

# Comandos de reconhecimento de voz em soluções corporativas

Marcio Geovani Jasinski<sup>1</sup>, Rodrigo D'avila<sup>1</sup>

<sup>1</sup>Senior Sistemas S/A  
Departamento de Pesquisa Aplicada  
Blumenau – SC – Brasil

marcio.jasinski@senior.com.br, rodrigo.davila@senior.com.br

**Abstract.** *The use of voice recognition in mobile applications has been accelerated due to the advances obtained in recent years. However the use of voice as natural alternative for input in web enterprise systems is not explored. In this paper we have adapted a corporative application to use speech recognition through Web Speech API and regular expressions. We also conducted a qualitative research regarding voice usage on corporate application with voice command prototype. We show that regular expressions are a suitable solution with voice recognition engines in order to map voice into commands. Also, we show that users see voice recognition as an alternative although there is cultural barriers to overcome.*

**Resumo.** *O uso do reconhecimento de voz em aplicações móveis aumenta devido aos avanços que essa área obteve nos últimos anos. No entanto, a utilização da voz como alternativa mais natural em sistemas corporativos não é muito explorada. Neste trabalho, um sistema corporativo foi adaptado para utilizar reconhecimento de voz com a Web Speech API e expressões regulares. Realizamos uma pesquisa qualitativa sobre a opinião dos usuários na utilização do reconhecimento de voz no dia a dia. Demonstramos que o uso de expressão regular com engines de reconhecimento de voz é viável mesmo para aplicações web. Além disso, há uma indicação de que os usuários entendem o reconhecimento de voz como uma alternativa viável mas ainda resistem ao uso por questões culturais.*

## 1. Introdução

O uso de interfaces naturais em sistemas computacionais vem aumentando devido à evolução das tecnologias de toque, gesto e voz. No entanto, aplicações corporativas para *web* ainda utilizam como principal forma de entrada a combinação de teclado e *mouse*. Dentre as interfaces naturais, a voz oferece uma alternativa com liberdade para o usuário interagir com a aplicação enquanto utiliza as mãos para outra atividade.

Os avanços nas técnicas de reconhecimento de voz indicam viabilidade da tecnologia em diversas aplicações, sobretudo em dispositivos móveis. Segundo [Hearst 2011] alguns fatores contribuem para a demanda de interfaces com reconhecimento de voz nesses dispositivos: (1) por se tratar de um caminho natural ao uso da fala e (2) as interfaces de toque dificultam a escrita de textos longos.

Existem tecnologias de reconhecimento de voz maduras que não necessitam de treinamento como Apple Siri, Microsoft Cortana, Google Now. Três fatores contribuíram

para essa maturidade: o aumento da capacidade computacional, a quantidade de dados disponíveis para treinamento e finalmente a evolução dos algoritmos de aprendizado de máquina [Picheny 2015]. Sendo possível identificar padrões complexos de voz com baixo tempo de resposta utilizando dispositivos portáteis e navegadores *web*. Isso sugere que o uso da voz deve aumentar significativamente a medida que o tempo de resposta e a precisão melhoram a ponto de atender as necessidades reais dos usuários [Hearst 2011].

Embora aplicações que utilizam reconhecimento de voz sejam perceptíveis em *smartphones*, o mesmo não é percebido em aplicações corporativas. Algumas razões para a baixa inserção dessa tecnologia são: a) Limitações de portabilidade; b) Suporte exclusivo ao inglês; c) Dificuldade de adaptação do vocabulário corporativo.

Assim como os sistemas de reconhecimentos de voz evoluíram, a computação na nuvem para o meio corporativo também ganhou espaço e maturidade. Aplicações legadas no modelo cliente-servidor estão evoluindo para soluções *web* onde a usabilidade é considerada um fator de sucesso do produto [Haine 2012]. Dessa forma, acredita-se que um modelo de interação mais natural utilizando a voz pode aumentar a competitividade no mercado.

Embora existam *engines* de reconhecimento de voz maduras, segundo [Schnelle et al. 2005], soluções que de fato atendam a demanda do usuário são parte de um tema que ainda não está bem resolvido. Existem desafios para o uso fluente da voz no dia a dia, principalmente em aplicações *web* e com vocabulários complexos.

Nesse trabalho, foi utilizada a *engine* do Google através da *Web Speech API* para reconhecimento de voz visando construir uma solução que atenda a demanda dos usuários conforme levantado por [Schnelle et al. 2005]. A escolha se baseou no suporte ao português do Brasil e na portabilidade. Atualmente a API é limitada ao navegador Google Chrome, mas por seguir a especificação da W3C [Glen Shires 2012], a funcionalidade deve ser adotada pelos demais navegadores.

Esse trabalho traz uma proposta de mapeamento de comandos através de expressões regulares com objetivo de avaliar o reconhecimento de voz em uma aplicação corporativa para gestão de pessoas. As principais contribuições apresentadas neste trabalho são:

1. A indicação da viabilidade do mapeamento da voz para comandos de um sistema usando expressões regulares;
2. Os resultados de um experimento com 15 pessoas sobre uma aplicação corporativa adaptada ao reconhecimento de voz quanto ao índice de acerto dos comandos e a percepção dos usuários com a solução;
3. Identificação de desafios na utilização do Português do Brasil para navegação entre telas de uma aplicação corporativa;

## 2. Trabalhos Relacionados

A literatura sobre técnicas de reconhecimento de voz é abrangente e teve evoluções significativas desde o início do século 21 [Picheny 2015]. No entanto, a utilização e desempenho dos mecanismos mais modernos de reconhecimento como a API do Google para português do Brasil, é pouco explorada. A seguir, os principais trabalhos utilizando a abordagem de reconhecimento de comandos com a *Web Speech API* são discutidos.

Em [Skraba et al. 2014, Skraba et al. 2015] é demonstrado a utilização do reconhecimento de voz no controle de uma cadeira de rodas. A abordagem considera o *Web Speech API do Google* e possui uma limitação análoga ao trabalho atual com a necessidade de conexão com a internet. Porém, essa limitação é mais severa para a cadeira de rodas devido à mobilidade da mesma. Uma diferença em relação à presente proposta está na complexidade dos comandos, uma vez que os comandos de controle da cadeira de rodas são simples e curtos. Finalmente, há uma mudança significativa no objetivo uma vez que buscamos uma validação com usuários finais enquanto [Skraba et al. 2014, Skraba et al. 2015] visaram propor uma solução de *software e hardware*.

O trabalho realizado em [Kimura et al. 2015] é uma aplicação *web* para realizar testes de pronúncia para estudantes não nativos da língua inglesa. Os pesquisadores utilizam o *Web Speech API do Google* para salvar a transcrição e comparar com a resposta do material de ensino. Um desafio comum ao trabalho mencionado é a necessidade de comparar o resultado do reconhecimento de voz com um resultado esperado. A estratégia utilizada pelos autores de [Kimura et al. 2015] foi a comparação de *arrays* enquanto o presente trabalho utiliza expressões regulares. Finalmente, o trabalho não aborda o reconhecimento de voz em português do Brasil, o que caracteriza um desafio uma vez que as principais *engines* do mercado são desenvolvidos para o inglês.

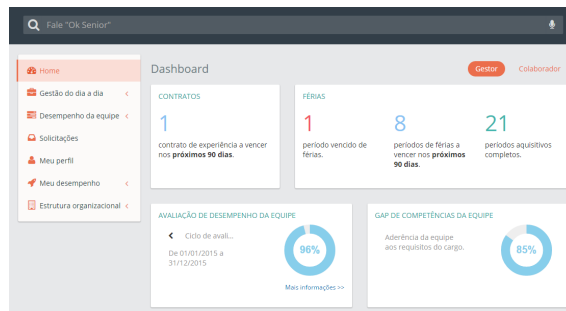
Um sistema autônomo para suporte a idosos é proposto por [Valencia-Redrován et al. 2014]. O sistema realiza o reconhecimento de voz com dois microfones e uma técnica de pré-processamento *fuzzy* para a melhorar a captação. A principal diferença em relação ao trabalho atual é o modelo de entrada de áudio que possui captação de dois microfones com tratamento antes do envio ao *Web Speech API*. No presente trabalho, utilizamos uma abordagem de captação simples para viabilizar a utilização em qualquer ambiente com um navegador e microfone.

A revisão da literatura indica que a utilização do *Web Speech API* é considerada pelos pesquisadores citados como uma alternativa viável a interfaces naturais. No entanto, os problemas encontrados e suas soluções quanto ao mapeamento de reconhecimento de voz para comandos é pouco detalhado e, portanto, esses tópicos são explorados neste trabalho.

### 3. Metodologia

O módulo de reconhecimento de voz foi desenvolvido sobre a interface de uma aplicação de gestão de pessoas conforme mostra a Figura 1. Uma característica comum em uma solução de reconhecimento de voz é a necessidade de um comando de ativação. De forma análoga ao chamar uma pessoa pelo nome para obter sua atenção, é necessário que o usuário indique explicitamente que deseja conversar com a aplicação. No presente trabalho, utilizamos um comando de voz próprio identificado como “*Ok Senior*”.

O diagrama de estados do reconhecimento de voz é apresentado na Figura 2. O sistema inicia esperando que o usuário fale a palavra de ativação - estado (a). Qualquer comando que não seja reconhecido pelo comando de ativação é ignorado. Uma vez que o comando de ativação é reconhecido, a aplicação entra em um modo de espera por um comando de negócio que visa uma troca de tela. A partir desse momento toda captura de voz será avaliada como um comando da aplicação - estado (b). A identificação da fala



**Figura 1. Tela principal da aplicação. As transcrições dos comandos de voz são apresentadas na parte superior da tela, onde está localizada um campo de entrada de dados com o ícone de microfone.**

dispara uma decisão do sistema no estado (c) com três transições possíveis:

- **Cancelar** - Caso o usuário não consiga lembrar o comando, essa opção permite voltar ao estado inicial (a), esperando novamente a frase de ativação.
- **Comando identificado** - Se o comando for identificado a aplicação vai para o estado (c) onde é feito o carregamento da tela, e depois volta ao estado (a).
- **Comando não identificado** - Podem ocorrer casos onde o comando não é identificado. Nesse caso a aplicação volta ao estado anterior (b) para que o usuário tente novamente.



**Figura 2. Diagrama de estados para o módulo de reconhecimento de voz**

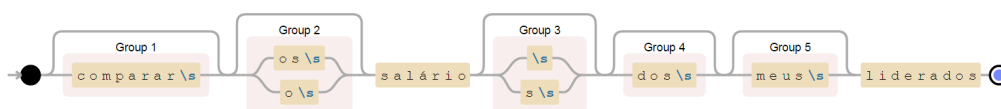
Neste trabalho, limitamos a utilização do reconhecimento de voz a interações de módulos da aplicação, sem possibilitar a edição de formulários ou a navegação dentro de uma página.

### 3.1. Arquitetura

A identificação do comando de voz é formada por dois componentes: o agente de reconhecimento de voz e o interpretador de comandos. O primeiro é responsável por realizar a transcrição do comando de voz para texto e utiliza o *Web Speech API*. A saída do texto é utilizada como entrada do interpretador de comandos que é responsável por verificar se o comando é válido. Nas seções a seguir, o funcionamento destes componentes são descritos com detalhes.

Para a identificação dos comandos, utilizamos expressões regulares com base nos comandos disponíveis nas Tabelas 1 e 2. Como o *Web Speech API* é baseado em eventos, a abordagem de expressões regulares permite avaliar um comando conforme o usuário fala para a aplicação. Cada resultado reconhecido é retornado com índices de confiabilidade que permite identificar o comando através do processo de transcrição para o interpretador de comandos.

As expressões regulares são construídas para permitir uma ou mais flexões por comando. Um exemplo pode ser visto na Figura 3. Esse método adiciona flexibilidade para palavras no singular ou plural e formas distintas de invocar um determinado comando. Entretanto, a abordagem traz um desafio em evitar colisões, ou seja cada comando precisa ter uma estrutura com no mínimo uma palavra que não se repita entre os outros comandos. Se uma transcrição conter mais de um comando, a primeira expressão onde a transcrição for válida é executada, não importando a ordem onde ela apareça no texto da transcrição.



**Figura 3. Exemplo de uma expressão regular para comparação do salário dos liderados.**

### 3.2. Experimento

O experimento foi conduzido com 15 pessoas em ambientes de menor ruído como salas de reunião utilizando equipamento de mercado: um fone/microfone de um celular da linha de celulares Samsung SII. Cada participante recebia um formulário com instruções e os comandos a serem ditados. O processo de reconhecimento foi dividido em duas etapas. A primeira, formada por comandos de estrutura exclusiva do português do Brasil, com variações entre plural e singular, preposições e formato da frase conforme a Tabela 1. A segunda etapa foi formada por um conjunto de comandos onde a formação das frases era mista entre o português do Brasil com siglas e palavras do inglês conforme a Tabela 2.

No formulário, o participante podia marcar quantas vezes tentou cada comando e assinalar se o mesmo foi reconhecido ou se houve desistência. O reconhecimento dos comandos não era obrigatório de forma que a desistência ficava sob critério do participante. Após o teste de reconhecimento, uma avaliação com quatro afirmações utilizando escala Likert era conduzida para coletar a opinião de cada participante sobre a precisão e aplicabilidade da solução.

## 4. Resultados

Os resultados dos experimentos permitem identificar que o índice de acerto dos comandos se mantém acima de 70% para a maioria dos comandos de estrutura simples conforme mostra a Tabela 1. A acerto de 65,5% do primeiro comando pode estar relacionado com a adaptação dos usuários ao iniciar o processo de reconhecimento. Embora os usuários fossem instruídos a fazer alguns testes, muitos optaram por iniciar o processo de reconhecimento de forma mais direta.

Durante o experimento, observamos que os comandos “8 - Contrato de resultados (71,4%)”, “12 - Escolaridade da equipe (50%)”, “18 - Perfil acadêmico (60%)” apresentaram a transcrição correta porém não redirecionaram à página correspondente. Essas anormalidades também foram observadas em outros comandos mas com menor frequência, os detalhes são discutidos na Seção 5.4.

**Tabela 1. Comandos implementados para avaliação de frases com palavras no singular e plural, com e sem preposições.**

| Núm. | Comando   | % de acerto |
|------|---|-------------|
| 1    | Comparar salário dos meus liderados                 | 65,2        |
| 2    | Evolução salarial da equipe                         | 93,8        |
| 3    | Comparar salários dos liderados                     | 78,9        |
| 4    | Evolução da equipe                                  | 88,2        |
| 5    | Salário dos liderados                               | 83,3        |
| 6    | Contrato de resultados da equipe                    | 93,8        |
| 7    | Plano de desenvolvimento individual da minha equipe | 93,8        |
| 8    | Contrato de resultados                              | 71,4        |
| 9    | Plano de desenvolvimento individual da equipe       | 100         |
| 10   | Nível escolaridade                                  | 88,2        |
| 11   | Solicitações dos colaboradores                      | 100         |
| 12   | Escolaridade da equipe                              | 50          |
| 13   | Meu perfil pessoal                                  | 88,2        |
| 14   | Meu perfil profissional                             | 100         |
| 15   | Meu perfil acadêmico                                | 78,9        |
| 16   | Perfil pessoal                                      | 93,8        |
| 17   | Perfil profissional                                 | 100         |
| 18   | Perfil acadêmico                                    | 60          |
| 19   | Contrato de resultado                               | 71,4        |
| 20   | Meu plano de desenvolvimento individual             | 93,8        |
| 21   | Minha equipe  | 75          |
| 22   | Holerite  | 93,8        |
| 23   | Plano de desenvolvimento individual                 | 93,8        |
| 24   | Meu holerite  | 100         |

Os comandos com mistura de idiomas e com siglas resultaram em maior oscilação quanto ao índice de acerto conforme mostra a Tabela 2. É possível identificar que o reconhecimento das palavras estrangeira isoladas é superior ao reconhecimento da mistura de idiomas pelo índice de acerto dos comandos “2 - Feedback (93,78%)” e “4 - Meus Feedbacks (65,2%)”. O comando “3 - Comparar GAP de competências (35,7%)” também teve um desempenho ruim parte pela mistura de idiomas e parte por erro de redirecionamento conforme discutido na Seção 5.4.

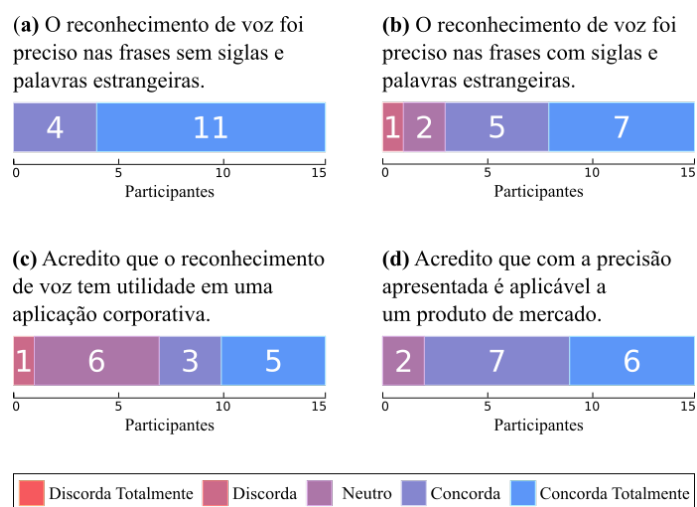
Após o experimento o candidato respondia um questionário baseado na escala de Likert, para avaliar de forma qualitativa o reconhecimento de voz no meio corporativo. Na afirmação (a) da Figura 4, 73,3% dos participantes concordam totalmente que o reconhecimento foi preciso nas frases da Tabela 1. No entanto, somente 46,6% dos participantes concordam totalmente com a precisão dos comandos com siglas e palavras estrangeiras

**Tabela 2. Comandos implementados para avaliação de frases com palavras em inglês e com siglas.**

| Núm. | Comando                      | % de acerto |
|------|------------------------------|-------------|
| 1    | GAP de competências          | 93,8        |
| 2    | Feedback                     | 83,3        |
| 3    | Comparar GAP de competências | 35,7        |
| 4    | Meus Feedbacks               | 65,2        |
| 5    | Meu PDI                      | 36,6        |
| 6    | Dashboard                    | 93,8        |

- Figura 4 - afirmação (b). Mesmo assim, é possível identificar uma posição positiva em relação ao índice de acerto considerando que a concordância parcial e total que obteve 100% para os comandos da Tabela 1 e 80% para os comandos da Tabela 2.

A percepção dos participantes sobre o uso da tecnologia em um produto é apresentado na Figura 4 - afirmação (c). Cerca de 40% dos participantes tiveram uma posição neutra referente a utilidade do reconhecimento de voz no meio corporativo onde 6% discordam da sua utilização, 20% concordam com o uso e 34% concordam totalmente. Já sobre a afirmação (d) não houve nenhuma discordância por parte dos participantes, sendo que 13% tiveram uma posição neutra, mas 47% concordam e 40% concordam totalmente.



**Figura 4. Resultados da avaliação qualitativa sobre o reconhecimento de voz no meio corporativo.**

## 5. Discussões

Nesta seção são discutidas as particularidades observadas com relação a o uso de reconhecimento de voz com expressões regulares bem como percepções as dos experimentos da pesquisa.

### 5.1. Transcrição em diferentes idiomas

O protótipo foi desenvolvido utilizando o português do Brasil como idioma padrão para reconhecimento da fala. No entanto, é comum que as aplicações tenham palavras

estrangeiras o que pode gerar problemas conforme observado por [Skraba et al. 2014, Skraba et al. 2015]. A inserção de palavras em inglês misturadas ao português trazem consequências distintas conforme a presença da palavra no vocabulário e pronúncia do locutor.

As palavras do inglês obtiveram índice de acerto aceitável conforme é possível ver na Tabela 2. Porém quando há mistura de idiomas, esse índice cai uma vez que o reconhecedor tenta compreender os fonemas como frases típicas do português. Por exemplo, a construção “*Meus feedbacks*” é frequentemente transcrito como “*Meu Sedex*”.

Outra constatação é que a pronúncia correta da palavra dentro do idioma nativo, neste caso o inglês, teve pouca influência na taxa de acerto. Os usuários que pronunciaram as palavras estrangeiras como se fossem do português tiveram a mesma taxa de acerto da pronúncia nativa. Esse efeito pode ocorrer devido à transcrição por fonemas e pela inserção de palavras estrangeiras no dicionário da API do Google.

## 5.2. Transcrição de siglas

No universo corporativo a adoção de siglas é comum como forma de simplificar a comunicação. Um analista de negócio utiliza a sigla PDI para Plano de Desenvolvimento Profissional. A transcrição de siglas se demonstrou um desafio uma vez que a frase “Meu PDI” foi frequentemente confundida com outras construções.

Embora a tabela de comandos utilizada no experimento tenha poucas siglas, observamos que com o crescimento do produto, as siglas são um desafio real para a transcrição. Experimentos isolados com as siglas FCA (Fato Causa Ação), ADE (Avaliação de Desempenho da Equipe), PDE (Plano de Desenvolvimento da Equipe) indicaram um baixo índice de acerto. A solução mais indicada para esses casos é evitar as siglas e usar a descrição por extenso.

## 5.3. Mapeamento de comandos

O reconhecimento de voz natural obtido exige um mapeamento constante aos comandos disponíveis na aplicação. Isso significa que nem sempre o usuário vai falar o comando com a mesma estrutura. Por isso é importante identificar e contornar situações onde o mapeamento pode ser otimizado para não frustrar a experiência do usuário:

- **Plural/Singular** - A sutileza que existe na pronúncia do plural pode impedir que algumas palavras sejam corretamente mapeadas da transcrição para a expressão regular. Assim, na expressão regular é importante remover essa diferença aceitando palavras no singular ou plural.
- **Flexibilizar preposições** - O reconhecimento nem sempre identifica as preposições de forma correta e alguns usuários tendem a falar de forma a pouco expressá-las. Assim, tornar as preposições opcionais ajuda o índice de acertos e a absorver as diferenças de comunicação entre usuários.
- **Remapear estruturas incomuns** - Por se tratar de um vocabulário restrito do ambiente corporativo, algumas estruturas podem ser incomuns para o reconhecimento natural esperado por uma *engine* genérica. Esse tipo de situação foi observado no comando de ativação conforme mostra a Figura 5.
- **Mapear sinônimos** - Utilizar diferentes expressões para um mesmo comando como em *holerite* e *folha de pagamento*. Essa abordagem permite a memorização

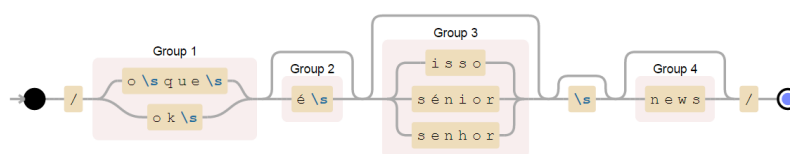


mais natural por diferentes usuários, sem necessidade de forçar que um só comando seja reconhecido.

Embora não tenha sido explorado no presente trabalho, o processamento de linguagem natural pode resolver alguns dos desafios de mapeamento. As palavras transcritas podem ser reduzidas para seus respectivos radicais (*stemming*) e as *stop-words* (preposições e artigos) eliminados o que facilita a identificação dos comandos de forma mais genérica.

#### 5.4. Anormalidades do ditado e das expressões regulares

Alguns comando falados pelos participantes tiveram um transcrição correta mas mesmo assim a aplicação não redirecionou para a respectiva tela. Assim, os participantes marcaram como sendo um erro de reconhecimento o que na verdade foi um erro de aplicação. Acredita-se que esse problema ocorre por pequenas variações dos comandos não estarem mapeados corretamente nas expressões regulares e pela falta do uso de um terminador na expressão regular (\$), identificado após o experimento.



**Figura 5. Montagem da expressão regular para frase de ativação. São verificadas todas as combinações de transcrições para a frase 'Ok Senior'.**

## 6. Conclusões e trabalhos futuros

O presente trabalho apresentou uma abordagem de reconhecimento de voz com a utilização de expressões regulares para o mapeamento de comandos. Os experimentos indicam que a tecnologia já atingiu maturidade suficiente para utilização em aplicações corporativas pela precisão em diferentes comandos. No entanto, como observado no decorrer do trabalho e também na literatura sobre reconhecimento de voz, a mistura de palavras de diferentes idiomas e o uso de comandos com siglas permanece um desafio a ser superado.

Embora a fala seja considerada uma interface mais natural que teclado e *mouse*, a sua utilização em ambientes corporativos pode demorar devido ao ruído do ambiente e até mesmo questões culturais. De forma qualitativa, observamos uma surpresa dos participantes quanto ao alto índice de acerto e quanto a velocidade da aplicação responder aos comandos. Apesar dos resultados positivos, muitos comentaram que ainda não se sentem confortáveis para utilizar uma interface por voz no dia a dia.

A maturidade da *engine* de reconhecimento de voz da plataforma avaliada confirma a revolução no tema do reconhecimento de voz indicada por diferentes autores. Atualmente, é possível oferecer soluções com alto nível de certo para aplicações móveis ou *web* sem demanda de treinamento. No entanto, o entendimento dos comandos pela aplicação não é semântico, ou seja, não há um entendimento do sistema em relação ao desejo do usuário. Esse tema pode ser explorado em futuros trabalhos, de forma que o agente reconhecedor utilize computação cognitiva para compreensão do desejo do usuário.

Durante os experimentos, foi observado uma melhora significativa em alguns comandos da aplicação a medida que o experimento foi repetido. Embora não tenhamos acesso ao funcionamento da *engine* de reconhecimento do Google, a percepção dos pesquisadores foi de um aprendizado sobre alguns termos onde o índice de acerto era baixo no início. No entanto, esse questionamento permanece em aberto pois não encontramos meios de comprovar tal hipótese.

Acreditamos que este trabalho traz contribuições pois indica não apenas a viabilidade do reconhecimento de voz em aplicações *web* com expressões regulares como também aponta os principais desafios do tema. O uso da fala é uma forma de trazer novos horizontes na construção de sistemas e permite a operação de sistemas sem ocupar as mãos do usuário final.

Finalmente, entendemos que o assunto ainda permite avanços como a utilização frases de ativação customizadas, utilização de síntese de voz e a criação de uma assistente pessoal voltado área corporativa capaz de responder perguntas com entendimento semântico.

## Referências

- Glen Shires, H. W. (2012). Web speech api specification. <https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>. Acesso: 2015-09-10.
- Haine, P. (2012). The ux revolution at successfactors. <http://www.successfactors.com/static/docs/successconnect/london/successfactors-philip-haine.pdf>. Acesso: 2015-09-10.
- Hearst, M. A. (2011). 'natural' search user interfaces. *Commun. ACM*, 54(11):60–67.
- Kimura, H., Hayashi, J., Demise, Y., Hasegawa, D., and Sakuta, H. (2015). The effects of listening agent in speech-based on-line test system. In *Global Engineering Education Conference (EDUCON), 2015 IEEE*, pages 366–370.
- Picheny, M. (2015). Ibm watson now brings cognitive speech capabilities to developers. <https://goo.gl/Pmzpb0>. Acesso: 2015-09-10.
- Schnelle, D., Lyardet, F., and Wei, T. (2005). Audio navigation patterns. In *EuroPLoP*, pages 237–260.
- Skraba, A., Kolozvari, A., Kofjac, D., and Stojanovic, R. (2014). Prototype of speech controlled cloud based wheelchair platform for disabled persons. In *Embedded Computing (MECO), 2014 3rd Mediterranean Conference on*, pages 162–165.
- Skraba, A., Kolozvari, A., Kofjac, D., and Stojanovic, R. (2015). Wheelchair maneuvering using leap motion controller and cloud based speech control: Prototype realization. In *Embedded Computing (MECO), 2015 4th Mediterranean Conference on*, pages 391–394.
- Valencia-Redrován, D., Gonzalez-Delgado, L., Robles-Bykbaev, V., Gonzalez-Delgado, N., and Panzner, T. (2014). Sa3m: An interactive robot to provide support for the elderly. In *Power, Electronics and Computing (ROPEC), 2014 IEEE International Autumn Meeting on*, pages 1–6.