử lý ngôn ngữ tự nhiên: Bộ xử lý này sẽ có nhiệm vụ phân tích tín hiệu âm thanh từ thiết bị nhận dạng giọng nói và xác định các lệnh điều khiển.

Bộ điều khiển robot: Bộ điều khiển này sẽ có nhiệm vụ chuyển đổi các lệnh điều khiển từ bộ xử lý ngôn ngữ tự nhiên thành các tín hiệu điều khiển động cơ của robot.

Dưới đây là một số giải pháp có thể được sử dụng để giải quyết bài toán điều khiển của đồ án này:

Đối với thiết bị nhận dạng giọng nói: Có thể sử dụng các thiết bị nhận dạng giọng nói thương mại sẵn có hoặc tự xây dựng thiết bị nhận dạng giọng nói dựa trên các thuật toán nhận dạng giọng nói.

Đối với bộ xử lý ngôn ngữ tự nhiên: Có thể sử dụng các thư viện xử lý ngôn ngữ tự nhiên sẵn có hoặc tự xây dựng bộ xử lý ngôn ngữ tự nhiên dựa trên các thuật toán xử lý ngôn ngữ tự nhiên.

Đối với bộ điều khiển robot: Có thể sử dụng các vi điều khiển hoặc bộ vi xử lý để xây dựng bộ điều khiển robot.

Kết luận Chương 1

Chương **TỔNG QUAN VỀ ĐIỀU KHIỂN ROBOT** đã trình bày các khái niệm cơ bản về điều khiển robot, bao gồm:

Định nghĩa, mục đích và các yêu cầu của điều khiển robot.

Các thành phần của hệ thống điều khiển robot.

Các phương pháp điều khiển robot.

Chương này cũng đã giới thiệu một số ứng dụng của điều khiển robot trong thực tế.

Kết luận chung của chương là điều khiển robot là một lĩnh vực nghiên cứu quan trọng trong lĩnh vực tự động hóa. Việc nghiên cứu và phát triển các phương pháp điều khiển robot mới sẽ góp phần nâng cao hiệu quả và khả năng ứng dụng của robot trong các lĩnh vực khác nhau.

Dưới đây là một số điểm nổi bật của chương:

Chương đã cung cấp một cái nhìn tổng quan về điều khiển robot, từ định nghĩa, mục đích, các yêu cầu, các thành phần, các phương pháp đến các ứng dụng.

Chương đã giới thiệu một số phương pháp điều khiển robot phổ biến, bao gồm điều khiển tuyến tính, điều khiển phi tuyến, điều khiển học tập và điều khiển dựa trên trí tuệ nhân tạo.

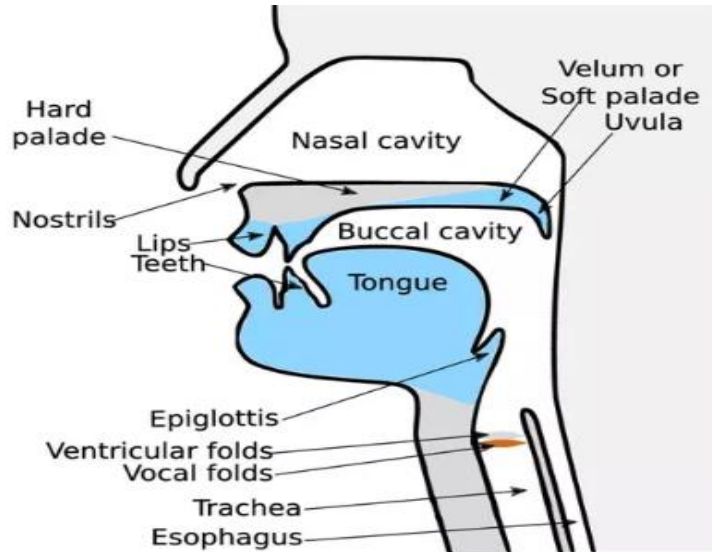
Chương đã đề cập đến các ứng dụng của điều khiển robot trong thực tế, bao gồm sản xuất, dịch vụ và giải trí.

Chương này là nền tảng quan trọng để tiếp tục nghiên cứu sâu hơn về các vấn đề liên quan đến điều khiển robot

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT VỀ ĐIỀU KHIỂN BẰNG GIỌNG NÓI

2.1. Kỹ thuật xử lý ngôn ngữ tự nhiên.

2.1.1. Nguyên lý hình thành tiếng nói.

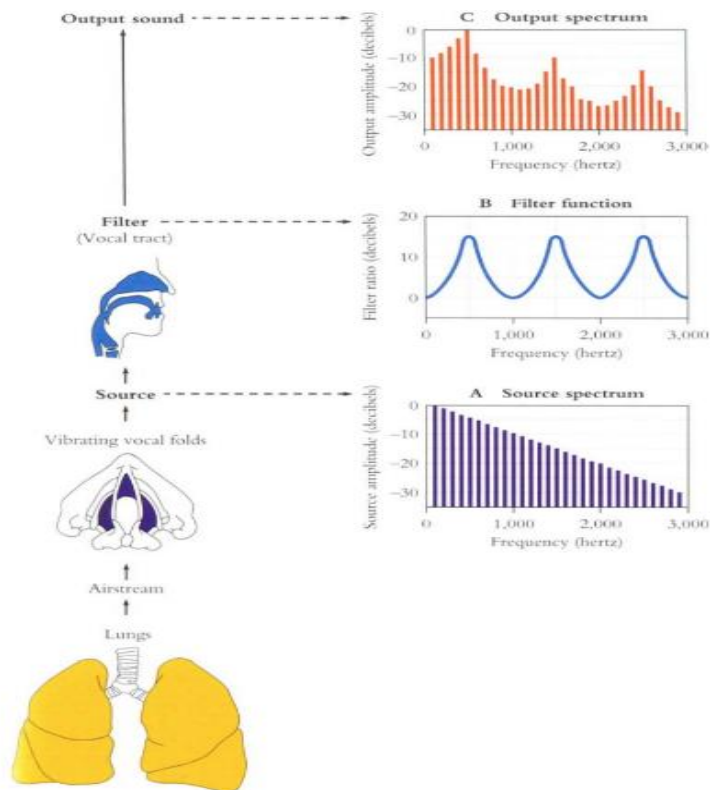


Hình 2.1. Cấu tạo miệng người.

Để tạo ra tiếng nói, luồng hơi được đẩy lên từ phổi tạo áp lực lên thanh quản. Dưới áp lực đó, các dây thanh âm mở ra, tạo ra một khoảng trống cho luồng hơi thoát qua. Khi luồng hơi thoát qua, áp lực giảm xuống khiến các dây thanh âm đóng lại. Việc đóng lại này lại khiến áp lực tăng lên và quá trình tái diễn. Các chu kỳ đóng/mở này liên tục tái diễn, tạo ra các tần số sóng âm với tần số cơ bản khoảng 125Hz với nam, 210Hz với nữ. Đó là lý do giọng của nữ giới thường có xu hướng cao hơn giọng nam. Tần số này gọi là tần số cơ bản (fundamental frequency, F0).

Tần số cơ bản là tần số cao nhất của các âm thanh được tạo ra bởi thanh quản. Tuy nhiên, để tạo ra tiếng nói, cần có thêm các âm thanh có tần số khác nhau. Các âm thanh này được tạo ra bởi các cơ quan khác trong hệ thống âm thanh, bao gồm vòm họng, khoang miệng, lưỡi, răng, môi, mũi.

Các cơ quan này hoạt động như một bộ cộng hưởng, giống như hộp đàn guitar. Bộ cộng hưởng này có tác dụng khuếch đại một số tần số và triệt tiêu một số tần số khác. Khả năng thay đổi hình dạng linh hoạt của các cơ quan này giúp tạo ra các âm thanh khác nhau, từ đó hình thành nên tiếng nói.



Hình 2.2. Cơ chế hình thành âm thanh.

Hình ảnh 2.2 mô tả cơ chế tạo ra âm thanh của tiếng nói. Theo cơ chế này, âm thanh của tiếng nói được tạo ra bởi sự cộng hưởng của các cơ quan trong hệ thống âm thanh. Các cơ quan này hoạt động như một bộ lọc, khuếch đại một số tần số và triệt tiêu một số tần số khác.

Tại phổ âm thanh của âm thanh phát ra, ta thấy có 3 đỉnh, lần lượt gọi là đỉnh F1, F2, F3. Các đỉnh này được gọi là formant. Giá trị, vị trí, sự thay đổi theo thời gian của các formant này đặc trưng cho các âm vị.

Trong các phương pháp nhận dạng tiếng nói truyền thống, người ta sẽ cố gắng tách thông tin về các formant ra khỏi tần số cơ bản (F0) rồi mới sử dụng thông tin này để nhận dạng. Âm vô thanh và hữu thanh (optional)

Trong quá trình tạo ra tiếng nói, có hai loại âm thanh là âm hữu thanh và âm vô thanh. Âm hữu thanh là những âm được tạo ra khi dây thanh âm rung động, còn âm vô thanh là những âm không được tạo ra khi dây thanh âm rung động.

Âm tiết

Âm tiết là đơn vị nhỏ nhất của ngôn ngữ có thể phát ra một cách độc lập. Một âm tiết có thể bao gồm một nguyên âm hoặc một nguyên âm kết hợp với phụ âm.

Trong tiếng Việt, âm tiết thường có một nguyên âm, có hoặc không có phụ âm đi kèm. Ví dụ, các từ "a", "bà", "cây", "chiếc" đều là một âm tiết.

Âm tiết có vai trò quan trọng trong ngôn ngữ, giúp phân biệt các từ và cấu trúc ngữ pháp. Ví dụ, các từ "bà" và "bá" chỉ khác nhau ở một âm tiết, nhưng có nghĩa hoàn toàn khác nhau.

Có thể phân loại âm tiết theo các tiêu chí sau:

Theo cấu tạo:

Âm tiết đơn: Là âm tiết chỉ bao gồm một nguyên âm.

Âm tiết kép: Là âm tiết bao gồm hai nguyên âm trở lên.

Theo vị trí của nguyên âm:

Âm tiết mở: Là âm tiết có nguyên âm ở cuối.

Âm tiết khép: Là âm tiết có nguyên âm ở đầu hoặc giữa âm tiết.

Theo vị trí của phụ âm:

Âm tiết bắt đầu bằng phụ âm: Là âm tiết có phụ âm ở đầu.

Âm tiết bắt đầu bằng nguyên âm: Là âm tiết có nguyên âm ở đầu.

Âm tiết kết thúc bằng phụ âm: Là âm tiết có phụ âm ở cuối.

Âm tiết kết thúc bằng nguyên âm: Là âm tiết có nguyên âm ở cuối.

Âm tiết là một đơn vị quan trọng của ngôn ngữ, cần được nghiên cứu và hiểu rõ để có thể sử dụng ngôn ngữ một cách chính xác và hiệu quả.

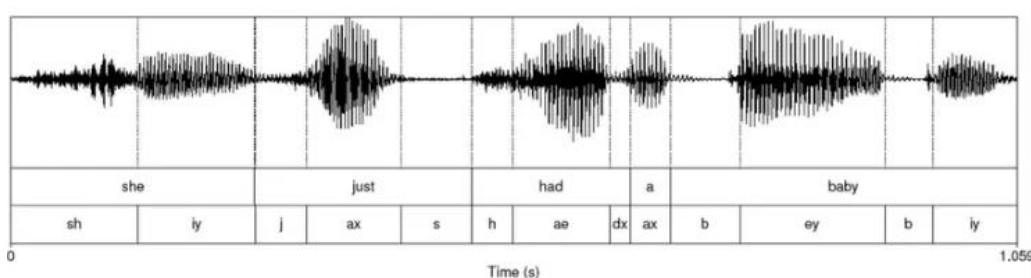
Âm vị và âm.

Vowels			
Phoneme	IPA Symbol	Graphemes	Examples
25	æ	a, ai, au	cat, plaid, laugh
26	eɪ	a, ai, eigh, aigh, ay, er, et, ei, au, a_e, ea, ey	bay, maid, weigh, straight, pay, foyer, filet, eight, gauge, mate, break, they
27	e	e, ea, u, ie, ai, a, eo, ei, ae	end, bread, bury, friend, said, many, leopard, heifer, aesthetic
28	i:	e, ee, ea, y, ey, oe, ie, i, ei, eo, ay	be, bee, meat, lady, key, phoenix, grief, ski, deceive, people, quay
29	ɪ	i, e, o, u, ui, y, ie	it, england, women, busy, guild, gym, sieve
30	aɪ	i, y, igh, ie, uy, ye, ai, is, eigh, i_e	spider, sky, night, pie, guy, sty, aisle, island, height, kite

Hình 2.3. Một số nguyên âm.

Âm vị: Tiếng anh là phoneme, trong nhiều loại ngôn ngữ, một kí tự/cụm kí tự (letter) trong các từ khác nhau có thể có nhiều cách phát âm khác nhau. Bảng chữ cái latin có 26 chữ cái nhưng có tới 44 phoneme.

Trong các bài toán Text to Speech (, người ta cần chuyển đổi từ dạng chữ viết sang dạng chuỗi các âm vị. Chữ tiếng Việt của chúng ta có tính tượng thanh cao hơn, có tính thống nhất cao giữa cách viết và cách đọc. Đó có thể là 1 thuận lợi của chúng ta khi làm việc với tiếng Việt.



Hình 2.4. Mức cường độ âm của một câu.

Âm: Tiếng anh là phone là sự hiện thực hoá âm vị. Cùng 1 phoneme nhưng mỗi người lại có 1 giọng đọc khác nhau, Ví dụ cùng từ "ba" nhưng giọng nam khác giọng nữ, giọng người A khác giọng người B. Để dễ phân biệt giữa "phoneme" và "phone" thì bạn có thể quan sát hình dưới đây. Hình ảnh mô tả câu "she just had a baby" được tách thành các âm vị (phoneme) ở hàng dưới và được hiện thực hoá thành các "phone" (hình ảnh các sóng âm thanh).

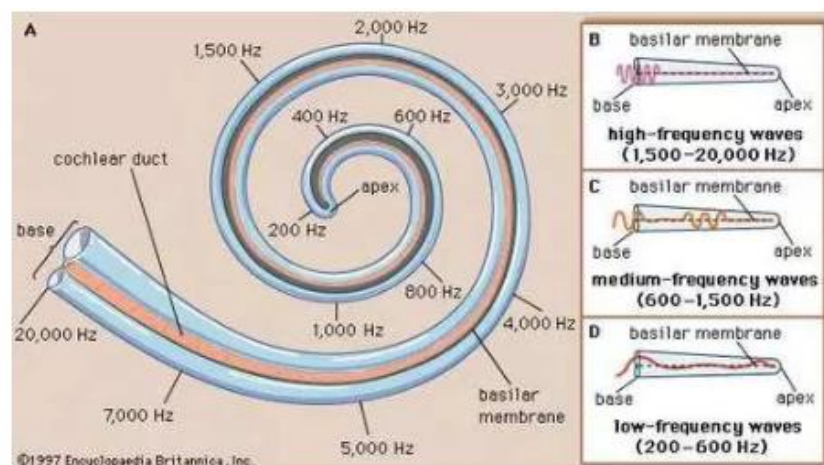
Trong lĩnh vực nhận dạng giọng nói, ta có tập dataset TIMIT - 1 tập các đoạn đọc được phiên âm và căn chỉnh (align) thời gian của 630 người Mỹ. Tập dữ liệu

được thu thập và annotate bởi các chuyên gia về ngữ âm học, từng âm được nghe và đánh dấu vị trí mở đầu và kết thúc rõ ràng.

2.1.2. Cơ chế hoạt động của tai.

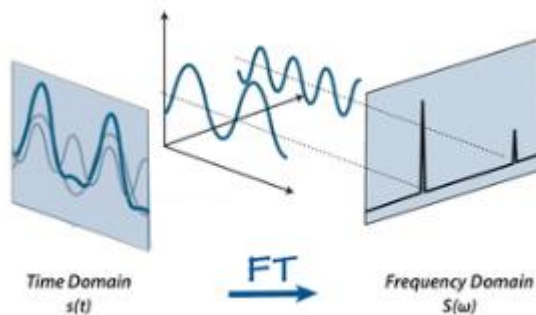
Âm thanh, tiếng nói mà chúng ta vẫn nghe hằng ngày là 1 pha trộn của rất nhiều sóng với các tần số khác nhau. Các tần số này thường nằm trong khoảng từ 20Hz -> 20000Hz. Tuy nhiên tai người (và các loài động vật) hoạt động phi tuyến tính, tức không phải rằng độ cảm nhận âm thanh 20000Hz sẽ gấp 1000 lần âm thanh 20Hz. Thường thì tai người rất nhạy cảm ở âm thanh tần số thấp, kém nhạy cảm ở tần số cao. Khi âm thanh truyền tới tai và đập vào màng nhĩ, màng nhĩ rung lên, truyền rung động lên 3 ba xương nhỏ: malleus, incus, stapes tới ốc tai. Ốc tai là 1 bộ phận dạng xoắn, rộng như 1 con ốc. Ốc tai chứa các dịch nhầy bên trong giúp truyền âm thanh, dọc theo ốc tai là các tế bào lông cảm nhận âm thanh. Các tế bào lông này rung lên khi có sóng truyền qua và gửi tín hiệu tới não bộ. Các tế bào ở đoạn đầu cứng hơn, rung động với các tần số cao. Càng sâu vào trong, các tế bào càng bớt cứng, đáp ứng các tần số thấp. Do cấu tạo ốc tai cùng số lượng các tế bào đáp ứng tần số thấp chiếm phần lớn khiến cho việc cảm nhận của tai người (và động vật) là phi tuyến tính, nhạy cảm ở tần số thấp, kém nhạy cảm ở tần số cao.

Trong xử lý tiếng nói, ta cần 1 cơ chế để map giữa tín hiệu âm thanh thu được bằng cảm biến và độ cảm nhận của tai người. Việc map này được thực hiện bởi Mel filterbank.



Hình 2.5. Tần số cơ bản tai có thể nghe.

2.1.3. Biến đổi Fourier.

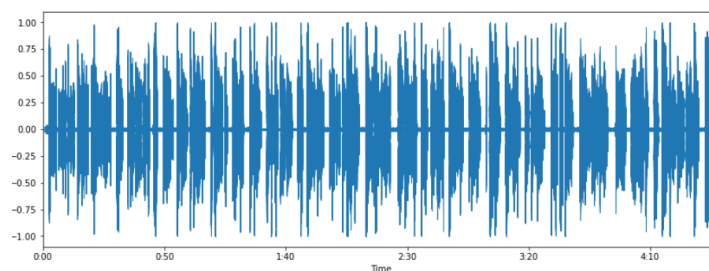


Hình 2.6. Biến đổi Fourier.

Âm thanh là một chuỗi tín hiệu thời gian, và hàm lượng thông tin trong đó thường không cao. Tuy nhiên, chúng ta có thể chuyển đổi chuỗi tín hiệu này sang miền tần số bằng cách sử dụng biến đổi Fourier.

Trong miền tần số, các thành phần tần số của tín hiệu được biểu diễn dưới dạng các hàm sin và cos. Điều này cho phép chúng ta lưu trữ thông tin âm thanh một cách hiệu quả hơn, vì chúng ta chỉ cần lưu trữ tần số, biên độ và pha của các thành phần tần số này.

Ví dụ, trong hình dưới đây, một đoạn âm thanh trong miền thời gian được kết hợp từ hai sóng tuần hoàn.



Hình 2.7. Đoạn âm thanh trong miền thời gian.

Hai sóng này có tính chất tuần hoàn, vì vậy chúng có thể được biểu diễn dưới dạng các hàm sin và cos. Thay vì phải lưu giá trị của các sóng này theo thời gian, chúng ta chỉ cần lưu lại tần số, biên độ và pha của chúng.

Điều này cho chúng ta một biểu diễn "tốt ưu" hơn cho đoạn âm thanh, vì nó chứa tất cả thông tin cần thiết để tái tạo đoạn âm thanh gốc.

Biến đổi Fourier là một công cụ quan trọng trong lĩnh vực xử lý tín hiệu. Nó được sử dụng trong nhiều ứng dụng khác nhau, bao gồm:

Nhận dạng âm thanh

Xử lý ảnh

Truyền thông

Lọc tín hiệu

Biến đổi Fourier:

Miền thời gian: Trong miền thời gian, các thành phần của tín hiệu được biểu diễn dưới dạng các giá trị theo thời gian. Đây là cách biểu diễn thông thường của tín hiệu âm thanh, nhưng nó không phải là cách biểu diễn hiệu quả nhất.

Miền tần số: Trong miền tần số, các thành phần của tín hiệu được biểu diễn dưới dạng các hàm sin và cos. Đây là cách biểu diễn hiệu quả hơn, vì nó cho phép chúng ta lưu trữ thông tin một cách gọn gàng hơn.

Biến đổi Fourier: Biến đổi Fourier là một phép toán toán học chuyển đổi một tín hiệu từ miền thời gian sang miền tần số.

Ứng dụng của biến đổi Fourier:

Biến đổi Fourier có ứng dụng rất lớn trong lĩnh vực xử lý tín hiệu. Một số ứng dụng cụ thể của biến đổi Fourier bao gồm:

Nhận dạng âm thanh: Biến đổi Fourier được sử dụng trong nhận dạng âm thanh để phân tách các thành phần tần số của âm thanh. Điều này giúp cho việc nhận dạng âm thanh trở nên dễ dàng hơn.

Xử lý ảnh: Biến đổi Fourier được sử dụng trong xử lý ảnh để phân tích các thành phần tần số của ảnh. Điều này giúp cho việc chỉnh sửa ảnh và phục hồi ảnh trở nên dễ dàng hơn.

Truyền thông: Biến đổi Fourier được sử dụng trong truyền thông để truyền tải dữ liệu hiệu quả hơn. Điều này giúp giảm nhiễu và tăng tốc độ truyền dữ liệu.

Lọc tín hiệu: Biến đổi Fourier được sử dụng trong lọc tín hiệu để loại bỏ các thành phần tần số không mong muốn. Điều này giúp cải thiện chất lượng của tín hiệu.

Công thức Fourier Transform cho hàm liên tục:

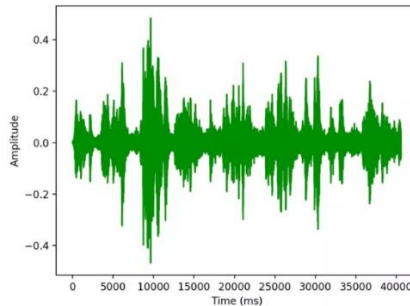
$$F(\omega) = \int f(x)e^{-2\pi i\omega} dx \quad (0.1)$$

2.1.4. Đặc trưng MFCCs.

Nguyên lý hoạt động.

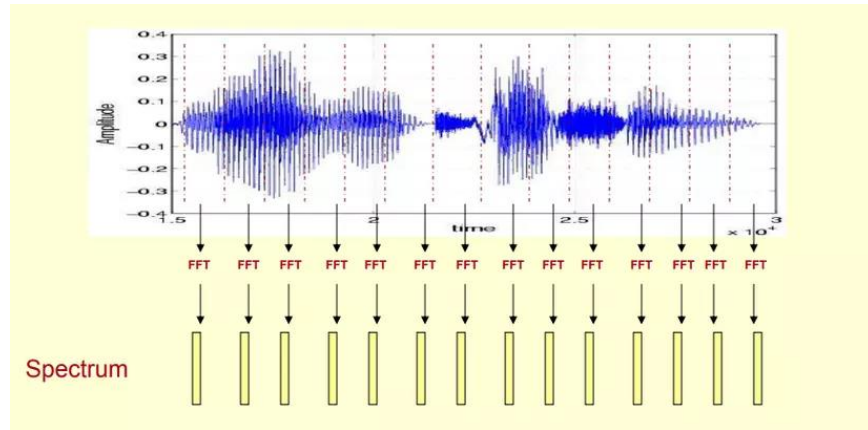
Âm thanh giọng nói là một tín hiệu âm thanh được tạo ra bởi các rung động của dây thanh âm trong thanh quản. Tín hiệu âm thanh có thể được biểu diễn dưới dạng hai chiều (x, y) , với x là thời gian đơn vị milliseconds (ms) và y là biên độ (amplitude).

Tín hiệu âm thanh được sinh ra trực tiếp từ bộ thu âm, do đó thường được gọi là tín hiệu âm thanh giọng nói (speech signal).



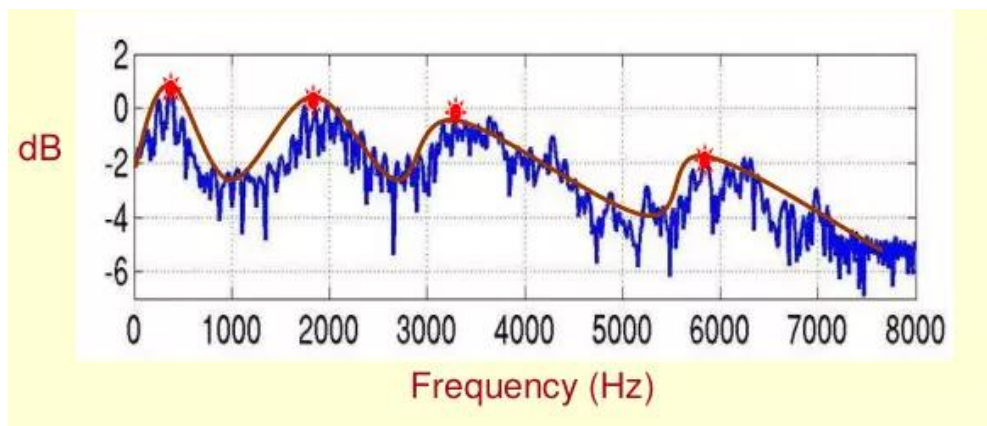
Hình 2.8 Mức năng lượng âm của một âm.

Để phân tích giọng nói, bước đầu tiên là chuyển đổi tín hiệu âm thanh giọng nói sang dạng phổ âm thanh (hay thường được gọi là spectrum) bằng cách sử dụng biến đổi Fourier nhanh (Fast Fourier Transform).



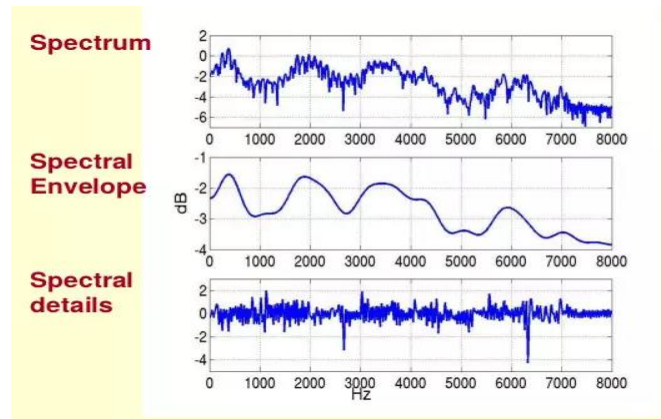
Hình 2.9. Biến đổi Fourier từ mức năng lượng âm.

Spectrum là kết quả của việc biến đổi và được biểu diễn dưới dạng hai chiều (x', y') với x' là tần số (Hz) và y' là cường độ (dB).



Hình 2.10. Phổ âm thanh.

Phổ âm thanh của một âm thanh giọng nói được biểu diễn dưới dạng hai chiều (f, s) , với f là tần số (Hz) và s là cường độ (dB). Các điểm màu đỏ trong phổ âm thanh là các formant (âm cộng hưởng), là nơi có các tần số áp đảo, mang đặc tính của âm thanh. Đường màu đỏ là spectral envelopes (đường bao phổ), là đường bao của các âm cộng hưởng. Mục tiêu chính của việc phân tích giọng nói là lấy được đường bao phổ này.

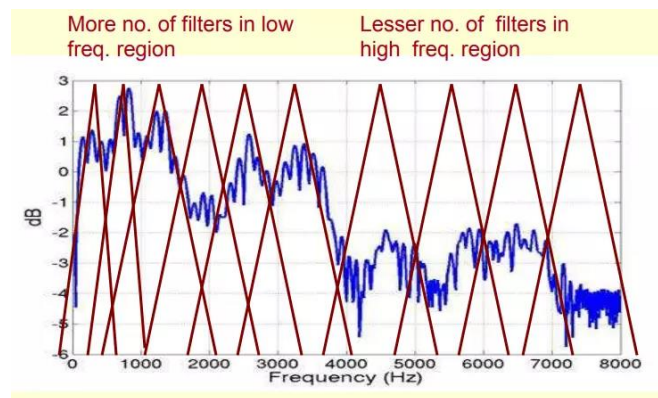


Hình 2.11. Lọc phổ âm thanh.

Để tách được $H[k]$, cần lấy logarithm của dạng phổ âm thanh và lấy phần ở tần số thấp (low frequency):

$$X[k] \leftrightarrow \log(X[k]) = H[k] * E[k] = \log(H[k] + \log(E[k])) \quad (0.2)$$

Tai người được hoạt động như một bộ lọc, chỉ tập trung vào một phần của đường bao phổ (spectrl envelopes) thay vì toàn bộ. Một bộ lọc tương tự tai người được gọi là bộ lọc tần số Mel (Mel-Frequency).



Hình 2.12. Cepstrum.

Sau khi áp dụng bộ lọc tần số Mel, ta sẽ sử dụng biến đổi Fourier ngược (Inverse Fast Fourier Transform) lên phổ âm thanh đã lấy logarithm:

$$IFFT(\log(X[k])) \leftrightarrow x[k] = IFFT(\log(H[k] + \log(E[k]))) = h[k] + e[k] \quad (0.3)$$

Trong đó, $x[k]$ được gọi là cepstrum vì biến đổi Fourier ngược (Inverse Fast Fourier Transform, IFFT) là nghịch đảo của biến đổi Fourier nhanh (Fast Fourier Transform, FFT), và cepstrum cũng là nghịch đảo của phổ âm thanh (spectrum).

Cepstrum bây giờ sẽ giống như tín hiệu âm thanh (speech signal), biểu diễn dưới dạng hai chiều (x'' , y'') (tần số, cường độ), nhưng giá trị sẽ khác nên người ta cũng gọi hai cột với tên khác là y'' là magnitude (không có đơn vị) và x'' là quefrequency (ms).

Và MFCCs cũng chính là các giá trị lấy từ Cepstrum này, thông thường người ta sẽ lấy 12 hệ số của y'' vì các hệ số còn lại không có nhiều tác dụng trong các hệ thống nhận diện âm thanh.

Tóm lại, trình tự lấy được thông tin từ âm thanh từ đặc trưng MFCCs sẽ là speech signal \rightarrow spectrum \rightarrow mel-freq filter \rightarrow cepstral.

2.1.5. Mô hình ẩn Markow.

Mô hình Markov ẩn (HMM) là một mô hình thống kê mô tả mối quan hệ giữa các trạng thái ẩn và các quan sát được. Trong các ứng dụng nhận dạng, trạng thái ẩn đại diện cho các âm vị hoặc từ trong một chuỗi âm thanh, còn quan sát được đại diện cho các mẫu âm thanh tương ứng.

Mô hình HMM có một số ưu điểm sau:

Khả năng biểu diễn đúng nhất sự liên tục của âm thanh theo thời gian: Mô hình HMM có thể mô tả sự chuyển đổi giữa các trạng thái ẩn theo thời gian, cho phép mô hình nắm bắt được các đặc điểm âm thanh dài hạn.

Khả năng mô hình hóa các hiện tượng đồng phát âm: Đồng phát âm là hiện tượng các âm vị lân cận ảnh hưởng lẫn nhau về mặt âm thanh. Mô hình HMM có thể mô hình hóa hiện tượng này bằng cách cho phép các xác suất phát xạ quan sát phụ thuộc vào các trạng thái ẩn trước đó.

Khả năng đưa ra kết quả nhận dạng bằng phương pháp thống kê: Điều này giúp xác suất tìm đúng từ là rất lớn và đáng tin cậy.

Tuy nhiên, mô hình HMM cũng có một số nhược điểm sau:

Giả thiết về bậc: Trong mô hình HMM, ta giả thiết rằng các xác suất chỉ phụ thuộc vào trạng thái hiện thời mà độc lập với quá khứ. Điều này không đúng trong

các ứng dụng tiếng nói, vì các âm vị thường bị ảnh hưởng bởi các âm vị trước đó. Hậu quả là mô hình HMM khó khăn trong mô hình hóa hiện tượng đồng phát âm.

Khả năng mở rộng: Mô hình HMM có thể trở nên phức tạp và khó huấn luyện khi số lượng trạng thái ẩn tăng lên.

Giải thích chi tiết hơn về các nhược điểm của mô hình HMM:

Giả thiết về bậc:

Giả sử ta có một chuỗi âm thanh gồm 3 âm vị, tương ứng với 3 trạng thái ẩn. Giả thiết về bậc trong mô hình HMM cho rằng xác suất của trạng thái ẩn thứ 2 chỉ phụ thuộc vào trạng thái ẩn thứ 1, và xác suất của trạng thái ẩn thứ 3 chỉ phụ thuộc vào trạng thái ẩn thứ 2. Điều này có nghĩa là xác suất của trạng thái ẩn thứ 2 không bị ảnh hưởng bởi trạng thái ẩn thứ 0, và xác suất của trạng thái ẩn thứ 3 không bị ảnh hưởng bởi trạng thái ẩn thứ 1.

Tuy nhiên, trong thực tế, các âm vị thường bị ảnh hưởng bởi các âm vị trước đó. Ví dụ, trong từ "bóng", âm vị "l" bị ảnh hưởng bởi âm vị "b". Do đó, giả thiết về bậc của mô hình HMM không đúng trong trường hợp này.

Khả năng mở rộng:

Mô hình HMM có thể trở nên phức tạp và khó huấn luyện khi số lượng trạng thái ẩn tăng lên. Điều này là do số lượng tham số của mô hình sẽ tăng theo cấp số nhân với số lượng trạng thái ẩn.

Ví dụ, nếu mô hình HMM có 10 trạng thái ẩn, thì mô hình sẽ có 100 tham số. Nếu mô hình HMM có 100 trạng thái ẩn, thì mô hình sẽ có 10.000 tham số.

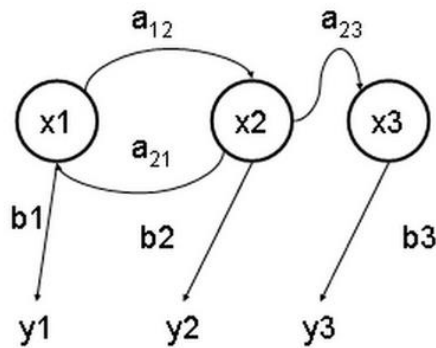
Để huấn luyện một mô hình HMM với nhiều trạng thái ẩn, cần có một lượng lớn dữ liệu huấn luyện. Nếu dữ liệu huấn luyện không đủ lớn, thì mô hình sẽ có thể bị quá khớp với dữ liệu huấn luyện, dẫn đến kết quả nhận dạng kém chính xác.

Để khắc phục các nhược điểm của mô hình HMM, có thể sử dụng các kỹ thuật sau:

Sử dụng các mô hình HMM nâng cao: Các mô hình HMM nâng cao có thể giải quyết được một số nhược điểm của mô hình HMM thông thường, chẳng hạn như mô hình HMM có bậc cao (HMM-GHMM) và mô hình HMM có độ dài trạng thái biến đổi (HMM-Viterbi).

Sử dụng các kỹ thuật học máy khác: Ngoài mô hình HMM, còn có nhiều kỹ thuật học máy khác có thể được sử dụng để nhận dạng tiếng nói, chẳng hạn như mạng nơ-ron nhân tạo (ANN) và học máy thống kê (SML).

Các chuyển tiếp trạng thái trong mô hình Markov ẩn:



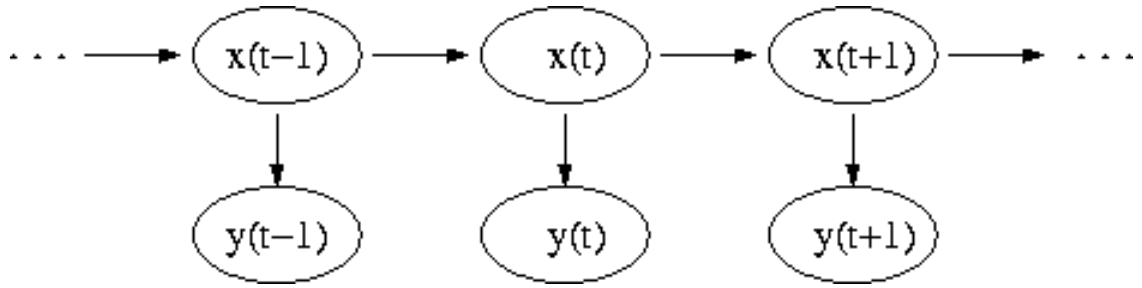
Hình 2.13. Chuyển tiếp trạng thái mô hình ẩn Markov.

Trong đó:

- x : Các trạng thái trong mô hình Markov
- a : Các xác suất chuyển tiếp
- b : Các xác suất đầu ra
- y : Các dữ liệu quan sát.

Sự tiến hóa của mô hình Markov.

Biểu đồ (Markov) dưới làm nổi bật các chuyển tiếp trạng thái của mô hình Markov ẩn. Nó cũng có ích để biểu diễn rõ ràng sự tiến hóa của mô hình theo thời gian, với các trạng thái tại các thời điểm khác nhau t_1 và t_2 được biểu diễn bằng các tham biến khác nhau, $x(t_1)$ và $x(t_2)$.



Hình 2.14. Biểu đồ Markov.

Trong biểu đồ này, nó được hiểu rằng thời gian chia cắt ra ($x(t)$, $y(t)$) mở rộng tới các thời gian trước và sau đó như một sự cần thiết. Thông thường lát cắt sớm nhất là thời gian $t = 0$ hay $t = 1$.

2.2. Mô hình điều khiển robot bằng giọng nói.

Mô hình điều khiển robot bằng giọng nói thường bao gồm các thành phần chính sau:

Thiết bị thu âm: Thiết bị này có nhiệm vụ thu âm các lệnh được phát ra bằng giọng nói. Thiết bị thu âm có thể là một microphone, một camera, hoặc một thiết bị thu âm chuyên dụng.

Hệ thống nhận dạng giọng nói: Hệ thống này có nhiệm vụ nhận dạng và phân tích các lệnh được phát ra bằng giọng nói. Hệ thống nhận dạng giọng nói thường sử dụng các thuật toán học máy để phân biệt các âm thanh khác nhau. Trong phạm vi bài toán này sẽ sử dụng trích xuất đặc trưng MFCCs và mô hình ẩn Markov để nhận diện được giọng nói.

Thiết bị kết nối không dây: là thiết bị truyền nhận dữ liệu từ máy tính đến robot khả năng sử dụng linh hoạt nên ta có thể dùng các thiết bị như bluetooth, wifi hoặc các thiết bị RF.

Bộ điều khiển robot: Bộ điều khiển này có nhiệm vụ chuyển các lệnh được nhận dạng thành các hành động để điều khiển robot. Bộ điều khiển robot có thể là một vi điều khiển, một máy tính, hoặc một hệ thống điều khiển chuyên dụng.

Cách thức hoạt động của mô hình điều khiển robot bằng giọng nói như sau:

Người dùng phát ra các lệnh bằng giọng nói.

Thiết bị thu âm thu âm các lệnh được phát ra bằng giọng nói.

Hệ thống nhận dạng giọng nói nhận dạng và phân tích các lệnh được phát ra bằng giọng nói.

Các lệnh được nhận dạng được chuyển thành các hành động để điều khiển robot.

Bộ điều khiển robot thực hiện các hành động đã được chuyển.

Ví dụ, nếu người dùng phát ra lệnh "Đi về phía trước", hệ thống nhận dạng giọng nói sẽ nhận dạng và phân tích lệnh này. Sau đó, hệ thống sẽ chuyển lệnh này thành một chuỗi số, ví dụ như "1, 1, 1, 1, 1". Chuỗi số này sẽ được chuyển đến bộ điều khiển robot. Bộ điều khiển robot sẽ sử dụng chuỗi số này để điều khiển robot di chuyển về phía trước.

Mô hình điều khiển robot bằng giọng nói có nhiều ưu điểm như:

Thuận tiện cho người dùng, không cần phải sử dụng các thiết bị điều khiển phức tạp.

Linh hoạt, có thể điều khiển robot từ xa.

Có thể sử dụng trong nhiều môi trường khác nhau.

Tuy nhiên, mô hình điều khiển robot bằng giọng nói cũng có một số hạn chế như:

Độ chính xác của hệ thống nhận dạng giọng nói phụ thuộc vào chất lượng của âm thanh thu được.

Mô hình có thể bị ảnh hưởng bởi tiếng ồn trong môi trường.

Mô hình có thể khó sử dụng trong môi trường có tiếng ồn lớn.

Kết luận chương 2

Chương CƠ SỞ LÝ THUYẾT VỀ ĐIỀU KHIỂN BẰNG GIỌNG NÓI tập trung vào việc biến đổi tín hiệu giọng nói, trong đó phương pháp đặc trưng MFCC (Mel-Frequency Cepstral Coefficients) được sử dụng để biểu diễn giọng nói dưới dạng các vector đặc trưng. Ngoài ra, chương cũng giới thiệu mô hình ẩn Markov, một phương pháp quan trọng trong việc nhận diện giọng nói và xử lý ngôn ngữ tự nhiên.

MFCC là một phương pháp biến đổi giọng nói thành một tập hữu hạn các đặc trưng có khả năng biểu diễn thông tin quan trọng về âm thanh. Điều này giúp giảm chiều dữ liệu và tạo ra biểu diễn tốt cho việc nhận diện và phân loại âm thanh. MFCC là kết quả của việc áp dụng các bước biến đổi trên tín hiệu âm thanh, bao gồm lọc Mel-Frequency, biến đổi Fourier, và lấy logarit tự nhiên, sau đó sử dụng các hệ số cepstral.

Mô hình ẩn Markov là một công cụ quan trọng trong nhận diện giọng nói và xử lý ngôn ngữ tự nhiên. Nó mô tả quá trình mà tín hiệu giọng nói chuyển đổi qua các trạng thái ẩn, mỗi trạng thái có một xác suất phát sinh dãy âm thanh cụ thể. Mô hình này cho phép mô phỏng và nhận diện các đơn vị ngôn ngữ như từng âm, từ, hoặc câu.

Kết luận, việc sử dụng đặc trưng MFCC và mô hình ẩn Markov là quan trọng trong việc biểu diễn và xử lý giọng nói. MFCC giúp biểu diễn tốt các thông tin quan trọng trong tín hiệu giọng nói, trong khi mô hình ẩn Markov cho phép mô phỏng và nhận diện các thành phần ngôn ngữ. Khi kết hợp, chúng làm cho việc điều khiển bằng giọng nói trở nên hiệu quả và đáng tin cậy hơn.

CHƯƠNG 3. MÔ HÌNH HÓA VÀ ĐIỀU KHIỂN ROBOT BẰNG GIỌNG NÓI

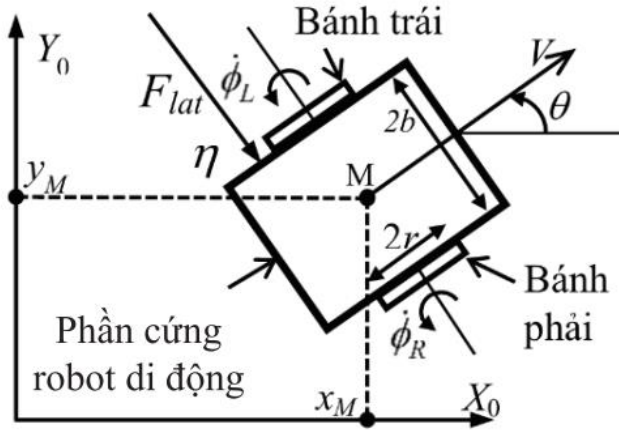
3.1. Mô hình toán học mobile robot.

3.1.1. Phân tích mô hình toán học.

Xét một robot di động bánh xe có trượt ngang, mô hình động học của xe được mô tả như sau:

$$\begin{aligned} r\dot{\phi}_R &= \dot{x}_M \cos \theta + \dot{y}_M \sin \theta + b\dot{\theta} \\ r\dot{\phi}_L &= \dot{x}_M \cos \theta + \dot{y}_M \sin \theta - b\dot{\theta} \\ \dot{\eta} &= -\dot{x}_M \sin \theta + \dot{y}_M \cos \theta \end{aligned} \quad (0.4)$$

Trong đó, η là độ trượt ngang của robot di động.



Hình 3.1. Robot di động bánh xe.

3.1.2. Hệ phương trình động lực học.

Động năng của thân robot di động:

$$K_M = \frac{1}{2} m_M (\dot{x}_M^2 + \dot{y}_M^2) + \frac{1}{2} I_M \dot{\theta}^2 \quad (0.5)$$

Trong đó, m_M là khối lượng của thân robot di động, I_M là mô men quán tính của thân này xung quanh trục thẳng đứng đi qua điểm M.

Động năng của bánh trái và bánh phải lần lượt là:

$$K_L = \frac{1}{2} m_w (r^2 \dot{\phi}_L^2 + \eta^2) + \frac{1}{2} I_w \dot{\phi}_L^2 + \frac{1}{2} I_D \dot{\theta}^2 \quad (3.3)$$

$$K_R = \frac{1}{2} m_w (r^2 \dot{\phi}_R^2 + \eta^2) + \frac{1}{2} I_w \dot{\phi}_R^2 + \frac{1}{2} I_D \dot{\theta}^2 \quad (0.6)$$

Tổng động năng của hệ là:

$$\begin{aligned} K &= K_M + K_L + K_R \\ &= \frac{1}{2} m_M (x_M^2 + y_M^2) + \frac{1}{2} m_w r^2 (\dot{\phi}_L^2 + \dot{\phi}_R^2) \\ &\quad + m_w \eta^2 + \frac{1}{2} I_w (\dot{\phi}_L^2 + \dot{\phi}_R^2) + \left(\frac{1}{2} I_M + I_D\right) \dot{\theta}^2 \end{aligned} \quad (0.7)$$

Trong đó, I_w và I_D lần lượt là mô men quán tính của bánh xe xung quanh trục quay và trục thẳng đứng.

Vì thế năng lượng của robot di động bằng 0, nên hàm Lagrange của nó là $L = K$.

Gọi véc tơ tọa độ Lagrange của robot di động là $q = [x_M, y_M, \theta, \eta, \phi_R, \phi_L]^T$, phương trình ràng buộc được biểu diễn theo dạng:

$$A(q)\dot{q} = 0 \quad (0.8)$$

$$A(q) = \begin{bmatrix} \cos \theta & \sin \theta & -b & 0 & -r & 0 \\ \cos \theta & \sin \theta & b & 0 & 0 & -r \\ -\sin \theta & \cos \theta & 0 & 1 & 0 & 0 \end{bmatrix} \quad (0.9)$$

Phương trình Lagrange của chuyển động của robot di động là:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) - \frac{\partial L}{\partial q} = u + A^T \lambda \quad (0.10)$$

Trong đó, $\lambda = [\lambda_1, \lambda_2, \lambda_3]^T$ là một véc tơ nhân tử Lagrange biểu diễn các lực ràng buộc của robot di động, u là véc tơ lực suy rộng tương ứng với các tọa độ suy rộng q . Bằng cách giải phương trình Lagrange, phương trình động lực học robot di động có dạng như sau:

$$M\ddot{q} = N_1 \tau + N_2 F_{lat} + A^T \lambda \quad (0.11)$$

Với $N_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$, $N_2 = [0 \ 0 \ 0 \ 1 \ 0 \ 0]^T$ là các ma trận đầu vào,

$$M = \begin{bmatrix} m_M & 0 & 0 & 0 & 0 & 0 \\ 0 & m_M & 0 & 0 & 0 & 0 \\ 0 & 0 & I_M + 2I_D & 0 & 0 & 0 \\ 0 & 0 & 0 & 2m_W & 0 & 0 \\ 0 & 0 & 0 & 0 & m_W r^2 + I_W & 0 \\ 0 & 0 & 0 & 0 & 0 & m_W r^2 + I_W \end{bmatrix} \quad (0.12)$$

$\tau = [\tau_R, \tau_L]^T$ là véc tơ đầu vào gồm mô men quay bánh phải, bánh trái, F_{lat} là lực đẩy tác động vào thân robot theo hướng ngang.

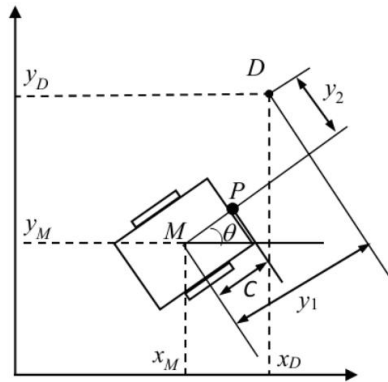
Gọi $v = [\dot{\phi}_R, \dot{\phi}_L]^T$, $S_1(q)$ và $S_2(q)$ là ma trận thỏa mãn phương trình sau:

$$\dot{q} = S_1(q)v + S_2(q)\dot{\eta} \quad (0.13)$$

$$\ddot{q} = \dot{S}_1(q)v + S_1(q)\dot{v} + \dot{S}_2(q)\dot{\eta} + S_2(q)\ddot{\eta} \quad (0.14)$$

$$\begin{aligned} \rightarrow \tau &= [S_1(q)^T M S_1(q)]\dot{v} + [S_1(q)^T M \dot{S}_1(q)]v \\ &+ [S_1(q)^T M \dot{S}_2(q)]\dot{\eta} + [S_1(q)^T M S_2(q)]\ddot{\eta} \\ \leftrightarrow m\dot{v} + b\dot{\eta}\omega &= \tau \end{aligned} \quad (0.15)$$

Trong đó, $m = S_1(q)^T M S_1(q) = \begin{bmatrix} \bar{m}_{11} & \bar{m}_{12} \\ \bar{m}_{21} & \bar{m}_{22} \end{bmatrix}$



Hình 3.2. Các biến đầu ra $y(x)$.

$$\bar{m}_{11} = \bar{m}_{22} = m_M \left(\frac{r}{2} \right)^2 + (I_M + 2I_D) \left(\frac{r}{2b} \right)^T + (m_W r^2 + I_W) \quad (0.16)$$

$$\bar{m}_{12} = \bar{m}_{21} = m_M \left(\frac{r}{2} \right)^2 - (I_M + 2I_D) \left(\frac{r}{2b} \right)^T \quad (0.17)$$

$$\begin{aligned} b\omega &= S_1(q)^T M S_2(q) \\ b &= \begin{bmatrix} -m_M \frac{r}{2} & -m_M \frac{r}{2} \end{bmatrix}^T \\ \omega &= \frac{r}{2b} (\dot{\phi}_R - \dot{\phi}_L) \end{aligned} \quad (0.18)$$

3.2. Xây dựng mô hình điều khiển bằng giọng nói.

3.2.1. Xử lý dữ liệu giọng nói.

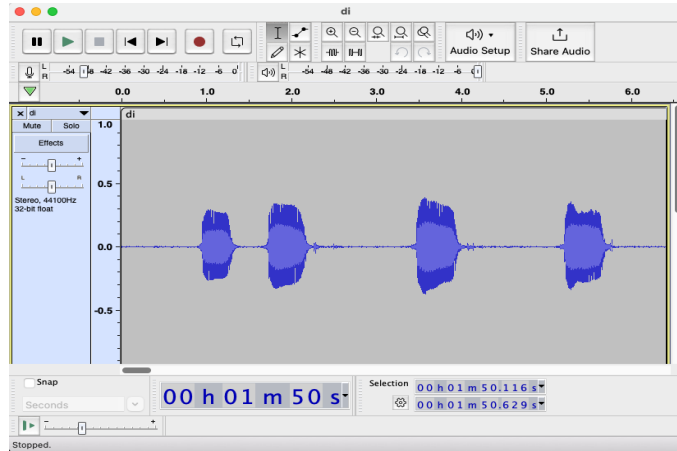
a. Thu thập dữ liệu.

Ghi âm bằng phần mềm Audacity với các âm thanh đầu vào có thông số như sau:

Định dạng mẫu của âm thanh là âm thanh 16 bit.

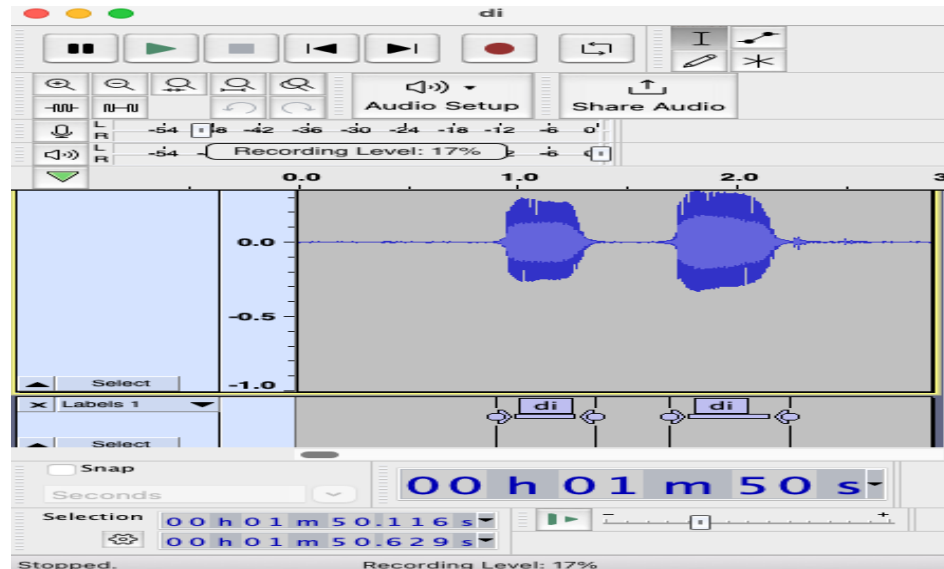
Số kênh âm thanh : 2 (Stereo).

Tần số lấy mẫu, số mẫu âm thanh được lấy mỗi giây : 22050 Hz.



Hình 3.3. Âm thanh được ghi bằng Audacity.

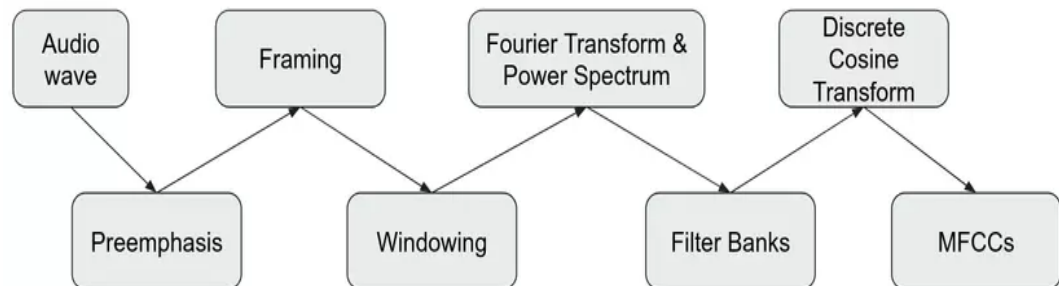
Gán nhãn cho từng âm thanh.



Hình 3.4. Gán nhãn âm thanh.

Dữ liệu sẽ bao gồm 7 label bao gồm: ‘di’, ‘thang’, ‘re’, ‘phai’, ‘lui’, ‘sil’.

b. Trích xuất đặc trưng MFCCs.



Hình 3.5. Quá trình trích xuất đặc trưng MFCCs.

Preemphasis.

Ta sẽ áp dụng công thức sau lên speech signal:

$$y(t) = x(t) - ax(t-1) \quad (3.17)$$

Có nhiều lý do để áp dụng preemphasis như:

Tránh vấn đề về số khi áp dụng FFT

Làm cân bằng tần số spectrum

Khuếch đại tần số cao (để lọc tần số thấp dễ hơn)

Nhưng mà preemphasis không bắt buộc sử dụng nữa vì FFT đã được cải thiện.

Framing.

Speech signal ở dạng liên tục theo từng ms, do đó khó để giải quyết nên người ta sẽ chia speech signal thành các frames.

Mỗi frame có kích thước khoảng 20-40 ms và chồng lên nhau (tức là từ đầu frame sau tới cuối frame trước) khoảng 10-15 ms.

Kết quả sẽ ở dưới dạng hai chiều (x, y) với x là độ dài frames và y là số lượng frames.

Windowing.

Một câu nói được phát ra là một chuỗi các âm vị. Do đó, tín hiệu nói là biến thời gian. Để trích xuất thông tin từ một tín hiệu, chúng ta cần phải chia tín hiệu thành các đoạn ngắn đủ, để, theo cách hiểu theo đúng, mỗi đoạn chứa chỉ một âm vị. Nói cách khác, chúng ta muốn trích xuất các đoạn ngắn đủ ngắn sao cho các đặc tính của tín hiệu nói không có sự thay đổi thời gian trong đoạn đó.

Windowing là một phương pháp cổ điển trong xử lý tín hiệu và nó đề cập đến việc chia tín hiệu đầu vào thành các đoạn thời gian. Biên của các đoạn sau đó trở nên rõ ràng dưới dạng những sự không liên tục, không phù hợp với tín hiệu thế giới thực. Để giảm thiểu ảnh hưởng của việc chia tín hiệu thành các đoạn đối với tính chất thống kê của tín hiệu, chúng ta áp dụng windowing cho các đoạn thời gian. Các hàm windowing là các hàm mượt, giảm dần đến không tại biên. Bằng cách nhân tín hiệu đầu vào với một hàm window, hàm windowing cũng giảm dần về không tại biên sao cho sự không liên tục tại biên trở nên không thể nhìn thấy. Windowing thay đổi tín hiệu, nhưng thay đổi được thiết kế để giảm thiểu ảnh hưởng của nó đối với thống kê của tín hiệu.

Do framing làm rời rạc hóa speech signal ta sẽ áp dụng một hàm gọi là Hamming Window để làm smooth các frames:

$$[n] = 0.54 - 0.46 \cos\left(N - \frac{12\pi n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad (3.18)$$

Trong đó N là độ dài frame.

Fourier Transform and Power Spectrum

Đây là bước ta chuyển speech signal thành spectrum, ta sẽ áp dụng công thức sau:

$$P = \frac{|FFT(x_i)|^2}{N_{FFT}} \quad (3.19)$$

Trong đó NFFT bằng 256 hoặc 512, x_i là frame thứ i của speech signal x .

Filter Banks.

Đây là bước ta áp dụng bộ lọc Mel-Frequency Filter.

Các phương trình sau dùng để chuyển giữa Hert (f) và Mel (m):

$$\begin{aligned} m &= 2595 \log_{10} \left(1 + \frac{f}{700} \right) \\ f &= 700 \left(10^{\frac{m}{2595}} - 1 \right) \end{aligned} \quad (3.20)$$

Discrete Cosine Transform and MFCCs

Đây là bước ta chuyển từ spectrum qua cepstrum, áp dụng DCT (1 dạng IFFT) lên kết quả của filter banks ta sẽ có được các MFCCs, sau đó lấy 12 hệ số.

c. Delta và delta-delta

Trong các tác vụ nhận dạng, chẳng hạn như nhận dạng âm vị hoặc phát hiện hoạt động giọng nói, một tính năng đầu vào cổ điển là hệ số cepstral tần số mel (MFCC). Chúng mô tả hình dạng đường bao quang phổ tức thời của tín hiệu giọng nói. Tuy nhiên, tín hiệu tiếng nói là tín hiệu thay đổi theo thời gian và có dòng không đổi. Mặc dù chúng ta mô tả lời nói trong ngôn ngữ học như những chuỗi âm vị được nối với nhau, nhưng tín hiệu âm thanh được mô tả chính xác hơn là một chuỗi chuyển đổi giữa các âm vị.

Quan sát tương tự cũng áp dụng cho các đặc điểm khác của giọng nói như tần số cơ bản (F0), mô tả giá trị tức thời. Tuy nhiên, việc phân tích hình dạng tổng thể của đường F0 thường mang lại nhiều thông tin hơn là giá trị tuyệt đối. Ví dụ, sự nhấn mạnh trong một câu thường được mã hóa với độ

tương phản cao-thấp rõ rệt trong F0 và các câu hỏi trong nhiều ngôn ngữ có đường viền F0 thấp-cao đặc trưng.

Một phương pháp phổ biến để trích xuất thông tin về những chuyển đổi như vậy là xác định sự khác biệt đầu tiên của các đặc điểm tín hiệu, được gọi là delta của một đặc điểm. Cụ thể, đối với một đặc trưng f_k , tại thời điểm k , delta tương ứng được xác định là: $\Delta_k = f_k - f_{k-1}$

Sự khác biệt thứ hai, gọi là delta-delta, tương ứng là $\Delta\Delta_k = \Delta_k - \Delta_{k-1}$

Các ký hiệu ngắn gọn phổ biến cho delta và delta-delta lần lượt là các đặc điểm Δ và $\Delta\Delta$. Sau đó, các tính năng trong công cụ nhận dạng thường được thêm vào các tính năng Δ và $\Delta\Delta$ của chúng để tăng gấp ba số lượng đặc trưng với chi phí tính toán rất nhỏ.

Một quan sát/giải thích tầm thường về các đặc điểm delta và delta-delta là chúng gần đúng với đạo hàm bậc nhất và bậc hai của tín hiệu. Theo ước tính của đạo hàm, chúng không đặc biệt chính xác, nhưng sự đơn giản của chúng có thể bù đắp cho điều đó. Vấn đề về độ chính xác là các bộ phân biệt có xu hướng khuếch đại nhiễu trắng, trong khi tín hiệu mong muốn vẫn không thay đổi. Do đó, đầu ra nhiều hơn tín hiệu ban đầu. Sự khác biệt được áp dụng hai lần trong tính năng delta-delta sao cho các vấn đề về nhiễu cũng được tích lũy.

Các đặc delta là các phép biến đổi tuyến tính của các đặc trưng đầu vào, sao cho nếu chúng được kết hợp với một lớp tuyến tính trong mạng thần kinh tiếp theo, hai lớp tuyến tính liên tiếp dư thừa. Tuy nhiên, việc sử dụng các tính năng delta vẫn có thể năng lại lợi ích trong việc hội tụ, Trong mọi trường hợp, các tính năng delta và delta-delta là một phần cổ điển của các thuật toán học máy. Chúng thành công vì tính toán đơn giản và thường mang lại lợi ích rõ ràng so với các tính năng tức thời.

d. Mô hình Markow nhận dạng giọng nói.

Huấn luyện mô hình Markov ẩn.

Bài toán: Với dãy huấn luyện O cần hiệu chỉnh các tham số của mô hình λ để cực đại hóa $P(O / \lambda)$. Ta có:

$$P(O, Q / \lambda) = \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) a_{q_2 q_3} \dots a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (3.21)$$

$$P(Q / \lambda) = \sum_Q P(O, Q / \lambda) = \sum_Q \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) a_{q_2 q_3} \dots a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (3.22)$$

Đặt $\alpha_t(i) = P(O_1, O_2, \dots, O_t, q_t = i / \lambda)$ và $\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T / q_t = i, \lambda)$, $1 \leq t \leq T$ với giá trị khởi tạo $\alpha_1(i) = \pi_i b_i(O_1)$ và $\beta_T(i) = 1$, $1 \leq i \leq N$

Định nghĩa công thức truy hồi $\alpha_{t+1}(j)$:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N a_t(i) a_{ij} \right] b_j(O_{t+1}) \text{ với } t = 1, 2, \dots, T-1 \quad (3.23)$$

Tương tự, định nghĩa công thức $\beta_t(i)$ cho tính toán ngược như sau:

$$\beta_t(t) = \left[\sum_{j=1}^N a_{ij} b_j(O_{t+1}) \right] \text{ với } t = T-1, T-2, \dots, 1 \quad (3.24)$$

Thuật toán tiền lùi Baum-Welch (Forward-Backward Baum-Welch Algorithm):

Bước 1. Xác định:

$$\gamma_t(i) = P(q_t = i / O, \lambda) = \frac{P(q_t = i, O | \lambda)}{P(O | \lambda)} = \frac{\alpha_t(i) \beta_t(i)}{P(O | \lambda)} \quad (3.25)$$

Bước 2: Xác định:

$$\begin{aligned} \xi_t(i, j) &= P(q_t = i, q_{t+1} = j / O, \lambda) \\ &= \frac{P(q_t = i, q_{t+1} = j, O | \lambda)}{P(O | \lambda)} = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)} \end{aligned} \quad (3.26)$$

Bước 3: Chỉnh tham số:

$$\begin{aligned}\bar{\pi} = \gamma_1(i); \bar{a}_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^T \gamma_t(i)}, \bar{b}_j(v_k) \\ &= \frac{\sum_{t=1, O_t=v_k}^T \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)}\end{aligned}\tag{3.27}$$

Bước 4: Nếu $P(O / \lambda_{new}) \leq P(O / \lambda_{old})$ thì kết thúc. Quay lại bước 1.

Nhận diện mô hình Markov ẩn.

Bài toán: Cho mô hình $\lambda = (A, B, \pi)$ và một dãy quan sát $O = (O_1, O_2, \dots, O_T)$. Cần tìm dãy trạng thái $Q = (q_1, q_2, \dots, q_T)$ để xác suất cực đại hóa $P(O, Q / \lambda)$.

Thuật toán Viterbi:

Bước 1. Gọi:

$$f(k, j) = \max_{\{q_t\}_{t=1}^k, q_k=j} P(O_1, O_2, \dots, O_k, q_1, q_2, \dots, q_k | \lambda)\tag{3.28}$$

Bước 2. Khởi tạo cơ sở quy hoạch động: $f(1, j) = \pi_j b_j(O_1)$.

Bước 3. Tính toán phương án f bằng công thức truy hồi:

$$f(k, j) = \max_{1 \leq i \leq N} (f(k-1, i) \cdot a_{ij} \cdot b_j(O_k))\tag{3.29}$$

Lưu vết:

$$Trace(k, j) = \arg \max_{1 \leq i \leq N} (f(k-1, i) \cdot a_{ij} \cdot b_j(O_k)), (k \geq 2)$$

Bước 4. Tìm dãy trạng thái tối ưu: $q_T = \arg \max_j f(T, j)$.

$$q_t = Trace(t+1, q_{t+1}), t = T-1, T-2, \dots, 1.\tag{3.30}$$

3.2.2. Chuyển đổi tín hiệu giọng nói thành tín hiệu điều khiển.

Giọng nói đầu vào sẽ được thu bằng micro của máy tính sau đó mô hình HMM đã được huấn luyện để nhận dạng. Đầu ra của mô hình sẽ là đoạn văn bản nhận dự đoán.

```

class CustomTokenizer:
    4 usages
    def tokenize(self, sentence, custom_phrases=None):
        tokenized_sentence = []
        if custom_phrases is None:
            custom_phrases = {"đi thang": "F", "re phai": "R", "re trai": "L", "đi lui": "B"}
        while sentence:
            found_custom_phrase = False
            for custom_phrase, replacement in custom_phrases.items():
                if sentence.startswith(custom_phrase):
                    tokenized_sentence.append(replacement)
                    sentence = sentence[len(custom_phrase):].strip()
                    found_custom_phrase = True
                    break
            if not found_custom_phrase:
                tokens = sentence.split(maxsplit=1)
                tokenized_sentence.append(tokens[0])
                if len(tokens) > 1:
                    sentence = tokens[1].strip()
                else:
                    break
            break
        return tokenized_sentence

```

Hình 3.6. Class chuyển đổi các nhãn dán dự đoán.

Mỗi kí tự điều khiển đều được lập trình để điều khiển động cơ PWM DC. PWM hay thay đổi độ rộng xung là một kỹ thuật cho phép chúng ta điều chỉnh giá trị trung bình của điện áp đến thiết bị điện tử bằng cách bật và tắt nguồn với tốc độ nhanh. Điện áp trung bình phụ thuộc vào chu kỳ xung, hoặc lượng thời gian tín hiệu BẬT so với lượng thời gian tín hiệu TẮT trong một khoảng thời gian quy định.

Kết luận chương 3

Trong chương 3, chúng tôi đã trình bày các mô hình toán học liên quan đến điều khiển robot di động, với sự tập trung đặc biệt vào việc xây dựng mô hình điều khiển bằng giọng nói. Để thực hiện điều này, chúng tôi sử dụng mô hình ẩn Markov chủ yếu, một lý thuyết đã được chứng minh hiệu quả trong việc xử lý và hiểu các tín hiệu giọng nói.

Chương 3 có ý nghĩa quan trọng không chỉ là bước đi từ lý thuyết sang thực tế mà còn đóng vai trò là nền tảng cho chương kế tiếp của đề án. Trong chương tiếp theo, chúng tôi sẽ thảo luận về kết quả chi tiết hơn và đề xuất hướng phát triển tiếp theo cho hệ thống điều khiển robot thông qua giọng nói, dựa trên cơ sở lý thuyết và mô hình đã được xây dựng trong chương này.

CHƯƠNG 4. THỰC NGHIỆM ĐIỀU KHIỂN ROBOT BẰNG GIỌNG NÓI

4.1. Mô hình thực nghiệm.

a. Arduino Uno

Arduino Uno là một board mạch vi điều khiển được phát triển bởi Arduino.cc, một nền tảng điện tử mã nguồn mở chủ yếu dựa trên vi điều khiển AVR Atmega328P. Với Arduino chúng ta có thể xây dựng các ứng dụng điện tử tương tác với nhau thông qua phần mềm và phần cứng hỗ trợ.



Hình 4.1. Mạch Arduino Uno R3.

Arduino UNO có thể được cấp nguồn 5V thông qua cổng USB hoặc cấp nguồn ngoài với điện áp khuyến dùng là 7-12V DC và giới hạn là 6-20V. Thường thì cấp nguồn bằng pin vuông 9V là hợp lí nhất nếu không có sẵn nguồn từ cổng USB. Nếu cấp nguồn vượt quá ngưỡng giới hạn trên, bạn sẽ làm hỏng Arduino UNO.

b. Động cơ DC.

Động cơ được sử dụng cho mô hình của đồ án này là động cơ DC. Động cơ này thường sử dụng làm các mô hình thí nghiệm hay bài tập lớn phục vụ cho học sinh, sinh viên.



Hình 4.2. Động cơ DC 5V.

Một vài thông số kỹ thuật:

Điện áp: 3 – 12VDC.

Số vòng quay: 125 vòng/phút tại điện áp 3VDC, 208 vòng/phút tại điện áp 5 VDC.

Momen xoắn: 800gfc.m.

Tỉ số truyền: 1:48.

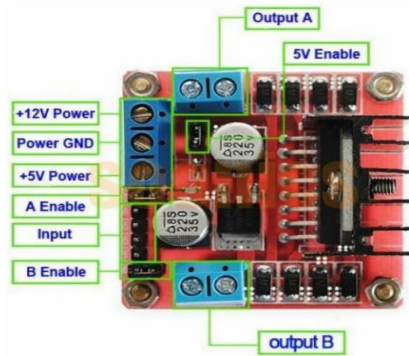
Kích thước: 64x19x22.6 (mm) (LxWxH).

Trọng lượng: 27g.

Động cơ DC 5V giảm tốc là một loại động cơ điện được thiết kế để có tốc độ quay chậm hơn so với tốc độ quay của động cơ DC thông thường. Nó thường được sử dụng trong các ứng dụng cần có lực xoắn lớn và tốc độ 30 quay chậm như trong các máy móc tự động hóa, máy in, máy quay phim hoặc trong các sản phẩm điện tử như robot, đồ chơi điện tử, ... Động cơ DC giảm tốc thường có một bộ giảm tốc tích hợp bên trong, giúp giảm tốc độ quay của trục động cơ và tăng lực xoắn. Bộ giảm tốc bao gồm các bánh răng và trục kết nối với động cơ và trục đầu ra. Khi động cơ quay, bánh răng sẽ xoay trục đầu ra, giúp cung cấp lực xoắn và động lực cho các ứng dụng cần sử dụng. Động cơ DC giảm tốc thường có nhiều loại kích thước và công suất khác nhau để phù hợp với các ứng dụng khác nhau. Nó được cấp nguồn bằng điện áp DC 5V thông qua các đầu cắm điện, đồng thời cũng có thể được kết nối với các mạch điều khiển để điều chỉnh tốc độ quay và hướng quay của trục động cơ

c. Mạch cầu L298N.

Mạch cầu L298N loại 1 có sẵn ốc gắn sử dụng IC điều khiển L298N có thể điều khiển hai động cơ một chiều hoặc 1 động cơ bước 4 pha. Có gắn tản nhiệt chống nóng cho IC, giúp IC có thể điều khiển với dòng định đạt 2A.



Hình 4.3. Mạch điều khiển động cơ L298N.

Mạch L298N gồm các chân:

12V power, 5V power. Đây là 2 chân cấp nguồn trực tiếp đến động cơ. Có thể cấp nguồn 9-12V ở 12V.

Power GND chân này là GND của nguồn cấp cho động cơ.

2 Jump A enable và B enable.

Gồm có 4 chân Input. IN1, IN2, IN3, IN4.

Thông số kỹ thuật:

Driver: L298N tích hợp hai mạch cầu H.

Điện áp điều khiển: +5 V ~ +12 V.

Dòng tối đa cho mỗi cầu H là: 2A (=> 2A cho mỗi motor).

Điện áp của tín hiệu điều khiển: +5 V ~ +7 V.

Dòng của tín hiệu điều khiển: 0 ~ 36mA.

Công suất hao phí: 20W (khi nhiệt độ T = 75 °C).

d. Module Bluetooth HC-05, HC-06.

Thông số kỹ thuật:

Điện áp hoạt động : +3.3VDC 30mA(hỗ trợ IC 5.0V)

Dòng điện khi hoạt động : Khi Pairing 30mA , sau khi pairing hoạt động truyền nhận bình thường 8mA.

Baudrate : 1200 ,2400 ,4800 ,9600(mặc định) ,19200 , 38400, 57600, 11520.

Dải tần hoạt động : 2.4GHz.

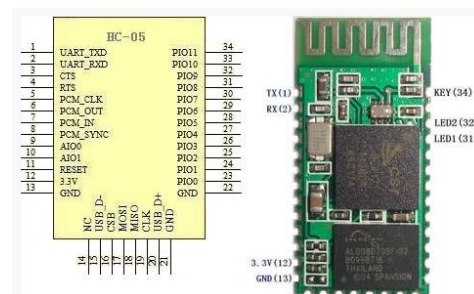
Kích thước : 26.9mm x 13mm x 2.2mm.

Giao tiếp : Bluetooth serial port.

Nhiệt độ làm việc : -20°C ~ +75°C.

Tốc độ : - Asynchronous : 2.1Mbps(Max)/160kbps.

Synchronous : 1Mbps/1Mbps.

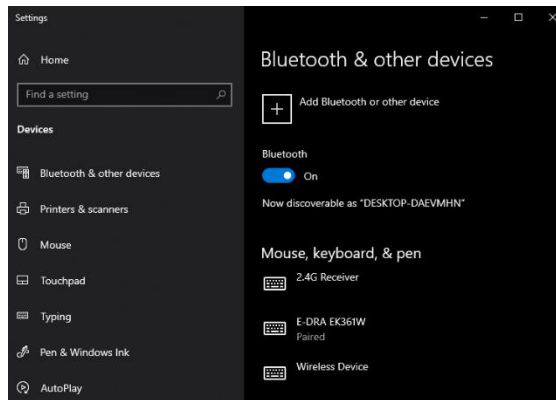


Hình 4.4. Module bluetooth HC-05.

- e. Kết nối không dây giữa module bluetooth và máy tính.

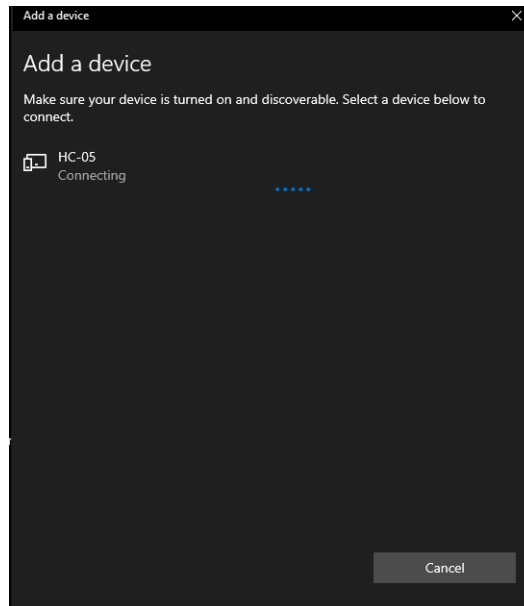
Window: Hệ điều hành windows 10

Ấn Window -> Gỡ Bluetooth & other devices.



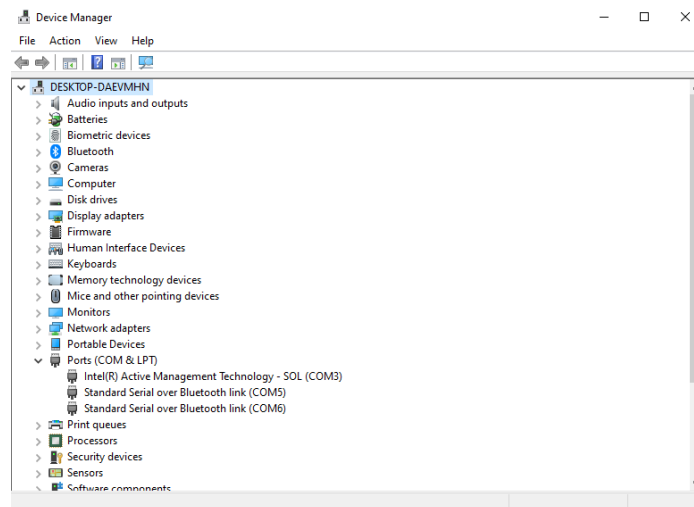
Hình 4.5. Cài đặt bluetooth.

Chọn Add Bluetooth or other device -> Bluetooth để dò tìm kết nối với HC-05.



Hình 4.6. Kết nối với bluetooth HC-05.

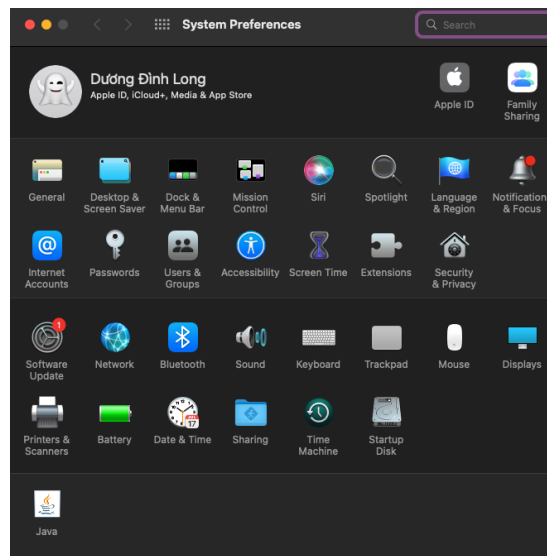
Kiểm tra việc tạo cổng ảo của bluetooth HC-05 để truyền nhận tín hiệu giữa máy tính và Arduino: ấn Window -> gỡ Device Manager -> chọn Ports. Nếu xuất hiện hai cổng ảo COM5 và COM6 là kết nối thành công, nếu không xuất hiện thì cần xem lại thiết bị máy tính hoặc bluetooth.



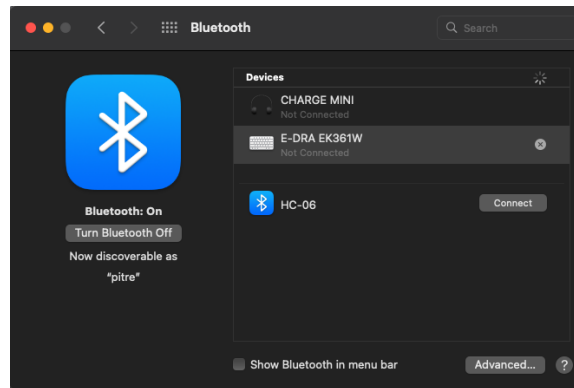
Hình 4.7. Kiểm tra cổng ảo

MacOs(hoặc Linux):

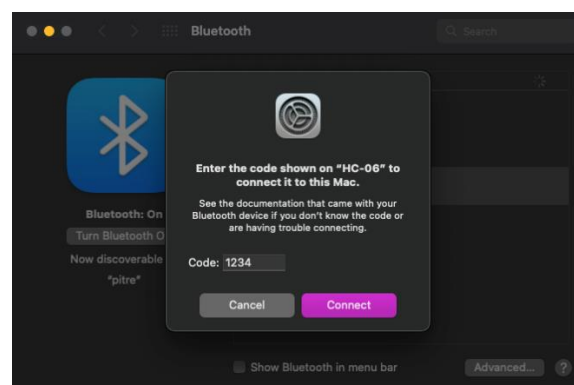
Vào System Preferences chọn Bluetooth để kết nối.



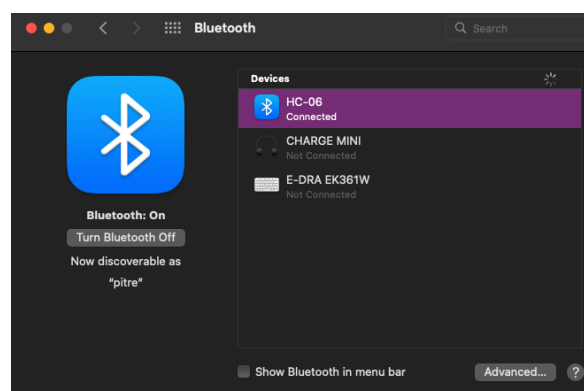
Hình 4.8. System Preferences.



Hình 4.9. Kết nối với bluetooth



Hình 4.10. Mật khẩu bluetooth là 1234.



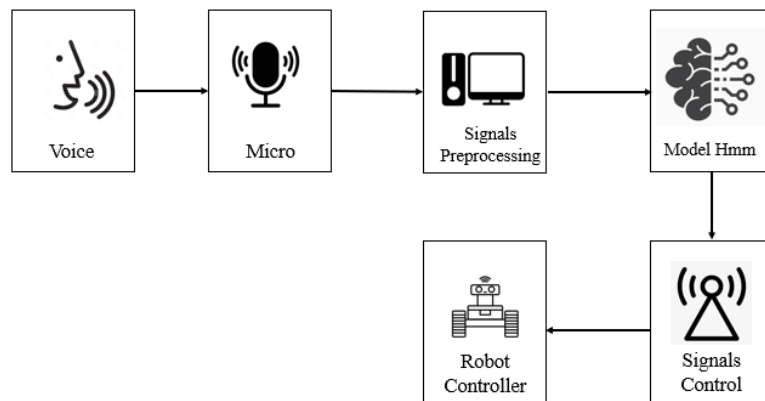
Hình 4.11. Kết nối thành công.


```
wabisabi -- ssh -- 70x21
(longdd) wabisabi@pitre ~ % ls /dev/cu.*
/dev/cu.Bluetooth-Incoming-Port /dev/cu.pci-serial22
/dev/cu.HC-86
(longdd) wabisabi@pitre ~ %
```

Hình 4.12. Kiểm tra cổng ảo.

Mô hình điều khiển thực tế.

Sử dụng mô hình Markov ẩn.



Hình 4.8. Mô hình điều khiển thực tế.

Để lưu trữ âm thanh nói và micro, sử dụng thư viện có sẵn của Python là PyAudio. Mỗi lần nói hay ra lệnh cho Mobile Robot sẽ nó trong vòng 4 giây (có thay đổi bất kì nhưng để đảm bảo tối ưu cũng như lượng câu lệnh ngắn thì 4 giây là con số vừa phải). Âm thanh đầu vào được mặc định các thông số như sau:

‘chuck = 512’: Số mẫu âm thanh được đọc hoặc ghi mỗi lần lấy mẫu.

Định dạng mẫu của âm thanh là âm thanh 16 bit.

‘channels = 1’: Số kênh âm thanh.

‘fs = 22050’: Tần số lấy mẫu, số mẫu âm thanh được lấy mỗi giây.

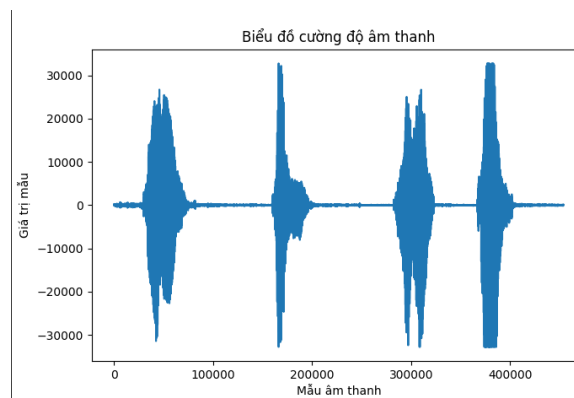
```
def record_audio_to_wav(duration=4):
    p = pyaudio.PyAudio()
    chunk = 512
    sample_format = pyaudio.paInt32
    channels = 2
    fs = 22050

    stream = p.open(format=sample_format,
                    channels=channels,
                    rate=fs,
                    frames_per_buffer=chunk,
                    input=True)

    frames = []
    recording = True
```

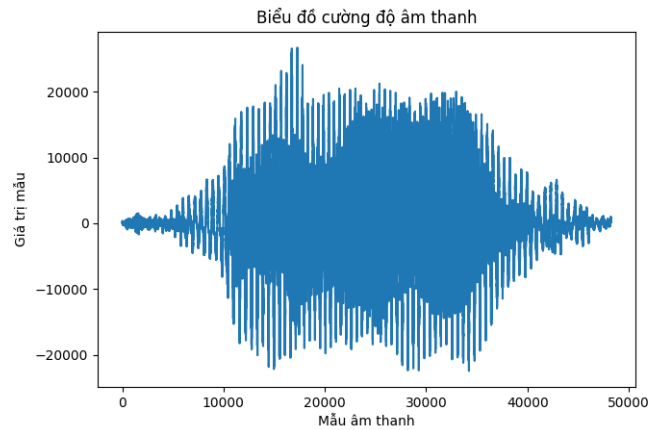
Hình 4.9. Code tách thu âm thanh từ micro.

Mô hình ẩn Markov được sử dụng là mô hình áp dụng monophone tức là xử lý trên một từ có âm thanh riêng biệt. Vì vậy, sau khi đã thu âm được giọng nói cần tách giọng nói thành các khoảng có giọng nói và im lặng. Khoảng im lặng gần như không có tác dụng trong việc điều khiển nên sẽ tách một file âm thanh thành các file nhỏ chứa âm thanh.



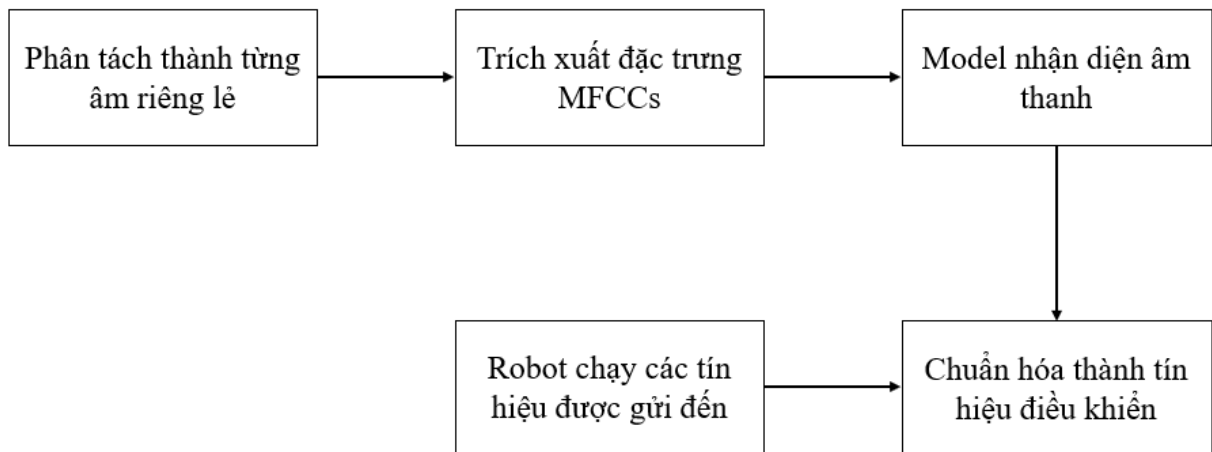
Hình 4.10. Biểu đồ cường độ âm thanh một câu.

Để tách một file âm thanh thành các file âm đơn lẻ (monophone) em sử dụng một thư viện có sẵn của Python là Pydub với các thông số cắt là những khoảng lặng có độ dài tối thiểu là 25 mili giây và ngưỡng âm lượng là -21dB. Có nghĩa một file âm thanh trong đó khoảng im lặng được định nghĩa là bất kỳ phần âm thanh nào có cấp độ dưới -21dB và độ dài tối thiểu của khoảng im lặng cho một phần chia là 25 mili giây và các khoảng im lặng này sẽ tách ra khỏi file âm thanh ban đầu.



Hình 4.11. Biểu đồ cường độ âm thanh của một từ sau khi tách.

Sau khi tách được từng từ trong một đoạn âm thanh thì tiến hành trích xuất đặc trưng từng âm thanh đã được tách rồi dùng mô hình đã đào tạo để nhận dạng từng âm thanh. Đầu ra sau khi nhận diện sẽ là các nhãn, tiếp tục chuẩn hóa các nhãn thành các tín hiệu điều khiển để truyền từ máy tính đến robot để robot thực hiện.



Hình 4.12. Mô hình thuật toán thực tế.

Sử dụng thư viện Python:

Bộ công cụ nhận dạng tiếng nói của Google, hay còn gọi là Google Speech Recognition, là một dịch vụ mạnh mẽ của Google Cloud Platform (GCP) cung cấp khả năng chuyển đổi giọng nói thành văn bản. Dịch vụ này được tích hợp sâu trong các ứng dụng của Google và cũng được cung cấp thông qua API để các nhà phát triển có thể tích hợp vào ứng dụng của họ.

Dưới đây là một số điểm nổi bật của Google Speech Recognition:

Chuyển đổi Giọng Nói thành Văn Bản: Dịch vụ này có khả năng chuyển đổi âm thanh từ giọng nói thành văn bản với độ chính xác cao. Điều này có thể hỗ trợ trong việc xây dựng ứng dụng giọng nói, hệ thống ghi chú thoại, và nhiều ứng dụng khác.

Đa ngôn ngữ: Google Speech Recognition hỗ trợ nhiều ngôn ngữ khác nhau, điều này làm cho dịch vụ phổ biến trên toàn cầu và có thể được tích hợp vào các ứng dụng đa ngôn ngữ.

Tích hợp dễ dàng: Bạn có thể tích hợp Google Speech Recognition vào ứng dụng của mình thông qua Google Cloud Speech-to-Text API. API này cung cấp một giao diện dễ sử dụng để gửi âm thanh và nhận văn bản đáp ứng.

Chất lượng và độ chính xác cao: Google sử dụng các thuật toán và mô hình máy học tiên tiến để cải thiện chất lượng và độ chính xác của việc nhận dạng tiếng nói.

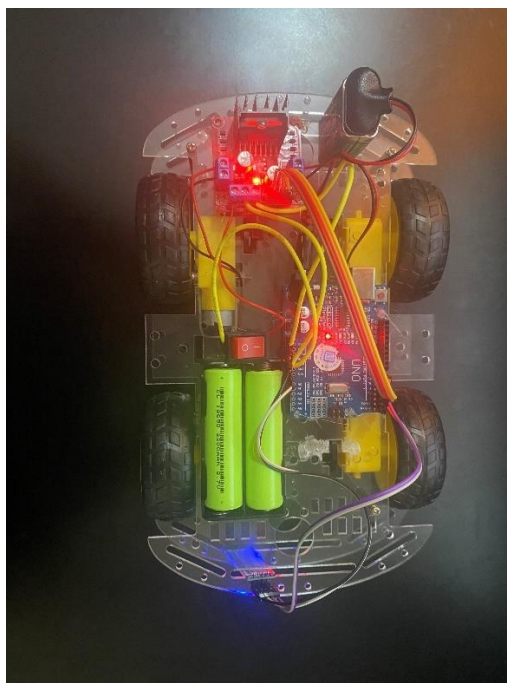
Phân loại cụm từ và dấu câu: Dịch vụ có khả năng phân loại cụm từ và dấu câu trong văn bản, cung cấp kết quả chính xác và dễ hiểu.

Tích hợp với công nghệ khác của Google Cloud: Google Speech Recognition có thể kết hợp tốt với các dịch vụ khác của Google Cloud Platform, như Google Cloud Storage, Google Cloud Pub/Sub, và nhiều dịch vụ khác nữa.

Và Google Speech Recognition cũng được tích hợp trong ngôn ngữ Python bằng thư viện `SpeechRecognition`.

Cách cài đặt, dễ dàng cài đặt bằng lệnh terminal : `'pip install SpeechRecognition'`.

4.2. Kết quả thực nghiệm điều khiển.



Hình 4.13. Mobile Robot.

Accuracy: 0.9961685823754789				
	precision	recall	f1-score	support
di	1.00	1.00	1.00	35
lui	1.00	1.00	1.00	36
phai	1.00	1.00	1.00	32
re	1.00	0.97	0.99	35
sil	0.98	1.00	0.99	57
thang	1.00	1.00	1.00	34
traoi	1.00	1.00	1.00	32
accuracy			1.00	261
macro avg	1.00	1.00	1.00	261
weighted avg	1.00	1.00	1.00	261

Hình 4.15. Kết quả huấn luyện mô hình trên tập dữ liệu.

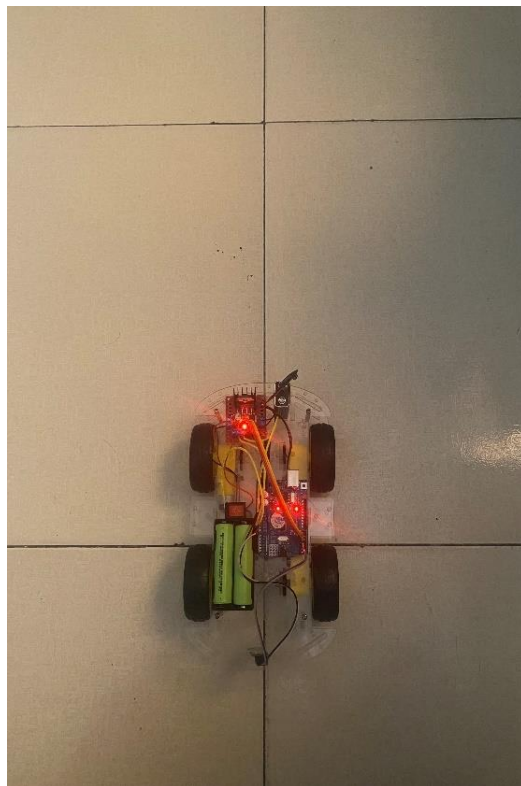
Khi nói điều khiển robot là ‘ đi thẳng’ máy tính sẽ nhận được và truyền tín hiệu như hình dưới:

```

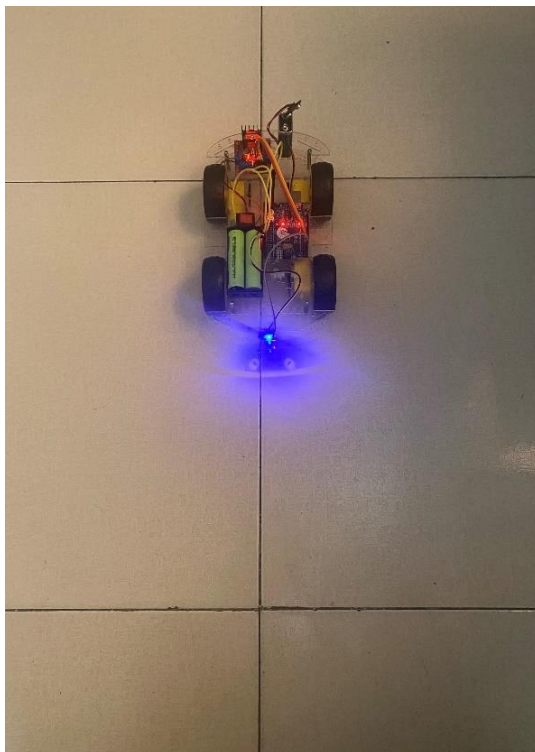
Lệnh dieu khien...
Text: di thang
Tín hiệu truyền cho robot là: F

```

Hình 4.16. Máy tính nhận âm thanh tín hiệu điều khiển.

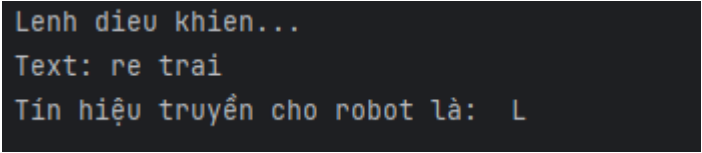


Hình 4.17. Vị trí ban đầu robot.



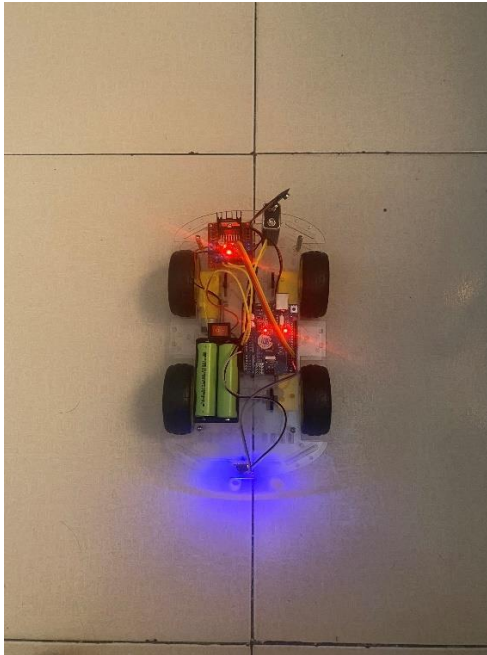
Hình 4.18. Vị trí sau khi đi thẳng.

Khi nói điều khiển robot là ‘rẽ phải’ máy tính sẽ nhận được và truyền tín hiệu như hình dưới:

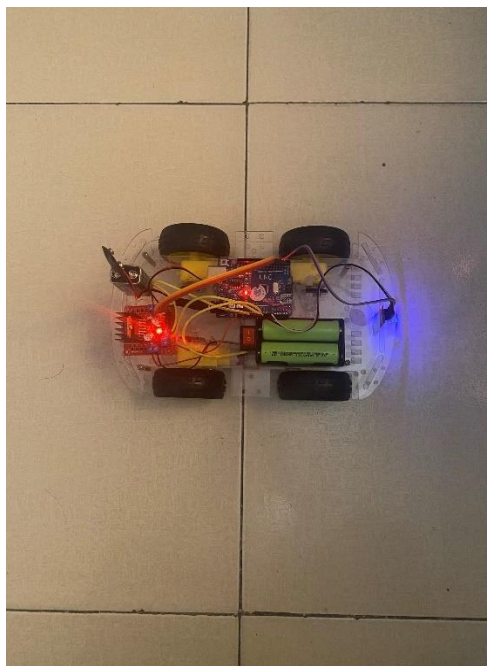


```
Lenh dieu khien...  
Text: re trai  
Tín hiệu truyền cho robot là: L
```

Hình 4.19. Máy tính nhận giọng nói chuyển tín hiệu điều khiển.



Hình 4.20. Vị trí ban đầu.



Hình 4.21. Vị trí sau khi rẽ trái.

Khi trong môi trường thực tế thì đây là liên kết đến video quay lại quá trình Mobile Robot chạy “<https://www.youtube.com/watch?v=Evpp4G80HC0>”. Có thể điều khiển các câu lệnh đơn lẻ như “đi thẳng”, “đi lùi”, “rẽ trái”, “rẽ phải”, hoặc có thể điều khiển lồng

ghép câu lệnh đơn lẻ với nhau như “ đi thẳng đi lùi rẽ trái”, “ đi thẳng rẽ phải đi lùi” . Với các câu lệnh ‘đi thẳng’, ‘đi lùi’ robot sẽ đi với khoảng cách 1,8 mét sẽ dừng lại còn với các câu lệnh “ rẽ trái”, ‘rẽ phải’ robot sẽ quay một góc 90 độ sang trái và sang phải so với vị trí ban đầu.

Liên kết github : https://github.com/longdd-pitre/Voice_Control_Mobile_Robot.git

Kết luận chương 4.

Chương 4 của đồ án đã trình bày kết quả của mô hình và thử nghiệm điều khiển mobile robot bằng giọng. Trong chương đã thực hiện các thử nghiệm với các lệnh đơn cũng như các lệnh lồng ghép nhau. Kết quả thử nghiệm cho thấy hệ thống điều khiển mobile robot bằng giọng nói hoạt động hiệu quả chính xác với những câu lệnh được đưa ra.

Tuy nhiên, vẫn còn một số hạn chế cần được khắc phục trong hệ thống điều khiển mobile robot bằng giọng nói. Một số hạn chế bao gồm:

Hệ thống có thể bị ảnh hưởng bởi tiếng ồn từ môi trường xung quanh.

Hệ thống có thể gặp khó khăn trong việc hiểu các câu lệnh phức tạp.

Chương 4 không chỉ là sự kết thúc của một chương trình nghiên cứu mà còn là sự bắt đầu cho những khám phá và phát triển tiếp theo. Việc thực nghiệm thành công đã cung cấp những cái nhìn quan trọng cho sự tiến bộ của công nghệ điều khiển bằng giọng nói và làm nền tảng cho những nghiên cứu tương lai trong lĩnh vực này.

TÀI LIỆU THAM KHẢO

- [1] Đặng Ngọc Đức, Lương Chi Mai, Nhận dạng từ có thanh điệu khác nhau trong tiếng Việt, Tạp chí Tin học và Điều khiển học, Số 2, trang 131-138, 2003.
- [2] T Hu, S Yang, F Wang, G Mittal (2002), A neural network for a non-holonomic mobile robot with unknown robot parameters, Proc. of the 2002 IEEE Int. Conf. on Robotics & Automation.
- [3] J. Schalkwyk, Hosom JP., Ed Kaiser, Khaldom Shobaki, CSLU HMM: The CSLU Hidden Markov Modelling Environment, Center of Spoken Language Understanding (CSLU), Oregon Graduate Institute of Science and Technology, 2000.
- [4] B.Yegnanarayana and S. Kishore. AANN: an alternative to GMM for pattern recognition. Neural Networks, pages 459–469, 2002.
- [5] M. W. Mak K. K. Yiu and S. Y. Kung. Environment adaptation for robust speaker verification, In Proc. of Eurospeech, pages 2973–2976, 2003
- [6] N Sidek, N Sarkar, SARKAR (2008), Dynamic modeling and control of nonholonomic mobile robot with lateral slip, Proc. of the 7th WSEAS Int. Conf. on Signal Processing, Robotics and Automation (ISPRA '08), University of Cambridge, UK.
- [7] Shrikanth Narayanan Soonil Kwon, Speaker change detection using a new weighted distance measure, In IEEE International Conference on Spoken Language Processing, Denver, USA, volume 4, pages 2537–2540, 2002.
- [8] Hong-Jiang Zhang Lie Lu and Hao Jiang, Content analysis for audio classification and segmentation, IEEE transactions on speech and audio processing, 10(7):504–516, 2002.
- [9] Luo, Y.; Mesgarani, N. TasNet: Time-Domain Audio Separation Network for Real-Time, Single-Channel Speech Separation. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 696–700.
- [10] Yu, D.; Kolbæk, M.; Tan, Z.H.; Jensen, J. Permutation Invariant Training of Deep Models for Speaker-Independent Multi-Talker Speech Separation. In Proceedings

of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 241–245.

[11] Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 1989, pp. 77, 257–286.

[12] CMU Sphinx. <https://cmusphinx.github.io/wiki/tutorialconcepts/>.

[13] Google ASR: <https://cloud.google.com/speech-to-text/docs>.

[14] Python SpeechRecognition: <https://pypi.org/project/SpeechRecognition>

[15] Coleman, J., & Pierrehumbert, J. Natural Language Processing in Lisp. University of Birmingham. <https://www.cs.bham.ac.uk/~pxc/nlp/NLPA-Phon1.pdf>