

Os efeitos do uso de um aplicativo baseado no reconhecimento automático de fala na pronúncia da língua inglesa

Daniel Reschke Pires¹
Luana Garbin Baldissera²

Resumo: Sistemas de Reconhecimento Automático de Fala (RAF) têm o potencial de prover feedback imediato e individualizado para estudantes de línguas estrangeiras. Com o objetivo de explorar esse potencial, esta pesquisa investigou o uso e os efeitos do uso de um aplicativo baseado no RAF na inteligibilidade da fala da língua inglesa. Os resultados dos pós-testes revelaram ganhos na inteligibilidade geral dos participantes e de palavras-alvo com vogais altas anteriores. Apesar das percepções predominantemente positivas sobre o aplicativo, alguns participantes enfrentaram dificuldades em utilizá-lo de forma adequada no contexto do ensino remoto.

Palavras-chave: Ensino de língua inglesa. Aprendizagem de Pronúncia. Reconhecimento Automático de Fala. Inteligibilidade.

Abstract: Automatic Speech Recognition (ASR) systems have the potential to provide immediate and individualized feedback for second language learners. With the goal of exploring this potential, this study sought to investigate the use and the effects of the use of an ASR-based application on the speech intelligibility of English learners. Intelligibility post-tests showed that there were significant gains in the overall intelligibility of the participants and on the target high-front vowels. In spite of the predominantly positive perceptions, some participants struggled to use it adequately in the remote learning context.

Keywords: Language teaching. Pronunciation learning. Automatic Speech Recognition. Intelligibility.

Resumen: Los sistemas de Reconocimiento Automático de Voz (RAH) tienen el potencial de proporcionar feedback inmediato e individualizado para estudiantes de un segundo idioma. Com el objetivo de explorar este potencial, este estudio buscó investigar el uso y los efectos de una aplicación basada en RAH em La inteligibilidad de estudiantes de inglés. Pruebas de

¹ Professor no Colégio de Aplicação da Universidade Federal de Santa Catarina (UFSC) e doutor pelo Programa de Pós-Graduação em Inglês – Estudos Linguísticos, da Universidade Federal de Santa Catarina (UFSC). <http://orcid.org/0000-0001-9970-7468> /E-mail: daniel.reschke.pires@gmail.com.

² Doutoranda em Inglês – Estudos Linguísticos e Literários, na Universidade Federal de Santa Catarina (UFSC). <http://orcid.org/0000-0001-9027-8951> E-mail: luanagbaldissera@gmail.com.



inteligibilidade mostraram melhoras significativas em la inteligibilidade de los participantes y en la inteligibilidade de las vocales frontales altas de la lengua inglesa. A pesar de las percepciones predominantemente positivas de la aplicación, a algunos participantes les costó usar la adecuadamente en el contexto del aprendizaje remoto.

Palabras clave: Enseñanza de idiomas. Enseñanza de la Pronunciación. Reconocimiento Automático del Habla. Inteligibilidade.

Introdução

Nas últimas décadas, pesquisadores da área de língua estrangeira/adicional (L2) têm entendido que o ensino de pronúncia deve ser guiado pelo princípio da inteligibilidade (DERWING; MUNRO, 2009; FIELD, 2005; GONÇALVES; SILVEIRA, 2014). De acordo com esses, aprendizes de uma L2 devem atingir alguns requerimentos fonológicos mínimos para ter uma comunicação bem-sucedida, o que pode acontecer mesmo que eles mantenham aspectos da sua primeira língua (SAITO, 2012). Nesse sentido, a instrução de pronúncia não busca emular a fala de falantes nativos da língua, mas procura auxiliar a compreensão entre falantes, sejam eles nativos ou não.

O ensino de pronúncia em L2 encontra desafios e resistências em sala de aula. No caso da língua inglesa, muitos professores não se sentem preparados para o ensino de pronúncia por não terem recebido treinamento adequado durante sua formação (BUSS, 2013; COSTA, 2016; HENDERSON *et al.*, 2012; THOMSON, 2012). Além do mais, a escassez de tempo e de recursos é apontada como uma limitação para o ensino de pronúncia (GILAKJANI; SABOURI, 2016; DERWING; MUNRO, 2009). Por outro lado, estudos mostram que os estudantes querem desenvolver sua pronúncia e apreciam receber feedback dos professores (COSTA, 2016).

Considerando os desafios supracitados, aplicativos para dispositivos móveis têm grande potencial para auxiliar o desenvolvimento da pronúncia, uma vez que podem ser utilizados quando, onde e por quanto tempo os aprendizes desejarem (KAISER, 2018). Com o auxílio do Reconhecimento Automático de Fala (RAF³), esses aplicativos são capazes de tornar sinais acústicos em palavras e, com a análise da fala, prover feedback

³ Do inglês *Automatic Speech Recognition* (ASR).

imediate e de qualidade para os aprendizes (HINCKS, 2015; KAISER, 2018). Finalmente, eles permitem que a pronúncia seja treinada de maneira privada e sem a pressão e o possível julgamento de outros aprendizes.

Apesar dessas potencialidades, sistemas que utilizam o RAF têm sido pouco explorados para o ensino de pronúncia no Brasil, visto que ainda não existem pesquisas que tenham investigado a viabilidade desses sistemas para um ensino de pronúncia guiado pelo princípio da inteligibilidade. Nesse sentido, este artigo apresenta os resultados de um projeto de pesquisa que teve como objetivo investigar o uso de um aplicativo móvel baseado no RAF, seus efeitos na inteligibilidade da pronúncia da língua inglesa, e as percepções dos participantes a seu respeito. Quatro perguntas de pesquisa guiaram esta investigação:

- 1) o uso de um aplicativo móvel baseado no RAF promove a inteligibilidade dos participantes?;
- 2) o uso do aplicativo promove a inteligibilidade das vogais /ɪ/ e /i/ da língua inglesa?;
- 3) como foi o uso do aplicativo pelos participantes no contexto do ensino remoto?;
- 4) quais são as percepções dos participantes a respeito do aplicativo e da aprendizagem móvel?

A seguir, apresentamos a fundamentação teórica desta pesquisa, na qual são abordados o RAF e seu uso para o ensino de pronúncia em língua estrangeira, bem como o conceito de inteligibilidade da fala e os estudos que a investigaram em contextos nacionais.

Reconhecimento automático de fala

O ensino de pronúncia em sala de aula de língua estrangeira enfrenta diversos

desafios. Um deles é que muitos professores se sentem despreparados para trabalhar com pronúncia porque não receberam treinamento adequado para isso (HENDERSON *et al.*, 2012; BUSS, 2013; THOMSON, 2012; COSTA, 2016). Mesmo quando se sentem aptos, o tempo em sala de aula costuma ser insuficiente para desenvolver todas as habilidades linguísticas. Considerando que, no Brasil, há em média 30,4 alunos por sala no ensino médio (INEP, 2019), o ensino de pronúncia é frequentemente deixado de lado, uma vez que demanda tempo, especialmente quando envolve correções individuais (PI-HUA, 2006; NAIR *et al.*, 2017).

Todavia, nas últimas décadas, as tecnologias de reconhecimento de voz vêm sendo exploradas para auxiliar o ensino de línguas, já que permitem que a fala humana seja reconhecida e avaliada com o suporte de sistemas de computação. De acordo com Neri, Cucchiarini e Strik (2003), os sistemas baseados em RAF geralmente possuem cinco fases: 1) reconhecimento da voz, que é quando os sinais de fala são convertidos em palavras usando modelos fonéticos e sintáticos; 2) pontuação, que é quando o sistema faz uma avaliação preliminar e menos aprofundada da pronúncia do usuário, comparando-a com modelos acústicos e temporais que são geralmente baseados em falantes nativos; 3) detecção de erros, na qual o sistema localiza onde estão erros na fala; 4) diagnósticos dos erros, que é quando o sistema identifica os *tipos* específicos de erros produzidos na fala do usuário e sugere formas de aperfeiçoá-la, baseando-se geralmente em erros comuns feitos por falantes não nativos; e 5) apresentação do *feedback*, a qual utiliza os dados coletados nas fases 2, 3 e 4 e os apresenta aos usuários. Os autores supracitados (NERI; CUCCHIARINI; STRIK, 2003) alertam que uma boa parte do aprendizado nesses sistemas depende da qualidade do *feedback* apresentado. Em outras palavras, ele precisa ser corretamente interpretado e incorporado pelos usuários.

Pesquisas vêm apontando diversos benefícios do uso de sistemas de RAF. Eles permitem, por exemplo, uma prática de pronúncia sem a pressão de falar em frente dos colegas de sala e dos professores, o que pode reduzir a ansiedade que os estudantes sentem (KAISER, 2018). Além disso, os estudantes podem praticar pronúncia no seu próprio ritmo e no horário que preferem, e ainda assim receber *feedback* instantâneo (MAHDI; KHATEEB, 2019).

Estudos já foram conduzidos com o objetivo de investigar os efeitos de sistemas

baseados no RAF na aprendizagem de língua estrangeira, entre eles Neri, Cucchiarini e Strik (2006), e Elimat e Abuseileek (2014). Alguns estudos investigaram sistemas de RAF desenvolvidos para dispositivos móveis, como *smartphones* e *tablets*. Liakin, Cardoso e Liakina (2015), por exemplo, investigaram o uso de um aplicativo de *smartphone* baseado em RAF na produção e na percepção da vogal francesa /y/. Enquanto um grupo recebeu instrução e *feedback* via aplicativo, outro recebeu instrução pelo aplicativo e *feedback* de um professor. O grupo controle, por sua vez, teve aulas de conversação em vez de prática de pronúncia. Os resultados da pesquisa mostram que o único grupo que teve ganhos significativos foi o que recebeu *feedback* e instrução baseada em RAF. Os autores entendem que esses ganhos são resultado do feedback imediato, da gamificação e das múltiplas oportunidades de aprendizado proporcionadas pelo aplicativo.

Mais recentemente, Fouz-González (2020) buscou explorar o potencial do aplicativo *English File Pronunciation* para o ensino de pronúncia da língua inglesa. Os participantes do estudo foram 52 espanhóis que faziam um curso de fonética da língua inglesa na Universidade de Múrcia. Os participantes utilizaram o aplicativo diariamente por cerca de 20 minutos durante duas semanas, nos momentos e locais que desejaram, para estudar os sons vocálicos (/æ a: ʌ ə/) e o contraste entre /s – z/. Para que todos os participantes pudessem utilizar o aplicativo e integrassem o grupo controle, eles foram divididos aleatoriamente em dois grupos. Durante as primeiras duas semanas, somente o primeiro grupo utilizou o aplicativo, enquanto o outro funcionou como um grupo controle. Ao final da segunda semana, ambos os grupos responderam os pós-testes. Após isso, o segundo grupo, também, utilizou o aplicativo por duas semanas e realizou outro pós-teste.

Os resultados de Fouz-González (2020) revelaram ganhos significativos para todos os sons trabalhados pelo aplicativo nos testes de percepção e de identificação, enquanto nos testes de produção, os efeitos do aplicativo só foram significativos para alguns sons. Em virtude desses resultados, o autor entende que pesquisas futuras não devem procurar por uma ferramenta perfeita, mas explorar como diferentes tecnologias podem facilitar diferentes tipos de instrução, e, no caso dos aplicativos, explorar suas possibilidades enquanto ferramenta de apoio para o desenvolvimento da pronúncia. Apesar desses resultados positivos, o uso de sistemas baseados no RAF para o ensino de pronúncia ainda carece de pesquisas que ajudem a compreender se e como esses

sistemas podem auxiliar no desenvolvimento da inteligibilidade entre falantes. Nesse sentido, caracterizamos a seguir o conceito de inteligibilidade e alguns estudos acerca desse tópico.

Inteligibilidade

No ensino de línguas, a inteligibilidade é, de acordo com abordagens mais tradicionais, um processo unidirecional no qual falantes não nativos de uma língua se esforçam para serem compreendidos por falantes nativos, que são quem decidem o que é inteligível ou não (BAMGBOSE, 1998). Todavia, abordagens mais atuais, que entendem que o inglês é uma língua global, consideram que os falantes nativos não são os únicos aptos a julgar o que é inteligível, e que a fala deles nem sempre é mais inteligível do que a de falantes não nativos. Assim, o conceito de inteligibilidade pode ser definido de forma ampla como o grau de compreensão de um enunciado por parte de um ouvinte (DERWING; MUNRO, 2009). Uma das formas mais comuns para obter um índice da inteligibilidade de um falante é a transcrição ortográfica da sua fala, a qual pode ser feita tanto por falantes nativos quanto por não nativos da língua alvo (MUNRO; DERWING; MORTON, 2006).

Apesar da inegável importância da inteligibilidade para o sucesso da comunicação oral (DERWING; MUNRO, 2005; FIELD, 2005), nem todos os aspectos fonológicos de uma língua têm o mesmo impacto na inteligibilidade. Como Derwing e Munro (2009) ressaltam, o tempo para o ensino de pronúncia em sala de aula é limitado, e se dedicado à instrução de aspectos que não afetam a inteligibilidade, outros que realmente impedem a compreensão estão sendo deixados de lado. Por isso, a identificação de aspectos que contribuem para a inteligibilidade tem recebido grande interesse nas pesquisas sobre pronúncia em L2.

Em se tratando da comunicação entre falantes que utilizam o inglês como língua

franca (ILF), Jenkins (2000) propõe o Lingua Franca Core⁴ (LFC), o qual considera as distinções oriundas da duração das vogais como um aspecto crucial para a inteligibilidade em ILF. Além disso, pesquisadores apontam que a instrução de pronúncia deve priorizar os problemas que provavelmente não irão desaparecer ao longo do aprendizado da língua sem intervenção explícita. Por exemplo, Munro, Derwing e Saito (2013) concluíram que, mesmo após sete anos de contato com a língua, as produções do som /I/ continuavam a ser ininteligíveis, enquanto a de outros sons vocálicos haviam se tornado quase 100% inteligíveis. Assim, faz-se necessário investigar como esses sons podem ser desenvolvidos, já que são necessários à inteligibilidade ao mesmo tempo que são desafiadores aos aprendizes. A ajuda de novas tecnologias, como o RAF, pode ser uma alternativa.

Embora tenha sido pouco investigada no Brasil (GONÇALVES; SILVEIRA, 2015), alguns estudos identificaram fatores que podem afetar a inteligibilidade de falantes e ouvintes brasileiros. Entre eles, estão a acentuação tônica (CRUZ, 2003) e a frequência de uso das palavras (BECKER, 2013), a vogal tensa /ɪ/ (GONÇALVES; SILVEIRA, 2014), epênteses que criam sílabas extras, e a africacão das oclusivas alveolares /t/ e /d/ (SILVEIRA; SILVA, 2018).

Com objetivo de investigar a inteligibilidade das vogais altas anteriores da língua inglesa produzidas por falantes brasileiros, os participantes de Gonçalves e Silveira (2014) gravaram sentenças contendo palavras com os sons /i/ e /ɪ/. As palavras foram transcritas por 32 ouvintes com 11 línguas-maternas diferentes. Os resultados do estudo mostraram que a vogal tensa /i/, presente em *beat* e *seat*, por exemplo, causou dificuldades para os ouvintes. Essas palavras foram produzidas com pistas acústicas com duração insuficiente para distingui-las, e foram incorretamente transcritas com a vogal frouxa. Além disso, transferências fonológicas do português brasileiro também diminuíram a inteligibilidade dos participantes. Isso aconteceu quando as oclusivas iniciais foram produzidas sem aspiração ou quando oclusivas alveolares foram palatalizadas (como quando a palavra *pit* foi transcrita como *bitch*). Considerando os

⁴ Cabe lembrar que, em relação ao LFC, Jenkins (2015) ressalta que este nunca teve a pretensão de ser um modelo monolítico ou um núcleo imutável, mas que se propunha a fornecer um pequeno número de características importantes para a inteligibilidade entre falantes.

resultados apresentados, Gonçalves e Silveira (2014) argumentam que a duração das vogais é uma pista acústica relevante, especialmente para abordagens instrucionais orientadas pela inteligibilidade. Como uma implicação pedagógica do estudo, os autores entendem que a instrução das vogais altas anteriores é necessária, uma vez que a distinção entre essas vogais afeta a inteligibilidade, e sugerem uma instrução que estimule estratégias de automonitoramento e a autocorreção.

Em Gonçalves e Silveira (2014), transferências fonológicas do português afetaram a inteligibilidade dos brasileiros. Já em Becker (2013), buscou-se investigar a inteligibilidade de falantes de inglês estrangeiros para ouvintes brasileiros. A autora não encontrou diferenças significativas na inteligibilidade da fala em inglês de falantes alemães, estadunidenses e chineses, enquanto os falantes japoneses foram significativamente menos inteligíveis. Já em Delatorre, Silveira e Gonçalves (2017), verbos no passado simples do inglês, produzidos por falantes nativos e não nativos da língua, foram mais inteligíveis para os ouvintes brasileiros quando produzidos por falantes cuja língua materna era similar ao português (como o espanhol). A inteligibilidade dos falantes nativos da língua inglesa foi menor do que a de falantes não nativos da língua devido a pouca experiência dos ouvintes em conversas com falantes nativos.

Os trabalhos discutidos nesta seção revelam um crescente interesse pelo estudo da inteligibilidade, pelos fatores que podem afetá-la, como transferências fonológicas e a língua materna dos falantes e dos ouvintes, e mostram a importância de uma instrução de pronúncia voltada para a inteligibilidade. Todavia, ainda não existem estudos que investigaram esse tipo de instrução em contexto nacional. Nesse contexto, o presente estudo tem o objetivo de investigar os efeitos do uso de um aplicativo baseado no RAF para o desenvolvimento da inteligibilidade da fala da língua inglesa. Apresentamos a seguir a metodologia empregada a fim de alcançar esse objetivo.

Metodologia

Nesta seção, são apresentados os participantes envolvidos na pesquisa, seguidos

pelos instrumentos e procedimentos para coleta e análise de dados.

Contexto e Participantes

Este artigo apresenta os resultados de um projeto de pesquisa realizado no Colégio de Aplicação da Universidade Federal de Santa Catarina (CA-UFSC), entre 01/04/2020 e 01/04/2022. Dois tipos de participantes fizeram parte da pesquisa: estudantes⁵ do Colégio de Aplicação da UFSC, os quais utilizaram o aplicativo e gravaram frases para a aferição da sua inteligibilidade em língua inglesa, e falantes de língua inglesa, os quais atuaram como os ouvintes da pesquisa.

Como a pesquisa envolve o uso de um aplicativo de *smartphone*, foram convidados estudantes dos anos finais do ensino fundamental e do ensino médio, uma vez que eles tendem a ter maior familiaridade com o uso de tecnologias para a aprendizagem do que estudantes dos anos iniciais do ensino fundamental.

Dos 34 estudantes que inicialmente aceitaram participar da pesquisa, 16 desistiram antes da realização dos testes de inteligibilidade. Os demais 18 receberam acesso ao aplicativo, porém somente 13 concluíram todas as etapas da pesquisa. Esses 13 participantes tinham entre 12 e 17 anos, com idade média de 15 anos, e haviam estudado inglês entre 1 e 11 anos, com média de 5 anos de estudo. Um teste⁶ revelou que o nível médio de proficiência dos participantes era A2 no Quadro Europeu Comum de Referência para Línguas, ou seja, tinham domínio básico da língua inglesa.

A inteligibilidade foi aferida por meio de transcrições ortográficas. Levando em conta o princípio da inteligibilidade, falantes da língua inglesa cuja língua materna não é o português foram convidados a participar na pesquisa como ouvintes. Um total de 13 pessoas aceitaram o convite, sendo 8 falantes nativos da língua inglesa, e 5 não nativos. Isso permitiu que cada um escutasse as frases de apenas um participante, evitando assim possíveis efeitos de repetição de tarefas. Todos os ouvintes responderam um breve questionário, no qual foi verificado que eles têm pouco ou nenhum conhecimento da

⁵ Como esta pesquisa teve a participação de seres humanos, este projeto foi submetido ao Comitê de Ética da UFSC em 14/04/2021 e aprovado em 04/05/2021, sob o CAAE 44097421.0.0000.0121.

⁶ O teste de proficiência foi feito pela plataforma ESL – *Language Studies Abroad*.

língua portuguesa e que têm pouca familiaridade com o inglês falado por brasileiros.

Instrumentos

Os seguintes instrumentos foram utilizados nesta pesquisa: 1) o aplicativo *ELSA Speak*; 2) pré-testes e pós-testes de inteligibilidade; 3) relatórios do painel para professores do *ELSA Speak*; 4) questionários sobre o uso e sobre as percepções dos participantes sobre a aprendizagem móvel e sobre o aplicativo. Devido à pandemia de COVID-19 e às medidas de isolamento social, todos os procedimentos para coleta de dados foram realizados *on-line*. Esses instrumentos são descritos a seguir, na ordem em que foram utilizados.

Primeiramente, foram realizados pré-testes para aferir o grau de inteligibilidade da fala dos participantes. Nestes testes, eles leram e gravaram, em seus próprios *smartphones*, uma sequência de 24 frases em língua inglesa, 20 delas contendo palavras com os sons vocálicos /i/ e /ɪ/ e que foram posteriormente instruídas pelo aplicativo durante a fase de instrução. Em 4 frases, havia distratores, ou seja, palavras que não possuem os sons vocálicos /i/ e /ɪ/. Como palavras menos frequentes tendem a serem menos inteligíveis (BECKER, 2013), apenas aquelas cuja frequência é considerada alta pelo Corpus Contemporâneo do Inglês Americano (COCA) foram utilizadas.

Após a conclusão dos pré-testes, os participantes puderam iniciar o uso do aplicativo *ELSA Speak*, o qual possui um sistema de RAF desenvolvido para reconhecer a fala de falantes não nativos e é capaz de dar *feedback* para palavras e sentenças (ELSA API, 2022). Para seu uso nesta pesquisa, os participantes receberam arquivos com instruções sobre como utilizá-lo apropriadamente. Mais especificamente, eles foram instruídos a: 1) assistir aos vídeos introdutórios disponibilizados pelo aplicativo; 2) completar as 38 lições da unidade 8, a qual trabalhava com os sons-alvo da pesquisa; 3) completar cerca de 10 unidades por semana; 4) concluir as 38 unidades dentro do período de um mês a partir da sua ativação. Apesar dessas orientações, os participantes tiveram liberdade para acessar o aplicativo quando desejassem, e completar quantas lições desejassem por semana, visto que um dos principais aspectos da aprendizagem móvel é a possibilidade de ter autonomia sobre o processo de aprendizagem (SHA *et al.*,

2012).

O *ELSA Speak* oferece um painel de controle para escolas e pesquisadores, o qual foi utilizado para coletar dados sobre o seu uso pelos participantes e verificar se eles seguiram as instruções que receberam, isto é, se completaram as 38 lições dentro do período de um mês e se completaram cerca de 10 lições por semana. Esses dados também serviram para garantir que eles haviam feito todas as atividades antes de realizar o pós-teste.

Após o término das 38 lições designadas, os participantes realizaram os pós-testes formados pelas mesmas frases dos pré-testes, e responderam um questionário via *Google Forms* para coletar dados sobre como utilizaram o aplicativo, o que incluiu perguntas sobre onde e como ele foi utilizado. Combinando esses dados com os dados do painel, foi possível investigar como se deu a aprendizagem móvel, uma vez que essa tem sido associada com a possibilidade de escolher quando, onde e como estudar (KHADDAGE *et al.*, 2016), com a exploração de curtos períodos de tempo para a aprendizagem (TRAXLER, 2007) e com a possibilidade de uma aprendizagem autorregulada (SHA *et al.*, 2012).

Nesse questionário, os participantes também puderam expressar suas percepções acerca do aplicativo e da experiência com a aprendizagem móvel. Mais especificamente, foram feitas perguntas sobre a qualidade do sistema de RAF e do feedback provido pelo aplicativo, e sobre as vantagens e desvantagens de aprender com ele.

Análise dos dados

As gravações foram enviadas aos ouvintes da pesquisa, os quais transcreveram ortograficamente todas as palavras pronunciadas pelos participantes. As transcrições foram então enviadas aos pesquisadores, os quais contabilizaram o número de palavras inteligíveis. Uma palavra foi considerada inteligível quando a transcrição dos ouvintes correspondia às frases originais.

As frases foram contabilizadas de três maneiras. Primeiramente, todas as palavras transcritas foram contabilizadas, o que forneceu uma medida da inteligibilidade geral dos participantes. Como cada palavra inteligível contabilizou um ponto, e como foram

gravadas 120 palavras em cada teste, a pontuação máxima individual era de 120, e de 1560 para os 13 participantes. Para as palavras-alvo, foram 20 pontos por participante, e 260 pontos para todos os 13, e para os pares mínimos das vogais /i/ e /I/, como *beat* e *bit*, foram 10 pontos individuais e 130 no total. Palavras com erros ortográficos por parte dos ouvintes, como *beatt* em vez de *beat*, foram consideradas inteligíveis.

Após as transcrições e análises desses testes, os estudantes receberam um relatório com as palavras inteligíveis, as não compreendidas e com os sons que lhes causavam maior dificuldade.

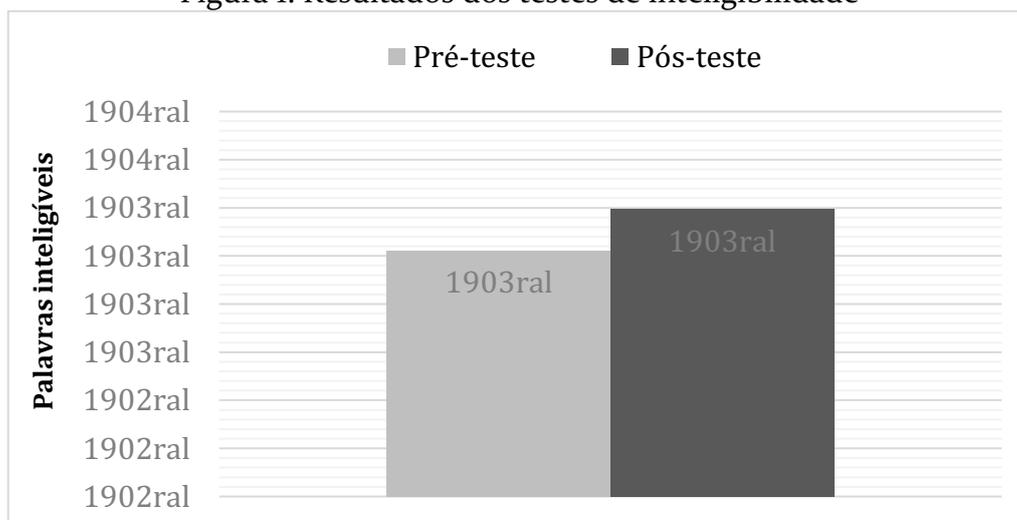
Resultados

A partir das perguntas de pesquisa, os resultados são apresentados a seguir.

O uso do aplicativo promoveu a inteligibilidade da fala dos estudantes?

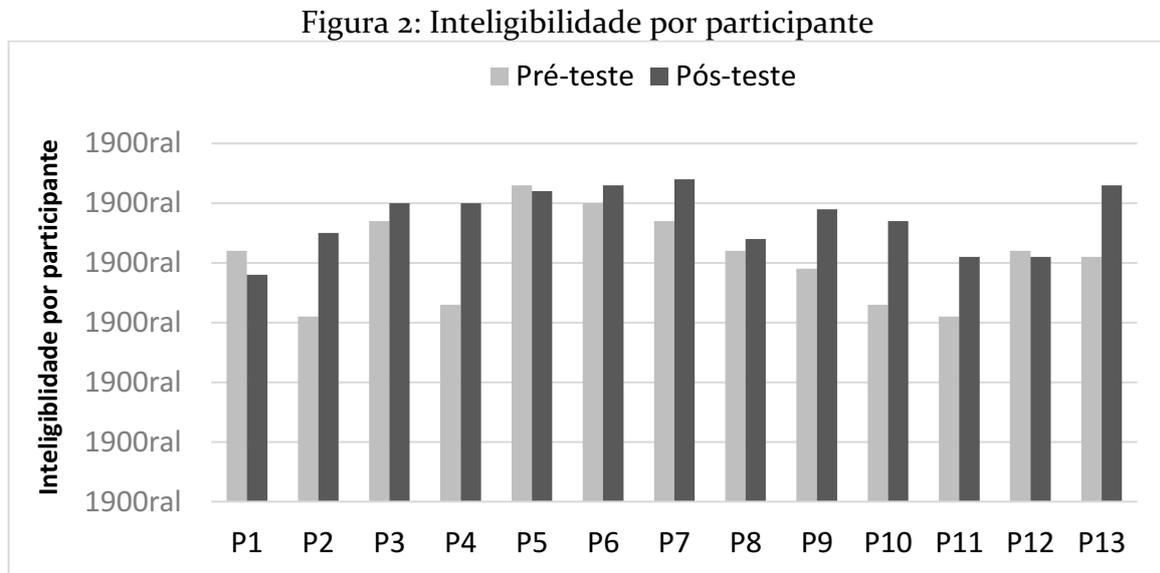
Um teste T pareado revelou uma diferença significativa entre os resultados do pré-teste (M = 100,8, DP = 7,6) e do pós-teste (M = 107, DP = 5,2), $t(11) = -3,13$, $p = 0,004$. A Figura 1 mostra o total de palavras inteligíveis em cada teste.

Figura 1: Resultados dos testes de inteligibilidade



Fonte: Os autores (2022).

Como pode ser visto na Figura 1, no pré-teste, 1311 (84%) das 1560 palavras pronunciadas pelos participantes foram corretamente transcritas pelos ouvintes, enquanto no pós-teste, 1397 (89,5%) palavras foram corretamente transcritas. Na Figura 2 são exibidos os resultados por participante.

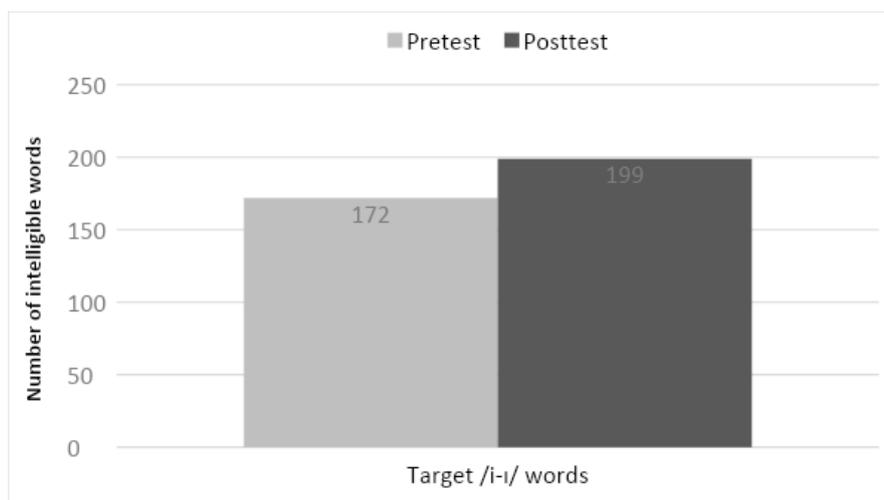


Esses dados indicam que o uso do aplicativo tornou a pronúncia em língua inglesa dos participantes mais inteligível, ou seja, um número maior de palavras foi corretamente compreendido pelos ouvintes após o uso do aplicativo *ELSA Speak*.

O uso do aplicativo promoveu a inteligibilidade de palavras com as vogais /ɪ/ e /i/ da língua inglesa?

Além dos efeitos do uso do *ELSA Speak* na inteligibilidade geral da pronúncia dos participantes, também foram analisados os efeitos na inteligibilidade de palavras com as vogais /ɪ/ e /i/ da língua inglesa, que foram as palavras-alvo da pesquisa. Um teste T pareado revelou uma diferença significativa entre os resultados do pré-teste ($M = 13.2$, $DP = 1.9$) e do pós-teste ($M = 15.3$, $DP = 1.6$); $t(12) = -4.06$, $p = 0,002$. A Figura 3 mostra o total de palavras inteligíveis em cada teste.

Figura 3: Número de palavras-alvo inteligíveis



Fonte: Os autores (2022).

Como pode ser visto na Figura 3, no pré-teste, 172 (66%) das 260 palavras-alvo pronunciadas pelos participantes foram corretamente transcritas pelos ouvintes, enquanto no pós-teste, o resultado obtido foi 199 (76,5%). Ao compararmos os ganhos percentuais, notamos que o uso do aplicativo teve um efeito maior nas palavras-alvo, as quais tiveram um ganho de 10,5%, enquanto um ganho de 4,5% foi observado em relação à inteligibilidade geral. A provável causa dessa diferença é que as palavras-alvo foram instruídas de forma explícita pelo aplicativo, com momentos em que os participantes as pronunciavam e recebiam *feedback* específico sobre elas, enquanto as demais requeriam que os estudantes observassem o feedback provido pelo sistema de RAF.

Como foi o uso do aplicativo pelos participantes no contexto do ensino remoto?

Um dos objetivos desta pesquisa foi compreender o uso do *ELSA Speak* pelos estudantes. Como este estudo foi conduzido durante o ano letivo de 2021, todos eles estavam inseridos na modalidade de ensino remoto. Por isso, não surpreende terem reportado o uso do aplicativo em casa, com apenas dois estudantes utilizando-o enquanto se deslocavam de um lugar a outro. Além disso, apenas um participante revelou ter feito outras atividades concomitantes, o que sugere que a maioria não foi distraída pelas opções de entretenimento dos seus smartphones.

Utilizando os dados do painel de controle do *ELSA Speak*, foi possível verificar quais estudantes seguiram as instruções para o seu uso adequado. Antes de usá-lo, eles foram instruídos a completar todas as 38 lições da unidade 8 durante o período de um mês, e a fazer cerca de 10 lições por semana. A tabela abaixo informa quais participantes seguiram essas recomendações e quantos minutos cada participante levou, em média, para completar cada lição.

Tabela 1: Conclusão das atividades pelos participantes

Participante	Completo a unidade 8 em mês?	Fez 10 lições por semana?	Ganhos no pós-teste de inteligibilidade?
1.	Não	Não	Não
2.	Sim	Sim	Sim
3.	Sim	Não	Sim
4.	Sim	Não	Sim
5.	Sim	Não	Não
6.	Sim	Não	Sim
7.	Sim	Não	Sim
8.	Sim	Sim	Sim
9.	Sim	Não	Sim
10.	Sim	Não	Sim
11.	Sim	Sim	Sim
12.	Não	Não	Não
13.	Sim	Não	Sim

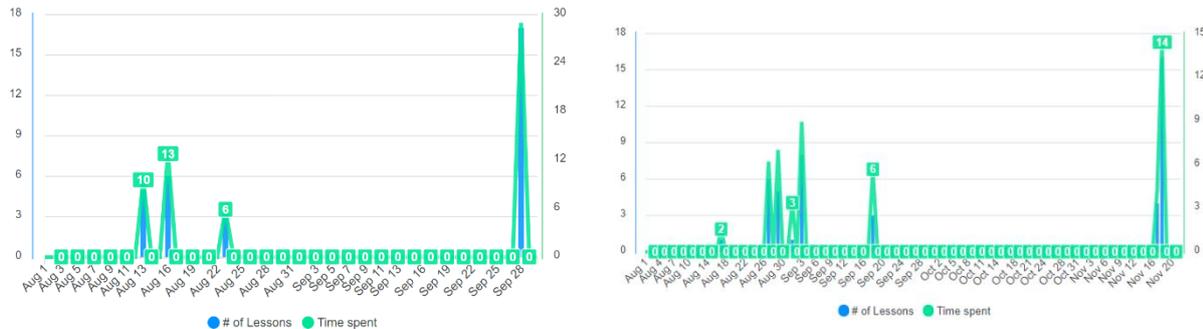
Fonte: Os autores (2022).

Como mostra a Tabela 1, 11 dos 13 participantes completaram a unidade em cerca de um mês. Todavia, apenas 3 deles seguiram a orientação de concluir 10 lições semanalmente. Entre os que não seguiram essa recomendação, alguns finalizaram todas as lições em um ou dois dias, enquanto outros fizeram apenas algumas inicialmente e necessitaram de diversos lembretes para concluí-las.

Ao compararmos os dados apresentados na tabela com os resultados individuais dos testes de inteligibilidade (Figura 2), percebemos que os dois participantes (P1 e P12) que não concluíram as atividades em um mês, nem fizeram cerca de 10 lições semanalmente, não tiveram ganhos no pós-teste de inteligibilidade. Como mostra a Figura 4, P1 completou 18 lições de 13 a 25 de agosto, mas passou mais de um mês sem

acessar o aplicativo. Do mesmo modo, o participante P12 fez seu primeiro acesso em 17 de agosto e completou 23 das 38 lições restantes até 18 de setembro. Depois disso, passou dois meses sem utilizá-lo, até completar as lições faltantes, em 17 e 18 de novembro.

Figura 4: Padrão de uso dos participantes 1 e 12



Fonte: Os autores (2022).

É provável que, na ocasião em que realizaram o pós-teste de inteligibilidade, esses participantes não lembravam mais o que haviam praticado ou estudado nas primeiras lições, o que pode ter influenciado nos resultados. Esse padrão de uso sugere que alguns participantes tiveram dificuldades para organizar seu processo de aprendizagem e manter sua motivação. As respostas à pergunta “Quão organizado foi o uso do aplicativo?” (Figura 5) revelam que esses participantes perceberam esta dificuldade.

Figura 5: Quão organizado foi o uso do aplicativo?



Fonte: Os autores (2022).

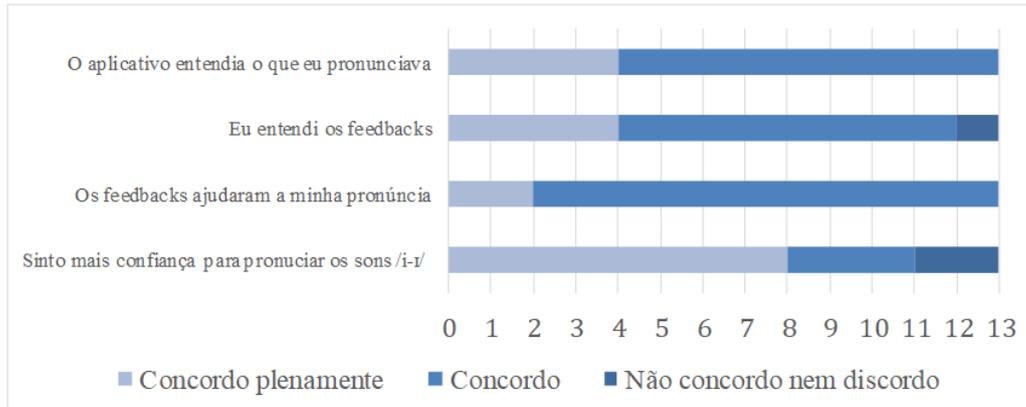
Como mostra a Figura 5, três participantes acreditaram que seu uso do aplicativo foi muito desorganizado ou apenas desorganizado, entre eles os participantes 1 e 12. Isto

é evidenciado pelo fato de terem afirmado que a maior desvantagem de aprender por meio dos dispositivos móveis foi sua falta de organização e dedicação (Figura 8). Resultados similares foram encontrados em Botero *et al.* (2019), em que estudantes universitários encontraram dificuldades para manter o uso de MALL em um contexto em que não era diretamente ligado às suas obrigações formais de aprendizagem, como foi no caso deste estudo.

Quais são as percepções dos participantes sobre o aplicativo e sobre a aprendizagem móvel?

Após completar a unidade 8, os participantes responderam um questionário no qual puderam compartilhar suas percepções sobre o *aplicativo* e sobre a experiência com a aprendizagem de línguas utilizando dispositivos móveis. Quatro destas perguntas foram respondidas com escalas de *Likert*, como mostra a Figura 6.

Figura 6: Percepções dos participantes sobre o aplicativo



Fonte: Os autores (2022).

Como podemos observar na Figura 6, os estudantes tiveram percepções positivas sobre o aplicativo. Todos concordaram que o sistema de RAF era capaz de compreender as palavras pronunciadas por eles. Isso indica que as tecnologias de reconhecimento de fala atuais são capazes de reconhecer a pronúncia de falantes não nativos de forma confiável, o que não era o caso nos anos 2000 (DERWING; MUNRO; CARBONARO,

2000). Apesar disso, 6 participantes reportaram ocasiões nas quais eles acreditam ter pronunciado uma palavra ou sentença corretamente, mas receberam *feedback* negativo mesmo assim. A respeito disso, eles disseram:

P2: Às vezes dava a impressão que eu tinha que mudar a entonação (tipo mudar a voz) pra ele contar como correto

P3: às vezes eu pronunciava de um jeito e dava como errado, e quando ia corrigir acaba pronunciando do mesmo jeito mas ia como certo.

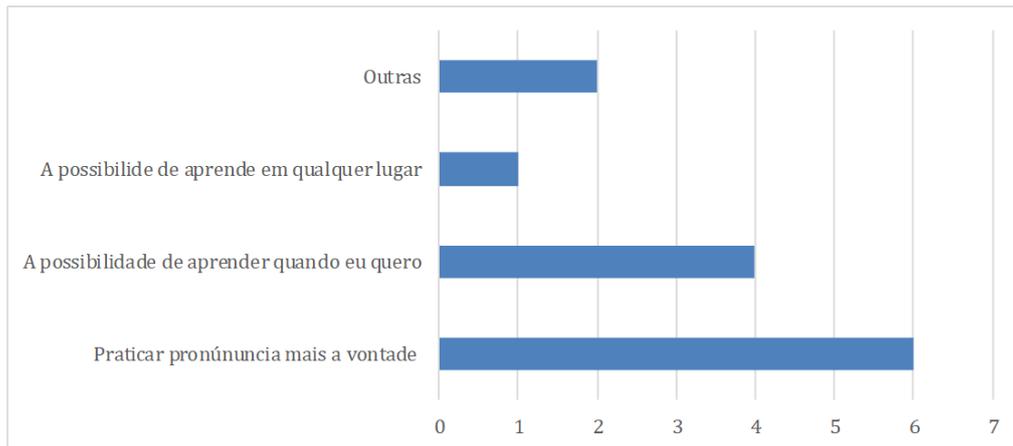
P8: Tentei falar a palavra "listeners" umas 30 vezes, mesmo acertando a pronúncia o app dizia que tava errado, aí comecei a tentar de várias maneiras diferentes e mesmo assim não aceitava

P9: Algumas vezes o app me pedia para dizer "lunch", eu escutava a palavra e então repetia exatamente o que eu ouvia, mas mesmo assim o app dizia que eu estava pronunciando errado.

P10: Se eu estou lembrado, eu respondi uma mesma frase dezenas de vezes e o aplicativo dava como errado, daí eu tentei responder copiando exatamente o app e deu errado, naquele momento eu estava com raiva já e falei de uma vez a frase, sem pensar muito. E deu certo...

Os participantes também puderam revelar suas percepções sobre as vantagens e desvantagens de aprender inglês utilizando um aplicativo de smartphone, como mostra a Figura 7.

Figura 7: Vantagens de aprender com o aplicativo

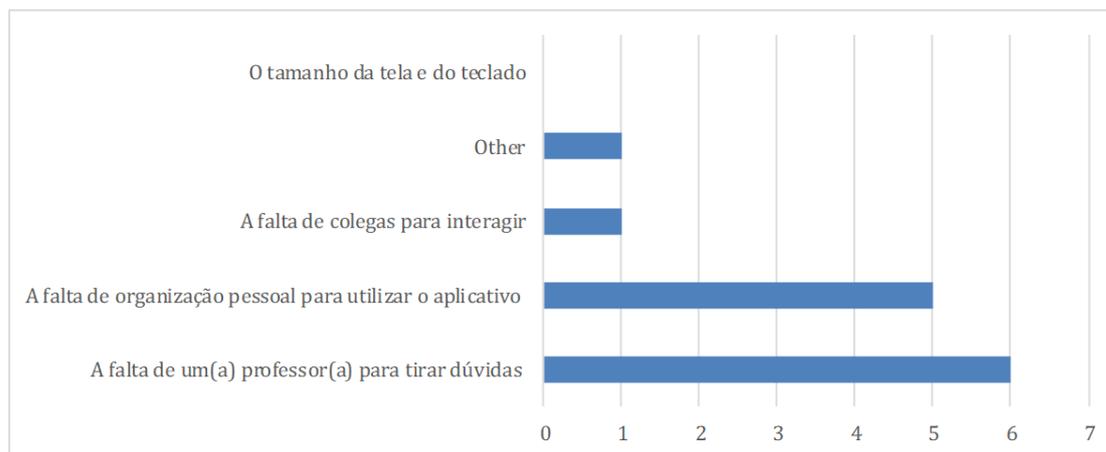


Fonte: Os autores (2022).

Para 6 participantes, a possibilidade de praticar a pronúncia mais à vontade foi considerada a maior vantagem. Esses resultados sugerem que o ensino de pronúncia baseado em RAF ajuda a reduzir a ansiedade que os estudantes sentem ao falar na frente de colegas, como também foi constatado por Aulia (2018) e Bashori *et al.* (2021). Nas aulas de língua estrangeira, é comum que estudantes se sintam ansiosos ao praticar a língua-alvo em situações em que os colegas da sala podem ouvi-los. Uma estratégia para deixar os alunos mais à vontade é pedir para que as atividades de pronúncia e conversação sejam feitas em pequenos grupos. Todavia, como isso envolve muitos estudantes falando ao mesmo tempo, um(a) professor(a) apenas não é suficiente para dar um retorno em relação à pronúncia de todos os estudantes imediatamente, especialmente quando eles têm diferentes níveis de proficiência e diferentes tipos de dificuldades com a língua inglesa. Nesse sentido, o *feedback* individualizado e imediato fornecido pelo aplicativo possibilita que os estudantes desenvolvam sua pronúncia com menor ansiedade e, ao mesmo tempo, recebam correções que são específicas as suas necessidades.

Segundo os participantes, as maiores desvantagens de aprender pronúncia pelo aplicativo foram a falta de um(a) professor(a) para tirar dúvidas e a falta de organização pessoal para utilizar a ferramenta de forma adequada, como mostra a Figura 8.

Figura 8: Desvantagens de aprender usando o aplicativo



Fonte: Os autores (2022).

Os participantes também puderam escrever sobre o que não gostaram sobre o aplicativo em si. Das 13 respostas, 3 mencionaram dificuldades de serem compreendidos e 1 mencionou a falta da turma para interagir. Embora a pergunta pedisse que apontassem coisas que não gostaram do aplicativo, alguns participantes relataram ter gostado da ferramenta como um todo:

P2: Gostei de tudo, adorei o app!

P5: Gostei de todas as ferramentas que usei.

P6: Acho que gostei de quase todo o aplicativo, o único problema foi quando ele não ouvia ou entendia o que estava falando e tinha que repedir algumas vezes.

P8: Sinceramente eu não tenho nada a dizer nessa questão

P11: Na realidade não teve algo que me fizesse não gostar do aplicativo adorei utilizar dessa plataforma para melhorar o meu conhecimento na língua inglesa.

Esses resultados sugerem que os participantes, cuja idade média é de 15 anos, têm dificuldades nesse modelo de ensino em que a capacidade de organização e autonomia

são essenciais. Nesse sentido, eles sentiram falta de um professor para auxiliá-los não somente com suas dúvidas de pronúncia, mas também para guiá-los em relação ao uso da ferramenta. Nenhum problema relacionado ao aplicativo em si foi mencionado nessa pergunta, o que indica que, atualmente, os maiores desafios para o uso destas tecnologias estão ligados às questões metodológicas e organizacionais, e não às tecnologias em si.

Considerações finais

Esta pesquisa buscou explorar e investigar uma nova possibilidade para o ensino de pronúncia em língua estrangeira, o Reconhecimento Automático de Fala (RAF). Mais especificamente, buscou-se verificar se o uso do *ELSA Speak*, um aplicativo baseado no RAF, poderia tornar a fala de estudantes da língua inglesa mais inteligível. Os resultados dos testes de inteligibilidade revelaram que a produção oral dos participantes ficou significativamente mais inteligível após utilizarem a ferramenta, tanto quando foram consideradas todas as palavras pronunciadas quanto quando consideradas apenas palavras-alvo com vogais altas anteriores. Além disso, notamos que os participantes que levaram mais de um mês para completar a unidade do aplicativo e que passaram muitas semanas sem utilizá-lo, não obtiveram ganhos nos pós-testes.

Em geral, os estudantes que participaram da pesquisa apreciaram a oportunidade de utilizar o aplicativo e explorar suas funcionalidades, e tiveram percepções majoritariamente positivas sobre o *feedback* fornecido e sobre o sistema de RAF. Apesar disso, relataram momentos em que acreditam que sua fala não foi compreendida, ou em que tiveram dificuldade em compreender o *feedback*. Considerando o contexto de ensino remoto que viviam, os estudantes sentiram falta da orientação dos professores e da interação com colegas, e apresentaram dificuldades para utilizar a ferramenta de forma organizada. Além disso, o fato de apenas 13 dos 34 participantes que iniciaram a participação na pesquisa a terem concluído indica que, apesar da grande motivação inicial, estudantes do ensino fundamental e médio necessitam de suporte para

manterem-se motivados, especialmente quando as atividades de aprendizagem não são parte da grade curricular.

A partir desses resultados, acreditamos que aplicativos como o *ELSA Speak* têm grande potencial para auxiliar estudantes de língua inglesa a se tornarem mais inteligíveis. Essas ferramentas podem ser muito valiosas quando consideramos o tempo limitado para desenvolver atividades de pronúncia em sala de aula e que, em muitos contextos, não é possível dar um retorno individualizado à pronúncia de cada estudante em sala de aula. Todavia, é necessário orientá-los em relação às suas potencialidades e limitações, e integrar seu uso ao currículo escolar para que sejam compreendidas como ferramentas de aprendizagem valiosas.

Os resultados desta pesquisa mostram-se positivos em relação à aplicabilidade de sistemas de RAF para o ensino e aprendizagem de línguas. Entretanto, ainda são necessários estudos que investiguem esses sistemas, especialmente com um número mais significativo de participantes. Devido às limitações causadas pelo ensino remoto, esta pesquisa teve um número reduzido de participantes, o que inviabilizou a criação de um grupo controle. Considerando as dificuldades relatadas em relação ao uso do aplicativo nesse estudo, futuras pesquisas podem, também, investigar o uso de sistemas baseados no RAF em situações com maior ou menor apoio de professores de língua.

Referências

AULIA, V. Reducing students' pronunciation anxiety by utilizing WhatsApp group in learning English consonant sounds. *ELT Worldwide*, [s. l.], v. 5, n. 2, p. 165-178, 2018.

BAMGBOSE, A. Torn between the norms: innovations in world Englishes. *World Englishes*, [s. l.], v. 17, n. 1 p. 1-14, 1998.

BASHORI, M. *et al.* Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. *System*, [s. l.], v. 99, n. 1, p. 1-16, 2021.

BECKER, M. R. **Inteligibilidade da língua inglesa sob o paradigma da Língua Franca**: Percepção de discursos de falantes de diferentes L1s por brasileiros. 2013. 256 f.

Tese (Doutorado em Letras) – Setor de Ciências Humanas, Letras e Artes, Universidade Federal do Paraná, Curitiba, 2013.

BOTERO, G. *et al.* Self-directed language learning in a mobile-assisted, out-of-class context: do students walk the talk? **Computer Assisted Language Learning**, [s. l.], v. 31, n. 1, p. 71-91, 2019.

BUSS, L. Pronunciation from the perspective of pre-service EFL teachers: An analysis of internship reports. In: J. LEVIS; K. LEVELLE (Eds.). **Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference**. Ames, IA: Iowa State University, 2013. p. 255-264.

COSTA, B. **Pronunciation Teaching Is Not a One-Size-Fits-All Endeavor: EFL Teachers' Beliefs and Classroom Practices**. 2016. 112 f. Dissertação (Mestrado em Inglês: Estudos Linguísticos e Literários) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2016.

CRUZ, N. C. An exploratory study of pronunciation intelligibility in the Brazilian learners' English. **The Specialist**, [s. l.], v. 24, n. 2, p. 155-175, 2003.

DELATORRE, F; SILVEIRA, R; GONÇALVES, A. The intelligibility of English verbs in the simple past tense: native and non-native speakers and Brazilian listeners. **Veredas Online**, [s. l.], v. 2, p. 57-77, 2017.

DERWING, T.; MUNRO, M.; CARBONARO, M. Does popular speech recognition software work with ESL speech? **TESOL Quarterly**, [s. l.], v. 34, p. 592-603, 2000.

DERWING, T.; MUNRO, M. Second Language Accent and Pronunciation Teaching: A Research-Based Approach. **TESOL Quarterly**, [s. l.], v. 39, n. 3, 2005.

DERWING, T.; MUNRO, M. Putting accent in its place: rethinking obstacles to communication. **Language teaching**, [s. l.], v. 37, n. 1, p. 476-490, 2009.

ELIMAT, A.K.; ABUSEILEEK, A.F. Automatic Speech Recognition Technology as an Effective Means for Teaching Pronunciation. **JALT CALL Journal**, [s. l.], v. 10, n. 1, p. 21-46, 2014.

ELSA API. How does ELSA API work? **ELSA – World leading Speech Recognition API**, 2022. Disponível em: <https://bit.ly/3LtOcS1>. Acesso em: 20 abr. 2022.

FIELD, J. Intelligibility and the Listener: The Role of Lexical Stress. **TESOL Quarterly**, [s. l.], v. 39, n. 3, p. 399-423, 2005.

FOUZ-GONZÁLEZ, J. Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. **Language Learning & Technology**, [s. l.], v. 21, n. 1, p. 62-85, 2020.

GILAKJANI, A. P.; SABOURI, B. N. Why Is English pronunciation ignored by EFL teachers in their classes? **International Journal of English Linguistics**, [s. l.], v. 6, n. 6, p. 195-208, 2016.

GONÇALVES, A; SILVEIRA, R. In search of speech intelligibility: the case of English high front vowels produced by Brazilian learners. **Linguagem & Ensino**, Pelotas, v. 17, n. 2, p. 361-390, 2014.

GONÇALVES, A; SILVEIRA, R. Intelligibility research in Brazil: empirical findings and methodological issues. **Horizontes de Linguística Aplicada**, [s. l.], n. 1, p. 51-81, 2015.

HENDERSON, A. *et al.* The English pronunciation teaching in Europe survey: Selected results. **Research in Language**, [s. l.], v. 10, n. 1, p. 5-27, 2012.

HINCKS, R. Technology and Learning Pronunciation. *In*: REED, M.; LEVIS, J (orgs.). **The Handbook of English Pronunciation**. Wiley Blackwell: Oxford, 2015. p. 504-519.

INEP. **Dados do Censo Escolar**: Ensino Médio brasileiro tem média de 30 alunos por sala, 2019. Disponível em: <http://portal.inep.gov.br/>. Acesso em: 21 dez.2022.

JENKINS, J. **The Phonology of English as an International Language**. Oxford, UK: Oxford University Press, 2000.

JENKINS, J. Repositioning English and multilingualism in English as a Lingua Franca. **Englishes in Practice**, [s. l.], v. 2, p. 49-85, 2015.

KAISER, D. Mobile-Assisted Pronunciation Training: The iPhone Pronunciation App Project. **IATEFL Pronunciation Special Interest Group Journal**, [s. l.], v. 58, n. 1, p. 38-52, 2018.

KHADDAGE, F.; MULLER, W.; FLINTOFF, K. Advancing Mobile Learning in Formal And Informal Settings via Mobile App Technology: Where to From Here, and How? **Educational Technology & Society**, [s. l.], v.19, n. 3, p. 16-26, 2016.

LIAKIN, D.; CARDOSO, W.; LIAKINA, N. Learning L2 pronunciation with a mobile speech recognizer: French /y/. **Computer Assisted Language Instruction Consortium**, [s. l.], v. 32, n. 1, p.1-25, 2015.

MAHDI, S. H.; KHATEEB, A. A. The effectiveness of computer-assisted pronunciation training: A meta-analysis. **Review of Education**, [s. l.], v. 7, n. 3, p. 733-753, 2019.

MUNRO, M.; DERWING, T.; MORTON, S. The mutual intelligibility of L2 Speech. **Studies in Second Language Acquisition**, [s. l.], v. 28, n. 1, p. 111-131, 2006.

MUNRO, M. J.; DERWING, T. M.; AND SAITO, K. English L2 vowel acquisition over

seven years. *In*: LEVIS, J; LEVELLE, K (Eds.). **Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference**. Ames, IA: Iowa State University, 2013.

NAIR, R. *et al.* Rethinking the teaching of pronunciation in the ESL classroom. **The English Teacher**, [s. l.], v. 35, p. 27-40, 2017.

NERI, A.; CUCCHIARINI, C.; STRIK, W. Automatic Speech Recognition for second language learning: How and why it actually works. **Speech works**, [s. l.], p. 1157-1160, 2003.

NERI, A.; CUCCHIARINI, C.; STRIK, W. ASR-based corrective feedback on pronunciation: does it really work? *In*: **Proceedings of Interspeech 2006- ICSLP, Ninth International Conference on Spoken Language Processing**, Pittsburgh, PA, EUA, 17-21 set. 2006.

PI-HUA, T. Bridging pedagogy and technology: User evaluation of pronunciation-oriented CALL software. **Australasian Journal of Educational Technology**, [s. l.], v. 2, n. 3, p. 375-397, 2006.

SAITO, K. Effects of Instruction on L2 Pronunciation Development: A Synthesis of 15 Quasi-Experimental Intervention Studies. **TESOL Quarterly**, [s. l.], v. 46, n. 4, p. 807-819, 2012.

SHA, L.; LOOI, C.-K.; CHEN, W.; ZHANG, B. H. Understanding mobile learning from the perspective of self-regulated learning. **Journal of Computer Assisted Language Learning**, [s.l.], v. 8, n.4, p. 366-378, 2012.

SILVEIRA, R.; SILVA, T. L2 Speech Intelligibility: effects of coda modification, degree of semantic information and listener's background. **Revista Brasileira de Linguística Aplicada**, [s. l.], v. 18, n. 3, p. 639-664, 2018.

THOMSON, R. ESL teachers' beliefs and practices in pronunciation teaching: Confidently right or confidently wrong? *In*: LEVIS, J; LEVELLE, K. (Eds.). **Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference**. Ames, IA: Iowa State University, 2012. p. 224-233.

TRAXLER, J. Defining, Discussing and Evaluating Mobile Learning: The moving finger writes and having writ . . . **The International Review of Research in Open and Distributed Learning**, [s.l.] v. 8, n. 2, p. 1-12, 2007.

Recebido em 15/05/2022.

Aprovado em 14/07/2022.