

Implementação de um Sistema de Reconhecimento de Voz em Protótipos para a Comunicação LIBRAS-Língua Portuguesa

Maria Alice Trinta¹, Vinicius Souza de Jesus¹, Carlos Eduardo Pantoja¹

¹ Centro Federal de Educação Tecnológica Celso Suckow (CEFET/RJ)
20785-220 – Rio de Janeiro – RJ – Brasil

atg200210@outlook.com, souza.vdj@gmail.com, pantoja@cefet-rj.br

Abstract. *This work presents a voice recognition system that was implemented to LuBRAS: a prototype that performs the translation between LIBRAS and the Portuguese language. The original architecture of LuBRAS was extended to add a mechanism to capture a spoken word and transform it into a written word, giving the signs that will be interpreted by the learning glove.*

Resumo. *Este trabalho apresenta um sistema de reconhecimento de voz que trabalha em conjunto com os protótipos de comunicação Língua Brasileira de Sinais (Libras)-Língua portuguesa conhecido como LuBRAS. A arquitetura original do LuBRAS foi estendida para que fosse possível adicionar um mecanismo de captura da palavra falada e a transformar em sinais que serão interpretados pela luva de aprendizado.*

Keywords: LIBRAS; portuguese language; communication glove; learning glove; Voice recognition system.

Palavras-chave: LIBRAS; língua portuguesa; luva de comunicação; luva de aprendizado; Sistema de reconhecimento de voz.

1. Introdução

A Tecnologia assistiva [Bersch 2008] é um termo referente ao conjunto de serviços e recursos que contribuem para propiciar ou aperfeiçoar habilidades funcionais a pessoas com deficiência, que ajuda a promover tanto independência, quanto a inclusão social. Na literatura é possível encontrar diversos trabalhos que visam melhorar a comunicação entre surdos e ouvintes, como por exemplo o *SensorLibras* [Tavares et al. 2010] que faz uma tradução da Língua Brasileira de Sinais (LIBRAS) para o português, e o *Hand-Talk* [Tenório et al. 2019] que traduz textos do português para a LIBRAS. Entretanto, estes trabalhos são unidirecionais funcionando apenas de ouvinte para surdo ou de surdo para ouvinte.

Para tratar tais limitações, o trabalho *LuBRAS* [Jesus et al. 2016], que foca na melhora da comunicação entre surdos e ouvintes, e também possibilita a promoção do ensino da LIBRAS, trabalhando assim de forma bidirecional. Porém, o protótipo era pouco uma vez que suas interfaces de interação destinadas ao ouvinte permitiam apenas inserção de sentença através de digitação, porém o protótipo não era prático uma vez que

suas interfaces de interação destinadas ao ouvinte permitiam apenas inserção de sentença através de digitação, podendo deixar a comunicação lenta. Logo, foi identificado que a implementação de um mecanismo de reconhecimento de voz nas interfaces trará mais praticidade e usabilidade aos protótipos.

Portanto, o intuito deste trabalho é implementar ao *LuBRAS* um sistema de reconhecimento voz que atua como uma alternativa de entrada na interface destinada ao ouvinte, tornando possível inserir sentenças tanto oralmente, quanto por meio da digitação, fazendo com que a tradução se torne mais prática, adaptável e eficiente. Para isso, a arquitetura original do fluxo de comunicação do *LuBRAS* foi estendida para permitir que a funcionalidade de reconhecimento de voz seja possível através de um sistema que identifique uma sentença a partir da produção sonora do usuário utilizando um captador de áudio juntamente de um *software* que fara o reconhecimento da fala. Como avaliação experimental, uma validação por prototipação a partir de um cenário e prova de conceito foram apresentados.

Este trabalho está estruturado da seguinte maneira: Na segunda seção estão presentes os referenciais teóricos; Na terceira seção é discutida a metodologia; A quarta seção traz a avaliação experimental; A quinta seção apresenta os trabalhos relacionados; A sexta seção apresenta a conclusão e os futuros trabalhos; Por fim são declaradas as referências utilizadas.

2. Referencial Teórico

Nesta seção são apresentados os conceitos e tecnologias que fazem parte do desenvolvimento deste trabalho e auxiliam na compreensão do protótipo de forma geral. Nos protótipos do *LuBRAS* são utilizados *softwares* que fazem a interface humano-hardware e estão embarcados em placas microcontroladas e em computadores com tamanho reduzido com capacidade para manter um sistema operacional e executar *software* com determinada robustez.

Esses sistemas que independem de periféricos são chamados de Sistemas Embarcados [Cunha 2007] que são construídos para atender a uma determinada aplicação, muitas vezes oferecido ao usuário como o responsável por um dispositivo completo e independente. Diferentemente dos computadores, que possuem capacidade de executar inúmeros aplicativos, os sistemas embarcados executam isoladamente, não tendo permissão ou capacidade, na maioria das vezes, para a execução de outras funcionalidades se não aquelas pré-determinadas em seu projeto.

Dessa forma, o *LuBRAS* utiliza uma abordagem híbrida onde o microcontrolador precisa tanto executar suas funções independentes quanto comunicar-se com o microprocessador, vice-versa, estabelecendo uma comunicação serial. E para estabelecer essa interface de comunicação foi utilizado o Javino [Lazarin and Pantoja 2015], uma biblioteca criada inicialmente para fazer a comunicação entre a linguagem de programação Java e a placa microcontrolada Arduíno, oferecendo uma comunicação de mão dupla, tanto do Java para o Arduíno quanto do Arduíno para o Java de maneira segura.

Os sistemas de reconhecimento de fala [Ishizaki 2018], ou sistemas de reconhecimento de voz, utilizados neste trabalho funcionam de maneira a converter um si-

nal acústico, capturado por um transdutor como um microfone, para um conjunto de palavras. O sinal de fala é obtido através da conversão de ondas sonoras em sinais elétricos. O Sistema de reconhecimento de voz que foi implementado ao dispositivo foi construído a partir de uma *Application Interface Program* (API) chamada *Java Speech* [Marangoni and Precipito 2017].

A *Java Speech* API (JSAPI) possui como principal característica ser totalmente proveniente da linguagem Java, apresentando duas funções: o reconhecimento da fala e a síntese da fala. O reconhecimento da fala permite ao computador detectar o áudio, processá-lo e transportá-lo de áudio para texto o que foi dito pelo usuário. Já a síntese da fala gera uma fala sintética que se baseia no texto fornecido por um usuário, sendo assim o processo inverso do reconhecimento da fala, chamado de *Text-to-speech* (Texto para fala).

3. Trabalhos relacionados

O trabalho *LuBRAS* [Jesus et al. 2016] é composto por duas luvas, a luva de aprendizado que é utilizada pelo ouvinte, e a luva de comunicação que é utilizada pelo surdo, e duas interfaces gráficas didáticas, uma destinada ao surdo e outra ao ouvinte, com o objetivo de melhorar a comunicação entre LIBRAS e língua-portuguesa.

Portanto, nesta seção são apresentados alguns trabalhos que também focam na comunicação entre LIBRAS e língua portuguesa. O *SensorLibras* [Tavares et al. 2010] possui uma arquitetura organizada em três elementos principais: o primeiro são duas luvas responsáveis pelo sensoriamento dos sinais, o segundo é um dispositivo que recebe via frequência os dados das luvas, e o terceiro é um cabo *USB* ligando o dispositivo anterior à um computador onde serão feitas as leituras de dados e a tradução propriamente dita. Porém, a tradução não é prática uma vez que o sistema, por não ser embarcado, precisa de periféricos como um computador à disposição. Já o *LuBRAS*.

Outro trabalho é o *HandTalk* [Tenório et al. 2019], um aplicativo que faz a tradução de um texto para a LIBRAS, possuindo tanto um módulo para a tradução de páginas *Web* quanto uma aplicação para celulares. Entretanto, o trabalho realiza somente a tradução de textos para LIBRAS e, além disso, é dependente de uma conexão com a internet para realizar a tradução de modo ideal, enquanto o *LuBRAS* faz traduções de ambas as vias e não depende de conexões externas para obtenção de dados.

O *LuBRAS* ao permitir interações somente através do *display*, tornava a comunicação lenta e limitada. Problema que foi solucionado com a implementação de um módulo de reconhecimento de voz, que tornou a comunicação mais natural e a interação com o dispositivo mais dinâmica.

4. Metodologia

Nesta seção será apresentada a metodologia referente à criação do sistema de reconhecimento de voz, suas estruturas e tecnologias, bem como sua atuação em conjunto das interfaces destinadas ao ouvinte presentes no protótipo.

O *LuBRAS* dispõe de duas interfaces gráficas, uma dedicada ao surdo e outra ao ouvinte, e dois protótipos de luvas. O primeiro protótipo de luva é chamado luva de

comunicação, é utilizada pelo surdo como um meio para interpretar seus movimentos em LIBRAS e traduzi-los para a língua portuguesa. A luva de comunicação pode ser controlada através da interface dedicada ao surdos. O segundo protótipo é chamado luva de aprendizado, é usada pelo ouvinte com a intenção de que a tradução da língua portuguesa para LIBRAS seja didática, ensinando ao ouvinte os movimentos necessários a se fazer para executar a sentença desejada em LIBRAS. A luva de aprendizado pode ser controlada a partir da interface dedicada ao ouvinte.

Existem também meio às interfaces gráficas, duas telas de interação que são utilizadas para fazer a tradução entre LIBRAS e língua portuguesa, sendo que uma delas faz a tradução de LIBRAS para língua portuguesa, montando as sentenças a partir de imagens dos sinais que representam letras. Já a outra interface gráfica faz a tradução de língua portuguesa para LIBRAS, inserindo uma sentença que, através de imagens, terá letra por letra demonstrada em LIBRAS no *display*. O dispositivo é totalmente vestível e dispensa qualquer conexão externa para obtenção de dados, como a *internet*, o que torna o trabalho totalmente móvel e independente.

Um fluxograma que vai do módulo de reconhecimento até as interfaces gráficas da luva será apresentado a fim de auxiliar na visualização do funcionamento do novo método de entrada.

4.1. O sistema de Reconhecimento de voz

Com o objetivo de tornar o *LuBRAS* mais diversificado e usual, foram implementados ao trabalho um dispositivo de captura de fala juntamente de um *software* auxiliar que fará o reconhecimento da fala, para que o áudio reconhecido se transforme em uma sentença que será traduzida para LIBRAS. O novo módulo é uma alternativa de entrada que funcionará juntamente da interface de controle da luva de aprendizado e da interface de tradução Língua portuguesa-LIBRAS.

A extensão da arquitetura consiste na adição da JSAPI juntamente de um captador de áudio, nesse caso um microfone, trabalhando com as interfaces destinadas aos ouvintes como uma alternativa para a entrada de informações que serão encaminhadas pelo restos da arquitetura como mostra o Fluxograma de Funcionamento (figura 1). A JSAPI foi escolhida principalmente por ser inteiramente proveniente da linguagem JAVA, pois o *LuBRAS* foi majoritariamente desenvolvido em JAVA. Em segundo, pois a JSAPI é totalmente embarcável e não necessita de conexão externa com *internet* para obtenção de dados.

Em funcionamento, o programa do reconhecimento de voz grava em uma *String* a palavra falada pelo usuário e disponibiliza através do *display* o que foi reconhecido, para que o usuário confirme e a arquitetura do *LuBRAS* faça a tradução. O novo método de entrada poderá ser utilizado tanto pela interface destinada ao controle da luva de aprendizado ou pela interface de tradução por inserção de texto, ambas interfaces dedicadas ao ouvinte. O *software* de reconhecimento foi embarcado juntamente da aplicação pré-existente do *LuBRAS*, assim como o microfone. Sendo assim, o módulo de reconhecimento de voz age juntamente das interfaces do dispositivo *LuBRAS* para uma usabilidade diversificada e interativa, tornando a comunicação mais fluida e natural.

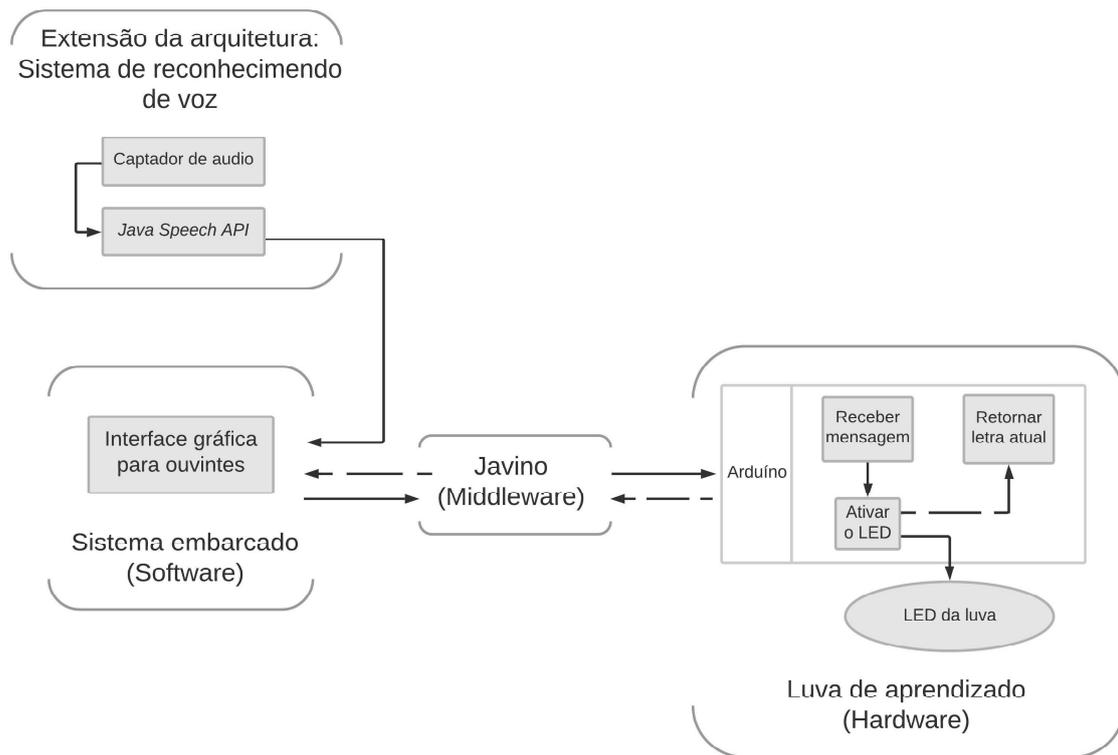


Figura 1. Fluxograma de Funcionamento da Extensão da Arquitetura

5. Prova de conceito

Esta seção tem como objetivo esclarecer o funcionamento do sistema de reconhecimento de voz, juntamente da luva de aprendizado, no meio prático como método de avaliação. Para exemplificar o funcionamento da luva de aprendizado e do sistema de reconhecimento de fala é apresentado uma prova de conceito.

A prova de conceito é apresentada através de uma situação presente em inúmeros estabelecimentos: O número limitado de funcionários que, frente a uma grande demanda, perde em qualidade de atendimento. Sem o comando de voz, utilizar somente as entradas para texto do *LuBRAS*, atrasaria o atendimento uma vez que seria preciso parar todas as ações para digitar as sentenças na interface gráfica da luva para os ouvintes.

Sendo assim, o usuário ouvinte com seu dispositivo *LuBRAS*, aciona o comando de voz através da interface gráfica para ouvintes e, em seguida, o usuário fala uma palavra em língua portuguesa. Essa palavra é reconhecida pelo sistema de voz e é retornada na tela do usuário para que haja a confirmação se a palavra identificada está correta ou não. Não havendo confirmação, o sistema de reconhecimento de voz é acionado novamente para uma nova identificação.

Assim que a palavra for confirmada pelo ouvinte, ela é transformada em um vetor de caracteres que é enviado através do *middleware* Javino para o microcontrolador Arduino onde é feita as associações entre os caracteres em língua portuguesa com os sinais em LIBRAS. Está associação é dada, transformando cada carácter em língua portuguesa

em sinais luminosos que indicam os movimentos necessários para fazer o sinal em LIBRAS referente a cada um destes caracteres.

6. Considerações Finais

O trabalho parte da ideia de inclusão, vendo como necessidade a melhora da comunicação entre surdos e ouvintes, promovendo também o ensino da LIBRAS. O *LuBRAS* busca ser um dispositivo de uso comum no dia-a-dia, por isso, nota-se a mobilidade do trabalho, a não necessidade de obtenção de dados externos via internet por exemplo, e a importância em facilitar o uso do dispositivo através de um *display* otimizado, simples e intuitivo. O objetivo da adição de um sistema reconhecimento de voz é justamente tornar o *LuBRAS* ainda mais prático, uma vez que possibilitou ao usuário uma interação mais flexível, rápida e natural com o dispositivo.

Para trabalhos futuros o *LuBRAS* pretende-se embarcar junto à interface de comunicação um sintetizador de voz, de modo que a palavra traduzida de LIBRAS para língua-portuguesa possa ser escutada pelo ouvinte. Além disso, é desejado desenvolver um protótipo mais robusto e resistente, pois iremos trocar a placa de arduíno por um circuito impresso próprio para cada protótipo com toda a parte de *hardware* embutida no circuito impresso.

Referências

- Bersch, R. (2008). Introdução à tecnologia assistiva. *Porto Alegre: CEDI*.
- Cunha, A. F. (2007). O que são sistemas embarcados. *Saber Eletrônica*, 43(414):1–6.
- Ishizaki, M. Y. (2018). Reconhecimento automático de palavras. B.S. thesis, Universidade Tecnológica Federal do Paraná.
- Jesus, V. S., Manoel, F. C. P. B., Pantoja, C. E., and Samyn, L. M. (2016). Lubras: Uma arquitetura de um dispositivo eletrônico para a comunicação libras-língua portuguesa utilizando o javino. *I Workshop de Pesquisa em Computação dos Campos Gerais (WPCCG)*.
- Lazarin, N. M. and Pantoja, C. E. (2015). A Robotic-Agent Platform for Embedding Software Agents using Raspberry Pi and Arduino Boards. In *9th Software Agents, Environments and Applications School*.
- Marangoni, J. B. and Precipito, W. B. (2017). Reconhecimento e sintetização de voz usando java speech. *Revista Científica Eletrônica de Sistemas de Informação (ISSN 1807-1872)*. Ano, 2.
- Tavares, J. E. d. R., Barbosa, J. L., and Leithardt, V. R. (2010). Sensorlibras: Tradução automática libras-português através da computação ubíqua. In *II Congresso Nacional de Pesquisas em Tradução e Interpretação de Libras e Língua Portuguesa, Universidade Federal de Santa Catarina-UFSC*.
- Tenório, R., Wanderlan, C., and Luz, T. (2019). Hand talk. *handtalk.com*. Accessed: 27 February 2019.