

UBIRATÃ KICKHÖFEL ALVES

PERCEPÇÃO AUDIOVISUAL DA FALA EM SEGUNDA LÍNGUA:

Uma discussão sobre o desenvolvimento das vogais do português brasileiro por aprendizes falantes de espanhol (L1)

Relatório de Pós-doutorado realizado na Universidade Estadual Paulista (UNESP), Faculdade de Filosofia e Ciências, Programa de Pós-Graduação em Ciências da Saúde e Comunicação Humana, Campus Marília-SP.

Supervisor(a): Prof. Dr. Larissa Cristina Berti

CNPq – Processo 101646/2024-0

Marília, SP

2026

PERCEPÇÃO AUDIOVISUAL DA FALA EM SEGUNDA LÍNGUA: UMA DISCUSSÃO SOBRE O DESENVOLVIMENTO DAS VOGAIS DO PORTUGUÊS BRASILEIRO POR APRENDIZES FALANTES DE ESPANHOL (L1)

RESUMO

Neste projeto, investigamos, por meio de experimentos de percepção das vogais do Português Brasileiro (PB) como segunda língua (L2), o papel das pistas auditivas e visuais na identificação desses segmentos por aprendizes hispanofalantes. Para isso, analisamos os efeitos de quatro condições de estímulo: (1) apenas auditiva (AO); (2) apenas visual (VO); (3) audiovisual congruente (AV+), na qual as pistas auditivas e visuais coincidem; e (4) audiovisual incongruente (AV-), na qual há conflito entre as pistas, com a informação visual remetendo a uma vogal distinta da expressa auditivamente (por exemplo, áudio /e/ e vídeo /ɛ/).

O estudo compreendeu dois experimentos conduzidos com o mesmo conjunto de estímulos, referentes às vogais médias do PB, apresentados nas quatro modalidades descritas. O Estudo I investigou a assimilação das vogais do PB às categorias vocálicas do catalão. Participaram 30 falantes bilíngues Catalão–Espanhol, residentes em Barcelona e região metropolitana, sendo 13 dominantes em espanhol (língua sem /ɛ/ e /ɔ/ em seu inventário fonológico) e 17 dominantes em catalão (língua que contempla todo o inventário vocálico do PB). Os participantes identificaram os estímulos do PB a partir das categorias /e/, /ɛ/, /o/ e /ɔ/ do catalão.

O Estudo II teve como foco a identificação das vogais por aprendizes hispanofalantes em processo de aquisição do PB. Participaram quatro grupos: (i) 13 aprendizes de nível básico, residentes no Brasil, sem instrução explícita de pronúncia; (ii) 13 aprendizes de nível avançado, também sem instrução; (iii) 16 aprendizes avançados, estudantes de Licenciatura em Língua Portuguesa como L2, residentes no exterior e com quatro semestres de instrução em Fonética e Fonologia do PB; e (iv) um grupo controle composto por 23 falantes nativos brasileiros.

Os resultados do Estudo I indicaram que os bilíngues com dominância em catalão apresentaram índices de assimilação corretos significativamente superiores aos dos dominantes em espanhol. Em ambos os grupos, não houve diferenças significativas entre as condições AO e AV+, cujos índices foram superiores aos de VO. Nos estímulos incongruentes, ambos os grupos priorizaram a informação auditiva, embora os dominantes em catalão tenham apresentado preferência significativamente maior pelo áudio.

No Estudo II, o grupo controle brasileiro apresentou alta acuidade nas condições AO e AV+, mas não em VO. Os aprendizes dos níveis básico e avançado sem instrução não diferiram estatisticamente entre AO, AV+ e VO. Em contraste, o grupo avançado com instrução explícita apresentou índices de acuidade significativamente mais altos em AV+ do que em AO e VO. Esses resultados indicam que o conhecimento explícito do sistema fonológico favorece a integração audiovisual na identificação vocálica. Quanto aos

estímulos incongruentes, embora os brasileiros tenham demonstrado preferência quase categórica pela pista auditiva, os demais aprendizes não exibiram preferência estatística clara, sugerindo indecisão e uso concomitante das pistas.

Em linhas gerais, os achados contribuem para o debate sobre modelos de percepção audiovisual da fala, oferecendo evidências empíricas compatíveis com abordagens que defendem a ponderação probabilística entre pistas auditivas e visuais, como o *Fuzzy Logical Model of Speech Perception* (FLMP; Massaro, 1998, 2004) e o *Bayesian Framework of Speech Perception* (Knill; Pouget, 2004; Ernst; Bühlhoff, 2004). Defendemos que esses modelos podem ser associados ao *Speech Learning Model* (Flege, 1995; Flege; Bohn. 2021), por esse modelo de L2 também operar com ponderação de pistas (ainda que unicamente acústicas, o que implica modificações adicionais ao modelo). A partir desses resultados, o projeto abre importantes caminhos de investigação referentes às bases epistemológicas dos modelos gerais de percepção audiovisual e dos modelos de percepção de sons da L2.

PALAVRAS-CHAVE: Percepção audiovisual da fala; Desenvolvimento fonético-fonológico de L2; Unidades representacionais; Modelos de percepção de L2; Fonologia de Laboratório.

1. INTRODUÇÃO

O presente relatório visa a apresentar os principais resultados e discutir o impacto do projeto “Percepção Audiovisual da Fala em Segunda Língua: Uma discussão sobre o desenvolvimento das vogais do Português Brasileiro por aprendizes falantes de Espanhol (L1)”, realizado junto ao Programa de Pós-Graduação em Ciências da Saúde e Comunicação Humana da UNESP – Marília..

O referido projeto teve origem a partir da constatação de que os principais modelos perceptuais de L2 vigentes, o *Speech Learning Model* (SLM – Flege, 1995; Flege; Bohn, 2021) e o *Perceptual Assimilation Model-L2* (PAM-L2 – Best; Tyler, 2007), não acomodam, em suas premissas e princípios, o modo como a formação de novas categorias de sons na L2 se dá a partir de um contexto audiovisual de informação. Tampouco tais modelos discutem, de forma explícita, o modo como a percepção audiovisual pode ser contemplada a partir dos primitivos representacionais previstos por tais modelos. Frente a essa lacuna, investigamos, por meio de experimentos de percepção das vogais do Português Brasileiro (L2), o papel das pistas auditiva e visual na identificação de tais segmentos vocálicos por parte de aprendizes hispanofalantes. Para isso, verificamos os efeitos referentes a quatro distintas condições de estímulo: (1) apenas auditiva (AO, *audio only*); (2) apenas visual (VO, *visual only*); (3) audiovisual congruente (AV+), na qual as pistas auditivas e visuais coincidiam; e (4) audiovisual incongruente (AV–), na qual as pistas auditivas e visuais entravam em conflito, com a informação visual remetendo a uma vogal diferente daquela expressa pela informação auditiva (por exemplo, informação auditiva remetendo a /e/ e informação visual referente a /ε/).

No que diz respeito ao âmbito teórico, o projeto buscou discutir as bases epistemológicas dos modelos de natureza dinâmica vigentes no cenário de L2 (SLM e PAM-L2), com relação às suas unidades representacionais basilares e às suas propostas de teorização acerca do processo de assimilação perceptual. Por sua vez, no que diz respeito à verificação empírica, o estudo contou com dois experimentos. Isso foi possível em função do período de visita pós-doutoral à *Universitat Pompeu Fabra* (Espanha), durante o mês de fevereiro de 2025, viabilizada pela Bolsa de Desenvolvimento Tecnológico e Inovação no Exterior Sênior - DES (Processo número [200111/2025-6]), outorgada pelo CNPq no âmbito do projeto maior de Pós-Doutorado Sênior. A referida visita possibilitou, para além do experimento originalmente previsto no projeto, a inclusão de um estudo do tipo *cross-language* (cf. Bohn, 2018), também chamado de

“assimilação perceptual” (cf. Best, 1995; Flege, 1995), com ouvintes “ingênuos” do PB (ou seja, ouvintes que não se encontravam em processo de aquisição da referida língua e que, além disso, não haviam recebido nenhuma forma de *input* na língua em questão).

Ambos os estudos foram conduzidos com o mesmo conjunto de estímulos, referentes às vogais médias do Português Brasileiro (PB). Conforme aponta a literatura da área (Machry da Silva, 2014; Feiden *et al.*, 2016; Alves; Santana, 2020; Alves *et al.*, 2020; Castro; Albuquerque; Gomes, 2020; Osborne, 2021; Santana, 2021), o desenvolvimento das vogais médias-baixas mostra-se difícil por aprendizes que têm o Espanhol como L1. Os estímulos foram produzidos por quatro falantes do interior do estado de São Paulo, da cidade de Marília (dois homens e duas mulheres), os quais repetiram cinco vezes cada uma das vogais /e/, /ɛ/, /o/, /ɔ/, bem como a vogal /a/ (usada como *catch trial*)¹. Após o registro das cinco repetições de cada vogal, todos os *tokens* foram analisados acusticamente quanto aos valores de F1, F2 e duração. O *token* mais próximo dos valores de referência reportados na literatura foi selecionado para servir como estímulo experimental a ser acrescentado ao instrumento perceptual.

Uma vez selecionadas as produções-alvo, os estímulos foram editados manualmente no *Software Filmora*, para que gerássemos arquivos com duração de dois segundos. Os clipes de vídeo foram recortados de modo a exibir a parte inferior do rosto (dos ombros ao nariz) dos locutores, com os olhos excluídos. Utilizando o mesmo *software* e com base nas gravações originais de áudio e vídeo, foram realizadas manipulações adicionais para criar estímulos nas modalidades AO, VO e AV incongruente.

Apesar de ambos os experimentos contarem com exatamente os mesmos estímulos (tanto em termos quantitativos quanto qualitativos), por se tratar de fenômenos diferentes investigados em cada experimento (assimilação por ouvintes ingênuos e identificação por aprendizes em diferentes graus de proficiência), a tarefa solicitada em cada um dos estudos foi distinta, conforme será visto a seguir. Para fins de organização, denominaremos cada um desses estudos como “Estudo I” (estudo *cross-language* realizado em Barcelona) e “Estudo II” (estudo com aprendizes de PB, previsto

¹ Conforme explica Silveira (2016), os *catch trials* são estímulos cuja identificação não deveria implicar dificuldades para os participantes. Dessa forma, um baixo desempenho na identificação desses estímulos sugere o não comprometimento do aprendiz com a tarefa. Assim, seriam excluídos do estudo os participantes hispanofalantes que obtivessem menos do que 80% de acuidade nos *catch trials*. Entretanto, em ambos os experimentos, nenhum participante apresentou índices de acuidade menores do que esse valor limiar.

originalmente no projeto), de modo a descrevermos suas metodologias e, posteriormente, seus resultados. Em ambos os estudos, verificamos efeitos referentes ao grupo de participantes e ao tipo de estímulo (AO, VO ou AV). No caso dos estímulos manipulados (AV-), investigamos se, dada a natureza incongruente do estímulo, os participantes o identificariam a partir da informação auditiva ou visual.

O Estudo I (*cross-language*) contou com a participação de 30 falantes bilíngues Catalão-Espanhol, residentes em Barcelona e sua região metropolitana, sendo que 13 reportaram apresentar dominância em Espanhol (língua que não apresenta /ɛ/ e /ɔ/ em seu inventário fonológico) e 17 em Catalão (língua que apresenta, em seu inventário fonológico, todas as vogais do inventário do PB). Após serem expostos às vogais do PB nas quatro modalidades de estímulo, os participantes deveriam identificar os estímulos a partir das categorias de /e/, /ɛ/, /o/, /ɔ/ do Catalão, de modo a evidenciar a assimilação das vogais do PB à referida língua. O estudo visava a verificar os índices de acuidade de assimilação em cada um dos grupos, indagando se haveria diferenças no desempenho de cada grupo em função da modalidade de estímulo.

O Estudo II (com aprendizes hispanofalantes em processo de aquisição do PB) solicitou que os aprendizes identificassem as vogais a partir das próprias categorias do PB. O estudo contou com quatro grupos: (i) um grupo de 13 aprendizes de nível básico de proficiência², residentes no Brasil e sem nenhuma instrução de pronúncia; (ii) um grupo de 13 aprendizes de nível avançado de proficiência, também sem instrução; (iii) um grupo de 16 aprendizes de nível avançado de proficiência, estudantes de cursos de Licenciatura de Língua Portuguesa como L2 e residentes no exterior, tendo todos eles recebido instrução (quatro semestres) sobre Fonética e Fonologia do PB; (iv) um grupo controle, composto por 23 participantes brasileiros e residentes no Brasil. A inclusão dos grupos (i) e (ii) permitiu verificar se os efeitos exercidos por diferentes tipos de estímulo se alteram em função do nível de proficiência do aprendiz. Por sua vez, a inclusão de um grupo de aprendizes que recebeu instrução mostrou-se importante por abrir caminhos para a discussão sobre a possibilidade de o conhecimento explícito acerca do sistema de sons exercer efeitos nos padrões de respostas, considerando-se cada um dos diferentes tipos de estímulo. Finalmente, o grupo controle (participantes brasileiros) mostrou-se importante, sobretudo, para a definição sobre como os falantes nativos da língua tratam os estímulos incongruentes (se a partir de uma preferência pelo estímulo de áudio ou pelo de vídeo).

² O nível de proficiência dos aprendizes foi aferido a partir do questionário de autoavaliação de Scholl e Finger (2013).

A partir da verificação dos resultados desses dois experimentos, consideramos possível não somente apontar empiricamente as diferentes tendências exibidas pelos grupos de participantes distintos, mas, também, discutir de que forma tais dados podem ser lidos a partir das concepções basilares de percepção audiovisual e dos modelos de percepção de L2 vigentes. No que segue, discutimos o impacto, os principais resultados teóricos e empíricos obtidos, bem como a lista de publicações e apresentações decorrentes do projeto.

2. IMPORTÂNCIA E IMPACTO DO PROJETO

Consideramos que este projeto contribuiu para preencher uma lacuna referente a estudos sobre percepção audiovisual caracterizados não somente pela descrição empírica, mas também pela discussão teórica. Tal discussão é instanciada a partir do debate sobre como os modelos perceptuais de L2 podem dar conta da multimodalidade do estímulo, considerando-se a natureza dos primitivos representacionais com os quais tais modelos operam. De nosso conhecimento, este foi o primeiro projeto, em contexto nacional, que explora a multimodalidade perceptual entre aprendizes de PB como L2. Em outras palavras, o presente projeto, ao se basear em modelos perceptuais contemporâneos, não somente buscou ampliar o rol dos sistemas de L2 investigados no contexto brasileiro de pesquisas, mas, de forma não menos importante, promoveu uma reflexão teórica acerca da viabilidade de aplicação de tais modelos para o entendimento da questão empírica abordada.

No que diz respeito ao âmbito teórico, cabe considerar que, mesmo no âmbito internacional, ainda são poucos os trabalhos que buscam discutir se e como modelos tradicionais de L2, como o *Speech Learning Model* (SLM) e o *Perceptual Assimilation Model-L2* (PAM-L2), conseguem (ou não) dar conta de dados audiovisuais. Frente a esse desafio, o projeto contribuiu para um avanço nas discussões da área ao promover um debate sobre as bases epistemológicas desses dois modelos, com vistas a associá-los a propostas teóricas mais amplas (da área da psicologia perceptual), referentes à percepção multimodal da fala. Para isso, foi necessário investigar, também, as bases epistemológicas de uma série de modelos de percepção audiovisual, tais como o *Fuzzy Logical Model of Speech Perception* (Massaro, 1998, 2004), o *Bayesian Framework of Speech Perception* (Knill; Pouget, 2004; Ernst; Bühlhoff, 2004), e a Perspectiva Amodal de Rosenblum

(2005, 2008). Conforme será possível verificar na seção seguinte (referente aos resultados), o projeto contribuiu para o estabelecimento de pontes teóricas entre os modelos de L2 e tais modelos multimodais (Alves; Berti; Reis, 2025).

Na discussão de possíveis alterações aos modelos perceptuais de L2, especial atenção foi dada ao *Perceptual Assimilation Model-L2* (Best; Tyler, 2007), que tem por base epistemológica o Realismo Direto (Gibson, 1966, 1979; Fowler, 1986, 1996; Gibson; Pick, 2000), ancorado em um primitivo gestual de Fonologia (Browman; Goldstein, 1986, 1992). Conforme será visto na seção de resultados, no que diz respeito ao modelo em questão, propomos a discussão de novas *affordances* (cf. Gibson, 1966, 1979) referentes ao aprendizado multimodal de L2. Com tal proposição, permitimos que o modelo passasse a acomodar contextos de sala de aula e de treinamento perceptual (não previstos originalmente em Best e Tyler, 2007), além de dar conta, de uma maneira mais harmoniosa, da natureza multimodal da fala. Nesse sentido, ainda que os autores do PAM-L2 alegassem a possibilidade de o modelo dar conta de informação visual (cf. Best; Tyler, 2004), até então não havia sido proposta nenhuma adaptação formal ao modelo, seja a partir da associação do primitivo gestual a uma concepção multimodal maior (como por nós realizado em Alves, Berti e Reis, 2025), seja a partir de previsões decorrentes da própria associação do modelo ao Realismo Direto (tal como a discussão por nós proposta em Reis e Alves, 2026, publicada na Revista *Ecological Psychology*, um importante periódico internacional sobre temas de percepção voltados à escola do Realismo Direto).

No que diz respeito ao plano empírico, o projeto proposto mostrou-se inovador por investigar os efeitos de uma série de variáveis preditoras na percepção multimodal, tais como o nível de proficiência dos aprendizes (básico e avançado) e o tipo de modalidade de estímulo fornecido (somente áudio, somente vídeo, e áudio e vídeo – nas modalidades congruente e incongruente). É importante ressaltar o caráter de ineditismo das verificações empíricas realizadas, uma vez que o projeto constituiu o primeiro trabalho, no cenário nacional, a investigar os efeitos do tipo de modalidade de estímulo em dados de L2. Além disso, ainda que efeitos da multimodalidade da fala na percepção de L2 já tenham sido investigados no cenário internacional (Reisberg *et al.*, 1987; Hazan *et al.*, 2002, 2006; Hardison, 2005; Navarra; Soto-Faraco, 2007; Hirata; Kelly, 2010; Springer, 2021; Reyes; Hazan, 2021), ainda são poucos os estudos voltados à investigação do papel de estímulos incongruentes na L2 (Sekiya; Burnham, 2008; Wang *et al.*, 2008; Chen; Hazan, 2009; Hazan *et al.*, 2010; Barros-Loscertales *et al.*, 2013; Marian *et al.*, 2018).

Para além desse fato, conforme já dito anteriormente, o projeto configurou-se como o primeiro estudo sobre modalidade da fala com foco no Português Brasileiro.

Também no que diz respeito ao aspecto empírico, conforme já dito na Introdução, o escopo do projeto foi expandido com a atribuição da Bolsa de Desenvolvimento Tecnológico e Inovação no Exterior Sênior - DES (Processo número [200111/2025-6]), a qual nos permitiu, ao longo do mês de fevereiro de 2025, desenvolver pesquisas na *Universitat Pompeu Fabra* (Barcelona). Essa oportunidade possibilitou a realização do Estudo I, de assimilação perceptual (*cross-language study*, cf. Bohn, 2018) das vogais do PB ao Catalão (vejam-se detalhes adicionais na Introdução). Ressaltemos, também, o ineditismo desse recorte de pesquisa, uma vez que não conhecemos, no cenário internacional, pesquisas que tenham associado, em um único estudo, os eixos de (i) percepção audiovisual; (ii) assimilação perceptual e (iii) graus de dominância no bilinguismo. Os resultados desse estudo serão apresentados na seção seguinte.

Ainda no que diz respeito à etapa empírica, os resultados do Estudo II (estudo de aquisição com aprendizes hispanofalantes) também se mostraram de grande ineditismo e relevância para o cenário brasileiro de investigações. Consideramos que uma das principais contribuições foi o mapeamento dos estímulos incongruentes nos dados de ouvintes brasileiros, verificação empírica até então não realizada em contexto nacional. Para além desse fato, conforme constataremos nos resultados, a verificação dos efeitos de dois níveis de proficiência (básico e avançado) também possibilitou achados importantes acerca do papel da multimodalidade, ao demonstrar que os efeitos dos diferentes tipos de informação (AO, VO, AV+ e AV-) variam em função do nível de proficiência do aprendiz.

Cabe mencionar que o Estudo II expandiu o recorte previsto no projeto, uma vez que, para além dos participantes hispanofalantes residentes no Brasil (sem instrução de pronúncia), o estudo também incluiu a participação de aprendizes residentes no exterior, alunos de curso de Licenciatura em Inglês, que haviam anteriormente recebido instrução de Fonética ao longo de quatro semestres de graduação. Conforme será visto na seção anterior, os resultados mostram-se de grande importância por se mostrarem distintos daqueles verificados entre os aprendizes de mesmo nível de proficiência que nunca haviam recebido instrução. Tais resultados abrem caminhos para uma futura discussão acerca da relação entre “consciência acerca dos sons da L2” e multimodalidade, discussão essa a ser ainda perseguida em nosso grupo de estudos.

Em suma, para além de seu ineditismo tanto em termos teóricos quanto empíricos (conforme evidenciado nesta seção), consideramos que o projeto tem sua principal relevância e impacto no fato de abrir caminhos para um novo tema de investigações no cenário nacional. Os resultados deste período de 12 meses de estudo instituem uma série de temas de pesquisa a serem ainda perseguidos, tais como o aprofundamento referente a possíveis premissas epistemológicas dos modelos, o papel da consciência sobre os sons na percepção multimodal, os efeitos do ensino de L2 com base na multimodalidade, dentre muitos outros.

3. PRINCIPAIS RESULTADOS OBTIDOS

Conforme dito na seção anterior, o presente projeto exerceu impacto tanto no plano teórico quanto no empírico. Dessa forma, a relação de resultados obtidos será organizada a partir desses dois âmbitos.

No que diz respeito ao âmbito teórico, obtivemos os seguintes resultados:

- Ao considerarmos os três modelos de percepção audiovisual analisados (*Fuzzy Logical Model of Speech Perception*, o *Bayesian Framework of Speech Perception*, e a Perspectiva Amodal), a partir de suas bases epistemológicas, esses foram agrupados em dois grupos: (i) modelos que consideram a integração de pistas audiovisuais e visuais a partir da ponderação de tais pistas (*Fuzzy Logical Model of Speech Perception*, o *Bayesian Framework of Speech Perception*) e (ii) modelos que consideram o percepto como uma unidade cujos componentes auditivo e visual se mostram inseparáveis entre si (Perspectiva Amodal). Tais resultados foram publicados em Alves, Berti e Reis (2025).

- Com base no estudo das bases epistemológicas dos modelos perceptuais de L2 e dos modelos gerais de percepção audiovisual, propusemos que, em função de sua natureza voltada à ponderação de pistas, o *Speech Learning Model* pode ser associado epistemologicamente aos modelos *Fuzzy Logical Model of Speech Perception* e *Bayesian Framework of Speech Perception*. Tal associação se dá, também, em função de tanto o modelo de L2 quanto os suprarreferidos modelos audiovisuais basearem suas previsões a partir da ponderação e “concorrência” entre diferentes fontes de informação (ainda que,

no SLM, tais pistas sejam unicamente acústicas, o que implica modificações adicionais ao modelo). Tais resultados foram publicados em Alves, Berti e Reis (2025).

- Na mesma linha anterior, propusemos que, por adotar uma visão gestual (Browman; Goldstein, 1986, 1992), o *Perceptual Assimilation Model-L2* pode ser associado epistemologicamente à Perspectiva Amodal de Rosenblum (2005, 2008), que prevê que tanto a informação visual quanto a auditiva são partes indissociáveis e inseparáveis, sem pensar numa “competição” entre essas pistas. Tais resultados foram publicados em Alves, Berti e Reis (2025).

- Ao contrário do previsto por Best e Tyler (2007), a partir de uma discussão baseada na premissa Realista Direta (Gibson, 1966, 1979; Fowler, 1986, 1996; Gibson; Pick, 2000), defendemos que o PAM-L2 se mostra apropriado para lidar com eventos instrucionais em L2, o que garante sua aplicação a contextos para além dos de imersão. Tais resultados foram publicados em Reis e Alves (2026).

- Visando a uma coerência epistemológica que permitisse ao PAM-L2 dar conta tanto de contextos de instrução quanto da própria multimodalidade da fala, propusemos construtos referentes a *affordances* (cf. Gibson, 1966, 1979), os quais, à luz do modelo, seriam ativados quando da percepção de novos sons. Partindo da distinção entre *affordances* de alta e baixa ordem (Stoffregen; Wagman, 2025), propusemos, para os contextos instrucionais, a *affordance* de alta ordem de “falabilidade” (*speakability*), a qual abarca as *affordances* de baixa ordem de “gesturabilidade” (*gesturability*) e “expressabilidade” (*expressability*). Tais resultados foram publicados em Reis e Alves (2026).

Com relação ao âmbito empírico, os resultados serão organizados em dois subgrupos. No primeiro, descreveremos os resultados do estudo “*cross-language*” (Bohn, 2018), de assimilação dos padrões do português ao catalão por ouvintes bilíngues dominantes em catalão ou espanhol (Estudo I). No segundo, descreveremos os resultados referentes ao estudo de aquisição (Estudo II), com aprendizes hispanofalantes de Português Brasileiro.

Listamos, a seguir, os resultados referentes ao Estudo I:

- Mesmo ouvintes bilíngues que apresentam o catalão como língua não dominante foram capazes de assimilar as vogais do português ao catalão com índices de acuidade altos, acima dos considerados apenas “chance” (AO = 61,54%; AV+ = 64,90%). Tais resultados foram apresentados em Alves e Berti (2025) e serão publicados em Alves e Berti (no prelo).

- No que diz respeito à língua dominante, os bilíngues com dominância em catalão (AO = 87,13; AV+ = 88,24) apresentaram índices de assimilação corretos significativamente maiores do que os com dominância em espanhol (AO = 61,54%; AV+ = 64,90%). Esses resultados foram apresentados em Alves (2025a) e serão publicados em Alves e Berti (no prelo).

- No que diz respeito aos índices de assimilação frente aos estímulos AO, VO e AV+, em ambos os grupos de bilíngues, não há diferenças significativas entre AO e AV+. Esses dois índices, por sua vez, são significativamente superiores aos de VO (catalão dominante = 65,07%; espanhol dominante = 42,05%). Esses resultados foram apresentados em Alves (2025a) e serão publicados em Alves e Berti (no prelo).

- No que diz respeito aos estímulos com informações de áudio e vídeo incongruentes, os participantes de ambos os grupos de bilíngues (dominância em catalão ou em espanhol) preferem o estímulo auditivo ao visual, porém em proporções diferentes, visto que o grupo com dominância em catalão apresenta resultados significativamente mais altos de preferência pelos estímulos de áudio (91,17%) do que os com dominância em espanhol (58,16%). Nesse sentido, os participantes com dominância em catalão apresentam valores semelhantes aos encontrados na literatura por crianças e adultos monolíngues (Assis; Berti, 2025). Esses resultados foram apresentados em Alves (2025a) e serão publicados em Alves e Berti (no prelo).

No que diz respeito ao experimento levado a cabo com aprendizes hispanofalantes de Português Brasileiro (Estudo II), cabe destacar que todos os resultados serão publicados em Alves, Berti e Soncin (no prelo). Lembremos que, no referido experimento, participaram quatro grupos: (i) aprendizes de nível básico de proficiência que nunca haviam recebido instrução sobre pronúncia; (ii) aprendizes de nível avançado

de proficiência que nunca haviam recebido instrução de pronúncia; (iii) aprendizes de nível avançado de proficiência que haviam previamente recebido instrução de pronúncia; (iv) grupo de brasileiros, falantes nativos do PB, os quais serviram de controle. Listamos, a seguir, os resultados:

- Falantes nativos do PB apresentaram altos índices de acuidade na identificação dos padrões AO (98,91%) e AV+ (97,28%), mas não VO (72,01%), o qual tende a apresentar índices significativamente mais baixos do que os demais.

- Ao contrário do verificado com os falantes nativos, o grupo de participantes de nível básico (sem instrução) não apresentou índices de acuidade diferentes entre os três tipos de estímulo, com valores de 42,31% frente a estímulos AO, 44,71% frente a AV+ e 49,04% frente a estímulos VO. Ainda que com padrões de acuidade mais altos, tampouco no grupo de aprendizes avançados (sem instrução) foram verificadas diferenças significativas entre os tipos de estímulo, com valores de 51,92% frente a estímulos AO, 58,17% frente a AV+ e 56,73% frente a estímulos VO.

- Frente aos três tipos de estímulo (AO, AV+ e VO), o grupo de aprendizes avançados que havia previamente recebido instrução apresentou índices descritivos e inferenciais diferenciados, com índices de acuidade referentes ao padrão AV+ (76,95%) significativamente maiores do que os encontrados nos padrões AO (69,14%) e VO (67,57%). A partir desses dados, concluímos que o conhecimento explícito sobre o sistema de sons contribui para maiores índices de acuidade referentes ao padrão AV+. Sem tal conhecimento, os índices de acuidade referentes aos padrões AO e AV+ não se mostram inferencialmente distintos, provavelmente porque os aprendizes tomam a informação auditiva como prioritária. Esses resultados podem ser especialmente interessantes para perspectivas de percepção multimodal que operam com a noção de ponderação de pistas (Holt; Lotto, 2006; Lehet; Holt, 2017), em associação com o *Speech Learning Model* (Flege, 1995; Flege; Bohn, 2021).

- No que diz respeito aos estímulos com informações de áudio e vídeo incongruentes (AV-), verificamos que o grupo controle, de falantes nativos do português, baseou suas identificações quase que categoricamente (96, 99%) a partir do estímulo auditivo, o que vai ao encontro da literatura da área (Assis; Berti, 2025).

- Com relação aos grupos de aprendizes de nível básico e avançado que nunca receberam instrução de pronúncia, há uma diferença entre os padrões dos dois grupos: os aprendizes de nível básico tendem a mostrar índices praticamente semelhantes de preferência entre os estímulos de áudio e vídeo, exibindo uma preferência de 52,42% pelos estímulos de áudio. Por sua vez, diferentemente do grupo de falantes nativos, o grupo de avançados sem instrução exibe uma preferência (58,45%) pelo estímulo de vídeo. Já o grupo de aprendizes avançados que previamente receberam instrução tende a apresentar índices mais próximos ao do grupo de aprendizes básicos sem instrução, com uma preferência de 51,56% pela informação de áudio. Os resultados sugerem que, independentemente do grupo de aprendizes, esses se sentem mais confusos frente a estímulos incongruentes, ora priorizando um tipo de informação, ora outra. Esse resultado marca uma grande diferença entre os aprendizes e os falantes nativos do PB, os quais, mesmo frente a estímulos incongruentes, priorizam a informação de áudio praticamente de forma categórica. Em termos pedagógicos, os resultados chamam a atenção para a necessidade de fornecimento de estímulos AV congruentes, especialmente quando associados à instrução explícita referente à produção articulatória dos sons.

4 PRINCIPAIS³ PRODUÇÕES BIBLIOGRÁFICAS DIRETAMENTE RELACIONADAS AO PROJETO DE PESQUISA (organizadas de acordo com o tipo de produção, em ordem cronológica crescente referente ao ano de publicação)

Artigos em Periódicos (já publicados)

ALVES, U. K.; BERTI, L. C.; REIS, F. S. L2 speech perception models: A theoretical discussion on the incorporation of visual information. *Gradus: Revista Brasileira de Fonologia de Laboratório*, v. 10, n. 1, p. 75-102, 2025.

REIS, F. S.; ALVES, U. K. Assessing the Compatibility of Perceptual Training with Direct Realist Assumptions: A Focus on the ‘Gesturability’ Affordance. *Ecological Psychology*, v. 38, n.1, p. 15-47, 2026.

³ Nesta seção, são apresentadas as produções relacionadas diretamente ao projeto. Não se trata, portanto, de uma listagem exaustiva de nossas produções ao longo do ano. A relação completa de trabalhos poderá ser vista em nosso CV Lattes.

Artigos em Periódicos (no prelo)

ALVES, U. K.; BERTI, L. C. Perceptual assimilation of Brazilian Portuguese mid vowels into Catalan: A multimodal study of bilinguals with Spanish or Catalan dominance. *Lengua y Habla*, no prelo.

Artigos em Periódicos (em elaboração)

ALVES, U. K.; BERTI, L. C.; SONCIN, G. Percepção audiovisual das vogais do PB: uma discussão sobre diferentes níveis de proficiência. *Revista Acadêmica UCMaule*, em elaboração

Apresentações de trabalho

ALVES, Ubiratã Kickhöfel; BERTI, Larissa Cristina. *Assimilação perceptual à língua menos dominante de indivíduos bilíngues: efeitos da multimodalidade do estímulo*. III Fonapliando. Florianópolis: Universidade Federal de Santa Catarina, 2025a. (Apresentação de Trabalho/Comunicação).

ALVES, Ubiratã Kickhöfel. Assimilação perceptual das vogais médias do Português Brasileiro ao Catalão: Um experimento multimodal com bilíngues Espanhol-Catalão. 39º Encontro Nacional da ANPOLL. Rio Branco: Universidade Federal do Acre, 2025b. (Apresentação de Trabalho/Comunicação).

ALVES, Ubiratã Kickhöfel. Talk Show: Medidas acústicas da fala - Uma abordagem quantitativa para avaliação e pesquisa. *Congresso Brasileira de Fonoaudiologia 2025*. São Paulo, 2025c. (Apresentação de Trabalho/Conferência ou palestra)

ALVES, Ubiratã Kickhöfel. Inteligibilidade e Compreensibilidade da fala em L2: Caracterização e operacionalização. *I Ciclo de Conferências em Tópicos de Análise Linguística*. Marília: Universidade Estadual Paulista Júlio de Mesquita Filho, 2025d. (Apresentação de Trabalho/Conferência ou palestra).

5 REFERÊNCIAS DO PROJETO CITADAS AO LONGO DESTE RELATÓRIO⁴

ALVES, U. K.; AQUINO, C.; BUSKE, A. C. S.; SILVA, I. F. Efeitos da instrução explícita de pronúncia na inteligibilidade local: um estudo sobre a identificação, por

⁴ Referências mencionadas na seção anterior (produções referentes especificamente ao projeto) não serão repetidas nesta seção.

ouvintes brasileiros, de vogais médias anteriores produzidas por um aprendiz argentino de português brasileiro. *Veredas*, v. 24, n. 3, p. 219-247, 2020.

ALVES, U. K.; SANTANA, A. M. Desenvolvimento das vogais orais tônicas do Português Brasileiro por um aprendiz argentino: Uma análise de processo via Teoria dos Sistemas Dinâmicos Complexos (TSDCs). *Estudos Linguísticos e Literários*, v. 67, p. 390-418, 2020.

ASSIS, M. F.; BERTI, L. C. Speech perception: auditory and visual cue integration in children with and without phonological disorders in voiceless fricatives. *Clinical Linguistics and Phonetics*, v. 39, n. 1, p. 21-37, 2025.

BARRÓS-LOSCERTALES, A.; VENTURA-CAMPOS, N.; VISSER, M.; ALSIUS, A.; PALLIER, C.; RIVERA, C. A.; SOTO-FARACO, S. Neural correlates of audiovisual speech processing in a second language. *Brain & Language*, v. 126, p. 253-262, 2013.

BEST, C. T.; TYLER, M. D. Nonnative and second-language speech perception: commonalities and complementarities. In: BOHN, O.-S.; MUNRO, M. J. (Eds.). *Language Experience in Second Language Speech Learning – in honor of James Emil Flege*. Amsterdam: John Benjamins Publishing Company, 2007, p. 13-34.

BOHN, O.-S. Cross-language and second language speech perception. In: FERNÁNDEZ, E. M.; CAIRNS, H. S. (Eds.), *The Handbook of Psycholinguistics*. Hoboken: Wiley, 2018, p. 213-239.

BROWMAN, C.; GOLDSTEIN, L. Towards an Articulatory Phonology. *Phonology Yearbook*, v. 3, p. 219-252, 1986.

BROWMAN, C.; GOLDSTEIN, L. Articulatory Phonology: an overview. *Phonetica*, v. 49, p. 155-180, 1992.

CASTRO, T. G.; ALBUQUERQUE, J. I. A.; GOMES, M. L. C. A produção das vogais médias do Português como Língua Adicional: Uma análise de dados preliminares de um aprendiz chileno via Sistemas Dinâmicos Complexos. *Revista Virtual de Estudos da Linguagem*, v. 18, n. 35, p. 110-146, 2020.

CHEN, Y.; HAZAN, V. Developmental factors and the non-native speaker effect in auditory-visual speech perception. *The Journal of the Acoustical Society of America*, v. 126, n. 2, p. 858-865, 2009.

ERNST, M. O.; BÜLTHOFF, H. H. Merging the senses into a robust percept. *Trends in Cognitive Sciences*, v. 8, n. 4, p. 162-169, 2004.

FEIDEN, J. A.; PEROZZO, R. V.; FINGER, I.; FONTES, A. B. A. Percepção de vogais médias altas e médias baixas do português brasileiro por falantes de espanhol rioplatense em tarefa de discriminação cateórica. In: ALVES, U. K. (org.). *Aquisição Fonético-Fonológica de Língua Estrangeira: Investigações Rio-Grandenses e Argentinas em Discussão*. Campinas: Pontes Editores, 2016, p. 85-103.

FLEGE, J. E. Second Language Speech Learning: Theory, findings, and problems. In: STRANGE, W. (Ed.). *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press, 1995, p. 233-277.

FLEGE, J. E.; BOHN, O.-S. The Revised Speech Learning Model (SLM-r). In: WAYLAND, R. (Ed.). *Second Language Speech Learning: Theoretical and Empirical Progress*. Cambridge: Cambridge University Press, 2021, p. 3-83.

FOWLER, C. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, v. 14, p. 3-28, 1986.

FOWLER, C. Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, v. 99, n. 3, p. 1.730-1.741, 1996.

GIBSON, E. J.; PICK, A. D. *An ecological approach to perceptual learning and development*. Oxford: Oxford University Press, 2000.

GIBSON, J. J. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin, 1966.

GIBSON, J. J. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 2015 [1979].

HARDISON, D. Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, v. 26, n. 4, p. 579-596, 2005.

HAZAN, V.; KIM, J.; CHEN, Y. Audiovisual perception in adverse conditions: Language, speaker and listener effects. *Speech Communication*, v. 52, n. 11-12, p. 996-1009, 2010.

HAZAN, V.; SENNEMA, A.; FAULKNER, A. Audiovisual perception in L2 learners. *7th International Conference on Spoken Language Processing*, p. 1685-1688, 2002.

HAZAN, V.; SENNEMA, A.; FAULKNER, A.; ORTEGA-LLEBARIAD, M.; IBA, M.; CHUNG, H. The use of visual cues in the perception of non-native consonant contrasts. *Journal of the Acoustical Society of America*, v. 199, n. 3, p. 1740-151, 2006.

HIRATA, Y.; KELLY, S. D. Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, v. 53, n. 2, p. 298-310, 2010.

HOLT, L. L.; LOTTO, A. J. Cue weighting in auditory categorization: implications for first and second language acquisition. *Journal of the Acoustical Society of America*, v. 119, n. 5, p. 3.059-3.071, 2006.

KNILL, D. C.; POUGET, A. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, v. 27, n. 12, p. 713-719, 2004.

LEHET, M.; HOLT, L. L. Dimension-based statistical learning affects both speech perception and production. *Cognitive Science*, v. 41, n. S4, p. 885-912, 2017.

MACHRY DA SILVA, S. *Aprendizagem fonológica e alofônica em L2: percepção e produção das vogais médias do Português por falantes nativos do Espanhol*. Tese (Doutorado em Letras). Porto Alegre: Pontifícia Universidade Católica do Rio Grande do Sul, 2014.

MARIAN, V.; HAYAKAWA, S.; SCHROEDER, S. R. Language experience changes audiovisual perception. *Brain Sciences*, v. 8, n. 85, 1014, 2018.

MASSARO, D. W. *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press, 1998.

MASSARO, D. W. From multisensory integration to talking heads and language learning. In: CALVERT, G.; SPENCE, C.; STEIN, B. (Eds.). *The Handbook of Multisensory Processes*. The MIT Press, 2004, p. 153-176.

NAVARRA, J.; SOTO-FARACO, S. Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, v. 7, n. 1, p. 4-12, 2007.

OSBORNE, D. M. Perception of Portuguese mid vowels by heritage speakers of Brazilian Portuguese. *Heritage Language Journal*, v. 18, n. 1, p. 1-32.

REISBERG, D.; McLEAN, J.; GOLDFIELD, A. Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In: DODD, B.; CAMPBELL, R. (Eds.), *Hearing by Eye: The Psychology of Lip-reading*. Mahwah: Erlbaum, 1987, p. 97-113.

REYES, Y. P.; HAZAN, V. English vowel perception by non-native speakers: Impact of audio and visual training modalities. *Onomázein – Journal of Linguistics, Philology and Translation*, v. 51, p. 111-136, 2021.

ROSENBLUM, L. D. The primacy of multimodal speech perception. In: PISONI, D. B.; REMEZ, R. E. (Eds.). *Handbook of speech perception*. Blackwell, 2005, p. 51-78.

ROSENBLUM, L. D. Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, v. 17, n. 6, p. 405-409, 2008.

SANTANA, A. M. *Desenvolvimento vocálico em um aprendiz multilíngue (L1: Espanhol; L2: Inglês; L3: Português): uma análise via Sistemas Dinâmicos Complexos*. Trabalho de Conclusão de Curso (Graduação em Letras). Porto Alegre: Universidade Federal do Rio Grande do Sul, 2021.

SCHOLL, A. P.; FINGER, I. Elaboração de um Questionário de Histórico da Linguagem para pesquisas com bilíngues. *Nonada: Letras em revista*, v. 2, p. 1-17, 2013.

SEKIYAMA, K.; BURNHAM, D. Impact of language on development of auditory-visual speech perception. *Developmental Science*, v. 11, n. 2, p. 306-320, 2008.

SILVEIRA, R. *The influence of pronunciation instruction on the perception and the production of English word-final consonants*. Florianópolis: Advanced Research in English Series, n. 11, 2016.

SPRINGER, H. *L2 listeners benefit from audiovisual information in the processing of Swedish vowels*. Dissertação de Mestrado (Language and Linguistics). Lund, Suécia: Lund University, 2021.

STOFFREGEN, T. A.; WAGMAN, J. B. Higher order affordances. *Psychonomic Bulletin & Review*, v. 32, n. 1, p. 1-30, 2025.

TYLER, M.; BEST, C. The Perceptual Assimilation Model: Early bilingual adults and developmental foundations. In: AMENGUAL, M. (ed.). *The Cambridge Handbook of Bilingual Phonetics and Phonology*. Cambridge: Cambridge University Press, 2024, p. 147-172.

WANG, Y.; BEHNE, D. M.; JIANG, H. Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America*, v. 124, n. 3, p. 1716-1726, 2008.