

Tecnologias de Interface de Voz Aplicadas no Desenvolvimento de Software para Ensino de Língua Inglesa

Thiago Alexsander da Costa Pereira¹, Denize Ricardi², André Luiz Satoshi Kawamoto¹

¹Departamento Acadêmico de Computação – DACOM
Universidade Tecnológica Federal do Paraná (UTFPR) – Campo Mourão, PR – Brasil

²Departamento Acadêmico de Humanidades – DAHUM
Universidade Tecnológica Federal do Paraná (UTFPR) – Campo Mourão, PR – Brasil

thiago.2014@alunos.utfpr.edu.br, {ricardi, kawamoto}@utfpr.edu.br

Abstract. *The English language is of great importance today, including in the area of Computing, where English terms are used as jargon, naming computer devices and components, among other things. Despite this, most of the Brazilian population, including computer professionals, does not know the English language. There are applications aimed at teaching a second language but mainly focused on grammar and listening, disregarding an essential aspect: pronunciation. The present work presents a Voice User Interfaces technology-based tool that helps learners retain vocabulary and correct pronunciation of technical computing terms.*

Resumo. *A língua inglesa tem grande importância na atualidade, inclusive na área de Computação, onde os termos em inglês são utilizados como jargões, nomeiam dispositivos e componentes do computador, entre outras coisas. Apesar disso, boa parte da população brasileira, incluindo os profissionais da Computação, não tem conhecimento da língua inglesa. Existem aplicativos voltados ao ensino de um segundo idioma, porém focados principalmente em gramática e listening, desconsiderando um aspecto muito importante: a pronúncia. O presente trabalho apresenta uma ferramenta baseada em tecnologias para interfaces de voz que auxilia aprendizes na retenção de vocabulário e pronúncia correta de termos técnicos da Computação.*

1. Introdução

Desde a Primeira Guerra Mundial a língua inglesa começou a se tornar uma das mais faladas tornando-se, com o advento da globalização, num idioma compreendido por pessoas em quase todos os lugares do planeta. Por conta disso, o inglês é a principal língua utilizada em publicações de artigos visando divulgação científica e tecnológica [Graddol 2000].

No que diz respeito à Computação, há uma forte relação com a língua inglesa desde o início de sua popularização. COBOL, concebida para ser facilmente lida, autodoocumentada e manutenível e uma das primeiras linguagens de alto de nível a surgir possuía mais de trezentas palavras reservadas originadas da língua inglesa. Mesmo as linguagens de programação criadas mais tarde, em países não falantes da língua inglesa, ainda utilizam palavras reservadas em inglês, por exemplo Python, Ruby e Lua, criadas respectivamente, na Holanda, Japão e Brasil. Além do uso em linguagens de programação,

diversas empresas de software contratam profissionais dispersos ao longo do planeta, e utilizam a língua inglesa na comunicação entre seus funcionários.

Não obstante, profissionais brasileiros apresentam dificuldade em adquirir inglês como língua estrangeira. Isso se torna mais claro, em grande parte, quando pronunciam termos na língua inglesa e cometem erros como a adição de fonemas, sílabas enfatizadas incorretamente, adição de sílabas extras e utilização de fonemas do português para pronunciar palavras em inglês [Assis 2007].

Nesse sentido, os avanços da tecnologia têm proporcionado a popularização de aplicativos e serviços Web voltados à aprendizagem de um segundo idioma. Melhores processadores, associados à maior capacidade de armazenamento viabilizaram versões *mobile* de aplicativos que anteriormente existiam apenas para Desktop [Godwin-Jones 2011]. Além disso, tecnologias utilizadas em Interfaces baseadas em Voz disponibilizadas em nuvem permitem o acesso a seus recursos em praticamente qualquer lugar.

O presente trabalho apresenta a Pronum-Master, uma ferramenta que visa a retenção da pronúncia correta de termos técnicos em inglês da área da Computação. A Pronum-Master utiliza tecnologia de reconhecimento de fala para identificar a articulação correta de termos, e se baseia em uma versão modificada de um algoritmo de *Spaced Repetition* – uma técnica que contribui para a fixação de novos itens na memória [Kang 2016, Neri et al. 2003, Chukharev-Hudilainen and Klepikova 2016].

O restante desse trabalho está organizado da seguinte maneira: na Seção 2 são apresentados os conceitos teóricos fundamentais para esse trabalho; em seguida, na Seção 3 são apresentados alguns trabalhos relacionados que utilizaram tecnologias de reconhecimento de voz para ensino de idiomas; na Seção 4 é descrito o método adotado para o desenvolvimento do trabalho e, na Seção 5 apresentamos a ferramenta criada; finalmente, na Seção 6, as considerações finais e perspectivas de trabalhos futuros.

2. Fundamentação Teórica

Aqui apresentamos os conceitos de *Spaced Repetition* e da Web Speech API, utilizados na implementação da Pronum-Master.

2.1. Spaced Repetition

Spaced Repetition (repetição espaçada) é uma técnica que consiste em estabelecer curtos períodos de estudo distribuídos ao longo do tempo, baseada em um efeito chamado de “curva de esquecimento” primeiramente observado por [Ebbinghaus 2013], e depois verificado por outros autores no contexto do aprendizado de um segundo idioma [Bloom and Shuell 1981, Atkinson 1972].

A curva de esquecimento representa o declínio de retenção dos termos ao longo do tempo, ou seja uma informação recentemente aprendida tende a ser esquecida caso não seja constantemente revisada [Ebbinghaus 2013]. A Figura 1 apresenta a curva de esquecimento relacionada a uma taxa de retenção de 60% a 100% em um período de 7 dias.

De acordo com essa Figura, a taxa de retenção de um conteúdo no mesmo dia em que se aprende (dia 0) é de 100%. Essa taxa decai nos dias seguintes, podendo alcançar

Curva de esquecimento típica para informações recém aprendidas

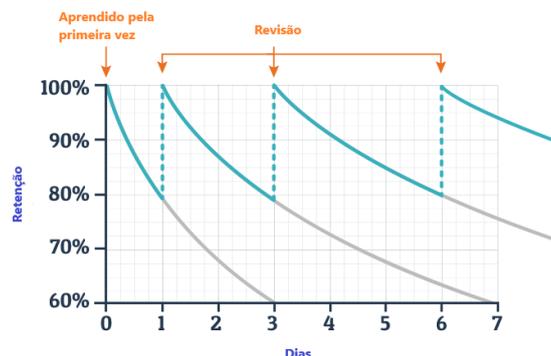


Figura 1. Curva de esquecimento.

60% do conteúdo estudado em três dias. Entretanto, revisar o material eleva a taxa de retenção de volta para 100% e estende o tempo necessário para decair novamente aos mesmos 60%, suavizando a curva de esquecimento. Assim, a retenção de termos estudados pode ser estendida ao longo do tempo, desde que aconteçam revisões periódicas [Ebbinghaus 2013].

É comum associar o uso de *flashcards* (cartões de estudo) a *Spaced Repetition*. Esses cartões geralmente possuem informações em ambos os lados. Na frente, uma imagem, fórmula, texto, associado ao item que se deseja aprender e no verso, as informações (respostas) a respeito deste item.

Quando se automatiza, por meio de um computador, o cálculo do intervalo entre as repetições, seja por meio de uma heurística ou a partir de um algoritmo, temos um *Spaced Repetition System* – SRS. Utilizar um SRS é vantajoso pois determinar a intermitência entre as revisões pode sobrecarregar o aluno devido ao número de *flashcards* que ele pode ter. Assim, prioriza-se os estudos em vez do cálculo dos intervalos.

Diversos algoritmos SRS têm sido propostos para obter um intervalo ideal entre as repetições. Estes algoritmos podem ser simples, como apenas adicionar ou remover um período de tempo específico entre a intermitência dos itens, ou mais complexos, como gerar um intervalo ótimo entre revisões, adaptável ao usuário, forçando que revisões aconteçam antes do item ser esquecido [Lindsey et al. 2014, Metzler-Baddeley and Baddeley 2009, Tabibian et al. 2019].

2.2. Web Speech API

Web Speech API é uma iniciativa do *World Wide Web Consortium* (W3C) para fornecer *Speech Recognition* e *Speech Synthesis* (Reconhecimento e Síntese de Fala) para Aplicações Web por meio de JavaScript. A API suporta comandos curtos de voz e frases contínuas que o usuário pode configurar utilizando uma lista de gramática para definir quais palavras o sistema deve reconhecer [Natal et al. 2020].

Quando acionada a partir de um navegador web comum, essa tecnologia envia um arquivo de áudio para um servidor remoto, que processa e devolve o texto correspondente ao áudio enviado. Assim, é necessário acesso à Internet para que funcione. Tem como vantagem a gratuidade, e funcionar tanto em dispositivos Android quanto em aplicações

Web executadas em navegadores baseados no Google Chrome.

3. Trabalhos Relacionados

Diversos trabalhos utilizam software de reconhecimento de voz e que se baseiam, ou não, em *Spaced Repetition* para ensinar, melhorar ou avaliar a pronúncia dos usuários.

Hincks (2003) investigou se o uso de uma ferramenta de *Speech Recognition* ajudaria na evolução da pronúncia de Inglês em um grupo de estudantes de meia idade. Os estudantes tiveram acesso a uma cópia do software *Talk to Me* (Auralog) e, como suplemento, um curso de 200 horas em inglês técnico. O progresso dos estudantes foi comparado com um grupo de controle que não utilizou o programa. Os resultados indicaram que o uso do programa melhorou a pronúncia dos estudantes que tinham fortes sotaques [Hincks 2003].

Ainda no mesmo ano, uma revisão da literatura buscou avaliar a usabilidade de ferramentas de *Speech Recognition* como forma de treinar pronúncia, além de analisar e discutir o estado-da-arte dessa tecnologia à época. Como conclusão, entre outras coisas, pontua que uma ferramenta ideal de treinamento de pronúncia deve possuir cinco fases (Reconhecimento, Pontuação, Detecção de erro e Feedback) [Neri et al. 2003].

Baturay et. al (2009) verificam os efeitos de uma aplicação web de *Spaced Repetition* sobre a retenção de vocabulário em estudantes de línguas estrangeiras. Usando uma aplicação chamada *WEBVOCLE*, os participantes foram apresentados a uma lista de itens a qual usavam na aplicação. Por meio de testes se determinava o desenvolvimento dos vocabulários adquiridos. Os resultados desse estudo mostram que o uso da aplicação demonstrou-se efetiva para o aumento na retenção de vocabulário entre os participantes [Baturay et al. 2009].

Keder (2009) conceitua *Spacing Effect* e descreve softwares que utilizam *Spaced Repetition* com o intuito de aprender algum conteúdo (SuperMemo, Anki, Mnemosyne e FullRecall). Uma aplicação baseada em *Spaced Repetition* foi criada para a Universidade de Masaryka (na República Checa), que permite, entre outras coisas: a criação de livros didáticos por usuários; a revisão de *flashcards* antes da data prevista; a exclusão de *flashcards* que não sejam do interesse dos usuários; a notificação do momento em que revisões devem ocorrer. Keder concluiu que a aplicação da ferramenta obteve sucesso, com a adoção por muitos alunos da Universidade, e propõe estudos adicionais para analisar os efeitos da utilização da ferramenta [Keder 2009].

Wallace (2016) apresenta a utilização do *Google Web Speech API* para transcrever falas dos estudantes e, em seguida compara o *script* original com o resultado da transcrição obtido pela API. Segundo a autora, a discrepância entre os dois *scripts* pode ser atribuída a erros de pronúncia. Wallace pontua também que o uso efetivo da API só é alcançado se o estudante usar um microfone adequado para clareza das palavras, e que o sotaque dos estudantes não deve diferenciar muito do modelo de reconhecimento da API. Os resultados indicam o benefício de usar *Google Web Speech API* para verificar a pronúncia [Wallace 2016].

A Pronum-Master se diferencia dos trabalhos citados por fornecer uma interface tanto para tutores quanto para estudantes, além de prover feedback para que ambos os atores do processo possam acompanhar a evolução. Além disso, ela adapta o algoritmo de

SR para permitir maior tolerância a erros dos usuários no que diz respeito a erros cometidos ao longo do tempo.

4. Método

O método adotado no projeto é apresentado na Figura 2. Inicialmente, foi feita uma revisão na literatura sobre métodos de aprendizagem de outra língua com o auxílio de computadores. Mais especificamente, ensino de língua inglesa para falantes de língua portuguesa. Ainda nessa fase, investigou-se diferentes tecnologias *Speech-to-text* e uma pequena prova de conceito foi implementada, a fim de verificar a possibilidade de sua utilização.

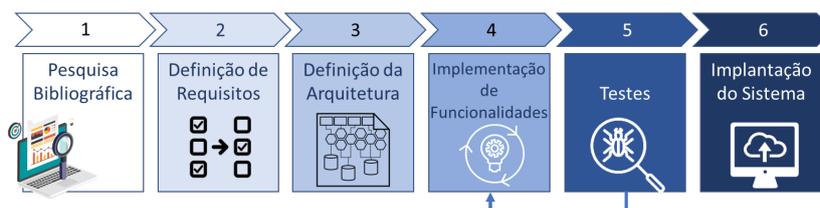


Figura 2. Método

Em seguida, foram definidos requisitos da aplicação e os termos a ser utilizados na versão inicial da ferramenta: palavras reservadas em linguagens de programação e termos da área de computação comumente utilizados.

Duas personas, recurso utilizado na área de Interação Humano-Computador para melhor descrever usuários típicos de um sistema, foram criadas para ilustrar usuários da ferramenta a partir da descrição do público-alvo:

Persona 1: Um professor de inglês em uma escola de idiomas, sempre preocupado em inovar e melhorar a dinâmica de sua aula para que os alunos aprendam de forma mais prática e eficiente.

Persona 2: um estudante de algum curso da área de Informática, que deseja melhorar sua pronúncia a fim de se comunicar melhor com colegas ou mesmo trabalhar remotamente em uma empresa de outro país.

A próxima etapa contou com a definição da arquitetura do sistema, representada na Figura 3. Usuários acessam a ferramenta por meio da internet. A interface Web acessa a aplicação, que se encontra hospedada em um servidor remoto e é integrada a um banco de dados usado para persistir os dados dos usuários, e que utiliza os serviços da Web Speech API para reconhecimento dos termos falados em inglês.

Para a fase de implementação, foi preciso definir o algoritmo de *Spaced Repetition* a ser utilizado na ferramenta. A escolha foi adaptar um algoritmo existente, originalmente proposto por Sebastian Leitner. O algoritmo de Leitner propõe separar *flashcards* em grupos, que são visitados em intervalos crescentes. Quanto maior o intervalo, maior é o fator de retenção do termo, ou seja, o item está melhor memorizado. Os *flashcards* do primeiro grupo são revisados com maior frequência, pois representam aqueles que o aprendiz comete mais erros. Toda vez um item é revisado, duas possibilidades podem acontecer: acerto ou erro. Em caso de acerto, esse item é movido para o próximo grupo (que será

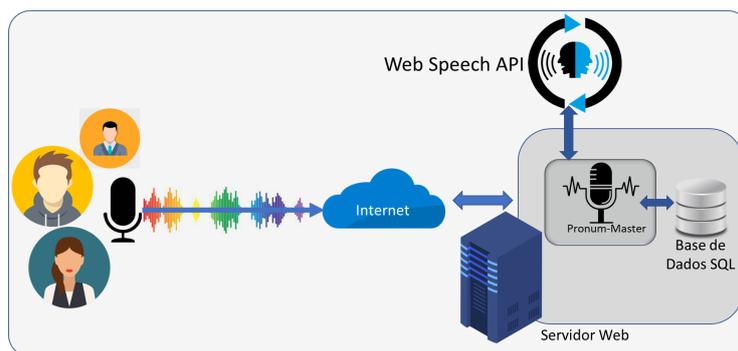


Figura 3. Arquitetura do sistema

visitado num intervalo maior); caso contrário, o item deve retornar para o primeiro grupo, independente do grupo em que se encontra no momento. A vantagem desse método é que o aprendiz pode focar nos cartões mais difíceis, que permanecem nos grupos iniciais [Leitner 1995].

A adaptação implementada emprega o mesmo conceito de *flashcards* agrupados, porém introduz maior tolerância em relação a erros cometidos pelos usuários (Figura 4). Nessa abordagem, o conjunto de *flashcards* é dividido em níveis (Iniciante, Intermediário, Avançado, Final), e cada nível em subgrupos. Para o primeiro nível, o intervalo para reapresentar um item respondido corretamente é dado em horas e é dobrado a cada acerto (quatro, oito e dezesseis horas). Para os demais, esse intervalo é dado em dias, triplicado a cada resposta correta, exceto o último subgrupo do nível “Final”.

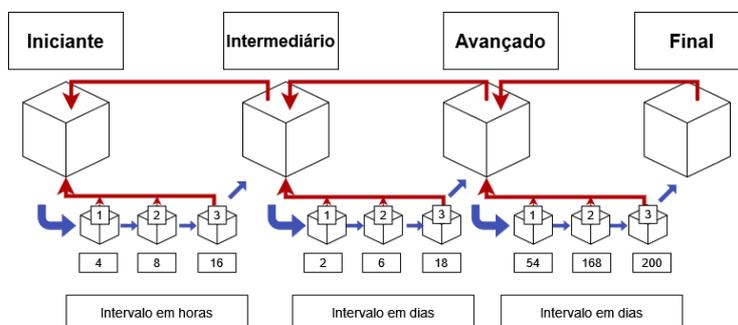


Figura 4. Algoritmo adaptado

Para cada revisão existem quatro possibilidades diferentes:

1. resposta correta para um item que se encontra no último subgrupo do nível “Final”;
2. resposta correta para um item em um subgrupo intermediário;
3. resposta incorreta para um item em um subgrupo intermediário;
4. resposta incorreta para um item que se encontra no primeiro subgrupo do nível “Iniciante”.

No primeiro caso, a pronúncia do item já foi memorizada, e por isso seu *card* pode ser dispensado. Se o *card* ainda não evoluiu até o máximo (caso 2), então ele é movido ou para o próximo subgrupo dentro do mesmo nível, ou para primeiro subgrupo do nível

seguinte. Na condição de pronúncia incorreta o *card* é movido para o subgrupo anterior. Se o item pertencer ao primeiro subgrupo de um nível, então ele deve ser movido para o nível anterior, com intervalo menor de repetição. Itens que se encontram no primeiro subgrupo do nível “Iniciante” (caso 4), que requerem repetições em intervalos menores e devem permanecer inalterados em caso de erro.

Convém salientar que, diferente do algoritmo Leitner original, erros cometidos em *cards* que se encontram em grupos mais avançados não implicam em retornar para o conjunto inicial de *cards*, o que provê maior tolerância a erros eventuais do usuário.

A última etapa tratou de implementar e implantar o sistema. Nessa etapa, novas funcionalidades foram gradualmente introduzidas de acordo com uma abordagem contínua que sugere, se necessário, corrigir imediatamente bugs e adaptações para problemas encontrados. Os testes nessa etapa visavam descobrir problemas como funcionalidades incorretas ou ausentes, erros de interface, erros de comportamento ou desempenho, erros na estrutura de dados adotada e inconsistências de início ou fim. O sistema resultante é apresentado na Seção seguinte.

5. Pronum-Master

Parte da interface inicial do sistema Pronum-Master é exibida na Figura 5. Esta interface apresenta ao usuário uma breve explicação da ferramenta. No canto superior esquerdo há um menu de navegação, destacado em vermelho na Figura. As funcionalidades disponíveis para os administradores são destacadas em amarelo. Esse menu permanece visível em todas as demais interfaces do sistema.



Figura 5. Interface da Aplicação e Menu

Administradores da ferramenta e tutores são capazes de criar novos *cards* e tópicos. A criação de um tópico requer apenas que se informe um nome e uma descrição. Para criar um *card* dentro de um tópico já existente, o administrador deve preencher os campos obrigatórios, sinalizados na interface: palavra, ou seja, o novo termo em inglês; a tradução do termo e o tópico ao qual esse card pertence. Entre os campos opcionais, é possível fornecer links para uma imagem representativa do item, e um vídeo do Youtube contendo a pronúncia correta (Figura 6).

Opcionalmente, administradores podem modificar *cards* existentes (por exemplo substituindo links para figuras ou vídeos), porém não é possível mover os *cards* para outros tópicos. Isso se justifica porque é necessário manter a consistência com as coleções de *cards* já totalmente apreendidas pelos usuários.

Além disso, administradores podem visualizar quais *cards* representam palavras com maior dificuldade para cada usuário. A Figura 7 exhibe essa interface, que contém um seletor onde se elege o usuário cujas informações serão exibidas. Ao escolher o usuário, os cinco *cards* para os quais ocorre a maior taxa de pronúncias incorretas são exibidos.

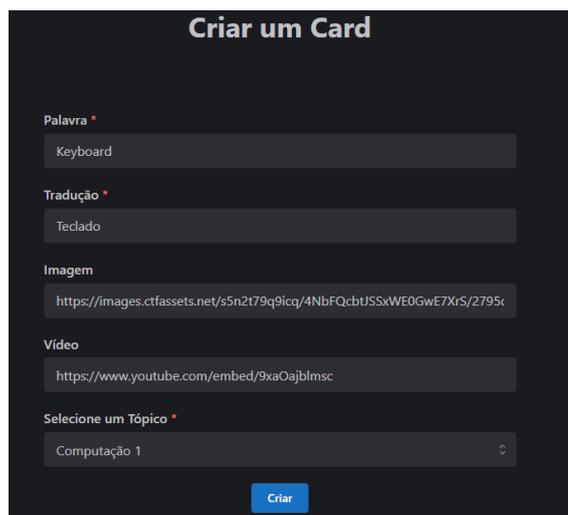


Figura 6. Interface de criação de *cards*

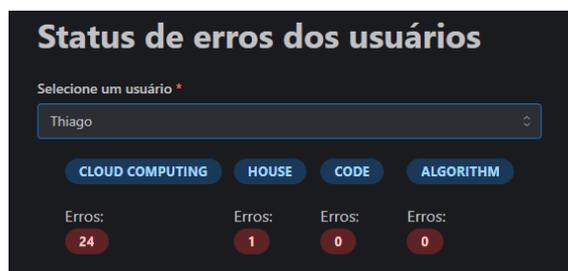


Figura 7. Interface status de erros dos usuários

Usuários comuns podem visualizar uma lista com todos os tópicos disponíveis para estudo. Dentro de cada tópico, o usuário visualiza todos os cards. Caso o usuário selecione aquele tópico, o conjunto de *cards* passa a integrar a coleção do usuário e ele inicia o processo treinamento utilizando repetição espaçada.

A Figura 8 apresenta uma visão do *card* referente ao termo *Cloud Computing*. Nesta interface são exibidos o vocabulário, a quantidade de erros e de acertos do usuário na pronúncia, a data da próxima revisão, a tradução, a imagem representativa do termo e o vídeo com a pronúncia correta.



Figura 8. Visão detalhada de um *cards*

Para revisar um tempo, a aplicação recupera do banco de dados todos os *cards* que o usuário possui. Apenas algumas informações do *card* aparecem como: Vocabulário,

imagem, quantidade de acertos e de erros do termo, como mostra a Figura 9.



Figura 9. Interface para Revisão de Termos

O reconhecimento de voz inicia quando o botão em verde “Revisar palavra” é clicado. Neste momento, um ícone substitui o botão, indicando que o usuário pode iniciar o processo de revisão (pronúncia) do termo. Caso o usuário não disser nada, ou se ocorrer alguma falha técnica, como por exemplo o microfone estiver desligado ou desconectado, a revisão é marcada como uma falha, e passa a ser contabilizada como uma resposta errada.

A Figura 10 apresenta um caso no qual o usuário fez a revisão de um *card* e a pronúncia não foi considerada correta. Nessa interface, a transcrição da palavra tal qual foi reconhecida pela *Web Speech API* é apresentada ao usuário para que ele saiba a razão de ter falhado na revisão. O mesmo *card* é mantido na interface, apenas com o vídeo contendo a pronúncia correta.

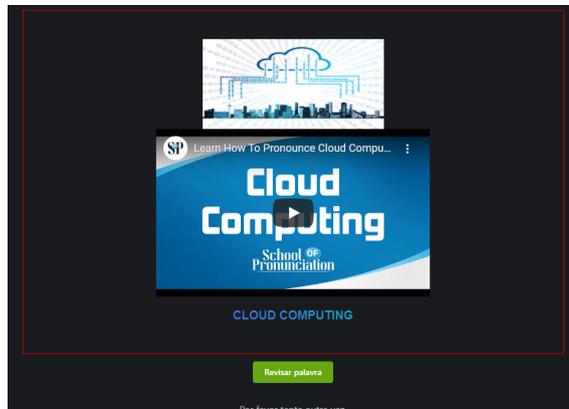


Figura 10. Interface em Caso de Falha

5.1. Discussão

Uma vez que a ferramenta é hospedada em um servidor Web e faz chamadas para a *Web Speech API*, é preciso que o usuário tenha um dispositivo com acesso à internet. É desejável, ainda que o usuário utilize um microfone de boa qualidade para suas revisões, a fim de evitar a captação de sons indesejados que possam atrapalhar o reconhecimento feito pela API.

Finalmente, a funcionalidade *Speech to Text* da *Web Speech API* reconhece palavras em Inglês padrão dos Estados Unidos. Dessa forma, a ferramenta considera corretos apenas os termos pronunciados neste padrão linguístico.

6. Considerações Finais e Trabalhos Futuros

Este trabalho apresentou a Pronum-Master, uma ferramenta baseada em *Speech-to-Text* e *Spaced repetition* para ajudar no ensino de pronúncia de palavras de língua inglesa, por meio de *flashcards* digitais. A ferramenta oferece, de forma gratuita, uma alternativa para que estudantes possam aprender e melhorar a pronúncia de palavras em inglês da área da computação. Para professores, ela prevê uma maneira de identificar possíveis dificuldades de pronúncia em seus alunos.

Por ser online, a ferramenta requer apenas um dispositivo com acesso à internet e microfone, e pode ser utilizada em qualquer horário, de acordo com a disponibilidade do usuário. Os *cards* podem ser divididos em diversos tópicos, de acordo com a estratégia de ensino do professor, que pode, ainda, criar tópicos específicos para diferentes turmas com diferentes interesses.

Trabalhos futuros incluem a utilização da ferramenta em ensino de termos técnicos de outras áreas, ou mesmo no ensino de pronúncia de palavras cotidianas da língua inglesa. Uma vez que a *Web Speech API* permite que a mesma seja configurada para reconhecer palavras em outros idiomas, é possível evoluir sua interface e implementar as adequações necessárias para viabilizar seu uso em diversos outros contextos, como por exemplo o ensino de Língua Portuguesa para estrangeiros, ou investigar a utilização de tecnologia de reconhecimento de voz em outras aplicações, como por exemplo profissionais da área de Fonoaudiologia no auxílio de pacientes a recuperar ou treinar a pronúncia correta de termos.

Outras possibilidades de trabalhos futuros visam a utilização e adaptação de outros algoritmos de *Spaced Repetition*. Pode-se tentar estimar qual intervalo é o mais adequado para a próxima revisão, a partir de parâmetros mensuráveis, como por exemplo o grau de confiança retornado pela *Web Speech API* a cada palavra testada para o reconhecimento. Uma segunda adaptação possível é comportar diversos algoritmos de *Spaced Repetition*, a fim de analisar qual apresenta melhores resultados no ensino de pronúncia de um segundo idioma.

Finalmente, melhorias na interface podem ser incorporadas de forma a aprimorar a usabilidade e a reforçar a motivação dos alunos durante o uso da ferramenta.

Agradecimentos

Os autores agradecem a Universidade Tecnológica Federal do Paraná – Campo Mourão pelo apoio financeiro (Edital DIRPPG-CM N° 01/2022).

Referências

- Assis, A. B. G. d. (2007). Adaptações fonológicas na pronúncia de estrangeirismos do inglês por falantes de português brasileiro. Master's thesis, Universidade Estadual Paulista (UNESP).
- Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of experimental psychology*, 96(1):124.
- Baturay, M., Yıldırım, S., and Daloğlu, A. (2009). Effects of web-based spaced repetition on vocabulary retention of foreign language learners. *Eurasian Journal of Educational Research (EJER)*.

- Bloom, K. C. and Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, 74(4):245–248.
- Chukharev-Hudilainen, E. and Klepikova, T. A. (2016). The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction: A double-blind study. *calico journal*, 33(3):334.
- Ebbinghaus, H. (2013). Memory: A contribution to experimental psychology. *Annals of neurosciences*, 20(4):155.
- Godwin-Jones, R. (2011). Mobile apps for language learning. *Language Learning & Technology*, 15(2):2–11.
- Graddol, D. (2000). *The future of English?: a guide to forecasting the popularity of the English language in the 21st century*. The British Council.
- Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL*, 15(1):3–20.
- Kang, S. H. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, 3(1):12–19.
- Keder, D. (2009). *Computer-assisted language learning using spaced repetition*. PhD thesis, Masarykova univerzita, Fakulta informatiky.
- Leitner, S. (1995). *So lernt man lernen: angewandte Lernpsychologie—ein Weg zum Erfolg*. Herder.
- Lindsey, R. V., Shroyer, J. D., Pashler, H., and Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological science*, 25(3):639–647.
- Metzler-Baddeley, C. and Baddeley, R. J. (2009). Does adaptive training work? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(2):254–266.
- Natal, André, S., Shires, G., and Jägenstedt, P. (2020). Web speech api. <https://wicg.github.io/speech-api/>. Accessed: June, 22, 2022.
- Neri, A., Cucchiaroni, C., and Strik, H. (2003). Automatic speech recognition for second language learning: how and why it actually works. In *Proc. ICPHS*, pages 1157–1160.
- Tabibian, B., Upadhyay, U., De, A., Zarezade, A., Schölkopf, B., and Gomez-Rodriguez, M. (2019). Enhancing human learning via spaced repetition optimization. *Proceedings of the National Academy of Sciences*, 116(10):3988–3993.
- Wallace, L. (2016). Using google web speech as a springboard for identifying personal pronunciation problems. In *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference, ISSN*, pages 2380–9566.