Ứng dụng nền tảng nhúng Beagleboard C4 trong điều khiển thiết bị bằng tiếng nói tiếng Việt

Application of Beagleboard C4 Platform in Controlling Devices using Vietnamese Voice Recognition System

Đào Trung Kiên, Phạm Thị Ngọc Yến và Nguyễn Thị Lan Hương

***Abstract**: In this paper, a device control system using voice recognition developed for the embedded platform Beagleboard C4 is introduced. The system is able to recognize voice commands in Vietnamese obtained via a microphone, then realizes the corresponding control tasks. Two applications are taken into experiment are a six-legged spider-like robot, and a fan. The whole system including voice capturing, voice recognition, and device control modules are developed on a single board, which is limited in size, resources as well as computing power.*

I. GIỚI THIỆU

Trong bài báo này, việc nhận dạng âm thanh được ứng dụng để xây dựng hệ thống điều khiển thiết bị bằng tiếng nói tiếng Việt. Dựa vào các tín hiệu âm thanh thu được, hệ thống nhận dạng sẽ nhận dạng câu lệnh của người điều khiển. Chương trình điều khiển sau khi tiếp nhận lệnh sẽ thực hiện việc điều khiển thiết bị tương ứng với câu lệnh được đọc.

Về mặt kỹ thuật, tiếng nói là tín hiệu phức tạp và có dư thừa thông tin. Để hệ thống nhận dạng tiếng nói hoạt động hiệu quả, các thông tin đặc trưng cần phải được trích ra từ tín hiệu tiếng nói. Quá trình này tạo ra một chuỗi các vector tham số, thường là các đặc trưng về phổ. Bộ trích chọn thông số đặc trưng bao gồm các hoạt động: số hoá, lọc tăng cường, phân chia tín hiệu thành các khung (20-30ms), nhân với hàm cửa sổ để giảm hiệu ứng không liên tục ở đầu và cuối khung tín hiệu, tính toán các thông số đặc trưng,… [1] Hiện nay kiểu vector đặc trưng MFCC (*Mel-frequency cepstrum coefficients*) được sử dụng nhiều trong các hệ thống nhận dạng, đặc biệt cho các ứng dụng di động [2]. Bên cạnh đó để giảm lượng tính toán, vector đặc trưng có thể được lượng tử hóa sử dụng các kỹ thuật lượng tử hóa vector [3].

Một hướng tiếp cận khác với bài toán nhận dạng tiếng nói là sử dụng phương pháp quy hoạch động theo thời gian (DTW - *dynamic time warping*), hay còn biết đến là thuật toán lập trình động. Phương pháp này được sử dụng để đo sự sai khác của hai mẫu tín hiệu tiếng nói, và thường được áp dụng cho các hệ thống nhận dạng từ rời rạc hoặc từ ghép [4], phù hợp cho hệ thống nhận dạng với số lượng từ nhỏ và phụ thuộc người nói [3].

Bên cạnh đó, mô hình chuỗi Markov ẩn (HMM - *hidden Markov model*) cũng được sử dụng phổ biến. Đây là mô hình thống kê được áp dụng trong nhận dạng tiếng nói từ những năm 1980. Hiện nay kỹ thuật HMM được sử dụng rộng rãi trong các hệ thống nhận dạng tiếng nói [5] và là HMM bậc 1. Tiếng nói có thể coi là tín hiệu ngẫu nhiên, do cùng một từ nhưng mọi người phát âm tạo ra tín hiệu khác nhau. Ngay bản thân một người khi phát âm một từ cũng tạo ra tín hiệu khác nhau. Với phương pháp DTW, hệ thống cần lưu trữ các mẫu thể hiện của tiếng nói cần nhận dạng. Điều này đòi hỏi số lượng mẫu lưu trữ lớn. Phương pháp này sẽ không hiệu quả khi số lượng từ cần nhận dạng lớn và tập người nói nhiều (cho ứng dụng độc lập người nói). Trong trường hợp này sử dụng HMM sẽ hiệu quả hơn, cho phép mô hình hóa một tập các thể hiện của một mẫu tiếng nói bằng một mô hình.

Trong nghiên cứu được mô tả trong bài báo này, các mô-đun nhận dạng và điều khiển đều được xây

dựng tích hợp trên một máy tính nhúng đơn Beagleboard C4. Hai ứng dụng thử nghiệm hệ thống được giới thiệu là robot nhện sáu chân, và quạt điều khiển bằng tiếng nói. Kết quả thử nghiệm cho thấy hệ thống hoạt động tốt, cho kết quả nhận lệnh chính xác và thời gian đáp ứng lệnh ngắn.

Các phần tiếp theo của bài báo được trình bày như sau. Cấu trúc tổng thể của hệ thống được giới thiệu trong Phần II. Tiếp theo, trong Phần III và Phần IV, hai ứng dụng điều khiển robot và quạt được trình bày. Cuối cùng, các kết quả thử nghiệm và đánh giá được phân tích và thảo luận trong Phần V.

II. CẤU TRÚC HỆ THỐNG VÀ HOẠT ĐỘNG

Trong nghiên cứu này, nền tảng Beagleboard C4 được sử dụng để thu nhận, xử lý âm thanh và điều khiển. Đây là bo mạch nhúng đơn nguồn điệp áp thấp (một chiều 5V) sản xuất bởi hãng Texas Instruments và Digi-Key, được trang bị bộ vi xử lý OMAP3530 (nhân ARM Cortex-A8 tốc độ 720 MHz), kèm theo một số cổng vào/ra chuẩn phổ biến: USB, JTAG, HDMI, SD/MMC, vào/ra audio, RS-232 và một cổng vào/ra mở rộng gồm 28 chân. Để có thể tương tác với hệ thống, hệ điều hành Ubuntu 10.10 phiên bản nhúng được cài đặt trên một thẻ nhớ SD. Đây là một phiên bản của Linux, hệ điều hành mã nguồn mở miễn phí. Chương trình cần xây dựng chạy trên hệ thống này được chia thành hai phần chính: (1) thu nhận và nhận dạng tín hiệu âm thanh từ microphone, và (2) điều khiển thiết bị ngoài qua các chân mở rộng của Beagleboard. Ngoài ra hệ thống cũng có khả năng thể hiện phản hồi với người dùng qua loa ngoài (tuỳ chọn). Nhờ khả năng lập trình dễ dàng, sử dụng hệ thống mã nguồn mở, khả năng kết nối và giao tiếp mạnh mẽ với các thiết bị ngoại vi, Beagleboard cho phép triển khai được các hệ thống điều khiển phức tạp mà tiêu thụ rất ít điện năng. Ngoài ra, bộ điều khiển có thể được thiết kế nhỏ gọn vì Beagleboard chỉ có kích thước $7{,}5\times7{,}5cm^2$.

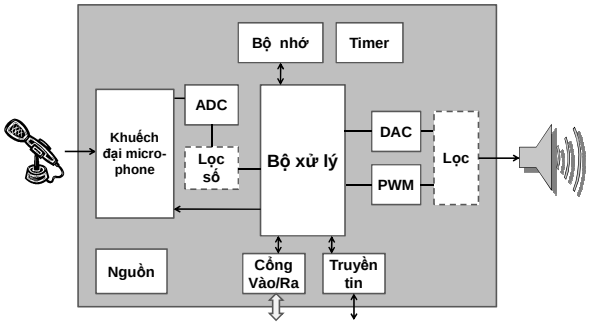
Hệ thống nhận các lệnh điều khiển từ microphone, và đưa đến bộ vi xử lý trung tâm để thực hiện quá trình nhận dạng lệnh điều khiển. Do việc tương tác giữa người và hệ thống được thực hiện trong môi trường bình thường (có nhiễu), nên để đảm bảo độ chính xác nhận dạng, tín hiệu tiếng nói từ microphone có thể cần đưa qua khâu lọc số để xử lý một phần nhiễu sau đó mới đưa vào bộ xử lý trung tâm. Bộ xử lý trung tâm sẽ thực hiện các công việc như nâng cao chất lượng tiếng nói (giải nhiễu), nhận dạng từ điều khiển, phân tích và thực hiện lệnh điều khiển, đưa ra phản hồi của hệ thống với người sử dụng thông qua tương tác bằng tiếng nói. Ngoài ra hệ thống cũng có khả năng điều khiển các thiết bị ngoại vi, trao đổi thông tin với hệ thống khác thông qua cổng vào/ra hoặc cổng truyền tin.

Có thể nói, với những yêu cầu và công việc cần thực hiện, khối xử lý trung tâm là "bộ não" của hệ thống nhúng tương tác bằng tiếng nói. Việc lựa chọn nền tảng phần cứng và phần mềm cho hệ thống nhúng tương tác bằng tiếng nói cần được xem xét cho từng ứng dụng cụ thể, đặc biệt cần chú ý đến các yếu tố ảnh hưởng như:

- Số từ cần nhận dạng của ứng dụng (quyết định dung lượng bộ nhớ khối xử lý trung tâm),
- Môi trường làm việc của hệ thống (xem xét sự ảnh hưởng của nhiễu môi trường (nhiều hay ít, tỉ số tín hiệu trên nhiễu) lên kết quả nhận dạng của hệ thống),
- Tính thời gian thực (quyết định khả năng tính toán của khối xử lý trung tâm).

Hiện nay, các hãng sản xuất cũng đưa ra các dòng sản phẩm chuyên biệt phục vụ cho các ứng dụng của công nghệ xử lý tiếng nói như các DSP dấu phẩy động cho các hệ thống tổng hợp hay nhận dạng. Dung lượng các bộ nhớ ROM, RAM cũng được tăng đáng kể nên các ứng dụng có thể mở rộng hơn.

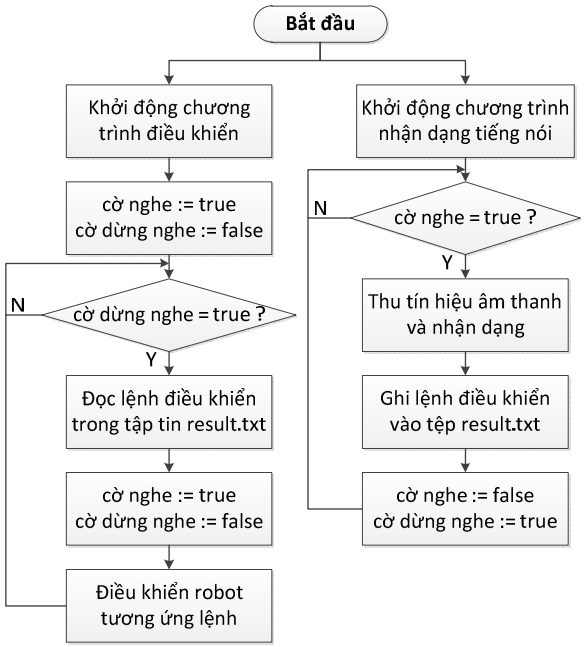
Hình 1 biểu diễn sơ đồ khối các mô-đun phần cứng cơ bản trong hệ thống nhúng trong các thiết bị giao tiếp (tương tác) bằng tiếng nói. Hệ thống nhận các lệnh điều khiển từ microphone, và đưa đến bộ xử lý trung tâm để thực hiện quá trình nhận dạng lệnh điều khiển. Hệ thống bao gồm các khối chức năng như sau:



Hình 1. Sơ đồ khối các mô-đun phần cứng hệ thống nhúng cho giao tiếp bằng tiếng nói

- Mô-đun giao tiếp người-máy: Hệ thống tương tác người-máy vận hành theo cơ chế hướng sự kiện có thể sử dụng ngắt cứng hoặc ngắt mềm, hoặc quét theo một chu kì nhất định. Việc nhận dạng được thực hiện liên tục, trong trường hợp chưa bắt đầu đối thoại, nếu đã bắt đầu đối thoại, hệ thống có thể cho phép/không cho phép nhận dạng để đảm bảo hạn chế nhận dạng nhầm do nhiễu môi trường. Mặt khác, tùy vào quá trình, trạng thái của thiết bị, tập lệnh nhận được sẽ bị hạn chế để tránh các hậu quả do nhận dạng nhầm dẫn đến điều khiển sai quá trình.
- Mô-đun nhận dạng: Vấn đề nhận dạng tiếng nói có thể được xem như là vấn đề phân loại mang tính thống kê (các lớp khác nhau), giống như hình thức nhận dạng mẫu cổ điển. Một lớp được định nghĩa là một tập các từ cho phép trong một từ điển đóng. Một tham số cho tín hiệu tiếng nói được chọn (như chuỗi các vector đặc trưng âm học), và một xác suất điều kiện MAP (*Maximum a Posteriori*).
- Mô-đun tổng hợp tiếng: Mô-đun nhận đầu vào là đoạn văn bản cần tổng hợp và tạo ra ở đầu ra là tín hiệu tiếng nói. Để có thể xử lý được đoạn văn bản đầu vào, chương trình sẽ cần phải có mô-đun phân tích văn bản. Vì mô-đun tổng hợp được phát triển trên ứng dụng nhúng, không cần phải đưa vào một văn bản, mà văn bản sẽ được đưa ra dưới dạng đã cấu trúc, do đó với mô-đun được xây dựng này đầu vào sẽ không phải là đoạn văn bản cần tổng hợp mà sẽ là file phân tích văn bản. Tệp tin này có được nhờ sử dụng mô-đun phân tích văn bản của mô-đun tổng hợp tiếng được phát triển trên PC.
- Mô-đun truyền tin: Căn cứ trên hệ thống thiết bị xử lý trung tâm của sản phẩm nhúng ta có thể sử dụng các cổng tryền tin theo mô thông qua bộ mã hóa tín hiệu (codec), qua cổng truyền tin USB hoặc cổng UART.

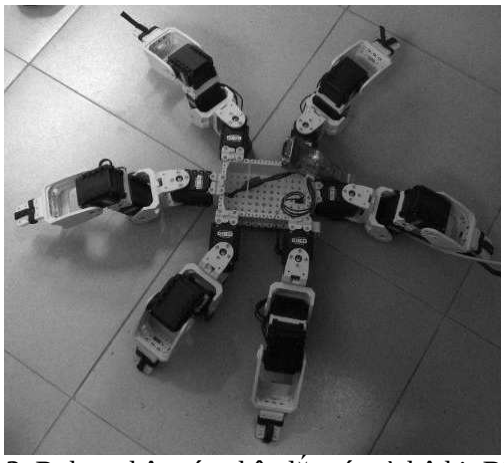
Sơ đồ điều khiển thiết bị bằng tiếng nói được giới thiệu khái quát như trong Hình 2. Việc điều khiển được thực hiện bởi hai mô-đun chạy song song. Mô-đun thứ nhất dùng để điều khiển thiết bị trong khi mô-đun thứ hai dùng để nhận dạng tiếng nói xác định lệnh điều khiển. Hai mô-đun này trao đổi thông tin với nhau thông qua *cờ_nghe*, *cờ_dừng_nghe* và tệp tin *result.txt*. Hai biến *cờ_nghe* và *cờ_dừng_nghe* là hai biến kiểu boolean có giá trị đúng hoặc sai để xác định sự kiện bắt đầu và kết thúc quá trình lắng nghe lệnh điều khiển. Kết quả nhận dạng được lưu vào tệp tin *result.txt* để chương trình điều khiển thiết bị đọc lệnh điều khiển và cho phép chương trình nhận dạng tiếp tục thu âm và thực hiện việc nhận dạng đồng thời thực hiện lệnh điều khiển.



Hình 2. Sơ đồ thuật toán điều khiển

III. ỨNG DỤNG ĐIỀU KHIỂN ROBOT NHỆN SÁU CHÂN

Việc sử dụng robot làm thay con người trong những môi trường có điều kiện khắc nghiệt, độc hại, nguy hiểm đã và đang là một xu hướng nổi trội trong lĩnh vực tự động hóa. Với việc điều khiển trực tiếp robot bằng tiếng nói, ứng dụng này đưa ra khả năng tích hợp năng lực phân tích, quan sát và trí thông minh của con người vào robot để thực hiện những tác vụ phức tạp. Việc điều khiển robot bằng tiếng nói cũng giúp cung cấp cho người dùng một phương thức giao tiếp dễ dàng và thuận tiện.



Hình 3. Robot nhện sáu chân lắp ráp từ bộ kit Bioloid

Bảng 1 Các đặc tính của động cơ AX-12

Đặc tính	Đơn vị đo
Điện áp	7V ÷ 10V (lý tưởng 9,6V)
Góc hoạt động	300°
Dòng cực đại	900mA
Tỷ lệ bánh răng	1/254
Lực giữ tối đa	12 (7V) ÷ 16,5kgf.cm (10V)
Vận tốc cực đại	223°/s (7V) ÷ 306°/s (10V)
Góc quay cực tiểu	0,35°
Khối lượng	55g
Thông tin phản hồi	Nhiệt độ, điện áp, tải, vị trí,…

Bảng 2 Các đặc tính của cảm biến AX-S1

Đặc tính	Đơn vị đo
Điện áp	7V ÷ 10V (lý tưởng 9,6V)
Dòng điện cung cấp	40mA
Độ phân giải	10 bit (1024 giá trị)
Khối lượng	37g
Thông tin phản hồi	Nhiệt độ, điện áp,…

Trong ứng dụng này, mô hình robot nhện sáu chân được lựa chọn để điều khiển cho phép ứng dụng trong những môi trường mấp mô, có nhiều vật cản nhỏ. Mô hình robot nhện sáu chân được lắp ráp từ bộ kit Bioloid [6] của hãng Robotis như trong 0. Robot nhện sáu chân này được tích hợp với hệ thống nhận dạng tiếng nói cài đặt trên hệ thống nhúng giúp cho việc xây dựng các ứng dụng điều khiển bằng tiếng nói được tích hợp gói gọn.

Việc lắp đặt robot nhện sáu chân được thực hiện với 18 động cơ AX-12, một cảm biến AX-S1, và một hộp điều khiển. Tất cả các thiết bị này và các chi tiết cơ khí lắp ráp được lấy từ bộ kit Bioloid. Mỗi chân của robot có 3 động cơ AX-12 tương ứng với 3 bậc tự do của chân nhện. Các đặc tính kỹ thuật của động cơ AX-12 và cảm biến AX-S1 được giới thiệu trong Bảng 1 và Bảng 2. Động cơ AX-12 thực tế là một mô-đun đóng gói bao gồm cả bộ bánh răng và mạch điều khiển ở bên trong. Do đó, động cơ AX-12 có thể sử dụng một cách dễ dàng trong các mô-đun lắp ghép mà không cần thêm mạch điều khiển công suất nào. AX-12 còn là một động cơ thông minh có thể phản hồi các thông tin về điều kiện làm việc hiện tại của động cơ như: điện áp, nhiệt độ, tải,… Khi lắp ráp cho robot nhện, các động cơ AX-12 được ghép nối tiếp với nhau và việc điều khiển từng động cơ được xác định nhờ vào ID quy định cho mỗi động cơ.

Việc điều khiển robot di chuyển được dựa trên nguyên lý di chuyển tam giác thay đổi. Với cách đi này, sáu chân của robot nhện được chia thành hai nhóm xen kẽ và di chuyển lệch pha nhau. Ở mỗi pha di chuyển, ba chân thuộc nhóm này nhấc khỏi nền thì ba chân của nhóm khác chạm nền để tạo thành một tam giác vững chắc và cân bằng nâng đỡ toàn bộ robot. Hai nhóm chân này sẽ luân phiên nhấc lên và tiếp đất để tạo nên bước đi nhịp nhàng cho robot.

Trong bài toán điều khiển robot nhện được giới thiệu bài báo này, các bước di chuyển được xây dựng bao gồm: tiến về trước, lùi ra sau, quay trái, và quay phải. Mỗi bước di chuyển được chia thành bốn pha nhỏ với sự thay đổi góc quay luân phiên của các nhóm

bộ ba chân nhện. Về cơ bản, bước di chuyển lùi ra sau sẽ bao gồm các pha như trong di chuyển tiến về trước nhưng với thứ tự của các pha được đảo ngược và các động cơ quay theo chiều ngược lại. Tương tự như vậy, quay phải sẽ bao gồm các pha di chuyển của quay trái với thứ tự ngược lại của mỗi pha và chiều quay đảo ngược của mỗi động cơ.

Để di chuyển tiến về trước, bộ ba chân nhện trong nhóm thứ nhất nhấc lên khỏi mặt đất trong pha thứ nhất. Trong pha thứ hai, bộ ba chân nhện thứ hai (đang chạm đất) đẩy thân nhện về phía trước. Ở pha thứ ba, bộ ba chân nhện thứ nhất tiếp đất và bộ ba chân nhện thứ hai được nhấc lên. Đến pha cuối cùng, các chân của bộ ba thứ nhất cũng đẩy thân nhện về phía sau trước. Nếu được yêu cầu tiếp tục đi đến thì bộ ba chân thứ nhất lại nhấc lên và bộ ba chân thứ hai lại chạm đất như ở pha thứ nhất. Chu trình này được lặp đi lặp lại tạo nên di chuyển liên tục của robot tiến về trước.

Để di chuyển quay trái, bộ ba chân thứ nhất của robot tiếp đất, giữ thăng bằng cho robot ở pha thứ nhất và thứ hai; trong khi đó, bộ ba chân thứ hai của robot chạm đất trong pha thứ ba và thứ tư. Ở pha thứ nhất, các chân của bộ ba thứ hai sẽ nhấc lên khỏi mặt đất, cho phép các chân của bộ ba thứ nhất tự do đẩy thân robot quay sang trái trong pha thứ hai. Các chân của bộ ba thứ nhất được điều chỉnh lại vị trí cho cân bằng với các chân của bộ ba thứ nhất trong pha thứ tư. Chu trình này khi thực hiện liên tiếp tạo nên di chuyển quay trái nhịp nhàng của robot. Để xác định các góc quay của động cơ tương ứng với các vị trí tiếp đất của robot trong mỗi bước đi, ta xây dựng bài toán động học thuận và giải bài toán động học ngược cho robot nhện sáu chân đã lắp ráp.

Tương tác để điều khiển robot có thể thực hiện bằng tiếng nói hoặc can thiệp bằng nút nhấn ở trên hộp điều khiển. Để điều khiển robot bằng tiếng nói, việc thu thập số liệu âm thanh được thực hiện với bộ thu nhận tín hiệu âm thanh không dây EM 300 G2 của hãng Sennheiser. Các điều kiện thu nhận âm thanh được tuân thủ với tần số lấy mẫu 8kHz, số bit lượng tử hoá 16bit, và truyền tin theo chuẩn sóng radio UHF tần số trong khoảng 518-554MHz.

Để tương tác điều khiển robot, người điểu khiển sẽ dùng tập các lệnh điều khiển được thiết kế sẵn trong Bảng 3. Khi nhận được một câu lệnh điều khiển trong tập lệnh điều khiển này, robot sẽ thực hiện tác vụ được định nghĩa tương ứng.

Bảng 3 Tập lệnh điều khiển robot nhện

Câu lệnh	Tác vụ thực hiện
tiến	Robot tiến về phía trước
robot tiến	
lùi	Robot lùi lại phía sau
robot lùi	
quay phải	Robot quay sang phải
robot quay phải	
quay trái	Robot quay sang trái
robot quay trái	
dừng	Robot dừng tất cả các tác vụ đang thực hiện, trở về trạng thái sẵn sang
robot dừng	

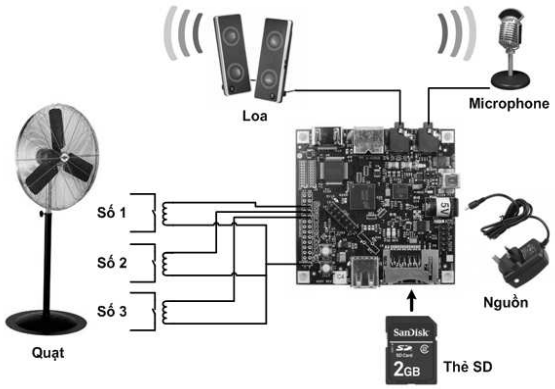
IV. ỨNG DỤNG ĐIỀU KHIỂN QUẠT

Hình 4 thể hiện sơ đồ các thiết bị chính của hệ thống được xây dựng. Bo mạch nhúng có thể được cung cấp nguồn chung với quạt, hoặc dùng nguồn độc lập. Hai cổng audio vào/ra được nối tương ứng với một microphone và một loa ngoài. Để bật/tắt 3 số của quạt bằng chương trình, 3 rơ-le điện được sử dụng tương ứng thay cho các nút bấm cơ có sẵn của quạt. Một đầu dây của mỗi rơ-le số 1, 2, 3 được nối với một chân cổng mở rộng của bo mạch nhúng, tương ứng là các chân 18, 20, 22, và đầu dây còn lại được nối với chân số 28. Sơ đồ các chân mở rộng của bo mạch nhúng được thể hiện trong Hình 5. Ngoại trừ hai chân 1, 2 với điện thế cố định, 27, 28 nối đất, và 25, 26 tương ứng với với RESET và REGEN, thì các chân đánh số nằm trong khoảng từ 3 đến 24 có thể được lập trình để điều khiển. Mỗi chân này có thể được điều khiển để xác lập ở một trong hai mức điện áp 0 hoặc 1.8V.

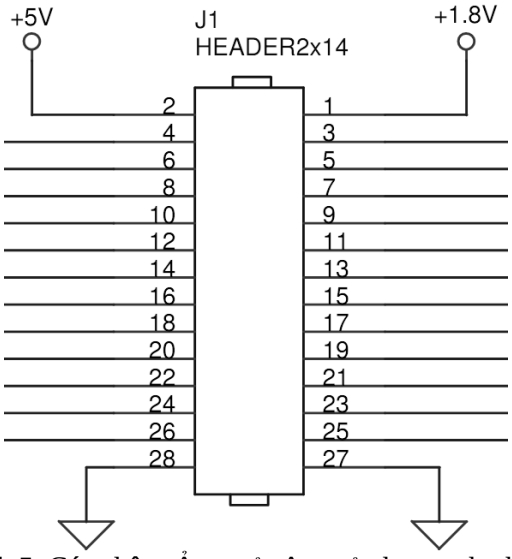
Hình 6 thể hiện sơ đồ điều khiển của hệ thống. Tín hiệu âm thanh từ người dùng sau khi được thu từ

microphone sẽ được xử lý và nhận dạng thành câu lệnh. Tương ứng với mỗi câu lệnh, chương trình chạy trên bo mạch nhúng sẽ điều khiển bật/tắt các rơ-le tương ứng để tăng/giảm hoặc bật/tắt quạt, đồng thời thể hiện phản hồi với người dùng qua loa (nếu được nối).

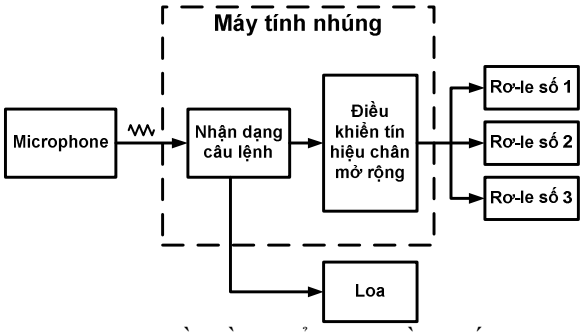
Chiếc quạt được điều khiển có ba chế độ chạy 1, 2, 3 lần lượt từ thấp đến cao. Ba rơ-le điện có đầu vào 5V được thiết lập tương ứng với ba chế độ này. Tại một thời điểm, sẽ chỉ có duy nhất một rơ-le được bật tương ứng với chế độ chạy của quạt, hoặc không có rơ-le nào bật nếu quạt đang tắt. Các câu lệnh được thiết kế như trong Bảng 4.



Hình 4. Sơ đồ thiết bị hệ thống điều khiển quạt



Hình 5. Các chân cổng mở rộng của bo mạch nhúng Beagleboard C4



Hình 6. Sơ đồ điều khiển quạt bằng tiếng nói

Bảng 4. Tập lệnh điều khiển quạt

Câu lệnh	Tác vụ thực hiện
quạt số một	Bật quạt số 1: bật rơ-le số 1 và tắt các rơ-le còn lại nếu đang bật.
quạt số hai	Bật quạt số 2: bật rơ-le số 2 và tắt các rơ-le còn lại nếu đang bật.
quạt số ba	Bật quạt số 3: bật rơ-le số 3 và tắt các rơ-le còn lại nếu đang bật.
quạt tắt	Tắt quạt: tắt hết các rơ-le.

Bảng 5. Mã GPIO của các chân cổng mở rộng

Chân	GPIO	Chân	GPIO
4	140	3	139
6	142	5	138
8	143	7	137
10	141	9	136
12	158	11	135
14	162	13	134
16	161	15	133
18	159	17	132
20	156	19	131
22	157	21	130
24	168	23	183

Hiệu điện thế các chân cổng mở rộng điều khiển được của Beagleboard được thiết lập ở điện áp 1.8V khi bật, và được kích lên 5V ở đầu ra để điều khiển rơ-le. Cổng mở rộng của bo mạch nhúng là các chân GPIO (*general purpose input/output*), mỗi chân tương ứng với một số GPIO được cho trong Bảng 5. Việc điều khiển trên Ubuntu đã được thiết lập trên hệ thống

tệp giả lập tại đường dẫn */sys/class/gpio*, với trình tự như sau:

- Ghi giá trị *N* vào tệp tin giả lập */sys/class/gpio/export* để giữ quyền điều khiển một chân có GPIO là *N*. Nếu thành công (khi chân đó chưa bị giữ bởi một chương trình nào khác), hệ thống sẽ tạo ra một thư mục giả lập */sys/class/gpio/gpioN* bao gồm các tệp tin *device*, *power*, *uevent*, *direction*, *subsystem*, và *value*.
- Ghi giá trị *N* vào tệp tin giả lập */sys/class/gpio/unexport* khi muốn giải phóng quyền điều khiển. Thư mục giả lập */sys/class/gpio/gpioN* và các tệp tin bên trong sẽ bị xoá.
- Ghi chuỗi *low* vào tệp tin giả lập */sys/class/gpio/gpioN/direction* khi muốn thiết lập điện thế của chân về mức thấp (0V).
- Ghi chuỗi *high* vào tệp tin giả lập */sys/class/gpio/gpioN/direction* khi muốn thiết lập điện thế của chân về mức cao (1.8V).
- Để đọc tín hiệu vào từ chân, trước hết cần ghi chuỗi *in* vào tệp tin giả lập */sys/class/gpio/gpioN/direction*, sau đó đọc tín hiệu từ tệp tin giả lập */sys/class/gpio/gpioN/value*.

Việc đọc/ghi vào các tệp tin giả lập tương tự như các tệp tin trên đĩa thông thường, có thể được thực hiện trên các ngôn ngữ lập trình, hay qua các lệnh shell của Linux.

Dùng chương trình lập trình sẵn trên bo mạch nhúng để điều khiển bật/tắt các rơ-le điện thay vì bấm trực tiếp các nút số của quạt thông thường. Người dùng ra lệnh cho chương trình bằng cách nói một trong các câu lệnh có sẵn vào microphone, như đã mô tả ở trên. Việc thu thập số liệu âm thanh được thực hiện với bộ thu nhận tín hiệu âm thanh không dây EM300G2 của hãng Sennheiser. Điều kiện thu nhận âm thanh được thiết lập với tần số lấy mẫu 8kHz, số bit của mẫu 16, truyền tin bằng sóng radio UHF tần số trong khoảng 518-554MHz.

V. CHỈ TIÊU KỸ THUẬT VÀ KẾT QUẢ THỬ NGHIỆM, ĐÁNH GIÁ

Hệ thống được kiểm chuẩn trong phòng ít nhiễu, hệ thống nhận dạng không phụ thuộc vào thông số môi trường như nhiệt độ, độ ẩm. Thiết bị được kiểm định với 12 người. Do đặc thù về sự thay đổi giọng nói của mỗi người, người thử nghiệm được lựa chọn là người có phương ngữ chuẩn miền Bắc, độ tuổi từ 20 đến 60 tuổi. Người kiểm tra không nằm trong bộ cơ sở dữ liệu để luyện hệ thống.

Đối với câu thử nghiệm, mỗi câu được đọc 10 lần với tốc độ nói bình thường. Căn cứ hướng dẫn đánh giá và trình bày độ không đảm bảo đo ĐLVN 131:2004. Sai số của thiết bị được đánh giá thông qua tỉ lệ sai tổng hợp của nhóm người kiểm tra. Tỉ lệ sai của mỗi người được tính theo công thức:

$$\gamma_i = N_s / N_{total} \qquad (1)$$

trong đó γ_i là sai số thử của người thứ *i*, N_s là số lần sai (không phân biệt câu/từ điều khiển), và N_{total} là tổng số lần thử.

Sai số của thiết bị được đánh giá theo công thức sai số bình quân quân phương (ước lượng phương sai) của nhóm người thử nghiệm:

$$\gamma_{tbi} = t_{st}\sqrt{\sum_{i=1}^{N}\gamma_i^2 \Big/ N(N-1)} \qquad (2)$$

trong đó γ_{tbi} là sai số thử nghiệm của thiết bị thứ *i*, *N* là số thử nghiệm, và t_{st} là hệ số student ($t_{st} = 1{,}80$ với mức tin cậy 90%, số bậc tự do $12-1=11$).

Các bước tiến hành theo trình tự như sau:

- Bước 1: Kiểm tra hoạt động kết nối của hệ thống với các phần tử điều khiển.
- Bước 2: Kiểm tra hệ thống theo từng đối tượng điều khiển. Kết quả thử nghiệm được ghi vào bảng kết quả.

- Bước 3: Ước lượng sai số của thiết bị theo công thức (1) và (2).

Bảng 6 Kết quả thử nghiệm

Ứng dụng	Sai số thử nghiệm (%)	Tỉ lệ nhận dạng đúng (%)
Robot nhện	5,59	93,15
Quạt	8,33	91,67

Cơ sở dữ liệu dùng để huấn luyện mô hình nhận dạng được xây dựng với tổng số 62.846 câu tiếng nói, tương đương với 32 giờ huấn luyện. Kết quả thử nghiệm với 12 người, như tổng kết trong Bảng 6, cho thấy tỉ lệ nhận dạng đúng là trên 90% và sai số thử nghiệm dưới 10% với cả hai ứng dụng robot nhện và quạt. Thời gian đáp ứng của việc nhận dạng là dưới 3 giây.

VI. KẾT LUẬN

Bài báo này trình bày một hệ thống tương tác người-máy giao tiếp bằng tiếng nói cài đặt trên bo mạch nhúng Beagleboard C4. Do việc tương tác giữa người và hệ thống được thực hiện trong môi trường bình thường (có nhiễu), nên để đảm bảo độ chính xác nhận dạng, tín hiệu tiếng nói từ microphone có thể cần đưa qua khâu lọc số để xử lý một phần nhiễu sau đó mới đưa vào bộ xử lý trung tâm. Bộ xử lý trung tâm sẽ thực hiện các công việc như nâng cao chất lượng tiếng nói (giải nhiễu), nhận dạng từ điều khiển, phân tích và thực hiện lệnh điều khiển, đưa ra phản hồi của hệ thống với người sử dụng thông qua tương tác bằng tiếng nói. Ngoài ra hệ thống cũng có khả năng điều khiển các thiết bị ngoại vi, trao đổi thông tin với hệ thống khác thông qua cổng vào/ra hoặc cổng truyền tin.

Trên cơ sở phân tích các yêu cầu, tính năng kỹ thuật chủ yếu của hệ tương tác, một nền tảng phần cứng và phần mềm cho hệ thống nhúng đã được xem xét và lựa chọn cho từng ứng dụng cụ thể. Hai sản phẩm ứng dụng là robot nhện và quạt điều khiển bằng tiếng nói đã được phân tích, thiết kế hệ thống điều khiển, và triển khai phương pháp tương tác bằng tiếng nói. Với mỗi sản phẩm ứng dụng, báo cáo cũng đã đưa ra phương pháp đánh giá chất lượng và độ tin cậy điều khiển. Kết quả cho thấy hệ thống tương tác giao tiếp người máy bằng tiếng nói trên hệ thống nhúng đã đáp ứng tốt các yêu cầu đặt ra.

LỜI CẢM ƠN

Công trình trình bày trong bài báo nằm trong khuôn khổ của Đề tài NCKH trọng điểm cấp nhà nước mã số KC.03.15/06-10 mang tên "*Nghiên cứu thiết kế chế tạo các thiết bị và hệ thống tự động hoá thông minh sử dụng tương tác người-máy bằng tiếng nói trong điều khiển*".

TÀI LIỆU THAM KHẢO

[1] PICONE, J.W., *Signal modeling techniques in speech recognition*, Proceedings of the IEEE, 81(9), 1215-1247, 1993.

[2] *Speech Processing, Transmission and Quality Aspects (STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithms*, ETSI ES 201 108.

[3] HUANG, X., *Spoken Language Processing: a Guide to Theory, Algorithm, and System Development*. Upper Saddle River NJ: Prentice Hall PTR, 2001.

[4] RABINER, L., SCHMIDT, C., *Application of dynamic time warping to connected digit recognition*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 377-388, 1980.

[5] BENESTY, J., *Springer Handbook of Speech Processing*. Springer, 2008.

[6] THAI, C.N., PAULISHEN, M., *Using Robotis Bioloid systems for instructional robotics*, Proceedings of IEEE Southeastcon, 300-306, 2011.

Nhận bài ngày: 29/11/2011

SƠ LƯỢC VỀ CÁC TÁC GIẢ

ĐÀO TRUNG KIÊN

Sinh năm 1981 tại Vĩnh Phúc

Tốt nghiệp đại học năm 2003 tại Đại học Cergy-Pontoise (Pháp), thạc sĩ năm 2005 tại Đại học Paris 6 (Pháp) về CNTT và nhận bằng tiến sĩ năm 2010 tại Đại học Dayeh (Đài Loan) về cơ khí và tự động hoá.

Hiện công tác tại Viện nghiên cứu MICA, Đại học Bách khoa Hà Nội.

Lĩnh vực nghiên cứu: môi trường tương tác và cảm thụ, hệ thống tự động

Điện thoại cơ quan: 0438683087, 0986128481

Email: trung-kien.dao@mica.edu.vn

PHẠM THỊ NGỌC YẾN

Sinh năm 1959

Hiện là Viện trưởng Viện nghiên cứu MICA, kiêm Trưởng Bộ môn Kỹ thuật đo và Tin học công nghiệp thuộc Viện Điện, Đại học Bách khoa Hà Nội.

Lĩnh vực nghiên cứu hiện nay: hệ thống tự động hóa điều khiển sử dụng tương tác người-máy, thiết bị đo cá nhân, môi trường cảm thụ.

Điện thoại: 0438683087

Email: ngoc-yen.pham@mica.edu.vn

NGUYỄN THỊ LAN HƯƠNG

VHiện là viện phó Viện Điện, đồng thời tham gia nghiên cứu tại Viện nghiên cứu MICA, Đại học Bách khoa Hà Nội.

Lĩnh vực nghiên cứu hiện nay: tổng hợp thiết bị đo, xử lý tín hiệu đặc biệt tín hiệu đo lường, môi trường thông minh, mạng cảm biến (không dây)

Điện thoại cơ quan: 0438683087

Email: lan-huong.nguyen@mica.edu.vn