

UMA PLATAFORMA DE BAIXO CUSTO COMANDADA POR VOZ PARA TECNOLOGIAS ASSISTIVAS COM PROGRAMAÇÃO EM PYTHON

V. C. A. Pimentel*, I. L. Barbacena*, S. E. N. Correia* e M. C. Dias**

*Programa de Pós-Graduação em Engenharia Elétrica, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, Brasil

** Coordenação do Curso Superior de Bacharelado em Engenharia Elétrica, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, Brasil
e-mail: victor.andrade@ifrn.edu.br

Resumo: Descreve o projeto de uma plataforma básica para o reconhecimento de comandos de voz em sistemas embarcados que proporciona uma redução do custo e do tempo de desenvolvimento de tecnologias assistivas controladas por voz. A plataforma consiste de um módulo de reconhecimento de voz (*Voice Recognition Module V2*) e uma plataforma de desenvolvimento *BeagleBone Black*. Inicialmente foi desenvolvida uma aplicação para o controle de um led RGB utilizando o sistema operacional Ubuntu Linux configurado com o ambiente de desenvolvimento para a linguagem Python. Destaca-se a potencialidade de se estender a utilização do dispositivo para numerosas aplicações que possam auxiliar pessoas com mobilidade reduzida.

Palavras-chave: Sistemas embarcados, reconhecimento de comandos de voz, tecnologias assistivas.

Abstract: Describes the design of a basic platform for recognizing voice commands in embedded systems to provide a reduction in costs as well as reduce the development time of assistive technologies controlled by voice. The platform consists of a voice recognition module (*Voice Recognition Module V2*) and a development platform *BeagleBone Black*. Initially an application was developed to control an RGB led using the Ubuntu Linux operating system configured with the development environment for the Python language. Highlights the potential to extend the use of the device for numerous applications that can assist people with disabilities.

Keywords: Embedded systems, recognizing voice commands, assistive technologies.

Introdução

A busca por maior independência e autonomia para pessoas com deficiência tem se apresentado como fator decisivo na melhoria da qualidade de vida desses indivíduos. O desenvolvimento de tecnologias assistivas visa prover melhoramentos na forma como se comunicam, em sua mobilidade, em passarem a ter um maior controle sobre o ambiente onde estão inseridos e aprimorar suas experiências de aprendizado e interação social [1].

Dispositivos e sistemas controlados por voz têm sido empregados em diversas soluções [2, 3, 4, 5, 6, 7, 8].

Neste artigo é apresentado um projeto de integração entre um módulo de reconhecimento de voz (*Voice Recognition Module V2*) [9] e uma plataforma de desenvolvimento para sistemas embarcados (*BeagleBone Black*) [10], ambos de baixo custo.

Pretende-se disponibilizar a arquitetura flexível e alta capacidade de processamento da referida plataforma para o tratamento e tomada de decisões, a partir de comandos por voz, permitindo a utilização dessa interface no desenvolvimento de sistemas que auxiliem no cotidiano de pessoas com mobilidade reduzida como, por exemplo, idosos, paraplégicos ou, até mesmo, tetraplégicos, proporcionando a esses indivíduos maior sensação de segurança, controle e independência com o uso de tecnologias assistivas abertas (*open source*) de baixo custo que ajudem em tarefas como controlar eletrodomésticos, iluminação ambiente, dispositivos de cuidados com a saúde, etc.

Materiais e métodos

A Figura 1 ilustra o diagrama em blocos da interface implementada, destacando-se que o módulo de reconhecimento de voz processa o áudio, enquanto a plataforma *BeagleBone Black* recebe os sinais digitais referentes a cada comando, podendo utilizá-los nas mais diversas aplicações.

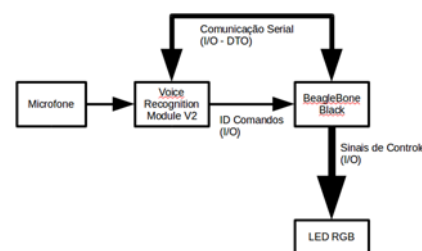


Figura 1: Diagrama em blocos da plataforma para comandos de voz.

Para validar o reconhecimento dos comandos de voz executados pela plataforma, foi montado um circuito básico para o controle das luzes de um *led RGB*, tendo sido produzido um vídeo demonstrativo do funcionamento do dispositivo, conforme referenciado na Figura 2.

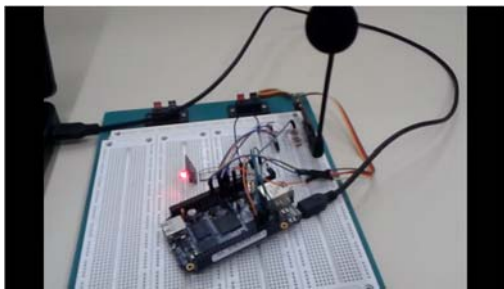


Figura 2: *Frame* do vídeo demonstrativo do funcionamento da plataforma [11].

BeagleBone Black – O *BeagleBone Black* se constitui numa plataforma de desenvolvimento que conta com uma vasta comunidade de suporte.

Dentre suas especificações, possui um processador AM335x 1GHz ARM® Cortex-A8, 512MB DDR3 RAM, memória flash interna de 4GB, acelerador gráfico 3D, dentre outras funcionalidades e recursos que ampliam seu potencial para o desenvolvimento de aplicações de sistemas embarcados [10]. Além disso, sua compatibilidade com diversos sistemas operacionais *open source* – como Debian, Ubuntu, Android e outros – e por se tratar de uma plataforma *open hardware*, apresenta uma alta flexibilidade quanto às ferramentas de desenvolvimento, podendo adicionar facilmente novos módulos de *hardware*, abrindo oportunidades para o desenvolvimento das mais diversas aplicações. Com alto potencial para sistemas ubíquos e pervasivos, ou ainda alinhados ao conceito de Internet das Coisas (*Internet of Things* – IoT).

A referida plataforma foi configurada com o sistema operacional Ubuntu Linux, que possui em sua instalação padrão as ferramentas de desenvolvimento para a linguagem Python.

Foi necessário ainda, a configuração de uma biblioteca para o acesso aos pinos de entrada e saída da plataforma *BeagleBone* utilizando a linguagem Python.

Com o ambiente adequadamente preparado, foi elaborado um código, desenvolvido em linguagem Python, para a recepção dos sinais provindos do módulo de reconhecimento de voz, que identificam cada comando e geram uma saída correspondente ao comando de voz acionado, de modo a controlar o acendimento ou desligamento das luzes do *led RGB*.

Voice Recognition Module V2 – É um módulo de reconhecimento de comandos de voz, dependente de locutor, de baixo custo, e que pode ser utilizado para controlar os sistemas num carro ou outros dispositivos eletroeletrônicos.

É capaz de armazenar até 15 comandos de voz, os quais são organizadas em 3 grupos. Assim, é necessário antes realizar o treinamento dos comandos de voz no referido módulo. Cada grupo de comandos só pode ser ativado individualmente, de modo que somente os 5 comandos do grupo atual serão disponibilizados.

Dessa forma, para o teste de validação da plataforma, foram gravados somente 5 comandos, conforme a Tabela 1.

Tabela 1: Comandos para validação da plataforma.

Comando de voz	Pino	Resposta esperada do sistema
/rubro/	O1	Acendimento da luz vermelha
/verde/	O2	Acendimento da luz verde
/azul/	O3	Acendimento da luz azul
/branco/	O4	Acendimento das luzes vermelha, verde e azul
/apaga/	O5	Desliga as luzes do led

A realização do treinamento do módulo, bem como ativação e desativação de seus modos de funcionamento é realizada enviando-se comandos via interface serial, conforme ilustrado na Figura 1. Além disso, através dessa mesma interface, é selecionado o grupo de comandos que deverá estar ativo num dado momento. Disponibiliza ainda a opção de que o grupo de comandos seja carregado a partir da utilização de pinos de seleção (GCH e GCL), conforme o manual [9].

As memórias internas onde os comandos são gravados durante o treinamento suportam elocuições de, no máximo, 1300ms, de modo que o sistema somente suporta comandos de palavras isoladas.

Por questões de diferença entre os níveis de tensão gerado pelas saídas do módulo de reconhecimento de voz (5V) e o nível suportado pelos pinos de entrada do *BeagleBone* (3,3V), foram utilizados divisores de tensão, evitando a queima das portas na plataforma. A Equação 1 apresenta o cálculo para dimensionamento dos divisores de tensão. Foram utilizados resistores comerciais de 330Ω e 1KΩ.

$$R_2 = \frac{-V_2 R_1}{V_2 - V_1} \quad (1)$$

Para o acionamento das luzes do *led RGB*, foram utilizados resistores limitadores de corrente (220Ω cada).

Testes realizados – De acordo com o manual do módulo *Voice Recognition Module V2* [9], sua precisão quanto ao reconhecimento de comandos num ambiente ideal é de 99%.

No entanto, mostra-se necessário sua verificação em um ambiente ruidoso, em que o sistema será naturalmente utilizado. Para tanto, foram realizados testes em um ambiente com nível de ruído na faixa de 45 a 65dB, com vistas a verificar a precisão do sistema quanto ao reconhecimento de comandos num contexto próximo de uma situação real de utilização.

Foram convidados 10 locutores para a os testes de validação, em seções individuais, sendo 5 locutores do sexo feminino e outros 5 do sexo masculino. Cada um

dos locutores realizou o treinamento do módulo de reconhecimento de voz com os comandos da Tabela 1.

Posteriormente, a execução do código implementado em linguagem Python realiza o envio de um comando no formato “*Head + Key*” [9], em hexadecimal, ao módulo de reconhecimento de voz, através da interface serial, fazendo com que o referido módulo importe o grupo de instruções de teste, definidas na etapa de treinamento, e permaneça no estado pronto para recebê-las.

Nesse estado de operação, cada locutor forneceu 20 repetições para cada um dos comandos de voz através do microfone, tendo sido realizada a execução de 1000 comandos por diferentes indivíduos, observando-se o comportamento do sistema, conforme a Tabela 1.

Resultados

A Tabela 2 apresenta a taxa de acertos por locutor, bem como a taxa média de acertos, relacionada à totalidade dos 1000 comandos executados.

Tabela 2: Resultados dos testes realizados

Locutor	Taxa de acertos (%)
Locutor 1	96%
Locutor 2	97%
Locutor 3	98%
Locutor 4	90%
Locutor 5	97%
Locutor 6	92%
Locutor 7	97%
Locutor 8	95%
Locutor 9	97%
Locutor 10	100%
Taxa média de acertos	95,9%

Discussão

Os resultados permitem concluir que, em ambientes ruidosos, a taxa média global de acertos (de 95,9%) é bastante alta, garantindo a validação do sistema. No entanto, durante a realização dos testes por cada locutor, ocorreram momentos isolados em que o comportamento desviou-se do esperado. Por exemplo, o acendimento do *led* verde para o locutor 5, não funcionou como esperado em 25% das vezes. Além disso, certa quantidade de vezes, o sistema permaneceu inoperante após ter sido fornecido o comando, representando 3,5% das vezes.

A escolha de comandos com sonoridade semelhante (homófonos) aumenta a ocorrência de acionamento indevido dos *leds*, podendo dois comandos resultar na mesma resposta. Há ainda relação entre a ocorrência de erros e o cansaço vocal do locutor, tendo sido as repetições dos comandos realizadas em sequência. Essas questões foram também discutidas em [12].

Não raras vezes, o sistema foi acionado acidentalmente, influenciado pelo ruído provindo de conversas de terceiros no ambiente de teste (situação próxima da real). Assim, as aplicações desenvolvidas utilizando o dispositivo proposto devem apresentar mecanismos que tratem questões de segurança, fornecendo modalidades de entrada diversas para ativar ou desativar o mecanismo comandado por voz [7].

A possibilidade de treinamento utilizando a língua português brasileiro permite sua ampla aceitação no país, uma vez que os usuários com mobilidade reduzida, como idosos ou indivíduos com lesão na coluna vertebral, paraplégicos ou tetraplégicos, tendem a preferir utilizar sua própria língua no controle por voz de dispositivos [5].

Um dispositivo ubíquo controlado por voz na língua árabe para ajudar pessoas com deficiência foi implementado em [5]. O autor apresenta 3 casos de teste utilizando o dispositivo projetado: o controle de mouse e teclado, o controle de cadeira de rodas e o controle de braço robótico industrial. A unidade de controle utilizada foi um microcontrolador Atmega8, da Atmel.

A utilização de microcontrolador é uma prática recorrente para esses dispositivos, conforme observado também em [2]. No entanto, a utilização de unidades de controle baseadas em computador, tendo um módulo externo específico responsável pela tarefa de realizar o reconhecimento de comandos por voz também vem sendo observada na literatura [3], onde a unidade de reconhecimento de voz se constitui numa placa que funciona como “*plug-in*” para o PC.

Essa prática tem o intuito de retirar o pesado processamento de áudio do computador (*Beaglebone Black*), de modo que os comandos de voz sejam utilizados somente para sinalizar tarefas, liberando a capacidade de processamento da plataforma para a utilização de recursos avançados, que podem exigir alta capacidade de processamento, como, por exemplo, a associação de diversas outras modalidades de entrada, permitindo o desenvolvimento de aplicações com interfaces multimodais [6, 13], de modo a aperfeiçoar a experiência dos usuários com mobilidade reduzida; como também o controle de diversos tipos de dispositivos utilizando protocolos de comunicação avançados, característica concernente ao caráter de uso multidisciplinar e estrutura heterogênea dos sistemas inteligentes de automação, podendo ainda envolver questões de segurança [8].

Nesse sentido, a grande flexibilidade e compatibilidade com as mais variadas tecnologias proporcionadas com a utilização da plataforma apresentada permitem que sua alta capacidade de processamento esteja disponível para as aplicações que demandem recursos avançados, fazendo com que os sinais enviados pela plataforma para o controle do acendimento e desligamento das luzes do *led RGB* possam ser adaptados para comandar os mais diversos tipos de dispositivos, otimizando o desenvolvimento rápido de aplicações inicialmente proposto. Além disso, o baixo custo viabiliza ainda mais sua utilização no

desenvolvimento de soluções acessíveis. A Tabela 2 apresenta os custos relacionados à plataforma.

Tabela 2: Custos dos componentes do projeto.

Item	Qtd.	Valor em 14/07/2014 (R\$)
<i>BeagleBone Black</i>	1	121,76
<i>Voice Recognition Module V2</i>	1	52,03
Resistor 220 Ω	3	0,30
Resistor 330 Ω	5	0,50
Resistor 1K Ω	5	0,50
Total		175,09

Assim, além de possibilitar o desenvolvimento de aplicações que envolvam maior complexidade, se atinge também uma redução nos custos dos dispositivos, o que é percebido quando comparado ao sistema apresentado por [4], em que, somente a interface de reconhecimento de voz, gera custos da ordem de \$329,98 dólares (R\$730,54 em 14 de Julho de 2014), o que sinaliza uma redução de gastos de cerca de 76%. Com relação ao dispositivo apresentado em [2], que tem uma previsão de custos da ordem de \$200 dólares (R\$ 447,20 em 04 de Setembro de 2014), o dispositivo aqui proposto apresenta-se 60% mais barato.

Conclusão

É apresentada uma interessante alternativa para o desenvolvimento rápido de tecnologias assistivas. O pequeno tamanho de seus componentes permite que seja desenvolvido um encapsulamento que possa ser adaptável a fixação em locais como uma cabeceira de cama ou adaptado a uma cadeira de rodas. O baixo custo dessa interface de reconhecimento de comandos de voz contribui para reduzir o preço final das aplicações que venham a ser desenvolvidas. Os resultados dos testes realizados apontam a confiabilidade necessária para sua adoção.

Convém estender a utilização dessa solução para aplicações com entradas multimodais para o controle de iluminação do ambiente, controle de aparelhos eletrodomésticos, integrando-os com mecanismos de segurança, bem como dispositivos domésticos de cuidados com a saúde (*Home Health Care*).

Referências

- [1] Bersch R. Introdução à Tecnologia Assistiva. Assistiva® Tecnologia e Educação. Porto Alegre, 2013 [cited 2014 Set 3]. Available from: http://www.assistiva.com.br/Introducao_Tecnologia_Assistiva.pdf.
- [2] Jiang H, Han Z, Success P, Robidoux S, Sun Y. Voice-Activated Environmental Control System for Persons with Disabilities. In: Proceedings of the IEEE 26th Annual Northeast Bioengineering Conference; 2000 Sep 08-09; Storrs, Connecticut,

Estados Unidos. 2000. p. 167-8.

- [3] Aguilera F, Ataaefar A, Brothers R, Castellano M, Ginart A, Grangeia G, La L, McCrae W, McKernan M, Moonsammy A, Noel Y, Padilha J, Patel B, Reyes R, Ryan J, Rydzak T, Shan J, Sinha T, Valatkavage J, Zelano JA, Members, IEEE. A Personal Computer Based Environmental Control System For The Disabled. In: 14th Annual International Conference of the Engineering in Medicine and Biology Society; 1992 Oct 29 – Nov 1; Paris, França. 1992. p. 1533-4.
- [4] Hall B, Molloy J. Installing A Voice Activated Environmental Control Unit For Under 500 Dollars. RESNA 26th International Annual Conference [internet]. 2003 Jun [cited 2014 Jul 14]. Available from: http://web.resna.org/conference/proceedings/2003/Papers/EA/Hall_EA.htm.
- [5] Shakir, U. Ubiquitous Arabic Voice Control Device To Assist People With Disabilities. In: 4th Conference on Intelligent and Advanced Systems; 2012 Jun 12 – 14; Kuala Lumpur, Malaysia. 2012. p. 333-8.
- [6] Vallés M, Manso F, Arredondo T, Pozo F. Multimodal Environmental Control System For Elderly And Disabled People. In: 18th International Conference of The IEEE Engineering in Medicine And Biology Society; 1996 Oct 31 – Nov 3; Amsterdam, The Netherlands. 1996. p. 516-7.
- [7] Lu Y, Chen Y. Prototyping Potential Control Systems to Assist Complete Quadriplegics. In: The Biomedical Engineering International Conference; 2012 Dec 5-7. p. 1-5.
- [8] Uzunai Y, Bicakci K. SHA: A Secure Voice Activated Smart Home For Quadriplegia Patients. In: IEEE International Conference on Bioinformatics and Biomedicine Workshops; 2007 Nov 2-4. p. 151-8.
- [9] Shen W. Voice Recognition Module V2. Elechouse [internet]. 2013 Mar [cited 2014 Jul 12]. Available from: <http://www.elechouse.com/elechouse/images/product/Voice%20Recognition%20Module/Manual.pdf>.
- [10] Beagleboard. BeagleBone Black. Beagleboard [internet]. 2014 [cited 2014 Jul 12]. Available from: <http://beagleboard.org/black>.
- [11] Pimentel V. Reconhecimento de Comandos por Voz. [internet]. 2014 Sep. Available from: <https://www.youtube.com/watch?v=53tq-M9TslQ>
- [12] Silva L. Reconhecimento de Voz Utilizando o Kit SR07 HM2007 [trabalho de conclusão de curso]. João Pessoa: Instituto Federal de Educação, Ciência e Tecnologia da Paraíba; 2013.
- [13] Hui P, Meng H. Latent Semantic Analysis for Multimodal User Input With Speech and Gestures. In: IEEE/ACM Transactions On Audio, Speech, And Language Processing, Vol. 22, n. 2, 2014 Feb. p. 417-29.