



IBILCE - Instituto de Biociências, Letras e Ciências Exatas
Campus de São José do Rio Preto - UNESP

MAYARA FERREIRA DE ASSIS

**INTEGRAÇÃO DE PISTAS AUDIOVISUAIS NA PERCEPÇÃO DE FALA DE
FRICATIVAS SURDAS:** comparação de desempenho em diferentes faixas etárias

São José do Rio Preto

2025

MAYARA FERREIRA DE ASSIS

**INTEGRAÇÃO DE PISTAS AUDIOVISUAIS NA PERCEPÇÃO DE FALA DE
FRICATIVAS SURDAS:** comparação de desempenho em diferentes faixas etárias

Exame de defesa da tese apresentada ao Programa de Pós-graduação em Estudos Linguísticos – UNESP – Campus de São José do Rio Preto, como parte das exigências para a obtenção do título de Doutora em Estudos Linguísticos.

Área de concentração: Análise Linguística.

Linha de pesquisa: Oralidade e Letramento.

Orientadora: Profa. Dra. Larissa Cristina Berti

Financiamento: O presente trabalho foi realizado com o apoio da Fapesp (Processo nº 2021/15214-2).

São José do Rio Preto

2025

A848i Assis, Mayara Ferreira de
Integração de pistas audiovisuais na percepção
de fala de fricativas surdas: comparação de
desempenho em diferentes faixas etárias / Mayara
Ferreira de Assis. -- São José do Rio Preto, 2025
103 f. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista
(UNESP), Instituto de Biociências Letras e Ciências
Exatas, São José do Rio Preto
Orientadora: Larissa Cristina Berti

1. Percepção multimodal. 2. Identificação sonora.
3. Idade. 4. Crianças. 5. Adultos. 1. Título.

Sistema de geração automática de fichas catalográficas da Unesp.
Dados fornecidos pelo autor(a).

IMPACTO CIENTÍFICO E SOCIAL

Em termos de impacto científico, destaca-se que a presente tese aborda um tema inédito na medida em que a trajetória do desenvolvimento típico da percepção audiovisual de fala ainda é inexistente no português brasileiro, sobretudo considerando a classe das fricativas. O potencial inovador desse trabalho recai na caracterização pioneira dos padrões perceptuais audiovisuais desta classe fonológica, contribuindo para o avanço do conhecimento sobre a integração audiovisual na língua portuguesa. Entre as produções derivadas desta tese, podem ser destacadas: “Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental”, apresentado no XIII Congresso Internacional da Abralín e no XV Seminário de Estudos Linguísticos da UNESP (SELin); “Análise da percepção multimodal em crianças e adultos: sempre há ganho visual?”, apresentado no Congresso Brasileiro de Fonoaudiologia; *“Integration of audiovisual cues in speech perception: A developmental analysis”*, apresentado no International Child Phonology Conference. Adicionalmente, houve a premiação de um trabalho derivado desta tese apresentado no Congresso Brasileiro de Fonoaudiologia. A abrangência da tese pode ser considerada internacional, uma vez que contou com a colaboração do Prof. Dr. Denis Burnham. Sua complexidade é alta por se inserir na fronteira do conhecimento entre a linguística, a psicologia experimental e a fonoaudiologia. O trabalho impacta diretamente sobre as teorias de percepção de fala à medida em que fornece dados que apoiam a natureza audiovisual e desenvolvimental da percepção da fala, especialmente no caso de fricativas. Também podemos destacar um potencial social em duas direções: na avaliação da percepção da fala em crianças e na intervenção de crianças com os chamados Transtornos dos Sons da Fala. Em termos de avaliação, o experimento proposto na tese pode ser adaptado para um instrumento de avaliação da percepção audiovisual em crianças. Em termos de intervenção, os resultados da presente tese mostram que as pistas visuais podem ser utilizadas ou enfatizadas nos planos terapêuticos de reabilitação fonológica em Transtorno de Sons da Fala.

MAYARA FERREIRA DE ASSIS

**INTEGRAÇÃO DE PISTAS AUDIOVISUAIS NA PERCEPÇÃO DE FALA DE
FRICATIVAS SURDAS:** comparação de desempenho em diferentes faixas etárias

Exame de defesa da tese apresentada ao Programa de Pós-graduação em Estudos Linguísticos – UNESP – Campus de São José do Rio Preto, como parte das exigências para a obtenção do título de Doutora em Estudos Linguísticos.

Área de concentração: Análise Linguística.

Linha de pesquisa: Oralidade e Letramento.

Data da defesa: 21/11/2025

Banca Examinadora:

Profa. Dra. Larissa Cristina Berti, professora vinculada ao Departamento de Fonoaudiologia da Unesp - Campus de Marília, docente do corpo permanente do Programa de Pós-Graduação em Ciências da Saúde e Comunicação Humana UNESP - Marília e docente do corpo permanente do Programa de Pós-Graduação em Estudos Linguísticos UNESP - SJRP.

1º Examinador: _____

Profa. Dra. Athany Gutierrez, professora da Universidade Federal da Fronteira do Sul - UFFS, vinculada ao programa de Pós-graduação Estudos Linguísticos.

2º Examinador: _____

Profa. Dra. Carmen Lúcia Barreto Matzenauer, professora voluntária na Universidade Federal de Pelotas – UFPEL, vinculada ao programa de Pós-Graduação em Letras.

3º Examinador: _____

Prof. Dr. Lourenço Chacon Jurado Filho, vinculado ao programa de Pós-Graduação em Estudos Linguísticos UNESP – SJRP e ao programa de Pós-Graduação em Ciências da Saúde e Comunicação Humana da Unesp - Marília.

4º Examinador: _____

Profa. Dra. Marcia Keske-Soares, professora da Universidade Federal de Santa Maria - UFSM, vinculada ao programa de Pós-Graduação em Distúrbios da Comunicação Humana.

Aos meus pais, **Rose** e **Adilson**.

Aos meus avós, **Ivanides** e **João**.

AGRADECIMENTOS

Aos meus pais, **Rose e Adilson**, por serem minha âncora e me oferecem cada ensinamento com muito amor e cautela. Amo vocês.

À minha orientadora, **Profa. Dra. Larissa Cristina Berti**, que me presenteou com sua orientação, desde a iniciação científica. Muito obrigada, seus ensinamentos contribuíram para além do meu desenvolvimento acadêmico.

Ao **Dr. Denis Burnham** e à **Dra. Antonia Götz**, pelas sugestões e colaborações que contribuíram para o aperfeiçoamento deste trabalho. Ao **MARCS Institute for Brain, Behaviour and Development**, Sydney, Austrália, pela atenção e prontidão durante minha visita técnica à Western Sydney University.

Ao Programa de **Pós-Graduação em Estudos Linguísticos da UNESP (IBILCE/ Câmpus de São José do Rio Preto)**, pela atenção e prontidão durante as dúvidas e dificuldades ao longo do doutorado.

Ao **corpo docente** do Programa de Pós-Graduação em Estudos Linguísticos da UNESP, pelas excelentes disciplinas ministradas.

Aos **membros do Laboratório de Análise Acústica e Articulatória**, especialmente à **Lídia da Silva**, pelo apoio na elaboração do instrumento de avaliação, e aos queridíssimos **Cássio Esperandino e Isabella Domingues**, pela parceria e amizade ao longo deste percurso. Sou grata a todos pelos enriquecedores momentos compartilhados.

Aos participantes voluntários deste estudo e a seus familiares. Se não fossem vocês, este trabalho não teria acontecido.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo financiamento concedido (Processo FAPESP nº 2021/15214-2), por ter acreditado no nosso trabalho e por ter possibilitado dedicação exclusiva para execução do mesmo.

Enfim, a todos que colaboraram direta ou indiretamente para a concretização deste estudo.

RESUMO

Introdução: A perspectiva teórica sobre a percepção de fala como um evento audiovisual, inaugurada por McGurk e Macdonald (1976), considera que, na percepção de fala, há a integração perceptual de pistas de ordem auditiva (informação acústica) e visual (informação articulatória). A partir desta perspectiva, estudos apontam que a percepção de fala em crianças seria pautada majoritariamente no componente auditivo, enquanto a percepção de fala para os adultos se daria através da integração audiovisual, refletida por um aumento perceptual da pista visual, ao longo do desenvolvimento. Dessa forma, a idade seria um fator influente na percepção de fala (Polka, 1994; Sekiyama *et al.*, 2003; Sekiyama; Burnham 2004). No entanto, a trajetória do desenvolvimento típico da percepção audiovisual de fala ainda é inexistente no Português Brasileiro, sobretudo considerando a classe das fricativas.

Objetivo: O presente estudo investiga a percepção audiovisual da fala de fricativas surdas em função da idade, em crianças na faixa etária de 4 a 12 anos e em adultos na faixa etária de 20 a 30 anos. Assumindo que a integração audiovisual na percepção de fala depende da experiência linguística (Hazan; Barret, 2000; Burnham *et al.*, 1991), do desenvolvimento das habilidades psicoacústicas (Schneider *et al.*, 1986) e visuais (Desjardins, 1997), as hipóteses deste trabalho foram assim definidas: espera-se que os adultos apresentem maior acurácia perceptual comparados às crianças; e que as crianças de maior faixa etária apresentem maior acurácia perceptual em comparação às crianças de menor faixa etária, marcado por efeito significativo para grupo etário, refletido pela maior porcentagem de acertos, maior ganho visual e maior ocorrência de Efeito McGurk (ilusão audiovisual). **Método:** Este estudo foi aprovado pelo CEP sob o nº 5.391.404 Participaram do estudo 75 sujeitos, falantes do Português Brasileiro (PB), com desenvolvimento típico de fala, divididos em cinco grupos etários: (G1) 4 a 5 anos; (G2) 6 a 8 anos; (G3) 9 a 10 anos; (G4) 11 a 12 anos; e (G5) 20 a 30 anos. Foi aplicado um teste de percepção audiovisual, que consistiu na apresentação dos estímulos silábicos /fa/, /sa/, /ja/, apresentados em quatro condições: auditiva (A), visual (V) e audiovisual congruente/incongruente (AV+) e (AV-). Foram analisadas as porcentagens de respostas corretas nas condições A e V, o ganho visual (AV+ vs A) e o efeito McGurk (tipos de respostas: auditiva, visual ou ilusão). **Resultados:** Considerando a análise comparativa entre as condições (A e V), a ANOVA de medidas repetidas mostrou efeito significativo para grupo etário ($F(4,70) = 27,066$, $p < ,001$) e

para a condição ($F(1,70) = 11,068$, $p = ,001$), bem como para a interação grupo $\times$ condição ($F(4,70) = 5,1495$, $p < 0,01$). A interação significativa foi investigada por meio de testes *Post hoc* de *Tukey*, que revelaram que a porcentagem de acertos na condição (A) foi significativamente maior do que na condição (V), mas apenas nas idades de 4–5 e 6–8 anos. Para as idades mais avançadas (9–10, 11–12 e 20–30 anos), não houve diferenças significativas entre as condições (A e V). Para análise do ganho visual, o teste *Kruskal–Wallis* mostrou uma diferença estatisticamente significativa entre os grupos etários ($H(4, N = 75) = 11,74$; $p = ,002$). O teste *Post hoc* de *Dunn* mostrou um ganho visual maior nos dois grupos mais jovens (4–6 e 6–8 anos). O ganho visual atingiu o pico aos 6–8 anos e, em seguida, reduziu-se significativamente aos 9–10 e 11–12 anos, antes de chegar próximo de zero no grupo adulto de 20–30 anos. Por fim, para análise do efeito McGurk, a ANOVA de medidas repetidas mostrou diferença significativa para relato perceptual ($F(2,140) = 392,53$, $p < 0,001$) e para interação entre grupo etário e relato perceptual ($F(8,140) = 3,5242$, $p < 0,001$). Essa interação significativa foi investigada por meio de análises *Post hoc* de *Tukey*, que mostrou um aumento no relato perceptual auditivo e uma diminuição no relato perceptual visual e de ilusão entre 4–5 e 6–8 anos ($p < 0,001$); e esses níveis altos de relato perceptual auditivo e baixos de relato perceptual visual e de ilusão foram mantidos ao longo das idades. **Conclusão:** A integração audiovisual na percepção de fricativas em falantes do português brasileiro (PB) já ocorre aos 4 anos de idade. Além disso, o maior ganho visual surge entre 4 e 8 anos, com pico entre 6 e 8 anos e declínio a partir desse ponto, sugerindo que crianças falantes do PB 4 a 8 anos superam possíveis obstáculos à percepção da fala e à aquisição da linguagem (a relativa dificuldade de perceber fricativas e a dificuldade associada em aprender as regras fonema-grafema das fricativas) ao recorrer à informação visual. O resultado é o uso otimizado das pistas visuais salientes e distintivas das fricativas do PB, em conjunto com pistas auditivas, e, uma vez superado esse obstáculo, observa-se uma redução da influência da pista visual ao longo das faixas etárias. **Contribuições:** Acredita-se que os achados deste estudo fornecem uma contribuição interdisciplinar, servindo como base teórica na área linguística e fonoaudiológica. Especificamente, no campo da linguística, os resultados contribuem para a fundamentação teórica que aponta para a natureza audiovisual e desenvolvimental da percepção da fala, especialmente no caso de fricativas que apresentam pistas auditivas semelhantes, mas pistas visuais distintas. Além disso, os achados sugerem que o conhecimento fonológico é

construído de forma heterogênea, abrangendo informações de diversas fontes. No que se refere à fonoaudiologia, os resultados da presente tese apresentam uma importante implicação clínica – que as pistas visuais podem ser utilizadas ou enfatizadas nos planos terapêuticos de reabilitação fonológica em Transtorno de Sons da Fala.

Palavras-Chave: Percepção multimodal, identificação sonora, idade; crianças; adultos.

ABSTRACT

Introduction: The theoretical perspective on speech perception as an audiovisual event, introduced by McGurk and MacDonald (1976), posits that speech perception involves the perceptual integration of auditory (acoustic information) and visual (articulatory gesture) cues. From this perspective, studies have shown that children's speech perception is predominantly based on the auditory component, whereas adults' speech perception relies on audiovisual integration, reflected in an increased perceptual use of visual cues throughout development. Thus, age is considered an influential factor in speech perception (Polka, 1994; Sekiyama et al., 2003; Sekiyama & Burnham, 2004). However, the typical developmental trajectory of audiovisual speech perception has not yet been documented in Brazilian Portuguese, particularly regarding the class of fricatives. **Objective:** This study investigates the audiovisual perception of voiceless fricatives as a function of age in children aged 4 to 12 years and adults aged 20 to 30 years. Assuming that audiovisual integration in speech perception depends on linguistic experience (Hazan & Barret, 2000; Burnham et al., 1991), as well as on the development of psychoacoustic (Schneider et al., 1986) and visual skills (Desjardins, 1997), the following hypotheses were proposed: adults are expected to show greater perceptual accuracy compared to children, and older children are expected to show greater perceptual accuracy than younger ones. This would be reflected in a significant age-group effect, indicated by higher percentages of correct responses, greater visual gain, and higher occurrence of the McGurk Effect (audiovisual illusion). **Method:** This study was approved by the Research Ethics Committee under protocol no. 56136022.6.0000.5406. Participants were 75 native speakers of Brazilian Portuguese (BP) with typical speech development, divided into five age groups: (G1) 4–5 years; (G2) 6–8 years; (G3) 9–10 years; (G4) 11–12 years; and (G5) 20–30 years. An audiovisual speech perception test was administered, consisting of syllables /fa/, /sa/, and /ja/ presented under four conditions: auditory (A), visual (V), and audiovisual congruent/incongruent (AV+) and (AV–). Percentages of correct responses in the A and V conditions, visual gain (AV+ vs. A), and McGurk effect (response types: auditory, visual, or illusion) were analyzed. **Results:** A repeated-measures ANOVA comparing the (A) and (V) conditions revealed significant effects for age group ($F(4,70) = 27.066$, $p < .001$) and condition ($F(1,70) = 11.068$, $p = .001$), as well as a significant group $\times$ condition interaction ($F(4,70) = 5.1495$, $p < .01$). Tukey's

post hoc tests indicated that the percentage of correct responses in the (A) condition was significantly higher than in the (V) condition, but only for the 4–5 and 6–8 age groups. For older groups (9–10, 11–12, and 20–30 years), no significant differences were observed between (A) and (V) conditions. The Kruskal–Wallis’s test for visual gain revealed a statistically significant difference among age groups ($H(4, N = 75) = 11.74$; $p = .002$). Dunn’s post hoc test indicated greater visual gain in the two youngest groups (4–6 and 6–8 years). Visual gain peaked at 6–8 years and then declined significantly at 9–10 and 11–12 years, approaching zero in the adult group (20–30 years). Finally, for the McGurk effect, repeated-measures ANOVA showed a significant difference for perceptual report ($F(2, 140) = 392.53$, $p < .001$) and for the interaction between age group and perceptual report ($F(8, 140) = 3.5242$, $p < .001$). Tukey’s post hoc analysis revealed an increase in auditory perceptual reports and a decrease in visual and illusion reports between 4–5 and 6–8 years ($p < .001$), and these levels were maintained across subsequent ages. **Conclusion:** Audiovisual integration in the perception of speakers of Brazilian Portuguese (BP) fricatives is already present by age 4. Moreover, the greatest visual gain occurs between 4 and 8 years of age, peaking between 6 and 8 years and declining thereafter. This suggests that children aged 4 to 8 who are speakers of Brazilian Portuguese may overcome potential challenges in speech perception and language acquisition (such as the relative difficulty of perceiving fricatives and learning fricative phoneme–grapheme correspondences) by relying on visual information. As a result, they optimize the use of salient and distinctive visual cues of BP fricatives alongside auditory cues, and once these challenges are overcome, a reduction in the influence of visual cues is observed across age groups. **Contributions:** The findings of this study offer an interdisciplinary contribution, serving as a theoretical basis for both linguistic and speech-language pathology fields. Specifically, in linguistics, the results support the theoretical framework highlighting the audiovisual and developmental nature of speech perception, especially for fricatives that exhibit similar auditory cues but distinct visual ones. Additionally, the findings suggest that phonological knowledge is constructed heterogeneously, integrating information from multiple sources. In speech-language pathology, the results have important clinical implications: visual cues can be used or emphasized in therapeutic plans for phonological rehabilitation in Speech Sound Disorders.

Keywords: multimodal perception, sound identification, age; children; adults.

LISTA DE QUADROS

Quadro 1 - Levantamento bibliográfico de estudos sobre integração audiovisual na percepção de fala, considerando a faixa etária como fator influente.....	54
Quadro 2 - Caracterização dos sujeitos participantes do estudo.	58
Quadro 3 - Condições de apresentação x estímulos presentes no teste de percepção audiovisual	60

LISTA DE TABELAS

Tabela 1 - Porcentagem de acerto relativo às condições auditiva (A) e visual (V) em comparação aos cinco grupos avaliados.....	64
Tabela 2 - Porcentagem de ganho visual em comparação aos cinco grupos avaliados	66
Tabela 3 - Estímulo audiovisual incongruente: preferência perceptiva entre os grupos avaliados	67

LISTA DE FIGURA E GRÁFICOS

Figura 1 – Teste de percepção audiovisual.....	62
Gráfico 1 - Comparação com base na interação entre grupos e condições de apresentação (auditiva (A) e visual (V))	65
Gráfico 2 - Comparação do ganho visual entre idades	66
Gráfico 3 - Estímulo audiovisual incongruente: comparação dos tipos de resposta entre os grupos.....	67

LISTA DE ABREVIATURAS E SIGLAS

A	Auditivo
AV+	Audiovisual Congruente
AV-	Audiovisual Incongruente
CEES	Centro de Estudos da Educação e Saúde
DP	Desvio Padrão
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FFC	Faculdade de Filosofia e Ciências
G	Grupo
H	Hipótese
PERCEFAL	Instrumento de avaliação da identificação de contrastes fonológicos
PB	Português Brasileiro
UNESP	Universidade Estadual Paulista
V	Visual

SUMÁRIO

APRESENTAÇÃO	18
1 TEORIAS AUDITIVAS E MOTORAS DE PERCEPÇÃO DE FALA	23
1.1 TEORIA AUDITIVA DE PERCEPÇÃO DE FALA	23
1.2 TEORIAS MOTORAS DE PERCEPÇÃO DE FALA	29
2 TEORIA AUDIOVISUAL DA PERCEPÇÃO DE FALA.....	344
2.1 PERCEPÇÃO AUDIOVISUAL DE FALA: INTRODUÇÃO AO TEMA.....	344
2.2 LÍNGUA COMO FATOR INFLUENTE NA PERCEPÇÃO AUDIOVISUAL	38
2.3 IDADE COMO FATOR INFLUENTE NA PERCEPÇÃO AUDIOVISUAL	43
3 OBJETIVOS E HIPÓTESES	56
4 METODOLOGIA	57
4.1 ASPECTOS ÉTICOS	57
4.2 CONSTITUIÇÃO DA AMOSTRA	57
4.3 TESTE DE PERCEPÇÃO DE FALA – ESTÍMULOS	59
4.4 TESTE DE PERCEPÇÃO DE FALA – PROCEDIMENTO EXPERIMENTAL....	60
4.5 FORMA DE ANÁLISE	62
5 RESULTADOS	64
5.1 DESEMPENHO UNISSENSORIAL (A E V):	64
5.2 DESEMPENHO AUDIOVISUAL	65
6 DISCUSSÃO.....	69
6.1 DESEMPENHO UNISSENSORIAL (A E V)	69
6.2 DESEMPENHO AUDIOVISUAL.....	71
7 CONSIDERAÇÕES FINAIS.....	76
REFERÊNCIAS	77
ANEXO A – PARECER CONSUBSTANCIADO DO CEP.....	85
ANEXO B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (RESPONSÁVEIS PELAS CRIANÇAS DE 4 A 12 ANOS)	88
ANEXO C - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (PARA PARTICIPANTES ADULTOS DE 20 A 30 ANOS).....	92
ANEXO D - TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO (PARA CRIANÇAS DE 4 A 6 ANOS DE IDADE)	96

ANEXO E - TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO (PARA CRIANÇAS DE 8 A 10 ANOS DE IDADE).....	98
ANEXO F - TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO (PARA CRIANÇAS DE 11 A 12 ANOS DE IDADE).....	100

APRESENTAÇÃO

Este trabalho, apresentado como requisito para a defesa de tese do Programa de Pós-Graduação em Estudos Linguísticos, está relacionado, por um lado, à formação acadêmica que venho desenvolvendo desde a minha Iniciação Científica (IC) e, por outro, aos dados empíricos observados na prática clínica. As pesquisas que desenvolvi até o momento ocorreram entre junho/2017 e maio/2018 no âmbito Bolsas no Brasil – Iniciação Científica, com o apoio da FAPESP (Processo nº 17/04968-0) e outubro/2019 e abril/2021 no âmbito Bolsas no Brasil – Mestrado, também com o apoio da FAPESP (Processo nº 19/12749-2).

O projeto de Iniciação Científica (IC) – “Desempenho perceptivo-auditivo em crianças com e sem transtorno fonológico na classe das oclusivas” – deu início a minha investigação voltada à percepção de fala em nível segmental, em crianças com transtorno fonológico (doravante TF), mais especificamente, na investigação da relação entre percepção de fala e TF. Naquela ocasião, (a) comparamos o desempenho perceptivo-auditivo em crianças com e sem TF em tarefa de identificação sonora, envolvendo os fonemas oclusivos do Português Brasileiro, e identificamos a similaridade (ou não) do padrão de erro apresentado pelas crianças com e sem TF na tarefa de identificação das oclusivas. Os resultados mostraram que crianças com TF apresentavam pior acurácia perceptual em relação às crianças sem alteração e, ainda, maior tempo de resposta para os acertos. Além disso, os erros envolvendo o ponto de articulação entre as oclusivas foram os mais frequentes em ambos os grupos. O fato de as crianças com TF apresentarem pior desempenho perceptual em relação ao grupo controle foi mais um indício de que crianças com TF também poderiam apresentar dificuldades no domínio perceptivo-auditivo (Cabbage, 2016). Esse estudo foi publicado na revista CoDAS (Qualis A3 na área da Linguística), <https://doi.org/10.1590/2317-1782/20202019248>.

Em paralelo ao desenvolvimento da IC, iniciei o estágio de Fonética e Fonologia Clínica. Durante os atendimentos e supervisões, observei o quanto as estratégias de fornecimento da pista visual — relacionadas ao modo e ponto de articulação dos fonemas — beneficiavam as crianças, especialmente na identificação fonêmica e no aprimoramento da produção. Os resultados obtidos na pesquisa, associados aos dados empíricos da prática clínica, desencadearam um desejo de aprofundar a investigação sobre as dificuldades no domínio perceptivo de crianças com Transtorno

Fonológico (TF), a fim de analisar se a natureza dessas dificuldades estaria restrita ao desempenho perceptivo-auditivo ou se envolveria também uma dificuldade no processamento da pista visual (parâmetros articulatórios). Esse percurso me levou ao meu interesse atual de pesquisa – Percepção de fala a partir da perspectiva teórica audiovisual, com desenvolvimento iniciado no mestrado, a partir da execução do projeto intitulado “Percepção multimodal das fricativas desvozeadas em crianças com e sem transtorno fonológico”. Cabe destacar que o direcionamento do estudo do mestrado para a análise das fricativas decorreu da frequente ocorrência de processos fonológicos envolvendo essa classe, observada durante a prática clínica (Patah, 2008; Namasivayam e Hansen-Brown, 2020). O objetivo desse estudo, portanto, foi comparar o desempenho de crianças com e sem TF na tarefa de percepção audiovisual envolvendo pistas auditivas e visuais em fricativas desvozeadas.

Os resultados indicaram que crianças com TF apresentaram pior desempenho na tarefa de integração de pistas auditivas e visuais comparativamente às crianças com desenvolvimento fonológico típico, observado tanto pela menor acurácia quanto pelo maior tempo de reação. Adicionalmente, tanto crianças com TF quanto crianças com desenvolvimento fonológico típico tiveram maior dificuldade na identificação do estímulo visual. Na análise dos estímulos incongruentes, ambos os grupos privilegiaram a informação auditiva. Esse estudo foi publicado na revista *Clinical Linguistics & Phonetics*, (Qualis A2 na área da Linguística), <https://doi.org/10.1080/02699206.2024.2328792>.

Mais uma vez, os resultados levaram a novos questionamentos, dos quais destaco os resultados referentes ao desempenho perceptual das crianças com desenvolvimento fonológico típico. Na análise da condição audiovisual incongruente, verificamos maior porcentagem de preferência perceptual para pista auditiva do estímulo. No tocante ao tempo de reação, o grupo apresentou maior tempo de resposta para condição visual. Assim, pergunta-se: a pouca influência da pista visual e a preferência perceptual pelo estímulo auditivo seriam indícios de um desempenho perceptual padrão do PB? Ou poderia ser mais um dado de que a percepção audiovisual modifica-se ao longo do desenvolvimento etário?

Embora verifiquemos, em âmbito internacional, a presença de estudos sobre o desenvolvimento da integração de pistas audiovisuais em diferentes faixas etárias, não há, até onde o levantamento bibliográfico possibilitou chegar, estudos que discorram a trajetória típica de desenvolvimento da percepção audiovisual para a

classe das fricativas. A classe fonológica das fricativas apresenta algumas peculiaridades que a torna interessante para análise sob a perspectiva audiovisual.

Os fonemas fricativos do Português Brasileiro (PB) apresentam características acústicas e/ou articulatórias comuns que os identificam como pertencentes à mesma classe. A classe das fricativas é composta pelos seguintes fonemas: /f/, /v/, /s/, /z/, /ʃ/ e /ʒ/. Esta classe tem como característica o ruído de turbulência, desencadeado a partir de uma constrição parcial entre os articuladores (durante a sua produção), permitindo que o fluxo de ar passe continuamente pela cavidade oral. À medida que o ar passa por esse canal estreito, ocorre uma turbulência significativa, produzindo o ruído característico associado ao fluxo turbulento. (Cagliari, 1981). Esse ruído abrange uma faixa de frequência de 1.200 Hz a 8.000 Hz, tornando as fricativas os sons de maior frequência e menor intensidade do português brasileiro (Russo; Behlau, 1994).

Ao analisar o impacto do sinal acústico na percepção auditiva das fricativas, observa-se que sua faixa de frequência (1.200 Hz a 8.000 Hz) não é ideal para percepção, uma vez que frequências acima de 5.000 Hz recebem pouca ou nenhuma amplificação pelo sistema auditivo (Johnson, 1997). Além disso, o ruído turbulento característico dessa classe gera efeitos de mascaramento que dificultam ainda mais a percepção. Diante desses desafios, pistas puramente acústicas podem não fornecer informações inequívocas para a identificação de fricativas individuais ou a discriminação de fonemas dentro dessa classe, sugerindo que pistas adicionais, não acústicas, podem ser particularmente importantes para a percepção das fricativas.

Uma pista não acústica poderosa é a informação visual da fala transmitida pelos lábios, boca e regiões faciais circundantes. No estudo pioneiro, hoje clássico, da percepção audiovisual da fala, McGurk e MacDonald (1976) demonstraram que não apenas a pista visual auxilia a percepção da fala, mas que a percepção envolve o processamento integrativo de pistas auditivas e visuais, que juntas permitem a recuperação simbólica da informação recebida.

Considerando a lacuna de estudos na temática da percepção de fala audiovisual para classe das fricativas e seu desenvolvimento ao longo das faixas etárias, desenvolvemos a presente tese intitulada “Integração de pistas audiovisuais na percepção de fala de fricativas surdas: análise desenvolvimental”.

O estudo objetiva investigar a percepção audiovisual da fala de fricativas surdas em função da idade, em crianças na faixa etária de 4 a 12 anos e em adultos na faixa

etária de 20 a 30 anos. A tese a ser defendida, portanto, é a de que a percepção é de natureza audiovisual, tendo a língua e a idade como fatores influentes.

Para responder ao objetivo de modo satisfatório, a presente tese foi estruturada em sete capítulos. O primeiro capítulo – Teorias auditivas e motoras de percepção da fala – apresenta um panorama geral das perspectivas teóricas sobre a percepção de fala. Para melhor apresentação das informações literárias, esse capítulo é didaticamente dividido em: (1.1) Teoria auditiva de percepção de fala; (1.2) Teorias motoras de percepção de fala.

O segundo capítulo – Teoria audiovisual da percepção de fala – apresenta o referencial teórico adotado para a execução do presente estudo. Inicialmente, apresentamos os principais estudos que contribuíram para a consolidação da perspectiva teórica da percepção de fala como um evento audiovisual. Em sequência, apresentamos o estado da arte e revisão teórica da literatura que aponta a língua e a idade como fatores influentes na percepção de fala a partir da perspectiva audiovisual. Esse capítulo foi estruturado da seguinte maneira: (2) Teoria audiovisual de percepção de fala; (2.1) Percepção audiovisual de fala: Introdução ao tema; (2.2) Língua como fator influente na percepção audiovisual; (2.3) Idade como fator influente na percepção audiovisual.

O terceiro capítulo – Objetivos e Hipóteses – informa ao leitor os objetivos e hipóteses do trabalho que nortearam a execução do estudo.

O quarto capítulo – Metodologia – especifica os métodos adotados para execução do estudo. Esse capítulo foi estruturado da seguinte maneira: (4.1) Aspectos éticos; (4.2) Constituição da amostra; (4.3) Teste de percepção de fala - Estímulos (4.4) Teste de percepção de fala – Procedimento experimental; (4.5) Forma de análise.

O quinto capítulo – Resultados – apresenta os resultados referentes aos achados quanto às análises descritivas e inferenciais correspondentes aos dados do teste de percepção aplicado. Os dados foram apresentados em valores absolutos e percentuais, em gráficos e tabelas. Com base nos resultados de cada objetivo específico proposto, este capítulo foi estruturado da seguinte maneira: (5.1) Desempenho unissensorial (A e V); e (5.2) Desempenho audiovisual.

O sexto capítulo – Discussão – apresenta discussões para o resultado de cada objetivo específico através de um diálogo com a literatura. Assim, o capítulo foi estruturado em três subcapítulos: (6.1) Desempenho unissensorial (A e V); (6.2) Desempenho audiovisual; (6.3) Análise do Efeito McGurk.

Por fim, o último capítulo da tese – Considerações finais – expõe a conclusão final do estudo e suas contribuições. Além disso, são reportadas as limitações e lacunas observadas ao longo do desenvolvimento deste trabalho, que possibilitarão o desenvolvimento de novos estudos.

Aproveito esta seção de apresentação como um prólogo para elucidar a concepção de desenvolvimento adotada nesta tese. A concepção de desenvolvimento aqui adotada tem enfoque no percurso, não descartando os fenômenos internos (biológico) e externos (experiências sociais) que podem provocar deslocamentos. Assim, compreendo o desenvolvimento da percepção de fala como um processo dinâmico e gradual de refinamento das habilidades perceptuais envolvidas na identificação e interpretação dos sons da fala ao longo do ciclo etário, que pode ser desencadeado também pelas próprias mudanças na condição do perceber. Assim, o desenvolvimento da percepção de fala corresponde a uma reorganização funcional dos mecanismos perceptuais, influenciada por fatores linguísticos, cognitivos, sensoriais e sociais (Liberman; Mattingly, 1985; Massaro, 1998), os quais se modificam ao longo das diferentes faixas etárias.

1 TEORIAS AUDITIVAS E MOTORAS DE PERCEPÇÃO DE FALA

1.1 TEORIA AUDITIVA DE PERCEPÇÃO DE FALA

A percepção de fala refere-se ao processo pelo qual os ouvintes captam e interpretam os sons produzidos na fala, reconhecendo-os como unidades linguísticas e promovendo a compreensão do conteúdo comunicativo.

Uma das perguntas primordiais que permearam o desenvolvimento de estudos sobre a percepção de fala refere-se à forma pela qual os ouvintes extraem a informação necessária para a compreensão de fala. Este questionamento impulsionou o desenvolvimento de estudos a partir da perspectiva perceptual de fala, já no século XIX (Borden *et al.*, 1994; Berti, 2008).

Tais pesquisas intensificaram-se somente no século XX a partir dos avanços tecnológicos, como, por exemplo, o desenvolvimento do espectrógrafo, inaugurado na década de 1940. Este equipamento analógico mensura com ferramentas gráficas (gráfico de três dimensões: amplitude, frequência e tempo) os parâmetros acústicos dos sons da fala (Berti, 2008; Cantoni; Oliveira; Nevado, 2022).

Com o advento do espectrógrafo, alguns autores (Fant, 1962; Lindblom, 1963; Ohman, 1964) analisaram os parâmetros de produção de fala a partir de dados espectrais (frequência, intensidade e duração) e outros dados acústicos, levantando a hipótese de que o cérebro utilizaria os mesmos parâmetros para a percepção de fala (Fant, 1967).

Os primeiros estudos que investigaram a percepção de fala consideraram apenas a percepção de parâmetros acústicos e, especificamente, os autores atribuíram que a informação do sinal acústico seria essencial e suficiente para a percepção de fala (Morton; Broadbent, 1967; Stevens; Blumstein, 1978).

O conjunto de estudos que considera o processo de percepção de fala como sendo primordialmente auditivo se concentra na perspectiva teórica denominada *Teoria Auditiva de Percepção de Fala* (Berti, 2008).

Um dos percursos dessa posição teórica foi Fant (1967). Este autor considera que a audição exerce papel primordial para a percepção de fala, de modo que, essencialmente, as habilidades auditivas (por exemplo, detecção, discriminação

e identificação de contrastes acústicos) seriam suficientes para o desenvolvimento da categorização de fonemas, sílabas e palavras que constituem a língua, isto é, os ouvintes não precisariam se referir à produção (processamento do gesto articulatório) para perceberem a fala. Assim, a teoria auditiva da percepção de fala caracteriza-se pela informação acústica como fonte única e determinante na percepção de fala, destacando que a informação acústica se refere à onda sonora (onda mecânica) que transporta os parâmetros sonoros da fala, e estes são captados pelos ouvidos e transformados pela cóclea em impulsos elétricos, interpretados pelo cérebro.

Posteriormente, novos estudos foram desenvolvidos com base na *Teoria Auditiva de Percepção de fala*, podendo ser divididos em três perspectivas: (a) dois estágios nos processos perceptuais; (b) identificação por atributos secundários; (c) percepção dos sons por outros animais (Nishida, 2012).

Os dois estágios nos processos perceptuais objetivam estabelecer, experimentalmente, a relação entre percepção de fala e memória (Pisoni, 1973; Pisoni; Tash, 1974; Carney *et al.*, 1977; Oden e Massaro, 1978). Segundo Fujisaki e Kawashima (1970), haveria uma diferença no processamento perceptual de fala, a depender do grupo de sons investigado. Diante das diferenças na maneira como são percebidos os sons da fala, propõe-se que haveria uma Percepção Contínua (ou auditiva) e uma Percepção Categórica (ou Fonética). Por exemplo, as pistas acústicas das vogais são marcadas por pistas acústicas salientes caracterizadas pela maior duração e pela alta intensidade; esses dados acústicos permitiriam a comparação entre as propriedades acústicas recebidas com as propriedades acústicas armazenadas na memória de curto prazo. No entanto, as pistas acústicas das consoantes são caracterizadas pela menor duração e menor intensidade quando comparadas às vogais. Os dados acústicos das consoantes implicariam na necessidade de um codificador, e a percepção das consoantes estaria no estágio de percepção categórica, isto é, os sujeitos utilizariam um código/engrama fonético ao invés dos níveis mais baixos da audição (Fujisaki; Kawashima, 1970; Pisoni, 1973).

A identificação por atributos secundários propõe que a fala seria percebida com o auxílio de atributos secundários. Os autores Cole e Scot (1974) e Stevens e Blumstein (1978 *apud* Silva, 2024, p. 28) sugerem a existência de três tipos de pistas envolvidas na percepção de fala: pistas invariantes; pistas condicionadas pelo contexto e pistas promovidas pelo envelope da forma de onda. As pistas invariantes

referem-se às pistas acústicas de um fonema isolado, ou seja, as pistas invariantes seriam as responsáveis por fazer um ouvinte perceber os sons individualmente. Por sua vez, as pistas condicionadas pelo contexto possibilitariam a percepção dos sons em um ordenamento temporal da fala. Por fim, o envelope da forma da onda possibilitaria a percepção dos aspectos prosódicos da fala (Cole; Scot, 1974, *apud* Silva, 2024, p. 28). Deste modo, para cada subtipo de pistas existiriam pistas acústicas específicas (características espectrais, intensidade, frequência) que permitiriam o resgate representacional da informação acústica recebida.

A percepção dos sons por outros animais é uma perspectiva encontrada nos trabalhos de Kuhl e Miller (1975) e Lotto, Kluender e Holt (1997). Os estudos foram desenvolvidos a partir da seguinte pergunta: a percepção de fala é uma habilidade específica da espécie humana? Referindo-se a esta questão, Kuhl e Miller (1975) investigaram se as chinchilas seriam capazes de perceber a diferença entre /t/ e /d/ em padrão silábico simples (CV). Os estímulos foram apresentados em diferentes condições: sílabas /t/ e /d/ sem alteração de padrões acústicos; /t/ e /d/ produzidos por diferentes falantes; e /t/ e /d/ em contextos vocálicos e com alteração do VOT (*voice-onset-time*). O experimento foi realizado com chinchilas treinadas, chinchilas não treinadas e com humanos. O desempenho apresentado por todos os animais e pelos humanos foram similares. Esses resultados apontam para o fato de que tais distinções realizadas pela percepção da fala não são “privilegio” de humanos (Nishida, 2012 *apud* Silva, 2024, p. 29).

Com base nos pressupostos teóricos percorridos anteriormente, diversos estudos objetivaram investigar o desenvolvimento perceptual de fala considerando a aquisição de parâmetros acústicos (Elimas *et al.*, 1971; Fernald; Simon, 1984; Huang *et al.*, 1997; Lecanuet, 1998).

Lecanuet (1998) descreve que o sistema auditivo humano é funcional entre a 22ª e 24ª semana gestacional e evolui durante o último trimestre de gestação, no qual o feto responde comportamental, elétrica e neuroquimicamente na presença de som. Em pesquisas análogas realizadas em ovelhas, pesquisadores observaram que, dentro do útero, ouve-se apenas estrutura prosódica (rítmica e entoacional), mas não os segmentos (estruturas fonéticas da língua) (Huang *et al.*, 1997). Mariani *et al.* (2023) mostraram em seu estudo que a atividade eletrofisiológica de recém-nascidos apresenta aumento nas correlações temporais de longo alcance após a estimulação

com fala, particularmente na língua ouvida durante o período pré-natal, indicando o surgimento precoce da especialização cerebral para a língua nativa.

Após o nascimento, o neonato humano parece estar biologicamente pré-sintonizado e predisposto a processar os sons da fala estudos indicam que o neonato já possui, habilidades auditivas essenciais semelhantes às de adultos, incluindo aprendizagem estatística e extração de regras a partir de um input de fala variável. Assim, o cérebro neonatal está preparado para categorizar sons, detectar fronteiras de palavras, aprender novos vocábulos e separar fluxos de fala — em resumo, para adquirir linguagem de forma rápida e eficiente a partir do input linguístico cotidiano (Kujala et al., 2023). Além disso, o neonato parece preferir o som da fala a outros sons, mas, principalmente, a fala materna, caracterizada pelo seu tempo lento, tom alto e entoação hiperbólica (Fernald; Simon, 1984; Grieser; Kuhl, 1988).

Já com um mês de vida, os bebês conseguem discriminar muitos dos contrastes fonéticos utilizados nas línguas do mundo e, especialmente nesta faixa etária, conseguem discriminar os contrastes acústicos fundamentais para o reconhecimento segmental.

Aos quatro meses de vida, os bebês são capazes de: detectar a mesma vogal com diferentes propriedades acústicas; reconhecer o mesmo elemento em diferentes ritmos de fala; detectar variações nos padrões entoacionais; reconhecer a mesma sílaba em diferentes enunciados (Jusczyk; Pisoni; Mullenix, 1992); e demonstrar sinais de categorização dos sons da fala, já que detectam as mesmas estruturas vocálicas e consonantais em diferentes contextos.

Com sete meses de vida, os bebês identificam padrões acentuais em sua língua, além de utilizarem traços prosódicos para distinguir palavras estrangeiras de palavras pertencentes à língua materna. No entanto, dos sete aos dez meses de idade, há redução na habilidade de detectar contrastes estrangeiros, e os bebês passam a se especializar perceptualmente nos contrastes fonéticos entre vogais e consoantes da sua língua (Werker; Tess, 1984; Jusczyk; Houston; Newsome, 1999). Portanto, essa faixa etária é o marco para a passagem de um ouvinte universal para um ouvinte linguístico, quando o bebê perde a capacidade perceptiva universal para se especializar em sua língua materna.

Em resumo, pode-se dizer que, do ponto de vista da percepção de fala, uma criança não aprende uma língua, ela esquece todas as outras (Penido; Rothe-Neves 2013, p. 120).

Para as faixas etárias entre um e quatro anos de idade, há poucos estudos desenvolvidos objetivando a investigação da percepção de fala. Considerando o Português Brasileiro (PB), alguns estudos se destacam, como os de Berti e Roque (2013) e Berti (2017).

Berti e Roque (2013) objetivaram: (1) verificar o desempenho perceptivo-auditivo das crianças em tarefa de identificação do contraste fonológico, envolvendo todas as vogais tônicas do PB; (2) identificar os contrastes vocálicos que apresentam maior ou menor grau de dificuldade de identificação perceptivo-auditiva; e (3) verificar se a idade se correlaciona com a acurácia perceptivo-auditiva. Participaram do estudo 66 crianças entre 5 e 6 anos de idade, e uma tarefa de identificação foi realizada usando o instrumento PerceVog (que avalia a identificação entre vogais tônicas), PerceFAL (Berti, 2011). A tarefa consistiu na apresentação de um estímulo acústico e, posteriormente, na escolha da gravura correspondente à palavra apresentada, dentre duas possibilidades de gravuras. Como critério de análise, considerou-se acurácia perceptivo-auditiva e o tempo de reação. Os resultados demonstraram que não há domínio efetivo na identificação das vogais tônicas pelas crianças, na faixa etária entre cinco e seis anos. As autoras também apontam que o tempo de reação durante a execução da tarefa sugere diferentes graus de similaridade entre os estímulos. Ou seja, quanto maior a diferença acústica entre dois pares de estímulos, mais rápida é a resposta dos sujeitos; e, ao contrário, quanto menor a diferença acústica entre os estímulos, maior é o tempo utilizado pelas crianças para a tomada de decisão.

Além da constatação da presença de diferentes graus de similaridade perceptivo-auditivo entre as vogais, as autoras indicam que vogais anteriores apresentaram maior similaridade do que as posteriores. As vogais periféricas (que ocupam os extremos do triângulo vocálico) parecem servir como pontos de ancoragem na tarefa de percepção, na medida em que a tendência que prevaleceu nos erros foi a das vogais menos para as mais periféricas. Adicionalmente, a idade das crianças correlacionou-se positivamente com a acurácia perceptivo-auditiva, sugerindo que, à medida que a idade avança, há melhora na habilidade de identificar os contrastes vocálicos.

Berti (2017) verificou a aquisição perceptual do sistema consonantal do PB do ponto de vista perceptivo-auditivo, em crianças entre quatro e cinco anos de idade. A autora investigou o desempenho perceptivo-auditivo de crianças em tarefa de

identificação de contrastes fônicos do PB. Especificamente, buscou verificar se uma determinada classe fônica apresentava maior ou menor dificuldade de identificação. Foram investigadas 66 crianças (de ambos os sexos), entre 4 e 5 anos de idade. Quatro tarefas de identificação foram testadas usando o instrumento PerceFAL (Berti, 2011). O PerceFAL é composto por um subconjunto de quatro experimentos: (a) PerceVog, que avalia a identificação entre vogais tônicas; (b) PerceOcl, que avalia a identificação entre oclusivas; (c) PerceFric, que avalia a identificação entre fricativas; (d) PerceSon, que avalia a identificação entre sonorantes. Neste instrumento, estão envolvidas, preferencialmente, palavras dissilábicas paroxítonas, familiares às crianças, contendo os 19 fonemas consonantais do PB na posição acentuada. A tarefa consistiu na apresentação do estímulo acústico e, posteriormente, na escolha da gravura correspondente à palavra apresentada, dentre duas possibilidades de gravuras. Como critério de análise, consideraram-se três pontos: 1) a acurácia perceptivo-auditiva; 2) o tempo de reação dos erros e acertos; e 3) o reconhecimento de padrões com o uso do escalonamento multidimensional e matriz de similaridade. Os resultados demonstraram que o desempenho perceptivo-auditivo de crianças ocorreu gradualmente, e é dependente da classe fônica. A maior acurácia na identificação dos contrastes fônicos parece obedecer a seguinte ordem: vogais, sonorantes, oclusivas e fricativas. O tempo de reação das respostas corretas foi menor do que o das respostas incorretas (exceto para a classe das vogais). No interior da classe das vogais, o parâmetro ântero-posterior parece exercer um importante papel na saliência perceptual. Para a classe das obstruintes e sonorantes (nasais e líquidas), as pistas acústicas que marcam o vozeamento (no caso das obstruintes) e o modo de articulação (no caso das sonorantes) são mais robustas perceptualmente do que as pistas de ponto de articulação.

Considerando os estudos citados (Berti; Roque, 2013; Berti, 2017), é possível inferir, a partir do ponto de vista teórico – *Teoria Auditiva da Percepção de Fala* –, que, para as faixas etárias entre quatro e cinco anos e entre cinco e seis anos de idade, a aquisição perceptual de fala ainda estaria em desenvolvimento para o PB, ou seja, para esta faixa etária, as crianças não adquiriram completamente as características fonético-acústicas dos segmentos de sua língua nativa.

Em síntese, os estudos desenvolvidos dentro da perspectiva teórica – *Teoria Auditiva de Percepção de fala* – forneceram informações relevantes quanto ao funcionamento e aquisição perceptual de fala, do ponto de vista acústico. Destaca-se

que, nesta perspectiva, a percepção é o equivalente à audição. No entanto, um problema surge por acaso quando Liberman e colegas (Liberman, 1957) tentavam encontrar o que havia de distintivo entre as categorias fonêmicas.

Para isso, eles buscavam (com a ajuda de um aparelho – o *Pattern Playback* – que convertia espectrogramas desenhados à mão em som) as pistas acústicas que diferenciavam /d/, /b/ e /g/; por exemplo, os autores observaram que as transições de formantes de cada fonema variavam radicalmente de acordo com a vogal seguinte. Ou seja, não foi observado um “formante fixo”. Os autores concluíram que não há pistas acústicas invariantes que identifiquem diretamente cada fonema. O mesmo fonema aparece com diferentes características acústicas dependendo do contexto (coarticulação, vizinhança fonêmica).

A falta de invariância do sinal acústico fomentou novas questões acerca da forma como a fala é percebida, indicando a necessidade de considerar outras pistas além das acústicas. Surge, então, uma nova corrente de estudos em busca do invariante perceptual, a qual direciona a investigação para a influência da articulação na percepção da fala (Liberman, 1957; Oden; Massaro, 1978; Liberman; Mattingly, 1985; Fowler, 1996). Dessa forma, emerge a Teoria Motora da Percepção da Fala, que será abordada com mais detalhes na próxima seção.

1.2 TEORIAS MOTORAS DE PERCEPÇÃO DE FALA

A perspectiva teórica de base articulatória para a percepção de fala engloba duas teorias: *Teoria Motora da Percepção de Fala* (Liberman; Mattingly, 1985) e *Teoria do Realismo Direto da Percepção de Fala* (Fowler, 1996). Segundo estas teorias, existem variações do sinal acústico, ou seja, o dado acústico por si só seria insuficiente para caracterizar o processo de percepção de fala como predominantemente auditivo. Deste modo, as teorias citadas tomam a articulação como o referencial de análise.

A *Teoria Motora da Percepção de Fala* (Liberman; Mattingly, 1985) preconiza a presença de dois objetos perceptuais: objeto proximal da percepção de fala e objeto distal da percepção de fala. O objeto proximal seriam as ondas sonoras captadas pelo aparelho auditivo; já o objeto distal refere-se ao gesto fonético pretendido (ou o montante de comandos neurais de um gesto que um movimento articulatório periférico realiza). O gesto fonético pretendido (objeto distal da percepção de fala) permitiria o resgate dos atributos do sinal acústico a partir do plano motor (presença ou ausência

de fechamento das pregas vocais, elevação ou não do véu palatino, oclusão labial, posicionamento de língua).

Para esta teoria, haveria a percepção do sinal acústico e o resgate de seus atributos motores (relação entre sinal acústico e gesto fonético pretendido), permitindo, assim, uma percepção de fala efetiva. Por exemplo, o fone [b] é resultante dos seguintes gestos fonéticos: fechamento de pregas vocais, elevação do véu palatino e oclusão labial; o fone [m], por sua vez, é resultante do fechamento de pregas vocais, abaixamento do véu palatino e oclusão labial. Segundo os autores, para essa teoria, perceber a fala é perceber um padrão específico de gestos fonéticos pretendidos subjacentes ao sinal acústico (Liberman; Mattingly, 1985: p. 238).

Assim, o gesto fonético pretendido seria o estatuto simbólico (categorias de representação) da percepção de fala.

O primeiro ponto estabelecido na *Teoria Motora da Percepção de Fala* destaca que o primitivo da percepção seriam os gestos fonéticos pretendidos (objeto distal da percepção de fala). A noção de gestos fonéticos pretendidos surge em virtude dos efeitos da coarticulação – gestos motores sobrepostos –, e estes seriam “representados no cérebro como comandos motores invariáveis para execução de movimentos dos articuladores em certas configurações linguisticamente significativas” (Liberman; Mattingly, 1985, p. 237). Os autores postularam a existência de um módulo fonético no cérebro, diferente do módulo auditivo, que permitiria a percepção dos gestos fonéticos pretendidos, incorporando a competência motora da percepção de fala.

O segundo ponto estabelecido na *Teoria Motora da Percepção de Fala* coloca a produção e a percepção da fala intimamente ligadas, uma vez que os movimentos articulatorios implicam em mudanças no formato do trato vocal, e, por consequência, gera-se um sinal acústico resultante. Portanto, o movimento articulatorio do trato vocal implica em consequências acústicas, sendo ambas captadas na percepção de fala.

Em resumo, segundo a *Teoria Motora da Percepção de Fala*,

a percepção da fala não é explicada por princípios que se aplicam à percepção de sons em geral, mas deve, ao contrário, ser vista como uma especialização para os gestos fonéticos. A incorporação de uma conexão biologicamente baseada entre a percepção e a produção, esta especialização impede os ouvintes de tomar o sinal como sons ordinários, mas permite que eles usem a relação sistemática, ainda especial, entre sinal e gesto para perceber o gesto. A relação é sistemática porque resulta de uma dependência de fidelidade entre os gestos, movimentos articulatorios, formatos de

cavidade oral e sinal acústico. Ela é especial porque ocorre apenas na fala (Liberman; Mattingly, 1985, p. 240-241).

A *Teoria do Realismo Direto de Percepção de Fala* (Fowler, 1996) tem como primitivo de análise o gesto articulatório e este assume um estatuto fonético e fonológico (engloba características acústica, articulatória e representacional – simbólica).

Deste modo, diferentemente da *Teoria Motora de Percepção de Fala (TM)* (Liberman; Mattingly, 1985),

o gesto articulatório não teria lugar num módulo especializado, capaz de fazer um cálculo contextualizado da inversão acústico-articulatória a fim de identificar os comandos motores subjacentes ao sinal de fala” (Albano, 1990 *apud* Berti, 2008, p. 47- 48).

Para a *Teoria do Realismo Direto da Percepção de Fala* (Fowler, 1996), o ouvinte é ao mesmo tempo falante e ouvinte, assim não há necessidade de traduzir o sinal acústico em gestos articulatórios (no sentido empregado na TM); ao contrário, os ouvintes extraem a informação sobre como a fala é produzida (dados acústicos e motores), resgatando, assim, o gesto articulatório.

Considerando os pressupostos citados, à luz da perspectiva teórica *Realismo Direto da Percepção de Fala* (Fowler, 1996), os ouvintes recuperam atributos articulatórios (movimento de lábio, língua, véu palatino, pregas vocais) a partir da informação acústica, não havendo dicotomia entre ambos, já que a unidade perceptiva que permitiria essa análise perceptual direta seria o gesto articulatório (unidade mediadora entre aspectos físicos e simbólicos).

Em síntese, a *Teoria Motora da Percepção da Fala* (Liberman; Mattingly, 1985) e a *Teoria do Realismo Direto* (Fowler, 1986, 1996) compartilham o pressuposto de que os gestos articulatórios desempenham papel central na percepção da fala, contudo, divergem quanto à natureza do acesso perceptual a esses gestos. Na perspectiva motora, a percepção da fala envolve o mapeamento do sinal acústico em representações motoras internas, de modo que o ouvinte reconhece as intenções articulatórias do falante por meio de um mecanismo especializado que vincula percepção e produção. Já sob a ótica do realismo direto, a percepção é concebida como um processo não mediado, no qual os gestos articulatórios são diretamente especificados no sinal acústico e, portanto, acessíveis ao ouvinte sem a necessidade de inferências ou representações intermediárias. Assim, enquanto a Teoria Motora adota uma abordagem mediacionista, que pressupõe a existência de um sistema

perceptivo específico para a fala, a Teoria do Realismo Direto fundamenta-se em uma concepção ecológica da percepção, segundo a qual os eventos articulatórios são percebidos diretamente a partir das informações disponíveis no ambiente acústico.

Embora as Teorias Motoras da Percepção da Fala (Liberman; Mattingly, 1985; Fowler, 1986, 1996) tenham representado um marco ao propor a interdependência entre os processos de produção e percepção, seu escopo permaneceu restrito ao domínio acústico-articulatório. Essa limitação resultou em uma lacuna teórica no que se refere à integração das múltiplas modalidades sensoriais envolvidas na percepção de fala, em especial a modalidade visual, questão que ganhou destaque a partir do estudo de McGurk e MacDonald (1976). Nesse estudo, os autores, diante das duas principais correntes teóricas vigentes à época — a auditiva e a motora —, buscaram investigar se a natureza da percepção de fala seria predominantemente auditiva ou articulatória.

De modo surpreendente, os autores observaram que, quando os participantes eram expostos a sinais auditivos e visuais incongruentes (por exemplo, ouvindo /ba/ e vendo a articulação de /ga/), a resposta perceptual resultante não correspondia a nenhum dos estímulos isoladamente, mas a uma fusão entre ambos (no exemplo, /da/). Esse fenômeno, intitulado posteriormente como efeito McGurk, revelou que a percepção de fala é fortemente influenciada pela informação auditiva e articulatória que, de modo integrado, proporcionam o resgate simbólico da informação apresentada.

A partir dos achados de McGurk e MacDonald (1976), levantou-se a questão de que a percepção de fala constitui um processo multissensorial, no qual a informação visual (gesto articulatório) exerce influência direta e sistemática sobre o processamento auditivo (parâmetros acústicos). Esse achado provocou uma mudança de paradigma nos estudos da área, deslocando o foco das explicações unissensoriais, centradas apenas em mecanismos auditivos ou motores, para abordagens que reconhecem a fala como um evento audiovisual integrado.

As teorias audiovisuais da percepção de fala, desenvolvidas a partir desse novo enquadramento, propõem que o sistema perceptual humano é sensível às regularidades entre os gestos articulatórios visuais e seus correspondentes acústicos, processando-os de forma conjunta para otimizar a extração de informações linguísticas. Nessa perspectiva, a percepção de fala resulta da integração intermodal entre pistas acústicas (como frequência, intensidade e duração) e pistas visuais (como

movimentos articulatorios), sendo essa combinação fundamental para a robustez e a estabilidade perceptual.

Na seção seguinte, serão abordados os pressupostos desta perspectiva teórica – *Teoria Audiovisual de Percepção de Fala*.

2 TEORIA AUDIOVISUAL DE PERCEPÇÃO DE FALA

2.1 PERCEPÇÃO AUDIOVISUAL DE FALA: INTRODUÇÃO AO TEMA

A perspectiva teórica da percepção de fala como um evento audiovisual foi inaugurada de maneira “acidental” por McGurk & MacDonald (1976). Diante das duas principais correntes teóricas vigentes à época — a auditiva, que defendia que a percepção de fala seria exclusivamente determinada pelo sinal acústico, e a motora, que postulava que a percepção estaria intimamente ligada aos gestos articulatórios do falante —, os autores buscaram investigar se a natureza da percepção de fala seria predominantemente auditiva ou articulatória. Durante seus experimentos, eles observaram uma possível integração entre informações visuais e auditivas que acarretaram em percepções distintas do que era efetivamente apresentado, sugerindo que a percepção de fala não se limitaria ao domínio auditivo ou motor. Esses achados não apenas desafiaram os modelos predominantes na época, mas também abriram caminho para uma compreensão mais ampla do processamento audiovisual na percepção de fala, mostrando que o indivíduo ativa simultaneamente múltiplas fontes de informação sensorial durante a percepção de fala. A seguir, será apresentado de modo detalhado o estudo vanguardista de McGurk & MacDonald (1976) que impulsionou a perspectiva teórica – *Teoria Audiovisual da Percepção de Fala*.

O estudo de McGurk & MacDonald (1976) teve como objetivo investigar a natureza da percepção de fala. Participaram do estudo crianças (3 a 8 anos) e adultos (18 a 40 anos), divididos em 3 grupos etários: (G1) crianças de 3 a 5 anos; (G2) crianças de 7 a 8 anos; e (G3) adultos de 18 a 40 anos. O estudo desenvolveu-se por meio de um teste de percepção de fala, no qual estímulos de natureza auditiva e visual foram obtidos a partir de gravação de áudio e vídeo de sílabas de padrão silábico simples: /ba/; /ga/; /ka/; /pa/. A partir das gravações, foram elaboradas quatro condições de apresentação dos estímulos: (1) estímulos auditivos, caracterizados pela gravação em áudio da produção vocal das sílabas-alvo; (2) estímulos visuais, caracterizados pela gravação em vídeo da produção das sílabas-alvo, sem áudio associado; (3) estímulos audiovisuais congruentes, caracterizados pela apresentação simultânea dos estímulos auditivo e visual (leitura labial) de modo correspondente (Exemplo: áudio [ba] e vídeo correspondente à produção articulatória de [ba]); e (4) estímulo audiovisual incongruente, caracterizado pela combinação simultânea de

informações auditivas e visuais que não correspondem entre si (Exemplo: áudio [ba] e vídeo correspondente à produção de [ga]).

Como resultado, os autores observaram que, para as condições auditiva e audiovisual congruente, os participantes relataram, frequentemente, as sílabas corretamente. No entanto, na condição audiovisual incongruente, para as combinações [áudio /ga/ - vídeo /ba/] e [áudio /pa/ - vídeo /ka/], os participantes relataram um “terceiro estímulo” não correspondente nem ao estímulo auditivo nem ao estímulo visual. Especificamente, para a combinação [áudio /ga/ - vídeo /ba/], o relato perceptual foi /da/; enquanto para a combinação [áudio /pa/ - vídeo /ka/], o relato perceptual foi de /ta/.

Ainda, segundo os mesmos autores, o relato de um “terceiro estímulo” diante da condição experimental audiovisual incongruente seria fruto de uma ilusão perceptual, também denominada Efeito McGurk, decorrente da integração das pistas auditivas e visuais, durante a percepção de fala.

A partir dos achados inovadores apresentados por McGurk & MacDonald (1976), Oden e Massaro (1978) propuseram o *Modelo Fuzzy da Percepção de Fala* e o *Modelo Lógico da Percepção Difusa (FLMP)*, com o objetivo de investigar como diferentes fontes de informação (auditiva e visual) seriam avaliadas e integradas na percepção de fala.

O modelo Fuzzy (Oden; Massaro, 1978) postula que a percepção de fala ocorre por meio de três operações principais: avaliação, integração e decisão. Nesse modelo, o sistema sensorial transforma um evento físico — dados acústicos e articulatórios — em traços perceptuais, isto é, propriedades específicas de cada segmento. Esses traços são, então, integrados e comparados com engramas armazenados na memória, representações cognitivas de sons ou palavras. Dessa forma, sob a ótica deste modelo, a categorização e o resgate perceptual envolvem a combinação de todos os traços disponíveis (Ferreira-Silva; Pacheco; Cagliari, 2017).

O *Modelo Lógico da Percepção Difusa (FLMP)*, proposto por Oden e Massaro (1978), é uma extensão do modelo Fuzzy que descreve de forma mais detalhada como múltiplas fontes de informação são combinadas durante a percepção. Nesse modelo, a percepção resulta da integração probabilística e graduada de pistas sensoriais e cognitivas, cada uma avaliada em termos de sua compatibilidade com categorias ou hipóteses perceptuais. O FLMP postula que o sistema perceptual não toma decisões binárias, mas realiza uma combinação contínua e ponderada dos dados disponíveis,

calculando o grau de apoio que cada pista oferece a uma possível interpretação. Assim, a escolha perceptual emerge do processo de normalização de informações difusas, produzindo uma resposta coerente com o padrão de ativação mais forte. Essa abordagem permite explicar fenômenos como a integração audiovisual na percepção de fala, mostrando como o sistema perceptual lida de maneira flexível com a ambiguidade presente nos estímulos linguísticos.

A *Teoria Audiovisual da Percepção de Fala* considera que a percepção de fala ocorre através da integração entre as informações advindas das vias sensoriais, auditiva (percepção do sinal acústico) e visual (percepção dos movimentos articulatórios) que, de modo integrado, proporcionam o resgate simbólico da informação recebida.

Atualmente, os estudos objetivando investigar as bases neurobiológicas da integração audiovisual da fala utilizam de medidas neurofisiológicas. Neste segmento, alguns estudos (Ojanen, 2005; Sekiyama et al., 2003) indicam quais regiões do cérebro estão envolvidas na percepção audiovisual da fala. Ojanen (2005), por exemplo, utilizou a ressonância magnética funcional, magnetoencefalografia e métodos comportamentais para estudar os mecanismos neurocognitivos envolvidos na percepção de fala. Os autores verificaram mediante apresentação de estímulos audiovisuais, ativação do lobo temporal (área cerebral responsável pela percepção auditiva) e ativação de áreas corticais motoras envolvidas na produção da fala. Algumas dessas regiões, segundo os autores, fazem parte do *sistema de neurônios espelho*.

Os *neurônios-espelho* constituem um sistema cortical ativado a partir da realização e observação de uma ação. Este sistema foi descrito inicialmente por Rizzolatti e colaboradores (1998), em macacos de raça Rhesus. Estes pesquisadores demonstraram que alguns neurônios da área F5, localizada no lobo frontal, eram ativados quando o animal realizava um movimento com uma finalidade específica (como pegar uvas passas) ou quando o animal observava outro indivíduo (macaco ou humano) realizando a mesma tarefa. Um padrão cerebral semelhante foi observado em humanos. Alguns estudos (Buccino, Binkofski e Riggio, 2004) mostraram que durante a percepção de fala são ativados os neurônios da área de Broca localizados no hemisfério esquerdo, região cerebral responsável pelo planejamento articulatório. No entanto, os neurônios-espelho são ativados somente quando a ação observada

faz parte do repertório de ações do ser humano. Diante deste achado, é possível inferir que a influência da pista visual se faz presente na percepção de fala de modo associado à percepção auditiva. Estudos demonstram ativação cerebral tanto do lobo temporal quanto do lobo frontal sugerindo assim uma possível integração destas duas fontes de informação (auditiva-visual) na percepção de fala.

Sekiyama et al. (2003) investigaram a ativação cerebral envolvida na percepção audiovisual da fala por meio da ressonância magnética funcional e tomografia por emissão de pósitrons (PET). Em cada experimento, os sujeitos foram solicitados a identificar as seguintes sílabas (/ba/;/da/;/ga/), apresentadas auditivamente, visualmente e audiovisualmente de forma incongruente em duas condições de inteligibilidade (alta X baixa) determinada pela razão sinal-ruído. Os dados de imagem do cérebro mostraram ativações específicas para os estímulos auditivos unimodais (no córtex temporal) e estímulos visuais (no V5-MT), região cerebral designada por ter sido identificada, por alguns autores, em localização médio-temporal, correspondente no cérebro humano à primeira região relacionada à percepção do movimento (Pereira et al., 2003). Para os estímulos audiovisuais bimodais, a ativação no córtex temporal esquerdo estendeu-se mais posteriormente em direção à área visual específica na condição de baixa inteligibilidade. A comparação direta entre as condições audiovisuais (alta X baixa inteligibilidade) mostrou aumento de ativações na parte posterior do sulco temporal superior esquerdo, indicando sua forte relação com a influência visual. Os autores discutiram que é provável que as regiões cerebrais – córtex temporal e médio-temporal (V5- MT) – estejam envolvidas na percepção audiovisual (cross-modal) de fala.

Embora o fenômeno de integração de pistas audiovisuais na percepção de fala seja algo notadamente reconhecido na literatura, há fatores que interferem nesse processo, a saber: a língua (Sekiyama; Tohkura, 1993), a condição clínica do sujeito (Desjardins; Werker, 1996; Dodd *et al.*, 2008) e a idade.

Assim, quando se fala em percepção audiovisual, há necessidade de se considerar diferentes aspectos, já que alguns fatores, como língua e idade, podem interferir no modo de integração de pistas audiovisuais, na percepção de fala. Nas próximas seções, serão tratados esses aspectos.

2.2 LÍNGUA COMO FATOR INFLUENTE NA PERCEPÇÃO AUDIOVISUAL

O estudo desenvolvido por McGurk & MacDonald (1976) demonstrou que a percepção de fala ocorreria por meio da integração audiovisual. Posteriormente, diversos estudos, majoritariamente desenvolvidos em falantes de língua inglesa, replicaram a metodologia, e seus resultados corroboram os achados prévios de McGurk & MacDonald (1976) (Dodd; Hermelin, 1977; Summerfield, 1979, 1983; Massaro; Cohen, 1983; Massaro *et al.*, 1986; Dekle; Fowler; Funnell, 1992).

Na década de 1990, torna-se em voga um novo questionamento: a integração audiovisual na percepção de fala teria um estatuto universal ou seria dependente da língua?

Sekiyama e Tohkura (1991) investigaram a percepção audiovisual de fala em 10 falantes do japonês como língua nativa, de idade entre 22 e 27 anos. Os estímulos de fala apresentados foram os seguintes: (/ba/; /da/; /ga/; /pa/; /ta/; /ka/; /ma/; /na/; /ra/ e /wa/), produzidos por uma falante japonesa. As gravações de áudio e vídeo foram realizadas separadamente e, posteriormente, sincronizadas, resultando em 100 estímulos audiovisuais congruentes e incongruentes. Os estímulos foram apresentados, individualmente, para os participantes e estes foram solicitados a escrever o estímulo percebido.

Durante o estudo piloto, os autores verificaram poucas ocorrências do Efeito McGurk, ou seja, os participantes avaliados não perceberam os estímulos incongruentes como resultantes de uma fusão de informações audiovisuais, de modo que o relato perceptual correspondia majoritariamente à pista auditiva do estímulo incongruente. Posteriormente, duas condições de fala foram comparadas: (1) estímulo audiovisual com adição de ruído; e (2) estímulo audiovisual sem adição de ruído. Mais uma vez, foram verificadas poucas ocorrências do Efeito McGurk na condição de estímulo sem adição de ruído. Diferentemente, na condição com adição de ruído, verificou-se maior ocorrência do efeito, sugerindo fusão de informações visual e auditiva dos estímulos de fala apresentados. Os autores concluíram, portanto, que falantes nativos do japonês integraram com maior frequência as pistas audiovisuais durante a percepção de fala quando o componente auditivo do estímulo não estava perfeitamente inteligível.

Em estudo posterior, Sekiyama e Tohkura (1993) replicaram a metodologia usada em Sekiyama e Tohkura (1991). Os autores investigaram 20 sujeitos falantes do japonês como língua materna e, adicionalmente, 20 sujeitos falantes do inglês como língua materna. Os participantes tinham idade entre 20 e 30 anos. Os estímulos apresentados consistiram em dez sílabas japonesas (/ba/; /da/; /ga/; /pa/; /ta/; /ka/; /ma/; /na/; /ra/ e /wa/) e seus representantes para a língua inglesa. Os estímulos gravados com sílabas japonesas e inglesas foram apresentados para todos os participantes, e estes foram solicitados a escrever o que perceberam após a apresentação de cada estímulo audiovisual.

Considerando os estímulos gravados com sílabas da língua japonesa, os resultados mostraram uma diferença significativa no desempenho perceptual entre os participantes das duas línguas investigadas, sendo que os falantes de japonês, quando comparados aos falantes de inglês, não perceberam os estímulos incongruentes como resultantes de uma fusão de informações audiovisuais, de modo que o relato perceptual, frequentemente, correspondeu à pista auditiva do estímulo incongruente.

Para os estímulos gravados com sílabas da língua inglesa, não foram verificadas diferenças significativas no desempenho perceptual entre os falantes da língua japonesa, comparados aos falantes da língua inglesa. Ambos os grupos avaliados perceberam os estímulos incongruentes como resultantes de uma fusão de informações audiovisuais. Dessa forma, para os falantes da língua japonesa, o contato com uma língua não nativa implicou em aumento na integração audiovisual.

Os autores concluíram que, frequentemente, os falantes do japonês apresentam menor grau de integração audivisual quando comparados aos falantes do inglês; aparentemente, os falantes de japonês como língua materna apresentam maior dominância perceptual da pista auditiva, durante a percepção de fala.

Nos anos seguintes, novos trabalhos foram desenvolvidos em diferentes línguas com o objetivo de comparar o desempenho perceptual de fala em tarefa audiovisual, considerando a língua como um fator influente (Fuster-Duran, 1996; Hayashi; Sekiyama, 1998; Sekiyama; Burnham, 2008; Chen; Hazan, 2009).

Fuster- Duran (1996) investigou a integração audiovisual na percepção de fala entre línguas, especificamente entre o espanhol e o alemão. Para isso, selecionou 120 participantes entre 20 e 30 anos de idade, dividindo-os em 4 grupos: 1) composto por 30 falantes nativos do alemão, que seriam testados com estímulos em língua

nativa; 2) composto por 30 falantes nativos do alemão, que seriam testados com estímulos em língua não nativa; 3) 30 falantes do espanhol, que seriam testados com estímulos em língua nativa; e 4) 30 falantes do espanhol, que seriam testados com estímulos em língua não nativa.

Os estímulos do experimento contemplaram sílabas em padrão simples (CV): /ba/, /ma/, /da/, /na/, /ga/, /Na/, /la/, /ra/, /rra/ e /Ra/), sílabas em padrão complexo (CCV): (/bra/, /dra/, /gra/, /bla/, /dla/ e /bla/) e palavras de ambas as línguas que contemplavam as sílabas selecionadas. Os estímulos foram gravados por uma falante nativa de cada língua, considerando quatro condições de apresentação: auditiva isolada, visual isolada, audiovisual congruente e audiovisual incongruente. Os resultados demonstraram que houve o Efeito McGurk para as duas línguas investigadas, ou seja, os participantes avaliados perceberam o estímulo incongruente como resultantes de uma fusão de informações audiovisuais, de modo que a magnitude da influência visual foi maior para estímulos não nativos do que para estímulos nativos. Os participantes ancoraram-se mais fortemente em pistas visuais, quando foram apresentadas palavras em língua não nativa.

Cabe ressaltar que se constataram diferenças no desempenho perceptual considerando as diferenças fonéticas e fonológicas de cada língua, por exemplo, o tepe /ra/ não ocorre em posição de início de palavra no espanhol. Dessa forma, os nativos do espanhol identificaram o estímulo como /gra/ em 50% das vezes, na condição auditiva isolada; quando o /ra/ foi sincronizado com /ba/, /ga/ ou /da/, os nativos do espanhol perceberam como /bra/ 80% das vezes, /dra/ 70% e /gra/ 67%.

Hayashi e Sekiyama (1998) compararam o desempenho perceptual de fala em tarefa de percepção audiovisual entre falantes do mandarim e japoneses. Participaram do estudo 40 sujeitos subdivididos em dois grupos: (G1) 17 falantes nativos de mandarim; e (G2) 23 falantes nativos de japoneses. Por meio de uma tarefa de identificação, os estímulos de fala – ([ba], [pa], [ma], [da], [ta], [na], [ga], e [ka] – foram apresentados em condições auditiva isolada, visual isolada, audiovisual congruente e incongruente, em japonês e em mandarim, para ambos os grupos.

Os resultados indicaram que, para as condições auditiva isolada, visual isolada e audiovisual congruente, os resultados encontrados foram semelhantes para ambos os grupos (90% de acerto). No entanto, na condição audiovisual incongruente, houve diferença entre os grupos: os sujeitos japoneses apresentaram maior frequência de ocorrência do Efeito MacGurk, ou seja, perceberam com maior frequência o estímulo

incongruente como resultante de uma fusão de informações audiovisuais quando comparados aos sujeitos chineses.

Os autores discutiram tal achado considerando o fato de o mandarim ser uma língua tonal, em que o significado da palavra não é determinado somente pela estrutura silábica, mas também pelo tom. Deste modo, tal característica acarretaria uma forte dependência da informação auditiva, resultando em menor percepção da pista visual. Assim, os falantes de mandarim apresentariam maior dominância perceptual da pista auditiva.

Seguindo a mesma temática de análise, Sekiyama e Burnham (2008) compararam a percepção audiovisual na fala entre as línguas japonesa e inglesa (inglês australiano). Participaram do estudo 96 sujeitos subdivididos em dois grupos: 1) 48 estudantes universitários monolíngues (24 falantes nativos do inglês e 24 falantes nativos do japonês), entre 18 e 29 anos; e 2) 48 crianças de 6 e 11 anos de idade (24 falantes nativos do inglês e 24 falantes nativos do japonês). Os estímulos apresentados em computador foram [ba], [da], [ga], em (1) condição auditiva isolada; (2) visual isolada; (3) audiovisual congruente; e (4) audiovisual incongruente. Os participantes foram solicitados a assistir e ouvir cada estímulo e indicar o que perceberam (pressionando um dos três botões para uma resposta: /ba/, /da/ ou /ga/). Os resultados mostraram que os falantes de japonês apresentaram menor ocorrência do Efeito McGurk. Além disso, os dados indicaram que a magnitude perceptual da pista visual foi baixa e equivalente para os japoneses e australianos de seis anos, e aumentou ao longo da idade para os participantes de língua inglesa, especialmente entre seis e oito anos, mas permaneceu o mesmo para os japoneses.

Os autores concluíram que os japoneses apresentaram menor influência da pista visual quando comparados aos australianos.

Chen e Hazan (2009) avaliaram o efeito da língua e idade na tarefa audiovisual de percepção de fala, em falantes do inglês e mandarim. Participaram do estudo crianças de 8 a 10 anos e adultos de 20 a 54 anos. Os estímulos de fala apresentados foram (/ba/, /da/, /ga/) produzidos por dois falantes de inglês e dois falantes de mandarim, em condição auditiva isolada, visual isolada, audiovisual congruente e incongruente, com e sem ruído associado. Para ambas as línguas investigadas, os resultados indicaram que os adultos, em comparação às crianças, apresentaram maior grau de integração audiovisual, refletido por maior frequência de ocorrência do Efeito McGurk. No que se refere ao efeito da língua, não houve diferença estatística entre os

grupos, ou seja, não houve diferença de desempenho para o processamento das pistas auditivas, visuais e audiovisuais.

Importante destacar que há diversos estudos desenvolvidos a partir desta perspectiva teórica em diferentes línguas na população com desenvolvimento típico de fala, como o finlandês (Sams *et al.*, 1998), italiano (Bovo *et al.*, 2009), francês (Fort *et al.*, 2012), alemão (Laarhoven *et al.*, 2018), espanhol (Pons *et al.*, 2013), e português brasileiro (Silva; Pacheco; Pagliari, 2017). Todos esses estudos não compararam a percepção audiovisual entre línguas, mas avaliam a percepção de fala sob a perspectiva audiovisual. Os resultados mostraram que a acurácia perceptual foi aumentada com a inserção da pista visual ao estímulo, de modo consistente entre os estudos supracitados, ocorrendo melhora na acurácia perceptual com acréscimo da informação visual. A seguir será resenhado o estudo de Silva *et al.* (2017) que, embora não faça uma comparação entre línguas, traz resultados interessantes quanto à percepção audiovisual para falantes do PB.

No estudo desenvolvido por Silva *et al.* (2017), os autores avaliaram o papel das informações auditiva e visual na percepção das fricativas do Português Brasileiro, quando apresentadas em condição incongruente. Participaram do estudo treze indivíduos com idade entre 25 e 45 anos. Os estímulos de fala apresentados foram palavras dissílabas (com estrutura silábica C1V1.C2V2, onde C1 é uma das fricativas /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, C2 é uma oclusiva e V1 e V2 são uma das vogais /a/, /i/ ou /u/), apresentadas em condição auditiva, visual e audiovisual. Os resultados demonstraram que os índices de identificação das fricativas foram, de maneira geral, maiores em condição de apresentação audiovisual. Os autores concluíram que as informações presentes no sinal visual das fricativas colaboraram para o aumento nos índices de identificação e discriminação perceptual desses sons, ou seja, a informação visual tem participação importante na percepção das fricativas do Português Brasileiro.

Em suma, os resultados dos estudos resenhados até aqui indicam que a integração audiovisual na percepção de fala teria um estatuto universal. No entanto, o grau de integração audiovisual varia a depender das características de cada língua (línguas de padrão tonal, entoacional, complexidade da estrutura silábica) e fatores extrínsecos, como o ruído e contato com língua não nativa, visto que há uma relação não linear entre a percepção de pistas auditivas e visuais, em que uma pode se sobrepor a outra a depender da língua e de fatores extrínsecos.

Nesta sessão, com base nos estudos apresentados, conclui-se que a língua é um fator influente na integração de pistas, durante a percepção de fala.

Assim como a língua, a idade tem sido apontada como um fator influente no modo como se integram pistas auditivas e visuais, na medida em que estudos apontam para um possível desenvolvimento da integração audiovisual, conforme o aumento da idade, na comparação entre bebês, crianças e adultos (Burnham *et al.*, 1991; Taitelbaum-Swed; Fostick, 2016; Chen; Hazan, 2007).

Na próxima seção, serão apresentados os estudos dentro da teoria audiovisual da percepção de fala, que consideram a idade como um fator influente na percepção de fala. O efeito da idade na percepção de fala é o ponto-chave da presente tese.

2.3 IDADE COMO FATOR INFLUENTE NA PERCEPÇÃO AUDIOVISUAL

Na literatura, a idade é reportada como um fator influente na integração de pistas audiovisuais na percepção de fala. De modo geral, os estudos convergem para uma relação de dominância entre pistas audiovisuais, de modo que, em crianças, a percepção de fala seria pautada, majoritariamente, no componente auditivo, enquanto em adultos se daria através da integração audiovisual, refletida por um aumento perceptual da pista visual ao longo do desenvolvimento (Hockley, 1994; Sekiyama *et al.*, 2003; Sekiyama; Burnham 2004).

Para melhor compreensão do desenvolvimento da percepção de fala a partir da teoria audiovisual, serão apresentados, a seguir, estudos realizados, considerando diferentes grupos etários: bebês, crianças e adultos.

Embora escassos na literatura, estudos em bebês demonstram que esta população é perceptualmente sensível aos estímulos audiovisuais na percepção de fala (Dodd, 1979; Kuhl; Meltzoff, 1984; Rosenblum; Schmuckler; Johnson, 1997; Lewkowicz; Hansen-Tift, 2012; Tomalski *et al.*, 2013; Pons; Bosch; Lewkowicz, 2015; Berdasco-Muñoz *et al.*, 2019). Na literatura, há relato sobre a investigação da percepção audiovisual em bebês, a partir de dois meses de vida até um ano de idade. Os estudos serão apresentados a seguir, conforme o desenvolvimento etário.

Dodd (1979) investigou a percepção audiovisual de fala em 12 bebês de 10 a 16 semanas de vida, aos quais foram apresentadas canções com sons da fala com movimentos labiais em congruência e fora de congruência (com atraso de 400 ms)

entre o gesto articulatório e estímulo auditivo). Dois observadores mediram a duração da fixação visual dos bebês para os estímulos apresentados. Os resultados mostraram que a duração da fixação visual foi menor diante da apresentação dos estímulos incongruentes em comparação à apresentação congruente. Esse achado indicou à autora que os bebês discriminam estímulos audiovisuais congruentes e incongruentes, durante a percepção da fala.

Resultado semelhante foi relatado no estudo desenvolvido por Kuhl e Meltzoff (1984), em que os autores investigaram as habilidades de percepção audiovisual de 64 bebês de 18 a 20 semanas de vida. Os autores elaboraram dois experimentos perceptuais. No experimento 1, os participantes assistiram a dois vídeos apresentados lado a lado, que correspondiam aos movimentos articulatórios das vogais /a/ e /i/, de forma sincronizada a um estímulo auditivo correspondente a uma das vogais. A fixação visual dos bebês foi registrada via vídeo e, posteriormente, analisada por um avaliador. Os resultados mostraram maior tempo de fixação visual para o vídeo que estava em congruência ao estímulo sonoro apresentado. Para o experimento 2, foram apresentados aos participantes dois vídeos lado a lado, que correspondiam aos movimentos articulatórios das vogais /a/ e /i/, de forma sincronizada com um estímulo auditivo. Contudo, o estímulo auditivo foi editado, precisamente, retirando-se a informação espectral contida na vogal, enquanto as características temporais foram mantidas. Os resultados mostram ausência de fixação visual significativa entre os estímulos. Os autores concluíram que os bebês reconhecem as correspondências audiovisuais contidas nas informações da fala, sugerindo que eles percebem as informações da fala de forma intermodal.

Os estudos vanguardistas na investigação da percepção audiovisual em bebês, desenvolvidos por Dodd (1979) e Kuhl e Meltzoff (1984), indicam que bebês de dois a cinco meses discriminam estímulos audiovisuais congruentes e incongruentes, mostrando que este grupo etário é perceptualmente sensível às pistas audiovisuais.

A partir dos seis meses de idade, gradualmente, os bebês iniciam atenção sustentada à boca de seu interlocutor, durante a percepção de fala. Tomalski *et al.* (2013) investigou 32 bebês ingleses dividindo-os em dois grupos: o primeiro contendo 16 bebês de 6 a 7 meses; e o segundo contendo 16 bebês de 8 a 9 meses. Os estímulos selecionados foram as sílabas /ba/ e /ga/ produzidas por uma falante do inglês, apresentados de forma congruente e incongruente.

Os participantes foram posicionados no colo dos pais, diante de um monitor com rastreador ocular. Os movimentos dos olhos foram analisados por áreas de fixação: boca, olhos e toda a face. A soma do tempo total de fixação do olhar foi calculada para cada participante. Os resultados mostraram que bebês de seis a nove meses aumentaram sua fixação do olhar para a boca do falante somente para a apresentação de estímulos audiovisuais incongruentes. Além disso, houve um efeito para a idade: quanto mais velho o bebê, maior o tempo de fixação visual para a boca.

Aos 12 meses de idade, os bebês diminuem sua fixação do olhar para a boca de seu interlocutor. Este dado é apresentado no estudo de Lewkowicz e Hansen-Tift (2012), em que rastrearam a fixação do olhar em bebês de 4 a 12 meses de idade, em aquisição do inglês, enquanto observavam e ouviam uma mulher recitando um monólogo em seu idioma nativo (inglês) ou não nativo (espanhol). Os resultados mostraram que os bebês de quatro a oito meses realizaram uma varredura visual dos olhos para a boca da interlocutora, com maior direcionamento ocular, independentemente da língua. Aos 12 meses, foi observado maior direcionamento ocular para os olhos da interlocutora, para os estímulos em língua nativa, mas não para os estímulos em língua não nativa.

Os autores observaram que o direcionamento ocular para a boca do interlocutor, em bebês de quatro a oito meses, permite que tenham acesso a pistas audiovisuais de fala de extrema importância para aquisição da língua materna. Aos 12 meses, quando se observou maior direcionamento ocular para os olhos da interlocutora, poderia refletir o crescente conhecimento da língua nativa que os “liberariam” para desviar a atenção para os olhos a fim de obter acesso a recursos sociais. Por este motivo, os bebês de 12 meses de idade não desviam a atenção para os olhos quando expostos à fala não nativa, porque o aumento do conhecimento da língua nativa e o estreitamento perceptivo a outras línguas tornam mais difícil o processamento da fala não nativa, exigindo, assim, ancoragem plena nas pistas auditivas e visuais.

Com achados semelhantes, Pons, Bosch e Lewkowicz (2015) compararam o desempenho de bebês monolíngues e bilíngues em tarefa de percepção audiovisual da fala. Para isso, selecionaram 60 bebês, que foram agrupados considerando suas faixas etárias: 4, 8 e 12 meses. Os bebês foram posicionados em um assento infantil diante de um monitor de computador. Os movimentos oculares foram registrados por um rastreador ocular. Os estímulos consistiram de um monólogo recitado por duas

atrizes, uma falante do catalão-espanhol e outra falante do inglês americano. Cada participante assistiu a dois vídeos, um em sua língua nativa e outro em língua não nativa.

Considerando o desempenho de bebês monolíngues, os resultados demonstraram que os bebês, aos 4 meses, fixaram mais o olhar para os olhos do falante; aos 8 meses, fixaram o olhar mais para a boca do falante, considerando estímulos de língua nativa e não nativa; e, aos 12 meses, fixaram mais o olhar para a boca do falante somente para os estímulos apresentados em língua não nativa. Considerando o desempenho de bebês bilíngues, os resultados indicaram que os bebês fixaram o olhar para os olhos e para a boca, igualmente, aos quatro meses; e fixaram mais o olhar para a boca do interlocutor, aos oito meses. Aos 12 meses, os bebês mantiveram uma alta fixação de olhar para a boca do falante para estímulos em língua nativa e não nativa.

Em síntese, os estudos desenvolvidos em bebês abrangem a faixa etária de 2 a 12 meses de idade, e os resultados demonstram que esta população, nos primeiros meses de vida, já é capaz de discriminar pistas audiovisuais em congruência e em incongruência, indicando, assim, que são sensíveis a pistas audiovisuais na percepção de fala. A partir dos 6 meses, há registro de que os bebês iniciam uma varredura visual para a face do interlocutor com maior fixação para região da boca; aos 12 meses, os bebês apresentam maior fixação visual para o olho de seu interlocutor.

Tais achados são justificados na literatura, considerando a aquisição de linguagem. O crescente conhecimento e contato com a língua nativa implicaria em uma habituação perceptual ao sistema fonético e fonológico nativo, o que justificaria a redução da fixação visual para a região da boca e maior fixação visual para o olho do interlocutor, observada aos 12 meses. Deste modo, esse grupo possivelmente integraria com maior velocidade de processamento as pistas audiovisuais.

Ademais, ao desempenho perceptivo-audiovisual em bebês, a literatura aponta um possível desenvolvimento no uso da integração de pistas, conforme avanço da idade (Hockley, 1994; Ross, 2011; Lalonde; Holt, 2016; Chen; Hazan 2009).

Polka (1994) indica em seu estudo que a percepção de fala no adulto é marcada por alto grau de integração audiovisual em comparação às crianças, especialmente a percepção de fala nas crianças entre cinco e oito anos, que seria marcada por baixa influência da pista visual e maior influência da pista auditiva. No entanto, com o avanço

da idade, a influência da informação visual aumenta, acarretando uma integração audiovisual em maior grau.

Os autores também investigaram o desempenho em tarefa de percepção audiovisual em diferentes faixas etárias. Os participantes deste estudo foram separados em cinco grupos: quatro grupos de crianças, sendo cada um composto por uma faixa etária (4 a 12 anos), e um grupo composto por adultos (18 a 28 anos). Foram testados em quatro diferentes condições: (1) auditiva isolada; (2); visual isolada; (3) audiovisual congruente; e (4) audiovisual incongruente. Os estímulos analisados foram os seguintes: /ba/, /va/, /ta/, /da/, e /ga/. Os resultados indicaram que a influência da informação auditiva diminuiu com a idade, enquanto a influência da informação visual aumentou com a idade. Além disso, um padrão de resposta semelhante ao adulto foi observado em apenas metade do grupo de crianças, na faixa etária entre 10 e 12 anos, sugerindo que a integração audiovisual continua a se desenvolver em idade posterior aos 12 anos.

Tal achado vai de encontro ao estudo de McGurk & MacDonald (1976), no qual crianças de três a cinco anos e de sete a oito anos obtiveram menor influência da pista visual na percepção de fala em comparação aos adultos. Os autores atribuíram tal achado a um aumento gradual (desenvolvimento) da influência visual, durante a infância até a idade adulta.

Adicionalmente, Sekiyama *et al.* (2003) investigaram o desempenho em tarefa de percepção audiovisual em falantes das línguas japonesa e inglesa. Foram analisados 48 adultos, na faixa etária entre 18 e 29 anos, e 16 crianças, na faixa etária de 6 anos, todos divididos em grupos a depender da sua língua nativa. O estudo envolveu dois grupos de adultos e dois grupos de crianças. Os estímulos investigados foram /ba/, /da/ e /ga/ produzidos por quatro falantes, um homem e uma mulher falantes do inglês e um homem e uma mulher falantes do japonês. Os estímulos foram editados de forma a apresentar somente pista visual, somente pista auditiva e pistas audiovisuais (congruentes e incongruentes), e apresentados sob diferentes níveis de inteligibilidade, testando-se estímulos com e sem adição de ruído. Todos os participantes foram instruídos a assistir e ouvir os estímulos e, posteriormente, indicar (pressionando um botão do computador, previamente acordado) o que haviam percebido dentre três possibilidade de resposta. A análise inferencial buscou verificar a interação entre as condições de apresentação, as línguas e as idades. Os resultados demonstraram que estímulos apresentados com adição de ruído exigiram maior grau

de influência visual – calculado pela diferença entre a porcentagem correta da condição AV+ e a porcentagem correta da condição AV- – de falantes da língua inglesa do que falantes da língua japonesa, independente da faixa etária investigada. Adicionalmente, os adultos falantes da língua inglesa demonstraram maior grau de influência visual, comparados aos adultos falantes da língua japonesa, para estímulos sem adição de ruído; enquanto, para o grupo de crianças, essa diferença não foi verificada.

Os autores concluem que existe uma diferença perceptual entre as línguas investigadas, que é marcada somente com o aumento da idade, sendo que falantes do inglês são mais influenciados pela pista visual, comparados aos falantes da língua japonesa. As crianças quando comparadas aos adultos apresentam menor grau de integração audiovisual, uma vez que são menos influenciadas pela pista visual.

Os estudos resenhados nesta seção demonstram que a integração audiovisual modifica-se ao longo das faixas etárias. No entanto, Um novo questionamento surge: haveria uma faixa etária específica para a qual a integração audiovisual atingiria o seu platô desenvolvimental? Com base em tal questionamento, Ross (2011) investigou o desenvolvimento da integração audiovisual, a partir dos cinco anos até a idade adulta, a fim de quantificar o aprimoramento da integração audiovisual ao longo das faixas etárias. Neste estudo, participaram 58 sujeitos falantes nativos do inglês com desenvolvimento típico de fala e linguagem, divididos em cinco grupos etários: (G1) 5 a 7 anos, (G2) 8 a 9 anos, (G3) 10 a 11 anos, (G4) 12 a 14 anos e (G5) 16 a 56 anos. Os participantes realizaram um teste de percepção de fala, constituído por 300 palavras monossilábicas de alta frequência, do ambiente cotidiano de uma criança. As palavras foram retiradas de um banco de dados psicolinguístico (Coltheart, 1981), gravadas por uma falante do sexo feminino e apresentadas para os participantes dos estudos em três condições: auditiva isolada (correspondente à gravação auditiva da palavra-alvo apresentada em conjunto com uma imagem estática da face do orador); visual isolada (correspondente à apresentação articulatória do locutor associada com ruído auditivo, mas sem qualquer estímulo de fala auditiva); audiovisual (correspondente à apresentação da gravação da locutora articulando as palavras-alvo). Os estímulos de fala foram apresentados em 15 blocos de 20 palavras, e cada palavra foi apresentada uma única vez em 50db, associada a seis níveis de ruído: 53db, 56db, 59db, 62db e 65db. Após a apresentação, os participantes foram orientados a reportar a palavra apresentada.

Como resultado, observou-se que crianças mais novas se beneficiaram da informação articulatória da fala. Entretanto, esse benefício foi significativamente inferior ao observado em adultos, corroborando as hipóteses levantadas. Os autores observaram que os adultos se beneficiaram significativamente mais com a pista visual do que crianças de 5 a 12 anos.

Para a condição visual, as crianças mais novas não apresentaram o pico de acurácia perceptual característico encontrado em adolescentes e adultos. Considerando a condição audiovisual, o desempenho foi sujeito à mudança substancial entre as faixas etárias de cinco anos até a idade adulta. Um ganho audiovisual máximo em ruídos ‘intermediários’ foi inferior em crianças em idade pré-escolar e, durante os primeiros anos escolares, continuaram a evoluir até a adolescência.

Já para a condição auditiva isolada, observou-se modesta melhora na acurácia perceptual com o aumento da idade, ou seja, houve pequena magnitude na diferença de desempenho para a condição auditiva isolada, entre crianças e adultos.

Os autores concluíram que o desenvolvimento da integração audiovisual não é uniforme, há picos de desempenho de 8 a 11 anos. Entretanto, somente entre as faixas etárias de 13 e 16 anos, o desempenho em tarefa de percepção audiovisual se assemelha ao padrão do adulto, sugerindo que a estabilização da integração audiovisual ocorreria a partir dos 13 anos de idade, corroborando o estudo de Barutchu et al. (2010).

Fort *et al.* (2012) compararam a influência da pista visual durante a percepção de fala, em 96 crianças (5 a 10 anos) e em 60 adultos falantes nativos do francês, com o desenvolvimento típico de fala e linguagem. Os participantes realizaram um teste de percepção de fala no qual foram apresentadas palavras e pseudopalavras em condição auditiva isolada e audiovisual congruente. Os estímulos do experimento foram mascarados com ruído branco. Como resultado, os autores reportaram melhora no desempenho perceptual para a condição audiovisual conforme o aumento da idade (o que sugere melhora na integração entre pistas auditivas e visuais, refletindo numa percepção de fala mais acurada). A partir deste resultado, os autores constataram um pico de integração audiovisual dos seis aos sete anos de idade. Para nenhuma faixa etária do grupo de crianças, o desempenho perceptual foi equiparado ao desempenho observado nos adultos. Deste modo, os autores registraram que, para os falantes de

francês, a integração audiovisual de crianças não estaria equiparada aos adultos até os dez anos de idade (faixa etária mais elevada no grupo de crianças do estudo).

Seguindo a mesma linha de pesquisa, Maidment (2015) investigou o desenvolvimento da integração audiovisual em crianças, com desenvolvimento típico de fala e linguagem, em idade de 4 a 11 anos. Participaram do estudo 84 sujeitos, 69 crianças (4 a 11 anos de idade) e 15 adultos (18 a 26 anos) falantes nativos do inglês. Os participantes realizaram um teste de percepção audiovisual. O corpus foi constituído por seis cores (azul, preto, rosa, vermelho, verde, branco) e oito dígitos (1 a 9 excluindo o 7), apresentados em uma frase veículo “Mostre ao cachorro onde está a [cor] [dígito]”. Após a apresentação de cada estímulo, os participantes foram orientados a indicar a cor e o dígito correspondente. Os estímulos foram apresentados em frames do terço inferior da face nas seguintes condições: auditiva, visual e audiovisual congruente. Ademais, foi adicionado ruído aos estímulos auditivos e audiovisuais. Na análise, realizaram-se comparações entre as condições auditiva e audiovisual, e calculado o ganho audiovisual através da subtração entre as condições auditiva e audiovisual ($A - AV$). Os resultados demonstraram que as crianças de 4 a 5 anos de idade necessitam de maior resolução espectral para identificar o estímulo-alvo em comparação às crianças de 6 a 11 anos e adultos. Observou-se também que o ganho visual aumentou em função da idade, concluindo-se que, a partir dos seis anos de idade, houve melhora na integração audiovisual durante a percepção de fala.

Os autores discutiram tais resultados, realizando um paralelo entre o início da alfabetização e o marco observado na melhora da integração audiovisual, aos seis anos. Verificaram que, no ambiente escolar, as crianças são expostas a condições sonoras desafiadoras (ruídos concorrentes). Porém, a partir desta faixa etária, o processamento bem sucedido das instruções do professor torna-se crucial para o sucesso acadêmico. Consequentemente, as crianças passariam a direcionar o foco atencional de modo mais consistente para articulação do professor, o que viabilizaria o aumento da integração audiovisual.

Lalonde e Holt (2016) investigaram a percepção auditiva e audiovisual, em crianças de 6 a 8 anos e em adultos de 18 a 31 anos, nos 3 níveis de processamento perceptual: detecção, discriminação e reconhecimento. Participaram do estudo 49 sujeitos falantes nativos do inglês, 24 crianças (6 a 8 anos) e 25 adultos (18 a 31 anos), que realizaram três testes perceptuais referentes aos três níveis de processamento (detecção, discriminação, reconhecimento).

Os estímulos do teste consistiram na apresentação de 22 palavras monossilábicas incorporadas a um intervalo de 2,37s de ruído, que antecederam sua apresentação. As palavras foram apresentadas na condição auditiva (informação acústica apresentada em conjunto com imagem da face congelada) e na condição audiovisual (vídeos da face e áudio apresentados de modo simultâneo). Os estímulos foram gravados por uma mulher falante nativa do inglês.

Ainda, de acordo com Lalonde e Holt, explicam-se os três níveis de processamento perceptual:

- Detecção, teste constituído pela apresentação de palavras monossilábicas entre ruído e apresentação de ruído isolado. Os participantes deveriam indicar “sim” para presença de palavra e “não” para ausência de palavra.

- Discriminação, teste em que cada conjunto para discriminação foi apresentado com 800 ms de intervalo entre cada palavra. Os participantes deveriam indicar “sim” para palavras iguais e “não” para palavras diferentes.

- Reconhecimento, teste em que as palavras-alvo foram apresentadas entre ruídos e, logo após, foram dadas sugestões referentes ou não à palavra-alvo. Os participantes deveriam julgar se as sugestões correspondiam ou não à palavra-alvo.

Importante destacar que, para cada teste, as palavras foram apresentadas em condições auditiva e audiovisual. Como resultado, os autores reportaram que o grupo de adultos apresentou maior influência visual (refletindo em maior grau de integração audiovisual) para a tarefa de reconhecimento em detrimento das tarefas de detecção e discriminação. Assim, para mecanismos perceptivos específicos, os adultos apresentaram maior grau de integração audiovisual.

Para o grupo de crianças, conservou-se o mesmo padrão de influência visual para todos os testes apresentados. Na comparação de desempenho entre o grupo de crianças e adultos, observou-se que as crianças mostraram menor influência visual para o teste de reconhecimento. A partir dos resultados, os autores concluíram que os adultos, em comparação às crianças, apresentaram integração audiovisual para todos os níveis de processamento avaliado (processamento inferior e superior); já o grupo de crianças apresentou menor grau de integração audiovisual para a tarefa de reconhecimento. Esta tarefa está relacionada a níveis perceptuais específicos, especialmente ao acesso lexical.

Em relação ao marco etário para integração audiovisual na percepção de fala, os autores afirmam que é importante considerar os níveis de processamento

perceptual, pois, para cada nível, há um grau de complexidade. Eles apontam que, para a faixa etária de crianças de seis a oito anos, a integração audiovisual ocorre em menor grau para níveis de processamento de maior complexidade, como a habilidade de reconhecimento (habilidade de maior complexidade, quando comparada às habilidades de detecção e discriminação), o que interferiria na acurácia perceptual. Portanto, o desempenho audiovisual na faixa etária de seis a oito anos não estaria equiparado ao padrão adulto.

Em 2018, Heikkilä *et al.* investigaram respostas comportamentais e neurais de crianças em idade escolar (7 a 11 anos) e em adultos (20 a 29 anos), em experimento de percepção audiovisual de fala. O rastreamento ocular foi realizado para os testes comportamentais e neurais. As perguntas de pesquisas foram assim definidas pelos autores: o processamento de fala incongruente é diferente do processamento de fala congruente? O processamento é diferente em crianças e em adultos?

Participaram do estudo 18 crianças de 7 a 11 anos de idade e 14 adultos de 20 a 29 anos de idade, falantes nativos do finlandês, com desenvolvimento típico de fala e linguagem. Os participantes realizaram um teste perceptual comportamental e neural, mediante apresentação dos estímulos silábicos /mi/ e /ni/.

O teste perceptual comportamental foi constituído por estímulos auditivo, visual, audiovisual congruente e incongruente (A/mi/V/ni/). Os estímulos foram previamente gravados por uma falante nativa do finlandês e apresentados com o registro de toda a face. Para a condição visual, o áudio foi associado à apresentação de toda a face com a boca fechada. Após a apresentação de cada estímulo, os participantes foram orientados a indicar a consoante percebida dentre duas opções de resposta: /mi/ e /ni/. Para o registro de atividade neural, realizou-se o eletroencefalograma. Os estímulos apresentados e as condições de apresentação foram as mesmas descritas anteriormente. Os participantes foram orientados a olhar para a tela, detectar personagens de desenho animado na parte superior do rosto e pressionar um botão em um dispositivo de resposta quando um aparecesse.

Considerando o experimento comportamental, os resultados indicaram diferença estatística entre grupos somente para a condição audiovisual incongruente. Os autores analisaram a força do Efeito McGurk através da porcentagem de respostas de /mi/ mediante apresentação do estímulo audiovisual (A/mi/V/ni/): quanto menos respostas ao componente auditivo, mais forte seria o Efeito McGurk.

Os resultados indicaram maior Efeito McGurk para o grupo de adultos, sugerindo que a pista visual influenciou mais a percepção de fala em adultos do que em crianças. Já para o experimento neural, mediante utilização de eletroencefalograma, os resultados mostram que, em crianças, o processamento neural da fala audiovisual difere em alguns aspectos em comparação com os adultos. Em adultos, os estímulos congruentes e incongruentes provocaram uma resposta de processamento semelhante; nas crianças, houve divergência de processamento entre os estímulos, especialmente para os estímulos congruentes, e a resposta foi mais acurada em comparação ao estímulo incongruente.

Os autores concluíram que os resultados refletem uma diferença maturacional para o processamento de estímulos audiovisuais em crianças (7 a 11 anos), em comparação aos adultos (20 a 29 anos). Destaca-se que este estudo não investigou a diferença de desempenho entre as faixas etárias de cada grupo, o que impede de se verificar a diferença de processamento perceptual ao longo das faixas etárias. Porém, é possível concluir que, para falantes do finlandês, a integração audiovisual não atingiu o seu platô desenvolvimental aos 11 anos de idade, ou seja, até esta faixa etária, a integração audiovisual não se equipara a do adulto.

Até aqui, é possível concluir que os estudos apresentam uma convergência de informações, principalmente no que se refere ao marco etário de seis anos de idade, em que ocorre aumento significativo da integração audiovisual. Porém, a literatura não aponta uma faixa etária específica na qual a integração audiovisual atingiria o seu platô desenvolvimental, assemelhando-se ao padrão-alvo adulto. Aparentemente, a partir dos 13 aos 16 anos, a integração audiovisual se assemelha ao padrão-alvo adulto, mas são necessários mais estudos.

A seguir, é apresentado o Quadro 1 com a síntese apenas dos artigos que compararam a integração audiovisual em diferentes faixas etárias, considerando sujeitos com desenvolvimento típico de fala e linguagem.

Quadro 1 - Levantamento bibliográfico de estudos sobre integração audiovisual na percepção de fala, considerando a faixa etária como fator influente

Autor/Ano	Número/ População	Grupo etário	Língua	Tarefa Perceptual	Presença de ruído	Idade para correspondência perceptual entre adultos e crianças
Hockley (1994)	61 sujeitos	4 - 12 18 - 28	Inglês Americano	Estímulo silábico /ba/, /va/, /ta/, /da/, apresentado em condições: A, V, AV+ e AV-. Teste de identificação.	Não	Não houve desempenho equiparado entre adultos e crianças.
Sekiyama <i>et al.</i> (2003)	64 sujeitos	6 18 - 29	Inglês e Japonês	Estímulo silábico /ba/, /da/, /ga/, apresentado em condição A, V, AV+ e AV-. Tarefa de identificação.	Sim	Não houve desempenho equiparado entre adultos e crianças.
Ross <i>et al.</i> (2011)	58 sujeitos	5 - 7 8 - 9 10 - 11 12 - 14 16 - 56	Inglês Americano	300 palavras monossilábicas apresentadas em condição: A, V, AV+. Teste de identificação	Sim	13 anos
Fort <i>et al.</i> (2012)	150 sujeitos	5 - 10 18 - 51	Francês	Palavra e pseudopalavra em condição: A e AV+. Teste de identificação.	Sim	Não houve desempenho equiparado entre adulto e criança. Pico de integração audiovisual 6 - 7 anos
Maidment <i>et al.</i> (2015)	84 sujeitos	4 - 11 18 - 26	Inglês Americano	Cores e dígitos, apresentados em uma frase veículo. Teste de identificação.	Sim	Não houve desempenho equiparado entre adulto e criança. Pico de integração audiovisual aos 6 anos.
Lalonde e Holt (2016)	49 sujeitos	6 - 8 18 - 31	Inglês Americano	22 palavras monossilábicas em condição: A e AV+.	Sim	Não houve desempenho equiparado entre adulto e criança.
Heikkilä <i>et al.</i> (2018)	32 sujeitos	7 - 11 20 - 29	Finlandês	Estímulo silábico /mi/ e /ni/, em condição: A, V, AV+ e AV-.	Não	Não houve desempenho equiparado entre adulto e criança.

				Registro neural via eletroencefalograma.		
--	--	--	--	--	--	--

Fonte: Elaborado pela autora.

Considerando que a língua constitui um fator relevante nos estudos sobre percepção de fala, aliada à lacuna existente em pesquisas sobre percepção audiovisual da classe das fricativas e seu desenvolvimento ao longo das diferentes faixas etárias em falantes do PB, a presente tese tem por objetivo investigar a percepção audiovisual da fala de fricativas desvozeadas em função da idade, em crianças na faixa etária de 4 a 12 anos e em adultos na faixa etária de 20 a 30 anos. A seção a seguir, serão apresentados os objetivos e as hipóteses do estudo.

3 OBJETIVOS E HIPÓTESES

Objetivo geral: Investigar a percepção audiovisual da fala de fricativas surdas em função da idade, em crianças na faixa etária de 4 a 12 anos e em adultos na faixa etária de 20 a 30 anos.

Objetivos específicos:

- descrever o processamento unissensorial auditivo e visual em função da idade;
- descrever o ganho visual na percepção de fala em função da idade;
- verificar se ocorre o Efeito McGurk (ilusão audiovisual) e, em caso afirmativo, pesquisar o efeito da idade.

Uma vez que a integração audiovisual na percepção de fala depende da experiência linguística (Hazan; Barret, 2000; Burnham *et al.*, 1991), do desenvolvimento das habilidades psicoacústicas (Schneider *et al.*, 1986) e visuais (Desjardins, 1997), as hipóteses deste trabalho foram assim definidas:

H1 - haveria diferença no desempenho perceptual referente ao processamento unissensorial entre os grupos. Espera-se que a acurácia perceptual auditiva e visual aumente com a idade;

H2 - esperam-se diferenças no ganho visual entre os grupos etários, marcadas pela maior porcentagem de ganho visual para adultos e crianças, a partir dos seis anos de idade, uma vez que estudos prévios apontam que o pico de ganho visual ocorreria nessa faixa etária (Fort *et al.*, 2012; Maidment *et al.*, 2015);

H3 - espera-se observar a presença do Efeito McGurk (ilusão audiovisual) e, ainda, uma diferença desse efeito entre os grupos etários. Particularmente, espera-se que os adultos e crianças de maior faixa etária apresentem maior porcentagem de relato perceptual de ilusão audiovisual.

A tese a ser defendida, portanto, é a de que a percepção é de natureza audiovisual, tendo a língua e a idade como fatores influentes.

4 METODOLOGIA

4.1 ASPECTOS ÉTICOS

Este é um estudo transversal prospectivo, aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Filosofia e Ciências – Unesp Campus de Marília, sob o Protocolo nº 5.391.404. Todos os responsáveis pelos participantes assinaram o Termo de Consentimento Livre e Esclarecido que lhes foi apresentado. Os menores de idade concordaram em participar da pesquisa, a partir da assinatura do Termo de Assentimento, elaborado de acordo com a faixa etária.

4.2 CONSTITUIÇÃO DA AMOSTRA

Participaram deste estudo 75 sujeitos, com idades entre 4 e 30 anos, todos falantes monolíngues do Português Brasileiro (PB), divididos em cinco grupos etários: grupo 1 (G1), 15 crianças entre 4 e 5 anos e 11 meses; grupo 2 (G2), 15 crianças entre 6 e 8 anos e 11 meses; grupo 3 (G3), 15 crianças entre 9 e 10 anos e 11 meses; grupo 4 (G4), 15 crianças entre 11 e 12 anos e 11 meses; e grupo 5 (G5), 15 adultos entre 20 e 30 anos e 11 meses. A definição dessas faixas etárias baseou-se em estudos prévios que indicam que o pico do ganho visual ocorre a partir dos 6 anos (Fort *et al.*, 2012) e que a estabilização da integração audiovisual tende a acontecer até os 13 anos (Barutchu *et al.*, 2010). Já o grupo adulto foi delimitado entre 20 e 30 anos, uma vez que essa faixa etária apresenta o melhor desempenho em termos de acurácia perceptual (Taitelbaum-Swed; Fostick, 2016).

Crianças de 4 a 12 anos foram recrutadas na Escola Municipal Prof. Sebastião Vayego de Carvalho, localizada em Ribeirão Pires/SP, enquanto os adultos (20 a 30 anos) foram recrutados estudantes na UNESP – Campus Marília. Os locais de recrutamento foram escolhidos por conveniência.

Todos os sujeitos passaram por avaliação da produção de fala a fim de garantir a ausência de qualquer alteração de fala. Realizou-se uma tarefa de nomeação, com o uso do Instrumento IAFAC - Avaliação de Fala para Análise Acústica, em contexto da vogal /a/ (Berti; Pagliuso; Lacava, 2009).

Os dados referentes às produções de fala foram submetidos ao julgamento perceptivo-auditivo do avaliador especialista, a partir da análise oitiva da produção de

fala. Após a análise da produção dos participantes, realizou-se o cálculo do PCC- R (Porcentagem de Consoantes Corretas-revisado) das amostras de fala dos sujeitos avaliados. Somente aqueles com valores de PCC-R acima de 85% foram incluídos na amostra, indicando ausência de qualquer dificuldade na produção dos sons da fala (Shriberg *et al.*, 1997). Para os adultos (20 a 30 anos), além da aplicação do teste IAFAC, foi investigada a história pregressa, incluindo a presença de queixas relacionadas à fala e à linguagem durante a infância.

Os sujeitos não passaram por exame auditivo e/ou oftalmológico, a inclusão foi pautada na ausência de queixa visual e/ou auditiva.

Portanto, seguindo os critérios de inclusão (sujeitos com desenvolvimento típico de fala) e exclusão (queixa audiológica, oftalmológica, neurológica, cognitiva e aletrações de fala), foram avaliados 90 sujeitos, porém 19 foram excluídos em virtude da presença de alterações na produção de fala (trocas fonológicas e distorções), queixas auditivas (zumbido, otite de repetição, perda auditiva), queixas visuais (miopia, astigmatismo). Assim, foram incluídos no estudo 75 sujeitos.

A seguir, no Quadro 2, segue a relação dos participantes presentes em cada grupo.

Quadro 2 - Caracterização dos sujeitos participantes do estudo

Grupo	Sujeito	Idade	Grupo	Sujeito	Idade	Grupo	Sujeito	Idade	Grupo	Sujeito	Idade	Grupo	Sujeito	Idade
G1	1	4a 0m	G2	1	6a 4m	G3	1	9a 4m	G4	1	11a m	G5	1	21a 0m
G1	2	5a 0m	G2	2	6a 4m	G3	2	9a 1m	G4	2	12a 1m	G5	2	20a 4m
G1	3	5a 10m	G2	3	6a 1m	G3	3	9a 7m	G4	3	11a 6m	G5	3	20a 9m
G1	4	5a 10m	G2	4	6a 3m	G3	4	9a 7m	G4	4	12a 2m	G5	4	22a 3m
G1	5	4a 8m	G2	5	7a 10m	G3	5	9a 8m	G4	5	11a 3m	G5	5	22a 11m
G1	6	5a 5m	G2	6	6a 4m	G3	6	9a 3m	G4	6	11a 0m	G5	6	21a 7m
G1	7	5a 4m	G2	7	8a 10m	G3	7	10a 11m	G4	7	12a 3m	G5	7	29a 8m
G1	8	5a 0m	G2	8	8a 0m	G3	8	10a 8m	G4	8	11a 5m	G5	8	22a 5m
G1	9	4a 9m	G2	9	8a 7m	G3	9	10a 4m	G4	9	12a 1m	G5	9	20a 0m
G1	10	5a 0m	G2	10	7a 7m	G3	10	9a 5m	G4	10	12a 0m	G5	10	20a 2m
G1	11	5a 0m	G2	11	7a 3m	G3	11	10a 2m	G4	11	12a 3m	G5	11	20a 3m
G1	12	5a 6m	G2	12	8a 2m	G3	12	10a 10m	G4	12	11a 2m	G5	12	21a 8m
G1	13	4a 9m	G2	13	7a 0m	G3	13	10a 5m	G4	13	11a 2m	G5	13	22a 0m
G1	14	5a 1m	G2	14	7a 6m	G3	14	10a 3m	G4	14	11a 7m	G5	14	21a 3m
G1	15	4a 10m	G2	15	7a 6m	G3	15	9a 9m	G4	15	11a 6m	G5	15	21a 3m

Fonte: Elaborado pela autora.

Todos os participantes selecionados para o estudo realizaram um teste de percepção audiovisual.

A seguir, será apresentado o teste de percepção aplicado.

4.3 TESTE DE PERCEPÇÃO DE FALA – ESTÍMULOS

Os estímulos investigados no teste de percepção audiovisual foram as fricativas surdas. A classe das fricativas do PB apresenta como características acústicas a alta faixa de frequência (1.200 Hz a 8.000 Hz), baixa intensidade e ruído turbulento (Eilers, 1977; Berti, 2017).

Ao analisar o impacto do sinal acústico na percepção auditiva, observa-se que a faixa de frequência das fricativas (1.200 Hz a 8.000 Hz) não é ideal para percepção, uma vez que frequências acima de 5.000 Hz recebem pouca ou nenhuma amplificação pelo sistema auditivo (Johnson, 1997). Além disso, o ruído turbulento característico dessa classe de sons gera efeitos de mascaramento que dificultam ainda mais sua identificação. Diante desses desafios, pistas puramente acústicas podem não fornecer informações suficientemente claras para a identificação das fricativas, sugerindo que pistas adicionais, de natureza não acústica, podem desempenhar um papel crucial na percepção dessas consoantes.

Essa questão se torna ainda mais relevante no caso das fricativas surdas - /f/, /s/ e /ʃ/ - que não apresentam o dado acústico associado à vibração das pregas vocais. A ausência dessa pista acústica torna essa classe de sons particularmente interessante para investigações perceptuais sob uma perspectiva audiovisual.

Assim, o teste de percepção audiovisual foi constituído por estímulos de natureza auditiva (arquivo de áudio) e visual (vídeo) referentes à produção das fricativas de: /f/, /s/ e /ʃ/ em contexto da vogal /a/. A vogal /a/ foi utilizada como veículo para padronizar a apresentação das fricativas desvozeadas. A apresentação dos estímulos em nível silábico pode favorecer uma análise perceptual mais precisa em nível fonêmico.

Os estímulos do teste foram obtidos a partir de gravações simultâneas de áudio e vídeo de um falante do sexo feminino, nativo do Português Brasileiro. Tais gravações foram editadas em computador, com o uso do *software Openshop*, de modo que cada estímulo foi salvo individualmente em formato .wav para os auditivos, e em formato .mp4 para os visuais.

A partir dos arquivos individuais de áudio e vídeo, os estímulos foram organizados da seguinte forma:

(1) auditivos, que se caracterizam por áudio que apresenta a produção vocal das sílabas-alvo;

(2) visuais, que se caracterizam pelo vídeo relativo à produção das sílabas-alvo;

(3) audiovisuais congruentes, que se caracterizam pela apresentação simultânea dos estímulos auditivo e visual de modo correspondente (Exemplo: áudio [fa] e vídeo correspondente à produção do [fa]);

(4) audiovisuais incongruentes, que se caracterizam pela apresentação simultânea incongruente dos estímulos auditivos e visuais, isto é, não ocorre correspondência entre eles (Exemplo: áudio [fa] e vídeo correspondente à produção de [a]).

Dessa forma, os estímulos foram combinados conforme síntese contida no Quadro 3.

Quadro 3 - Condições de apresentação x estímulos presentes no teste de percepção audiovisual

Condições de apresentação dos estímulos	Somente auditivo	Somente visual	Integração audiovisual	
			Congruentes	Incongruentes
Estímulos	[fa]; [sa] e [ja]	[fã]; [sã] e [jã]	Audio fa_Visual fa Audio sa_Visual sa Audio ja_Visual ja	Audio fa_Visual sa Audio fa_Visual ja Audio ja_Visual fa Audio sa_Visual sa Audio sa_Visual fa Audio sa_Visual ja

Fonte: Elaborado pela autora.

4.4 TESTE DE PERCEPÇÃO DE FALA – PROCEDIMENTO EXPERIMENTAL

O teste de percepção audiovisual refere-se a um teste de identificação, também designado tarefa de escolha forçada. Nesse tipo de teste, um estímulo é apresentado, e o informante deve identificá-lo, escolhendo uma dentre as opções oferecidas (Kluge *et al.*, 2013). Para sua elaboração, utilizou-se o *software OpenSesame* (Mathôt; Schreij; Theeuwes, 2012).

O teste foi composto por duas etapas distintas.

Na primeira etapa – fase treino –, apresentaram-se aos participantes todos os estímulos constituintes do teste: auditivo, visual, audiovisual congruente e

incongruente, por meio de uma apresentação em formato *PowerPoint*. Tal processo permitiu proporcionar familiaridade dos participantes com os estímulos.

Na segunda etapa – fase teste –, com o uso do *software OpenSesame*, apresentaram-se, respectivamente, os estímulos: (1) auditivo (A); (2) visual (V); (3) audiovisual congruente (AV+); e (4) audiovisual incongruente (AV-). Cada estímulo foi apresentado duas vezes e, logo após, foram disponibilizadas três possibilidades de resposta: FA, SA, CHA, e o participante deveria indicar qual foi o estímulo percebido. O estímulo seguinte foi apresentado após a resposta do participante. Ainda nesta etapa, os participantes foram dispostos confortavelmente diante da tela do computador (contendo o *software OpenSesame*), com fones acoplados aos ouvidos, com nível de intensidade acusticamente confortável.

Na condição auditiva de apresentação, os participantes ouviram o estímulo auditivo binauralmente em um nível de 50 dB, com ajustes realizados para mais ou para menos conforme a solicitação de cada participante. Durante a apresentação, era exibida na tela do computador uma imagem congelada do terço inferior da região orofacial, em posição neutra. Em seguida, os participantes precisavam decidir e indicar a sílaba correspondente ao estímulo auditivo apresentado, escolhendo entre três alternativas (FA, SA, CHA) dispostas na tela.

Os participantes foram orientados a teclar a resposta correspondente, após a apresentação do estímulo; cada estímulo auditivo foi apresentado duas vezes.

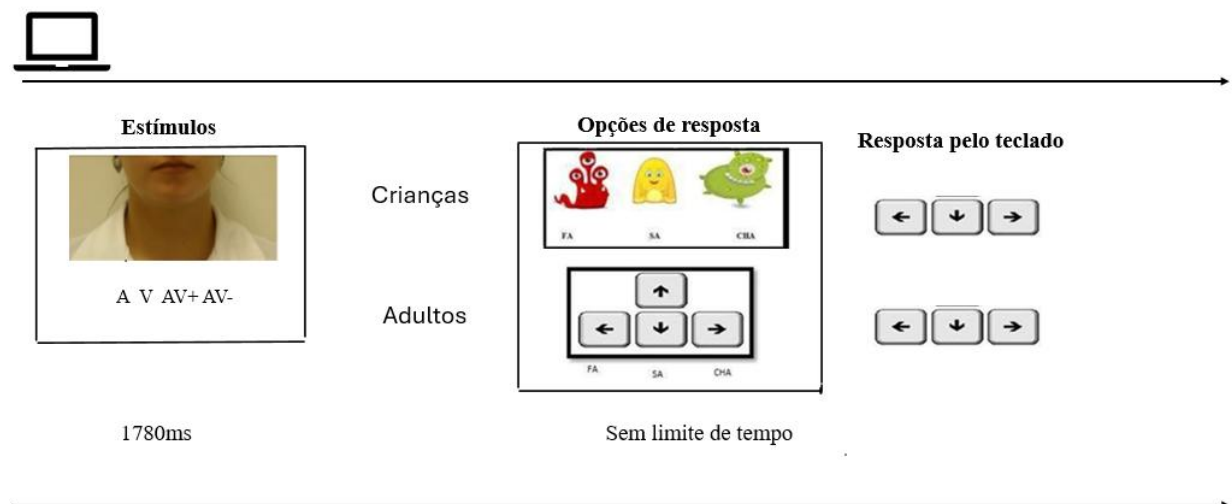
Na condição de apresentação visual, o estímulo foi apresentado apenas visualmente. Trata-se de gravação em vídeo, mostrando o movimento articulatorio correspondente à produção da sílaba-alvo, sem qualquer informação auditiva associada. Similarmente à etapa anterior, após a apresentação do estímulo visual, o participante deveria decidir e indicar, dentre três possibilidades de escolha (FA, SA, CHA) dispostas na tela do computador, qual a sílaba correspondente ao estímulo visual apresentado. Cada estímulo visual foi apresentado duas vezes.

Por fim, na condição audiovisual do teste, apresentaram-se os estímulos audiovisuais, caracterizados por conter tanto a informação auditiva quanto a informação visual, ou seja, uma gravação de vídeo relativa à produção das sílabas foi apresentada simultaneamente à gravação do áudio. Essa etapa foi constituída de estímulos congruentes (em que a informação visual corresponde à informação auditiva) e incongruentes (em que a informação visual não corresponde à informação auditiva). A tarefa foi realizada pelo participante, conforme instruções das etapas 1 e

2. Destaca-se que não foi dada nenhuma instrução explícita para que os participantes direcionassem a atenção à informação visual ou auditiva.

O modo (*design* gráfico) de apresentação para as possibilidades de resposta foi distinto, para aplicação em crianças e em adultos. Para aplicação em adultos, as possibilidades de resposta foram constituídas somente com a representação gráfica dos estímulos. A versão para aplicação em crianças contou com figuras de três personagens, nas quais cada personagem é associado a uma sílaba-alvo e sua representação gráfica correspondente. A Figura 1 ilustra o teste de percepção audiovisual.

Figura 1 - Teste de percepção audiovisual



Fonte: Elaborado pela autora

Em síntese, o teste foi constituído por um total de 15 estímulos: 3 auditivos, 3 visuais, 3 congruentes e 6 incongruentes, e cada estímulo foi apresentado duas vezes, totalizando 30 apresentações. As possibilidades de resposta (FA, SA, CHA) ocorreram após a apresentação de cada estímulo. A duração média do experimento foi de 15 minutos.

4.5 FORMA DE ANÁLISE

Foi realizado um tratamento estatístico descritivo e inferencial dos dados com o uso do *software IBM SPSS Statistics* (versão 2.2). Para análise dos dados, foram consideradas as seguintes medidas:

1. Acurácia (% de acertos) na tarefa de identificação, para estímulos auditivos, visuais e audiovisuais congruentes;

2. Ganho visual, calculado pela fórmula ($[AV+ - AO] \div [100 - AO]$), que permite avaliar o aumento relativo no desempenho audiovisual (AV) devido à adição da informação visual (Leybaert *et al.*, 2014).

3. Efeito McGurk (Ilusão audiovisual), calculada em todas as seis condições de estímulo AV- (Áudio **fa**_Visual **sa**, Áudio **fa**_Visual **ja**, Áudio **ja**_Visual **fa**, Áudio **ja**_Visual **sa**, Áudio **sa**_Visual **fa**, Áudio **sa**_Visual **ja**). Os relatos perceptivos foram classificados em três categorias: auditiva, visual e ilusão. A porcentagem de respostas em cada categoria foi registrada. (Exemplo: Áudio_**fa**_Visual_**ja** – auditiva [fa], visual [ja] e ilusão [sa]).

Para a análise inferencial, foram utilizados os testes Anova de Medidas Repetidas e Kruskal-Wallis. Adotou-se o nível de significância de $p \leq 0,05$.

5 RESULTADOS

5.1 DESEMPENHO UNISSENSORIAL (A E V)

Referindo-se ao primeiro objetivo específico deste estudo – investigar o processamento unissensorial auditivo e visual em função da idade –, será apresentada a análise de comparação entre as condições (A e V) a partir do Teste Anova de Medidas Repetidas, considerando-se a faixa etária como fator intergrupo, e como fator intragrupo, as condições de estímulos (auditivo e visual).

A Tabela 1, a seguir, mostra a porcentagem de acerto dos cinco grupos avaliados nas condições A e V.

Tabela 1 - Porcentagem de acerto relativo às condições auditiva (A) e visual (V) em comparação aos cinco grupos avaliados

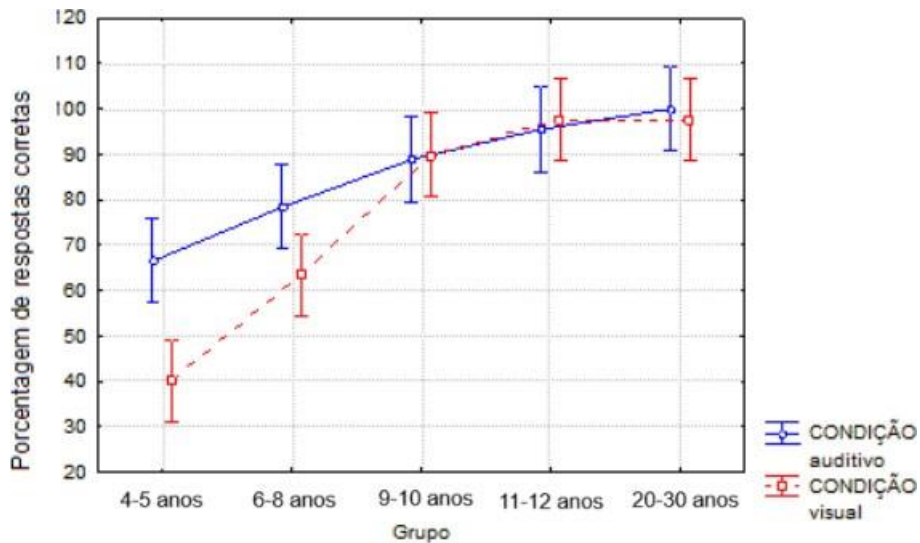
Grupos etários	% de acerto de A	% de acerto de V
	Média (DP)	Média (DP)
4-5 anos	66,5 (± 24,2)	40,0 (± 21,6)
6-8 anos	78,4 (± 23,2)	63,4 (± 23,7)
9-10 anos	88,9 (± 22,3)	89,9 (± 21,6)
11-12 anos	95,5 (± 9,9)	97,7 (± 6,0)
20-30 anos	100 (± 0)	97,7 (± 6,0)

Fonte: Elaborado pela autora.

Legenda: G1 - Grupo 1 (4 – 5 anos e 11 meses), G2 - Grupo 2 (6 – 8 anos e 11 meses), G3 – Grupo 3 (9 – 10 anos e 11 meses), G4 – Grupo 4 (11 – 12 anos e 11 meses), G5 – Grupo 5 (20 – 30 anos e 11 meses), A (estímulo auditivo), V (estímulo visual), DP (Desvio padrão).

O Gráfico 1 ilustra os resultados. A ANOVA de medidas repetidas mostrou efeitos principais significativos para grupo etário ($F(4,70) = 27,066$, $p < ,001$) e condição ($F(1,70) = 11,068$, $p = ,001$), bem como para a interação idade × condição ($F(4,70) = 5,1495$, $p < 0,01$). A interação significativa foi investigada por meio de testes *Post hoc* de *Tukey*, que revelaram que a porcentagem de acertos na condição A foi significativamente maior do que na condição V, mas apenas nas idades de 4–5 e 6–8 anos. Para as idades mais avançadas (9–10, 11–12 e 20–30 anos), não houve diferenças significativas entre as condições A e V.

Gráfico 1 - Comparação com base na interação entre grupos e condições de apresentação (auditiva (A) e visual (V))



Fonte: Elaborado pela autora.

Legenda: G1 - Grupo 1 (4 – 5 anos e 11 meses), G2 - Grupo 2 (6 – 8 anos e 11 meses), G3 – Grupo 3 (9 – 10 anos e 11 meses), G4 – Grupo 4 (11 – 12 anos e 11 meses), G5 – Grupo 5 (20 – 30 anos e 11 meses).

5.2 DESEMPENHO AUDIOVISUAL

Considerando o segundo e terceiro objetivo específico - descrever o ganho visual na percepção de fala em função da idade; e verificar se ocorre o Efeito McGurk (ilusão audiovisual) e, em caso afirmativo, pesquisar o efeito da idade entre os grupos etários -, será apresentada a análise do ganho visual e do efeito McGurk (ilusão audiovisual), de forma respectiva.

Ganho visual

A análise comparativa do ganho visual ocorreu a partir da aplicação da seguinte fórmula: $(AV+ - AO) \div (100 - AO)$. Um valor positivo indica a presença de ganho visual (ou seja, a pista visual foi perceptualmente relevante para a percepção da fala), enquanto um valor zero ou negativo indica a ausência de ganho visual (ou seja, a pista não foi perceptualmente relevante para a percepção da fala).

A Tabela 2, a seguir, mostra a porcentagem de ganho visual dos cinco grupos avaliados.

Tabela 2 - Porcentagem de ganho visual em comparação aos cinco grupos avaliados

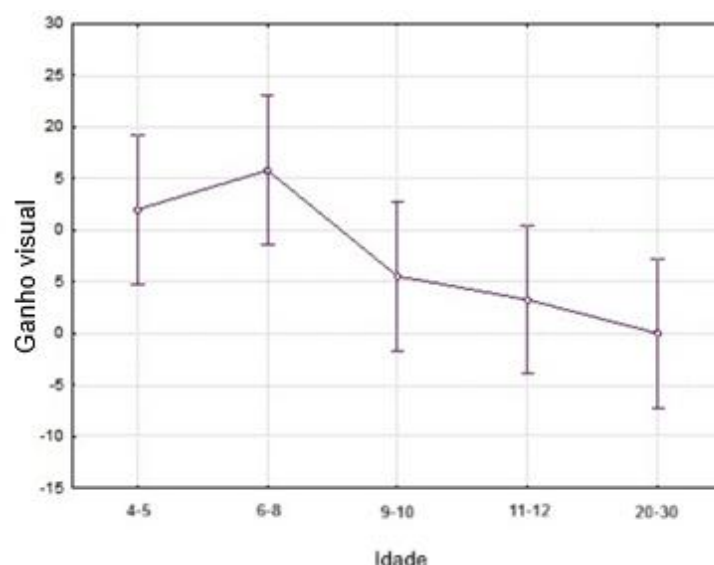
Grupos etários	Ganho visual (AV - A ÷ 100 - A) Média (DP)
4-5 anos	9,5 (± 20,0)
6-8 anos	10,9 (± 27,5)
9-10 anos	4,9 (±10,5)
11-12 anos	2,4 (± 9,4)
20-30 anos	-1 (± 0)

Fonte: Elaborado pela autora.

Legenda: G1 - Grupo 1 (4 – 5 anos e 11 meses), G2 - Grupo 2 (6 – 8 anos e 11 meses), G3 – Grupo 3 (9 – 10 anos e 11 meses), G4 – Grupo 4 (11 – 12 anos e 11 meses), G5 – Grupo 5 (20 – 30 anos e 11 meses); DP (Desvio padrão).

O ganho visual médio ao longo da idade está representado no Gráfico 2. A distribuição da amostra violou os pressupostos de normalidade, portanto, foi utilizado o teste não paramétrico de *Kruskal–Wallis*. Este revelou uma diferença estatisticamente significativa entre os grupos etários ($H(4, N = 75) = 11,74; p = ,002$). O teste *Post hoc* de *Dunn* mostrou um ganho visual maior nos dois grupos mais jovens (4-6 e 6-8 anos). O ganho visual atingiu o pico aos 6–8 anos e, em seguida, reduziu-se significativamente aos 9–10 e 11–12 anos, antes de chegar próximo de zero no grupo adulto de 20–30 anos.

Gráfico 2 - Comparação do ganho visual entre idades



Fonte: Elaborado pela autora.

Efeito McGurk (ilusão audiovisual)

Para investigar o efeito McGurk, as preferências perceptivas dos participantes foram calculadas para condição AV-, isto é, a incidência relativa dos três tipos de resposta (“auditiva”, “visual” e “ilusão”) para os seis tipos de estímulos AV- foi comparada entre os grupos etários.

A Tabela 3 e o Gráfico 3 apresentam as porcentagens dos três tipos de resposta (“auditiva”, “visual” e “ilusão”) entre os grupos etários.

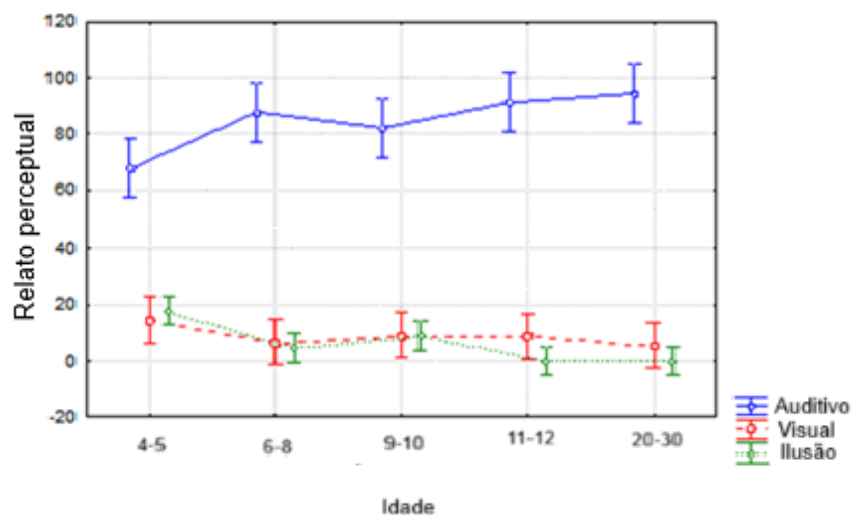
Tabela 3 - Estímulo audiovisual incongruente: preferência perceptiva entre os grupos avaliados

Média (% tipo de resposta)	Auditiva	Visual	Ilusão
	Média (DP)	Média (DP)	Média (DP)
4-5 anos	67,7 (±24,0)	14,5 (±13,8)	17,8 (±17,2)
6-8 anos	87,7 (±18,4)	6,7 (±12,2)	4,5 (±7,8)
9-10 anos	82,1 (±20,4)	8,9 (±8,9)	8,9 (±12,4)
11-12 anos	91,1 (±24,2)	8,9 (±25,2)	0,0 (±0,0)
20-30 anos	94,5 (±12,0)	5,5 (±12,0)	0,0 (±0,0)

Fonte: Elaborado pela autora

Legenda: DP – Desvio padrão

Gráfico 3 - Estímulo audiovisual incongruente: comparação dos tipos de resposta entre os grupos



Fonte: Elaborado pela autora.

A ANOVA de medidas repetidas não mostrou diferença significativa para grupo etário ($F(4,70) = 0,97$, $p = .430$), mas revelou diferença significativa para relato perceptual ($F(2,140) = 392,53$, $p < 0,001$) e uma interação significativa entre grupo etário e relato perceptual ($F(8,140) = 3,5242$, $p < 0,001$). Essa interação significativa foi investigada por meio de análises *Post hoc* de *Tukey*:

Tipo de resposta Auditiva: Houve um aumento significativo nas respostas auditivas entre 4–5 e 6–8 anos, e essa taxa mais alta aos 6–8 anos foi mantida em todas as idades subsequentes (9–10, 11–12 e 20–30 anos, todas as comparações $p < 0,001$).

Tipo de resposta Visual: Houve uma diminuição significativa nas respostas visuais de 4–5 para 6–8 anos, e essa taxa mais baixa aos 6–8 anos foi mantida em todas as idades subsequentes (9–10, 11–12 e 20–30 anos, todas as comparações $p < 0,001$). Além disso, houve algumas flutuações nas idades mais avançadas: verificou-se um aumento significativo na taxa de respostas dos 6–8 para 9–10 e 11–12 anos (ainda assim menor do que aos 4–5 anos), mas as respostas visuais diminuíram dos 9–12 anos para os adultos.

Tipo de resposta Ilusão: Houve uma diminuição significativa nas respostas de ilusão entre 4–5 e 6–8 anos, e essa taxa mais baixa aos 6–8 anos foi mantida em todas as idades subsequentes (9–10, 11–12 e 20–30 anos, todas as comparações $p < 0,001$). Além disso, houve um pequeno pico aos 9–10 anos (9–10 significativamente $> 6–8$, mas ainda $< 4–5$); depois dessa idade, as respostas de ilusão continuaram a diminuir aos 11–12 anos e na idade adulta.

Auditiva versus Influência Visual: Houve significativamente mais respostas auditivas do que visuais e de ilusão em todas as idades, ou seja, as respostas auditivas foram o tipo de resposta dominante ($p < 0,001$, para todas as comparações). Não houve diferença significativa entre respostas visuais e de ilusão em nenhuma idade ($p = .684$).

De forma simples, os resultados mostram que houve um aumento nas respostas auditivas e uma diminuição na influência visual (respostas visuais e de ilusão) entre 4–5 e 6–8 anos; e esses níveis altos (auditivos) e baixos (influência visual) foram mantidos ao longo das idades.

6 DISCUSSÃO

Neste capítulo, serão apresentadas as tendências encontradas neste estudo e suas possíveis hipóteses explicativas. Para tanto, será retomado o objetivo geral, que foi assim definido: Investigar a percepção audiovisual da fala de fricativas desvozeadas em função da idade, em crianças na faixa etária de 4 a 12 anos e em adultos na faixa etária de 20 a 30 anos.

As tendências encontradas e as possíveis hipóteses explicativas serão apresentadas, considerando os objetivos específicos, na seguinte ordem:

I - descrever o processamento unissensorial auditivo e visual em função da idade;

II - descrever o ganho visual na percepção de fala em função da idade;

III - verificar se ocorre o Efeito McGurk (ilusão audiovisual) e, em caso afirmativo, pesquisar o efeito da idade.

Para fins didáticos, a apresentação será estruturada em subtópicos, em conformidade aos objetivos.

6.1 DESEMPENHO UNISSENSORIAL (A E V)

Considerando-se o primeiro objetivo específico – descrever o processamento unissensorial auditivo e visual em função da idade – supunha-se que haveria diferença no desempenho perceptual referente ao processamento unissensorial entre os grupos. Esperava-se que a acurácia perceptual auditiva e visual aumentaria com a idade.

Os resultados confirmaram a primeira hipótese, a acurácia tanto na condição auditiva (A) quanto na condição visual (V) melhorou com a idade. O aumento no desempenho foi mais evidente nas idades mais jovens: os dois grupos etários mais novos (4–5 anos e 6–8 anos) apresentaram desempenho perceptual geral mais baixo (tanto em A quanto em V) em comparação com os grupos mais velhos, (9–10 anos, 11–12 anos) e adultos (20–30 anos). Além desse resultado, foi observado que as crianças dos dois grupos etários mais jovens, de 4 a 8 anos, apresentaram desempenho relativamente baixo na condição visual em comparação à condição auditiva, enquanto a partir de 9–10 anos e nas idades seguintes o desempenho para

condição visual melhorou, alcançando o mesmo nível do desempenho da condição auditiva.

Este resultado é consistente com estudos que investigam a percepção audiovisual da fala em função da idade (Polka, 1994; Sekiyama *et al.*, 2003; Sekiyama; Burnham, 2004). Especificamente, esses estudos sugerem uma tendência de desenvolvimento no uso da integração de pistas com o aumento da idade, de modo que, inicialmente, a percepção da fala seria guiada principalmente pelo componente auditivo, e gradualmente os indivíduos tornariam-se mais sensíveis às pistas visuais, permitindo uma integração audiovisual mais eficaz (Polka, 1994; Ross, 2011; Lalonde; Holt, 2016; Heikila, 2012).

Com base nos nossos resultados, pode-se inferir que, para falantes do PB, a percepção da fala está ancorada principalmente no componente auditivo (parâmetros acústicos) na faixa etária de 4 a 8 anos. A partir dos 9 anos, não foram observadas diferenças significativas entre os grupos etários ou entre as condições (A e V), sugerindo que, a partir dessa idade, a percepção de pistas auditivas (parâmetros acústicos) e visuais (parâmetros articulatórios) começa a ser percebida de forma equivalente.

Além disso, o dado que a acurácia perceptual para os estímulos unissensorial auditivo e visual demonstrou-se em 90% (9–10 anos) até 100% nos adultos, indicaria que as fricativas surdas do PB são identificadas com bastante facilidade com base apenas em informações visuais, ou seja, /fa/, /sa/ e /ja/ do PB seriam visualmente distintas.

Considerando fatores maturacionais e a experiência sociolinguística no desenvolvimento da linguagem, a literatura indica que, dos 0 aos 3 anos, estabelece-se o período de aquisição da língua; dos 3 aos 9 anos, ocorre um refinamento linguístico, incluindo expansão de vocabulário, maturidade articulatória, aumento da competência morfosintático-semântico-pragmática e melhora na inteligibilidade do discurso (Andrade, 1995; Martins; Andrade, 2008). Assim, a partir dos 9 anos, as habilidades linguísticas estariam suficientemente refinadas, de modo que a percepção auditiva e visual da fala se torna equivalente, aproximando-se do nível de 100% de acurácia identificado na idade adulta.

6.2 DESEMPENHO AUDIOVISUAL

Ganho Visual

Considerando-se o segundo objetivo específico – descrever o ganho visual na percepção de fala em função da idade – supunha-se que haveria diferenças no ganho visual entre os grupos etários, marcadas pela maior porcentagem de ganho visual para adultos e em crianças, a partir dos seis anos de idade, uma vez que estudos prévios apontam que o pico de ganho visual ocorreria nessa faixa etária (Fort *et al.*, 2012; Maidment *et al.*, 2015).

A análise comparativa do ganho visual entre os grupos etários não confirmou a segunda hipótese levantada, já que o ganho visual foi maior em crianças mais novas do que em crianças mais velhas e em adultos. Ele aumentou entre 4 e 5 anos, atingindo um pico entre 6 e 8 anos, e depois diminuiu significativamente entre 9–10 e 11–12 anos, antes de cair para cerca de zero nos adultos. Em síntese, o ganho visual foi mais pronunciado em crianças de 4 a 8 anos. Para examinar e discutir esse achado, será realizado um paralelo com as características perceptuais das fricativas e aquisição de linguagem.

Como mencionado inicialmente na seção de apresentação, as fricativas do PB são produzidas com uma constrição parcial entre os articuladores e, à medida que o ar passa por esse canal estreito, gera turbulência significativa, resultando em ruído de fricção (Cagliari, 1981) em uma faixa de frequência de 1.200 Hz a 8.000 Hz, tornando as fricativas os sons mais agudos e de menor intensidade do PB (Haupt, 2008). Essa faixa de frequência não é ideal para percepção (Johnson, 1997), e o ruído de turbulência dificulta ainda mais a percepção (Lawrence *et al.*, 2007).

Assim, parece que pistas puramente acústicas podem não fornecer informações inequívocas para apoiar a identificação ou discriminação das fricativas, sugerindo que pistas visuais podem ser especialmente úteis para a percepção das fricativas, especialmente em crianças pequenas. De fato, as fricativas são uma das últimas classes de sons que os bebês aprendem a discriminar (Eilers, 1977). Considerar essas dificuldades pode afetar a percepção de fricativas em crianças pequenas e aumentar sua dependência das pistas visuais distintivas que as fricativas oferecem, como ocorre neste caso.

Também foi observado que houve uma redução gradual no ganho visual entre 9 e 30 anos. Na aquisição da linguagem, a formação de categorias linguísticas requer o desenvolvimento de múltiplas conexões sinápticas (Zimmer, 2008), e pistas multisensoriais são essenciais para estabelecer caminhos interconectados para o processamento das informações recebidas. O conhecimento fonológico, por exemplo, ocorre por meio de pistas multissensoriais através do: conhecimento das características acústicas e perceptivas dos sons da fala (conhecimento perceptivo); conhecimento das características articulatórias dos sons da fala (conhecimento articulatório); conhecimento de nível superior sobre as maneiras como as palavras podem ser divididas em sons e como os sons podem ser combinados em palavras (conhecimento fonológico de nível superior); e conhecimento das maneiras como a variação na pronúncia pode ser usada para transmitir a identidade social (conhecimento socioindexical) (Munson; Edwards; Beckman; 2005).

Portanto, considerando a cronologia do desenvolvimento, crianças mais novas estão em uma fase ativa de aquisição, com alto grau de formação de conectividade neural para estabelecer categorias linguísticas (Benedict, 1979; Reed, 1992). Nesse sentido, pistas multisensoriais apoiariam a aquisição fonológica. Isso é consistente tanto com o aumento do ganho visual observado em crianças mais novas no início da escolarização quanto com a redução gradual do ganho visual na faixa etária de 9 a 30 anos. Com o aumento da idade, os indivíduos possuem um repertório linguístico mais amplo, permitindo acesso perceptivo mais direto pelo canal sensorial mais robusto: pistas auditivas (Dodd, 2008).

Efeito McGurk (Ilusão audiovisual)

Em relação ao terceiro objetivo específico – verificar se ocorre o Efeito McGurk (ilusão audiovisual) e, em caso afirmativo, pesquisar o efeito da idade – supunha-se que haveria presença do Efeito McGurk (ilusão audiovisual) e, ainda, uma diferença desse efeito entre os grupos etários. Particularmente, esperava-se que os adultos e crianças de maior faixa etária apresentassem maior porcentagem de relato perceptual de ilusão audiovisual.

Na análise do efeito McGurk (Ilusão audiovisual), a terceira hipótese de que o efeito McGurk (ilusão audiovisual) aumentaria com a idade não foi confirmada. Especificamente, o relato perceptual auditivo foi o mais frequente em todos os grupos

etários (4–30 anos), e observou-se um aumento do relato perceptual auditivo com a idade. Esse aumento foi geralmente linear, exceto por uma queda entre 6–8 e 9–10 anos, antes de voltar a crescer até a idade adulta.

Tanto o relato perceptual do estímulo visual quanto o de ilusão diminuíram acentuadamente entre 4-5 e 6–8 anos, e esses baixos níveis foram mantidos (apesar de pequenas flutuações) dos 6–8 anos até a idade adulta.

Comparando o relato perceptual do estímulo auditivo com o visual e de ilusão, observa-se que o canal sensorial preferencial é o auditivo. Isso está em consonância com estudos anteriores que demonstram a dominância do canal auditivo sobre o canal visual (Polka, 1994; Sekiyama *et al.*, 2003). Mesmo considerando que a classe das fricativas apresenta características espectrais desfavoráveis à sua percepção auditiva, o canal auditivo exerce uma influência perceptual mais forte em contextos ambíguos. Um aumento acentuado nas respostas auditivas entre 4-5 e 6–8 anos sugere que essas crianças mais novas são especialmente suscetíveis aos desafios acústicos intrínsecos das fricativas, como baixa intensidade e ruído turbulento (Eilers, 1977). Diante dessas dificuldades, as pistas visuais podem desempenhar um papel crucial no apoio à percepção de fricativas em crianças de 4 a 5 anos.

Considerando o relato perceptual do estímulo visual e de ilusão, as crianças de 4–5 anos apresentaram significativamente mais respostas visuais e de ilusão do que as de 6–8 anos, e essa tendência continuou para todas as idades posteriores, resultado que complementa o aumento do relato perceptual para o estímulo auditivo. Isso sugere que, para a classe das fricativas surdas, as pistas visuais desempenham um papel fundamental em crianças mais novas em idade pré-escolar.

A consideração conjunta dos três tipos de análise de desempenho (Unissensorial, Ganho Visual e efeito McGurk (Ilusão audiovisual)) fornece um panorama geral da percepção audiovisual da fala em fricativas desvozeadas em falantes do PB ao longo da idade. Considerando os dados de desempenho unissensorial, os dois grupos etários mais jovens, de 4 a 8 anos em conjunto, apresentaram um desempenho relativamente baixo na identificação de fricativas na condição (V) em comparação com a identificação em condição (A), mas, a partir dos 9–10 anos, o desempenho na condição (V) melhorou até atingir o mesmo nível do desempenho da condição (A), e essa equivalência se manteve com o avanço da idade. Tomados ao pé da letra, esses resultados unissensoriais poderiam levar à expectativa

de que crianças brasileiras pequenas pouco utilizam informação visual na identificação de fricativas. No entanto, não foi isso o indicado através das análises do ganho visual e efeito McGurk (ilusão audiovisual), os quais demonstraram-se mais elevados nas idades mais jovens (4 a 8 anos).

Esse dado indica a complexidade da percepção audiovisual, que seria mais complexa ou, pelo menos, diferente da soma dos índices perceptuais (A e V), como é observado no efeito McGurk original, em que o estímulo auditivo /ba/ sobreposto ao estímulo visual /ga/ é predominantemente percebido como “da” (McGurk; MacDonald, 1976), o que sugere a natureza integrativa da percepção de fala.

Nosso estudo demonstrou para as análises audiovisuais que o ganho visual nas respostas aos estímulos AV+ foi alto aos 4–5 anos, aumentou até um pico entre 6–8 anos, depois diminuiu acentuadamente aos 9–10 e 11–12 anos, antes de cair para perto de zero na idade adulta. A análise do efeito McGurk mostrou que tanto respostas visuais quanto ilusórias foi alta aos 4–5 anos, depois diminuiu acentuadamente entre 4–5 e 6–8 anos, mantendo-se nesse nível mais baixo (apesar de pequenas flutuações) dos 6–8 anos até a vida adulta. Assim, o uso da informação visual – seja para reforçar a percepção em A (ganho visual) em condições AV+ ou para interferir na percepção em A (influência visual) em condições AV- – na percepção audiovisual da fala já era elevado em crianças pequenas em idade pré-escolar. O que poderia causar isso?

É importante considerar o impacto do início da instrução de leitura e do desenvolvimento da alfabetização como um fator influente na percepção da fala. No início da leitura, as crianças são expostas aos grafemas (enunciação gráfica da língua), que têm uma relação com os fonemas (segmentos sonoros da língua), mas, como os grafemas são convenções arbitrárias, as relações grafema-fonema precisam ser aprendidas (Burnham, 2003; Horlyck; Reid; Burnham, 2012; Guilherme et al., 2023). Esse processo é difícil, independentemente da classe de fonema, mas essa dificuldade pode ser agravada considerando as regras ortográficas das fricativas surdas, especialmente /s/ e /ʃ/.

No português brasileiro, as fricativas surdas /s/ e /ʃ/ apresentam muitas possibilidades de grafemas, por exemplo: para /s/, <s> em “pasta”, <ç> em “caçar”, <x> em “texto”; e para /ʃ/, <x> em “xícara” e <ch> em “chapéu”. O alto ganho visual e o grande relato de ilusão audiovisual em crianças em idade pré-escolar de 4 a 5 anos, seguido de um aumento adicional do ganho visual após o início da escolarização, até os 6-8 anos, sugerem essa dificuldade; a introdução das crianças ao ambiente escolar

e o início da instrução de leitura e alfabetização impactariam a percepção da fala, de modo que, ao se depararem com a aprendizagem de associar grafemas arbitrários a fonemas de sons semelhantes, informações visuais mais distintas para esses fonemas seriam utilizadas. Dessa forma, pistas visuais exerceriam uma influência extra importante na aprendizagem das regras fonema-grafema (especialmente para as fricativas).

Observamos como limitações deste estudo a ausência de ruído adicional no teste de percepção, bem como a restrição da investigação da percepção audiovisual à classe das fricativas. Quanto ao estudo ser restrito à classe das fricativas, sugere-se, para estudos futuros, a investigação da percepção audiovisual para as demais classes fonológicas do PB, já que, em termos de aquisição perceptual, estudos apontam uma gradação de aquisição entre as classes fonológicas do PB (Berti, 2017). Analisar a percepção audiovisual, considerando as demais classes consonantais, acarretará ganhos para o entendimento da percepção audiovisual em função do inventário fonológico do PB, pouco investigado no panorama nacional e internacional.

Em relação ao uso de ruído nos testes de percepção audiovisual, é importante considerar que, na vida cotidiana, os indivíduos estão expostos a ambientes nos quais não há controle sobre sons concorrentes da fala. Consequentemente, as interações comunicativas ocorrem em contextos com múltiplos estímulos competindo com o sinal de fala. Dessa forma, seria relevante que estudos futuros considerassem a adição de ruído aos estímulos utilizados nos testes de percepção audiovisual.

7 CONSIDERAÇÕES FINAIS

Pode-se concluir que a integração audiovisual na percepção de fricativas em falantes do PB já ocorre aos 4 anos de idade. Além disso, o maior ganho visual surge entre 4 e 8 anos, com pico entre 6 e 8 anos e declínio a partir desse ponto. Nessa idade precoce, além do alto ganho visual, observa-se um desempenho perceptivo relativamente baixo na via unissensorial visual (V). Assim, na percepção audiovisual da fala, as informações auditivas e visuais são complementares, em vez de simplesmente somadas ($A + V$). Dessa forma, parece que os aprendizes jovens falantes do PB superam possíveis obstáculos à percepção da fala e à aquisição da linguagem (a relativa dificuldade de perceber fricativas e a dificuldade associada em aprender as regras fonema-grafema das fricativas) ao recorrer à informação visual. O resultado é o uso otimizado das pistas visuais salientes e distintivas das fricativas surdas em falantes do PB, em conjunto com pistas auditivas e, uma vez superado esse obstáculo, observa-se uma redução da influência da pista visual ao longo das faixas etárias.

Acredita-se que os achados deste estudo fornecem uma contribuição interdisciplinar, servindo como base teórica na área linguística e fonoaudiológica. Especificamente, no campo da linguística, os resultados contribuem para a fundamentação teórica que apoia a natureza audiovisual e desenvolvimental da percepção da fala, especialmente no caso de fricativas que apresentam sinais auditivos semelhantes, mas pistas visuais distintas. Além disso, os achados sugerem que o conhecimento fonológico é heterogêneo, abrangendo informações de diversas fontes. No que se refere à fonoaudiologia, os resultados deste experimento apresentam uma importante implicação clínica – que as pistas visuais podem ser utilizadas ou enfatizadas nos planos terapêuticos de reabilitação fonológica em Transtorno de Sons da Fala.

REFERÊNCIAS

- ANDRADE, M. S. Aquisição de linguagem. São Paulo: [Editora], 1995.
- BARUTCHU, A. *et al.* Audiovisual integration in noise by children and adults. **J. Exp. Child Psychol.**, v. 105, p. 38–50, 2010.
- BENEDICT, H. Early lexical development: Comprehension and production. **Journal of child language**, v. 6, n. 2, p. 183-200, 1979.
- BERDASCO-MUÑOZ, E.; NAZZI, T.; YEUNG, H. H. Visual scanning of a talking face in preterm and full-term infants. **Developmental Psychology**, v. 55, n. 7, p. 1353–1361, 2019.
- BERTI, L. C. Desempenho perceptivo-auditivo de crianças na identificação de contrastes fônicos. **Alfa: Revista de Linguística**, São José do Rio Preto, v. 61, p. 81-103, 2017.
- BERTI, L. C. Relação entre produção e percepção de fala: coerência com o parâmetro fonético-acústico. **Cadernos de estudos linguísticos**, v. 50, n. 1, p. 45-68, 2008.
- BERTI, L. C.; PAGLIUSO, A.; LACAVA, F. Instrumento de avaliação de fala para análise acústica (IAFAC), baseado em critérios linguísticos. **Revista Sociedade Brasileira de Fonoaudiologia**, São Paulo, v. 14, n. 2, p. 305-314, 2009.
- BERTI, L. C. PERCEFAL: instrumento de avaliação da identificação de contrastes fonológicos. **Audiology Communication Research**, v. 22, 2017.
- BERTI L. C. PERCEFAL – Instrumento de percepção de fala. *In: I Simpósio Internacional do Grupo de Pesquisa “Avaliação da Fala e da Linguagem” – Perspectivas Interdisciplinares em Fonoaudiologia*, Set 29-Out 01. Marília, 2011.
- BERTI, L. C; ROQUE, L. M. R. Desempenho perceptivo-auditivo de crianças na identificação de contrastes fonológicos entre as vogais tônicas. **CoDAS**. Sociedade Brasileira de Fonoaudiologia, v. 25, p. 534-541, 2013.
- BORDEN, G. J.; HARRIS K. S; LAWRENCE, J. R. **Speech Science Primer**. Physiology, Acoustics, and perception of Speech. Baltimore: Williams & Wilkins, 3ª edição, 1994.
- BOVO, R. *et al.* A. The McGurk phenomenon in italian listeners. **ACTA Otorhinolaryngologica Italica**, v. 29, p.203- 208, 2009.
- BUCCINO, G.; BINKOFSKI, F.; RIGGIO, L. The mirror neuron system and action recognition. **Brain and Language**, v. 89, p. 370-376, 2004.

BURNHAM, D. K.; EARNSSHAW, L. J.; CLARK, J. E. Development of categorical identification of native and non-native bilabial stops: Infants, children and adults. **Journal of Child Language**, v. 18, p. 231–260, 1991.

BURNHAM, D. Language specific speech perception and the onset of reading. **Reading and Writing**, v. 16, n. 6, p. 573-609, 2003.

CANTONI, M. M.; OLIVEIRA, B. G. D.; NEVADO, H. M. Introdução à análise acústica da fala com o Praat. **Texto Livre**, v. 15, 2022.

CARNEY, A. E.; WIDIN, G. P.; VIEMEISTER, N. F. Non categorical perception of stop consonants differing in VOT. **Journal of Acoustical Society of America**, v. 62, p. 961-970, 1977.

CABBAGE, K. et.al. Speech perception differences in children with dyslexia and persistent speech delay. **Speech Communication**, v. 82, p. 14-25, 2016.

CAGLIARI LC. Elementos de fonetica do portugues brasileiro. Campinas. Tese [Livre-Docência] - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem; 1981.

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, p. 858–865, 2009.

CHEN, Y.; HAZAN, V. Language effects on the degree of visual influence in audiovisual speech perception. **International Congress of Phonetic Sciences**, v. 6, n. 10, p. 2177- 2180, 2007.

COLE, R. A.; SCOT, B. Toward a theory of speech perception. **Psychological Review**, v. 81, n. 4, p. 348, 1974.

COLTHEART, M. The MRC psycholinguistic database. **Q. J. Exp. Psychol.**, v. 33A, p. 497–505, 1981.

DEKLE, D. J.; FOWLER, C. A.; FUNNELL, M. G. Audiovisual integration in perception of real words. **Perception & Psychophysics**, v. 51, p. 355-362, 1992.

DESJARDINS, R. N.; WERKER, J. F. 4-month-old female infants are influenced by visible speech. **International Conference of Infant Studies**, Providence RI, 1996.

DODD, B. *et al.* Perception of the auditory-visual illusion in speech perception by children with phonological disorders. **Clinical Linguistics & Phonetics**, v. 22, n. 1., p. 69-82, 2008.

DODD, B.; HERMELIN, Beate. Phonological coding by the prelinguistically deaf. **Perception & Psychophysics**, v. 21, p. 413-417, 1977.

DODD, B. Lip reading in infants: Attention to speech presented in-and out-of synchrony. **Cognitive psychology**, v. 11, n. 4, p. 478-484, 1979.

EIMAS, P. D. *et al.* Speech perception in infants. **Science**, v. 171, n. 3968, p. 303–306, 1971.

EILERS, R. E. Context-sensitive perception of naturally produced stop and fricative consonants by infants. **The Journal of the Acoustical Society of America**, v. 61, n. 5, p. 1321-1336, 1977.

FANT, G. Auditory patterns of speech. **Models for the perception of speech and visual form**, p. 111-125, 1967.

FANT, G. Descriptive analysis of the acoustic aspects of speech. **Logos**, v. 5, p. 3-17, 1962.

FERNALD, A.; SIMON, T. Expanded intonation contours in mothers' speech to newborns. **Developmental psychology**, v. 20, n. 1, p. 104, 1984.

FERREIRA-SILVA, A.; PACHECO, V.; CAGLIARI, L. C. A percepção auditiva e visual das fricativas do Português Brasileiro diante da manipulação do sinal acústico. **Scripta**, v. 21, n. 41, p. 165-183, 2017.

FOWLER, C. Listeners do hear sounds, not tongues. **The Journal of the Acoustical Society of America**, v. 99, n.3, p. 1730-1741, 1996.

FORT, M. *et al.* Audiovisual vowel monitoring and the word superiority effect in children. **International Journal of Behavioral Development**, v. 36, n. 6, p. 457–467, 2012.

FUJISAKI, H.; KAWASHIMA, T. On the modes and mechanisms of perception of speech sounds. **The Journal of the Acoustical Society of America**, v. 47, n. 1A_Supplement, p. 57-57, 1970.

FUSTER-DURAN, A. Perception of conflict in audio-visual speech: an examination across Spanish and German. *In*: STORK, D.G., HENNECKE, M. E. (eds.). **Speechreading by Humans and Machines. NATO ASI Series**, v. 150, Springer, Berlin, Heidelberg, p.135- 143, 1996.

GUILHERME, J. *et al.* Correlação entre características fonético- fonológicas da fala e características ortográficas da escrita em crianças com alterações fonológicas. **Organon**, Porto Alegre, v. 38, n. 76, 2023.

GRIESER, D. L.; KUHL, P. K. Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. **Developmental psychology**, v. 24, n. 1, p. 14, 1988.

HAUPT, C. As fricativas [s], [z], [ʃ] e [ʒ] do português brasileiro. **Letras & Letras**, Uberlândia, v. 24, n. 1, p. 59-71, 2008.

- HAYASHI, Y.; SEKIYAMA, K. Native- foreign language effect in the McGurk effect: a test with Chinese and Japanese. **Auditory-visual speech processing**, v. 4, p. 61-66, 1998.
- HAZAN, V.; BARRETT, S. The development of phonemic categorization in children aged 6–12. **Journal of Phonetics**, v. 28, n. 4, p. 377–396, 2000.
- HEIKKILÄ, J. *et al.* Neural Processing of Congruent and Incongruent Audiovisual Speech in School-Age Children and Adults. **Language Learning**, v. 68, p. 58-79, 2018.
- HOCKLEY, N. S. The development of audiovisual speech perception. Unpublished master's thesis, McGill University, 1994.
- HORLYCK, S.; REID, A.; BURNHAM, D. The relationship between learning to read and language-specific speech perception: maturation versus experience. **Scientific Studies of Reading**, v. 16, n. 3, p. 218-239, 2012.
- HUANG, X. *et al.* Temporary threshold shifts induced by low-pass and high-pass filtered noises in fetal sheep in utero. **Hearing Research**, v. 113, n. 1-2, p. 173-181, 1997.
- JOHNSON, K. Acoustic and auditory phonetics. London: Blackwell, 1997.
- JUSCZYK, P. W.; HOUSTON, D. M.; NEWSOME, M. The beginnings of word segmentation in English-learning infants. **Cognitive psychology**, v. 39, n. 3-4, p. 159-207, 1999.
- JUSCZYK, P. W.; PISONI, D. B.; MULLENIX, J. Some consequences of stimulus variability on speech processing by 2-month-old infants. **Cognition**, v. 43, n. 3, p. 253-291, 1992.
- KLUGE, D. C. *et al.* Percepção de sons de língua estrangeira: questões metodológicas e o uso dos aplicativos Praat e TP. **Revista Letras**, Curitiba, v. 88, n. 1, p. 171- 188, 2013.
- KUHL, P. K.; MILLER, J. D. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. **Science**, v. 190, n. 4209, p. 69-72, 1975.
- KUHL, Patricia K.; MELTZOFF, Andrew N. The intermodal representation of speech in infants. **Infant behavior and development**, v. 7, n. 3, p. 361-381, 1984.
- KUJALA, Teija; PARTANEN, Eino; VIRTALA, Paula M.; WINKLER, István. Prerequisites of language acquisition in the newborn brain. **Trends in Neurosciences**, v. 46, n. 9, p. 726–737, 2023.
- LAARHOVEN, T. *et al.* Audio-visual speech in noise perception in dyslexia. **Developmental Science**, v.21, n.1, p. 1–10, 2018.

LALONDE, K.; HOLT, R. F. Audiovisual speech perception development at varying levels of perceptual processing. **The Journal of the Acoustical Society of America**, v. 139, n. 4, p. 1713-1723, 2016.

LAWRENCE, D. Phonetic variation as a cue to regional identity: an experimental approach. Proceedings of the 2nd Postgraduate and Academic Researchers in Linguistics at York (PARLAY 2014), v. 2, p. 42–63, 2007.

LECANUET, J. Fetal responses to auditory speech stimuli. *In*: SLATER, A. (Org.). **Perceptual development – visual, auditory and speech perception in infancy**. Reino Unido: Routledge, 1998.

LEYBAERT, J. *et al.* Atypical audio-visual speech perception and McGurk effects in children with specific language impairment. **Frontiers in psychology**, v. 5, p. 422, 2014.

LEWKOWICZ, D. J.; HANSEN-TIFT, A. M. Infants deploy selective attention to the mouth of a talking face when learning speech. **Proceedings of the National Academy of Sciences of the United States of America**, v.109, n. 5, p. 1431–1436, 2012.

LIBERMAN, A. M.; MATTINGLY, I. G. The motor theory of speech perception revised. **Cognition**, v. 21, p.1-36, 1985.

LIBERMAN, A. Some results of research on speech perception. **The Journal of the Acoustical Society of America**, v. 27, n. 1, jan. 1957.

LINDBLOM, B. Spectrographic study of vowel reduction. **The journal of the Acoustical society of America**, v. 35, n. 11, p. 1773-1781, 1963.

LOTTO, A. J.; KLUENDER, K. R.; HOLT, L. L. Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). **The Journal of the Acoustical Society of America**, v. 102, n. 2, p. 1134-1140, 1997.

MAIDMENT, D. W. *et al.* Audiovisual integration in children listening to spectrally degraded speech. **Journal of Speech, Language, and Hearing Research**, v. 58, n. 1, p. 61-68, 2015.

MARTINS, J. A. de A.; ANDRADE, M. M. de. **Aquisição de linguagem**. São Paulo: Editora Cortez, 2008.

MARIANI, Benedetta *et al.* Prenatal experience with language shapes the brain. **Science Advances**, v. 9, n. 47, p. ead3524, 2023.

MASSARO, D. W.; COHEN, M. M. Evaluation and integration of visual and auditory information in speech perception. **Journal of Experimental Psychology: human perception and performance**, v. 9, n. 5, p. 753, 1983.

MASSARO, D. W. *et al.* Developmental changes in visual and auditory contributions to speech perception. **Journal of Experimental Child Psychology**, v. 1, p. 93–113, 1986.

MASSARO, D. W. Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale NJ: Erlbaum, 1987.

MASSARO, Dominic W. Perceiving talking faces: From speech perception to a behavioral principle. Mit Press, 1998.

MATHÔT, S.; SCHREIJ, D.; THEEUWES, J. OpenSesame: An open-source, graphical experiment builder for the social sciences. **Behavior Research Methods**, v. 44, n. 2, p. 314-324, 2012.

MCGURK, H.; MACDONALD, J. Hearing lips and seeing voices. **Nature**, Reino Unido, v. 264, n. 5588, p. 746-748, 1976.

MORTON, J.; BROADBENT, D. E. Passive versus Active Recognition Models or Is Homunculus Really Necessary? *In*: WATHEN-DUNN, W. (org.). **Models for the Perception of Speech and Visual Form**. Cambridge, Mass.: M.I.T. Press, 1967. p.103-110.

MUNSON, B.; EDWARDS, J.; BECKMAN, M. E. Phonological knowledge in typical and atypical speech–sound development. **Topics in language disorders**, v. 25, n. 3, p. 190-206, 2005.

NAMASIVAYAM, A. K.; HANSEN-BROWN, A. A. Speech motor control and developmental speech sound disorders: A review of the literature. **International Journal of Speech-Language Pathology**, v. 22, n. 3, p. 253-265, 2020.

NISHIDA, G. **Sobre teorias de percepção da fala**. 2012. 207f. Tese (Doutorado em Letras) – Programa de Pós-Graduação em Letras, Universidade Federal do Paraná, Curitiba, 2012.

ODEN, G. C.; MASSARO, D. W. Integration of featural information in speech perception. **Psychological Review**, v. 85, p. 172-191, 1978.

OJENAN, Ville; KETTUNEN, Ari; SALMELIN, Riitta. Neurocognitive mechanisms of audiovisual speech perception. Helsinki: Helsinki University of Technology, 2005.

ÖHMAN, S. E. G. Numerical Model for Coarticulation, Using a Computer-Simulated Vocal Tract. **The Journal of the Acoustical Society of America**, v. 36, n. 5_Supplement, p. 1038-1038, 1964.

PATAH, L. K. B. Aquisição das fricativas do Português Brasileiro: um estudo acústico-articulatório. 2008. 186 f. Dissertação (Mestrado em Linguística) – Universidade de São Paulo, São Paulo, 2008.

PENIDO, F. A.; ROTHE-NEVES, R. Percepção da fala em desenvolvimento: uma retrospectiva. **Verba Volant**, v. 4, n. 1, p. 117-40, 2013.

PEREIRA, J. R.; REIS, A. M.; MAGALHÃES, Z. Neuroanatomia funcional: anatomia das áreas activáveis nos usuais paradigmas em ressonância magnética funcional. **Acta Médica Portuguesa**, Porto, v. 16, p. 107-116, 2003.

PISONI, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. In: **Perception and Psychophysics**, v. 3, p. 253-260, 1973.

PISONI, D. B.; TASH, J. Reaction times to comparisons within and across phonetic categories. **Perception & Psychophysics**, v. 15, 285-290, 1974.

PONS, Ferran et al. Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment. **Journal of Child Language**, v. 40, n. 3, p. 687-700, 2013.

PONS, F; BOSCH, L.; LEWKOWICZ, D. J. Bilingualism modulates infants' selective attention to the mouth of a talking face. **Psychol Si**, v. 26, n. 4, p. 490- 498, 2015.

REED, V. A. Associations between phonology and other language components in children's communicative performance: clinical implications. **Australian Journal of Human Communication Disorders**, v. 20, n. 2, p. 75-87, 1992.

RIZZOLATTI, G.; ARBIB, M. A. Language within our grasp. **Trends in Neurosciences**, v. 21, p. 188-194, 1998.

ROSENBLUM, L. D.; SCHMUCKLER, M. A.; JOHNSON, J. A. The McGurk effect in infants. **Perception & Psychophysics**, v. 59, n. 3, p. 347-357, 1997.

ROSS, L. A. *et al.* The development of multisensory speech perception continues into the late childhood years. **European Journal of Neuroscience**, v. 33, n. 12, p. 2329-2337, 2011.

RUSSO, I.; BEHLAU, M. *Percepção da fala: análise acústica do português brasileiro*. São Paulo: Lovise, 1994.

SAMS, M. *et al.* McGurk effect in finnish syllables, isolated words, and words in sentences: effects of word meaning and sentence context. **Speech Communication**, v. 26, p. 75-87, 1998.

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

SEKIYAMA, K.; BURNHAM, D. Issues in the development of Auditory- visual speech perception: adults, infants, and children. **Eighth International Conference on Spoken Language Processing**, Korea, 2004

SEKIYAMA, K. *et al.* Development of auditory- visual speech perception in Japanese and English speakers. **Speech Processing**, St. Jorioz, France, p. 43-46, 2003.

SEKIYAMA, K. *et al.* Auditory-visual speech perception examined by fMRI and PET. **Neuroscience Research**, v. 47, n. 3, p. 277-287, 2003.

SEKIYAMA, K.; TOHKURA, Y. Inter-language differences in the influence of visual cues in speech perception. **Journal of Phonetics**, v. 21, p. 427-444, 1993.

SEKIYAMA, K.; TOHKURA, Y. McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. **Journal of the Acoustical Society of America**, v. 90, p. 1797-1805, 1991.

SILVA, A. F.; PACHECO, V.; CAGLIARI, L. C. A percepção auditiva e visual das fricativas do Português Brasileiro diante da manipulação do sinal acústico. **Scripta**, v. 21, n. 41, p. 165-183, 2017.

SILVA, L. M. **Integração de pistas audiovisuais em tarefa de percepção de fala em falantes do Português Brasileiro**. 2024. 123f. Tese (Doutorado em Estudos Linguísticos) – Programa de pós-Graduação em Estudos Linguísticos, Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto, 2023.

SCHNEIDER, P.; RUSSO, I.; BEHLAU, M. Percepção da fala: análise acústica do português brasileiro. São Paulo: Lovise, 1986.

STEVENS, K. N.; BLUMSTEIN, S. E. Invariant cues for place of articulation in stop consonants. **The Journal of the Acoustical Society of America**, v. 64, n. 5, p. 1358-1368, 1978.

SUMMERFIELD, Q. Audio-visual speech perception, lipreading and artificial stimulation. *In*: Hearing science and hearing disorders. **Academic press**. p. 131-182, 1983.

SUMMERFIELD, Q. Use of visual information for phonetic perception. **Phonetica**, v. 36, n. 4-5, p. 314-331, 1979.

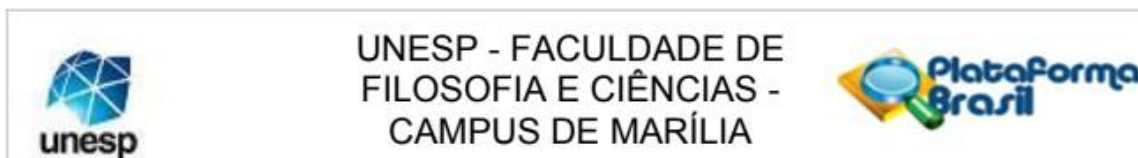
TAITELBAUM-SWEAD, R.; FOSTICK, L. Auditory and visual information in speech perception: A developmental perspective. **Clinical Linguistics & Phonetics**, v. 30, n. 7, p. 531-545, 2016.

TOMALSKI, P. *et al.* Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. **European Journal of Developmental Psychology**, v. 10, p. 611–624, 2013.

WERKER, J. F.; TESS, R. C. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. **Infant Behavior and Development**, v. 7, n. 1, p. 49-63, 1984.

ZIMMER, M. C. Cognição e aprendizagem de L2: uma abordagem conexionista. *In*: MACEDO, A. C.; FELTES, H.; FARIAS, E. M. (orgs.). **Cognição e Linguística: explorando territórios, mapeamentos e percursos**. Porto Alegre / Caxias do Sul: EDIPUCRS / EDUCS, 2008, p. 229-248.

ANEXO A – PARECER CONSUBSTANCIADO DO CEP



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental.

Pesquisador: MAYARA FERREIRA DE ASSIS

Área Temática:

Versão: 2

CAAE: 56136022.6.0000.5406

Instituição Proponente: UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 5.391.404

Apresentação do Projeto:

O projeto de pesquisa "Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental." sob CAAE: 56136022.6.0000.5406, apresentado para análise do CEP em 22/02/2022, apresenta contextualização e revisão teórica sobre as temáticas da pesquisa. De forma geral, está bem redigido, apresentando questões de pesquisa e procedimentos metodológicos que os pesquisadores consideram adequados para responder aos objetivos da pesquisa

Objetivo da Pesquisa:

O objetivo do presente estudo "é comparar o desempenho de crianças e adultos na tarefa de percepção multimodal envolvendo pistas auditivas e visuais nas diferentes faixas etárias. Assumindo que a percepção multimodal de fala se desenvolve gradativamente (BURNHAM et al., 1991) conforme o desenvolvimento das habilidades psicoacústica (SCHENEIDER et al., 1986; THREHUB et al; 1988), visuais (DESJARDINS, 1997; SEKIYAMA e BURHAM, 2008) e linguística (HAZAN e BARRETT, 2000), supomos: (I) que haja efeito significativo para grupo etário, especificamente, que o grupo de adultos apresente desempenho mais acurado e menos laborioso em comparação às crianças, já que, para os adultos, a integração de pistas audiovisuais durante a percepção de fala estaria desenvolvida (MASSARO, 1986; BRANDWEIN, 2011; TAITELBAUM-SWEAD, 2016); (II) que haja efeito significativo para as condições de apresentação, especificamente melhora no

Endereço: Avenida Hygino Muzzi Filho, 737, Prédio da Administração, Sala nº 20

Bairro: Campus Universitário

CEP: 17.525-900

UF: SP

Município: MARILIA

Telefone: (14)3402-1346

E-mail: cep.marilia@unesp.br



UNESP - FACULDADE DE FILOSOFIA E CIÊNCIAS - CAMPUS DE MARÍLIA



Continuação do Parecer: 5.391.404

desempenho da condição visual a partir dos seis anos de idade, refletido tanto pela maior acurácia quanto pelo menor tempo de reação."

Avaliação dos Riscos e Benefícios:

A descrição da coleta de dados não apresenta riscos para os participantes do estudo.

Os benefícios descritos são científicos, como: (1) fornecer informações inéditas sobre o desenvolvimento da integração de pistas na percepção dos contrastes entre as fricativas surdas no PB; (2) contribuir para a discussão, no campo da Linguística, sobre teorias uni vs multimodais de percepção de fala.

Comentários e Considerações sobre a Pesquisa:

A pesquisa será realizada dentro dos princípios de ética em pesquisa com seres humanos.

Considerações sobre os Termos de apresentação obrigatória:

Os documentos obrigatórios foram entregues e estão devidamente assinado pelos responsáveis.

O cronograma está adequado. O Termo de Consentimento Livre e Esclarecido e o termo de Assentimento estão redigidos seguindo a normativa vigente.

Conclusões ou Pendências e Lista de Inadequações:

Aprovado

Considerações Finais a critério do CEP:

O CEP da FFC da UNESP de MARÍLIA, em reunião ordinária de 28/04/2022, após acatar o parecer do membro relator previamente aprovado para o presente estudo e atendendo a todos os dispositivos das resoluções 466/2012, 510/2016 e complementares, bem como ter aprovado o Termo de Consentimento Livre e Esclarecido como também todos os anexos incluídos na pesquisa, resolve APROVAR a pesquisa "Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental."

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1902989.pdf	18/04/2022 10:05:13		Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	termos.pdf	16/04/2022 21:04:11	MAYARA FERREIRA DE ASSIS	Aceito
Declaração de	autorizacoes.pdf	16/04/2022	MAYARA FERREIRA	Aceito

Endereço: Avenida Hygino Muzzi Filho, 737, Prédio da Administração, Sala nº 20

Bairro: Campus Universitário

CEP: 17.525-900

UF: SP

Município: MARÍLIA

Telefone: (14)3402-1346

E-mail: cep.marilia@unesp.br



UNESP - FACULDADE DE
FILOSOFIA E CIÊNCIAS -
CAMPUS DE MARÍLIA



Continuação do Parecer: 5.391.404

concordância	autorizacoes.pdf	21:03:30	DE ASSIS	Aceito
Projeto Detalhado / Brochura Investigador	brochura.pdf	16/04/2022 21:03:00	MAYARA FERREIRA DE ASSIS	Aceito
Declaração de Instituição e Infraestrutura	infraestrutura.pdf	22/02/2022 19:38:35	MAYARA FERREIRA DE ASSIS	Aceito
Cronograma	cronograma.pdf	22/02/2022 19:37:38	MAYARA FERREIRA DE ASSIS	Aceito
Folha de Rosto	folha.pdf	22/02/2022 19:35:54	MAYARA FERREIRA DE ASSIS	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

MARILIA, 05 de Maio de 2022

Assinado por:
MEIRE LUCI DA SILVA
(Coordenador(a))

**ANEXO B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(RESPONSÁVEIS PELAS CRIANÇAS DE 4 A 12 ANOS)**



UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(RESPONSÁVEIS PELAS CRIANÇAS DE 4 A 12 ANOS)

Estamos desenvolvendo uma pesquisa na Universidade Estadual Paulista “Júlio de Mesquita Filho”, intitulada “Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental”, e gostaríamos que participasse da mesma. O objetivo deste estudo é comparar o desempenho de crianças e adultos na tarefa de percepção multimodal envolvendo pistas auditivas e visuais nas diferentes faixas etárias. Participar desta pesquisa é uma opção, e você poderá desistir a qualquer momento.

Destaco que a coleta das crianças será realizada na unidade escolar E.M. Prof. Sebastião Vayego de Carvalho.

Para condução deste estudo, seu filho(a) realizará um teste de identificação de fala, no qual será apresentado, via computador, um estímulo de fala. Após a apresentação, seu filho deverá indicar qual foi o estímulo percebido dentre três possibilidades de resposta.

Exemplo:

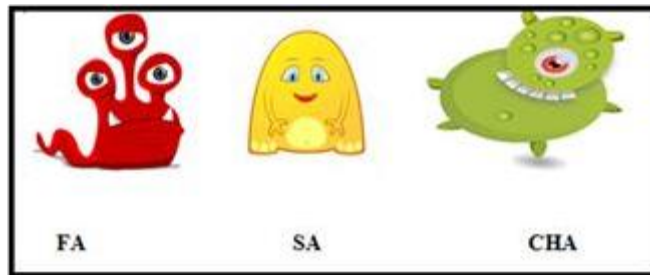
- Apresentação do estímulo de fala 

Figura 1: apresentação do estímulo



- *Frame* do estímulo de fala extraído do teste.

Figura 2: opções de escolha após apresentação do estímulo



Para a obtenção deste termo, gostaríamos que você soubesse do seguinte:

- A coleta das crianças será realizada na unidade escolar E.M. Prof. Sebastião Vayego de Carvalho;
- Participar deste projeto é uma opção, podendo decidir participar ou não;
- Caso você decida não autorizar a participação do seu(sua) filho(a) ou haja desistência a qualquer momento, não haverá prejuízo de qualquer natureza;
- A qualquer momento, você terá a liberdade de buscar, junto aos responsáveis pelo projeto, esclarecimentos de qualquer natureza;
- A participação nesta pesquisa é total e completamente isenta de qualquer ônus financeiro;
- Você receberá uma cópia deste Termo de Consentimento Livre e Esclarecido.

Caso autorize seu filho a participar deste projeto de pesquisa, gostaríamos de informar o seguinte:

- A) Será realizada triagem auditiva por meio de um Checklist, para verificar queixas audiológicas;
- B) A produção de fala será avaliada por meio de nomeação de figuras do “Instrumento de Avaliação de Fala para Análise Acústica” – IAFAC, em contexto da vogal /a/ (Berti, Pagliuso; Lacava, 2009), permitindo, assim, uma triagem rápida entre os participantes quanto à produção dos sons da fala;
- C) A avaliação é segura, não invasiva, não é dolorosa e não utiliza radiação;
- D) Os instrumentos utilizados neste estudo incluem computador (notebook) e fone de ouvido;
- E) A duração aproximada da coleta é de 15 minutos;

- F) Os senhores contribuirão para a descrição sobre o uso de integração de pistas audiovisuais na percepção de sons da fala;
- G) Apesar de os resultados deste estudo não beneficiarem diretamente os participantes, os dados coletados poderão auxiliar no desenvolvimento de programas de atuação fonoaudiológica, assim como aprimorar a avaliação dos pacientes com alterações de fala;
- H) Os resultados desta pesquisa poderão ser divulgados para fins científicos em revistas e congressos especializados na área, com garantia de que a identidade dos participantes será preservada.

Eu, _____, portador (a) do R.G. nº _____ responsável pela criança), autorizo a participação na pesquisa intitulada “Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental”. Declaro ter recebido as devidas explicações sobre a referida pesquisa e concordo que minha desistência poderá ocorrer em qualquer momento da pesquisa, sem quaisquer prejuízos físicos e/ou mentais. Declaro, ainda, estar ciente de que a participação é voluntária e que fui devidamente esclarecido(a) quanto aos objetivos e procedimentos desta pesquisa.

Autorizo,

Nome completo da criança: _____

Data de nascimento: ____/____/____

Data: ____/____/____

ASSINATURA DO RESPONSÁVEL

Responsáveis pela pesquisa:

Profa. Dra. Larissa Cristina Berti Fone: (14) 3402-1324. Endereço: Av. Hygino Muzzi Filho, 737, Caixa Postal 181, Departamento de Fonoaudiologia - Marília/SP - CEP: 17525- 900 larissa.berti@unesp.br

Mayara Ferreira de Assis Fone: (11) 98542-3541. Endereço: Av. Hygino Muzzi Filho, 737, Mirante, Marília - SP, CEP- 17525-900 mayara.assis@unesp.br

ANEXO C - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (PARA PARTICIPANTES ADULTOS DE 20 A 30 ANOS)



UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(Para participantes adultos de 20 a 30 anos)

Estamos desenvolvendo uma pesquisa na Universidade Estadual Paulista “Júlio de Mesquita Filho”, intitulada “Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental”, e gostaríamos que participasse da mesma. O objetivo deste estudo é comparar o desempenho de crianças e adultos na tarefa de percepção multimodal envolvendo pistas auditivas e visuais nas diferentes faixas etárias. Participar desta pesquisa é uma opção, e você poderá desistir a qualquer momento.

Destaco que a coleta de adultos (20 a 30 anos) será realizada no Laboratório de Análise Acústica – LAAC, localizado no Centro de Estudos da Educação e da Saúde — CEES, Unesp, Campus de Marília.

Para a condução deste estudo, você realizará um teste de identificação de fala no qual será apresentado, via computador, um estímulo de fala. Após a apresentação, você deverá indicar qual foi o estímulo percebido dentre três possibilidades de resposta.

Exemplo:

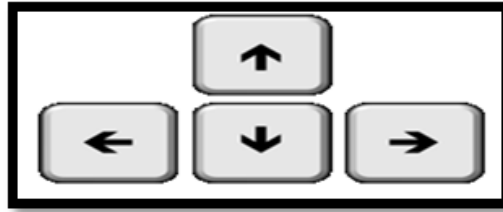
- Apresentação do estímulo de fala 

Figura 1: apresentação do estímulo



- *Frame* do estímulo de fala extraído do teste.

Figura 2: opções de escolha após apresentação do estímulo.



Para a obtenção deste termo, gostaríamos que você soubesse do seguinte:

- A coleta dos participantes adultos (20 a 30 anos) será realizada no Laboratório de Análise Acústica – LAAC, localizado no Centro de Estudos da Educação e da Saúde — CEES, Unesp, Campus de Marília;
- Participar deste projeto é uma opção sua, você pode decidir participar ou não;
- Caso você decida não participar ou desista a qualquer momento, não haverá prejuízo de qualquer natureza;
- A qualquer momento, você terá a liberdade de buscar, junto aos responsáveis pelo projeto, esclarecimentos de qualquer natureza;
- Sua participação nesta pesquisa é total e completamente isenta de qualquer ônus financeiro;
- Você receberá uma cópia deste Termo de Consentimento Livre e Esclarecido.

Caso aceite participar deste projeto de pesquisa, gostaríamos de informar o seguinte:

- A) Será realizada triagem auditiva por meio de um Checklist, para verificar queixas audiológicas;
- B) A produção de fala será avaliada por meio de nomeação de figuras do “Instrumento de Avaliação de Fala para Análise Acústica” – IAFAC, em contexto da vogal /a/ (Berti; Pagliuso; Lacava, 2009), permitindo, assim, uma triagem rápida entre os participantes quanto à produção dos sons da fala;
- C) A avaliação é segura, não invasiva, não é dolorosa e não utiliza radiação;
- D) Os instrumentos utilizados neste estudo incluem computador (notebook) e fone de ouvido;
- E) A duração aproximada do teste de percepção multimodal de fala é de 15 minutos;

- F) Os senhores contribuirão para a descrição sobre o uso de integração de pistas audiovisuais na percepção de sons da fala;
- G) Apesar de os resultados deste estudo não beneficiarem diretamente os participantes, os dados coletados poderão auxiliar no desenvolvimento de programas de atuação fonoaudiológica, assim como aprimorar a avaliação dos pacientes com alterações de fala;
- H) Os resultados desta pesquisa poderão ser divulgados para fins científicos em revistas e congressos especializados na área, com garantia de que a identidade dos participantes será preservada.

Eu, _____,
portador (a) do RG _____, concordo em participar da pesquisa intitulada “Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental”. Declaro ter recebido as devidas explicações sobre a referida pesquisa e concordo que minha desistência poderá ocorrer em qualquer momento da pesquisa, sem quaisquer prejuízos físicos e/ou mentais. Declaro, ainda, estar ciente de que a participação é voluntária e que fui devidamente esclarecido(a) quanto aos objetivos e procedimentos desta pesquisa.

Autorizo,

Nome completo: _____

Data de nascimento: ____/____/____

Data: ____/____/____

Assinatura: _____

Responsáveis pela pesquisa:

Profa. Dra. Larissa Cristina Berti Fone: (14) 3402-1324. Endereço: Av. Hygino Muzzi Filho, 737, Caixa Postal 181, Departamento de Fonoaudiologia - Marília/SP - CEP: 17525- 900 larissa.berti@unesp.br

Mayara Ferreira de Assis Fone: (11) 98542-3541. Endereço: Av. Hygino Muzzi Filho, 737, Mirante, Marília - SP, CEP- 17525-900 mayara.assis@unesp.br

ANEXO D - TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO (PARA CRIANÇAS DE 4 A 6 ANOS DE IDADE)



Universidade Estadual Paulista "Júlio De Mesquita Filho"

TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO

(Para Crianças de 4 a 6 anos de idade)

- Avaliadora explicará oralmente para criança o que será realizado utilizando como apoio as figuras ilustrativas e os quadros informativos, representados abaixo.



Olá, estou te convidando para jogar no computador os "jogos da fala".

Nesses jogos queremos saber como você fala e escuta os sons!!

JOGO I

Você irá falar os nomes das figuras mostradas no computador!



JOGO II

Nesse jogo, você ouvirá um som!!!!



Depois irá apontar
para o som que
você ouviu!



QUERO PARTICIPAR



(~~~~~)

NÃO QUERO PARTICIPAR



()

Nome de quem obteve o assentimento: _____

Data __/__/__

Carimbo da digital do menor

Assinatura do pesquisador

--	--

ANEXO E - TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO (PARA CRIANÇAS DE 8 A 10 ANOS DE IDADE)



Universidade Estadual Paulista “Júlio De Mesquita Filho”

TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO

(Para Crianças de 8 a 10 anos de idade)

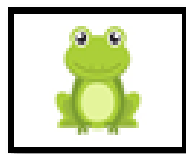
Olá!! Eu gostaria de saber se você quer participar do meu estudo. Seus pais concordaram com a sua participação, mas quero saber se você também aceita participar.

O que acontecerá comigo se eu aceitar?

Queremos saber como você percebemos os sons da fala, caso aceite participar faremos duas atividades.

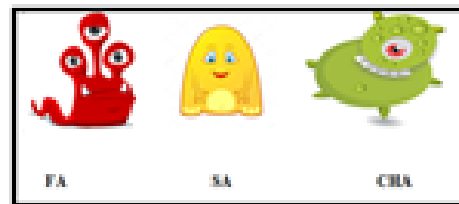
Atividade 1

Nessa atividade você irá ver figuras no computador e precisará falar o nome delas:



Atividade 2

Nessa atividade, será apresentado um som e você precisará apontar para o som escutado.



Fica tranquilo!!! Essas atividades não valem nota e seu nome será mantido em segredo. Você participará do estudo somente se quiser. Caso aceite, poderá desistir a qualquer momento.

Aceito participar

(_____)

Não aceito participar

(_____)

CONSENTIMENTO PÓS INFORMADO

Eu _____ aceito participar da pesquisa Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental. Entendi que posso dizer "sim" e participar, mas que, a qualquer momento, posso dizer "não" e desistir e que ninguém vai ficar com raiva de mim. Os pesquisadores tiraram minhas dúvidas e conversaram com os meus responsáveis. Recebi uma cópia deste termo de assentimento e li e concordo em participar da pesquisa.

Data: __/__/__

ANEXO F - TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO (PARA CRIANÇAS DE 11 A 12 ANOS DE IDADE)



Universidade Estadual Paulista "Júlio De Mesquita Filho"

TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO

(Para Crianças de 11 a 12 anos de idade)

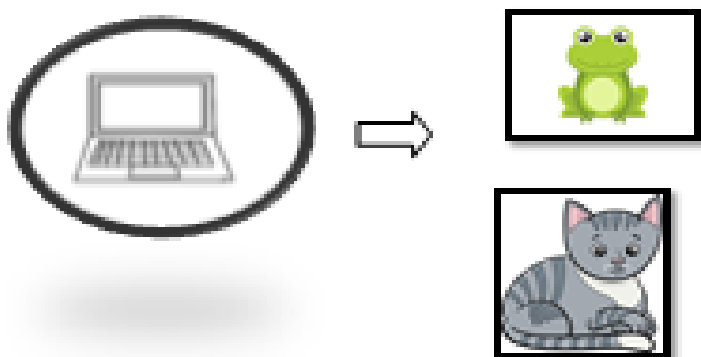
Você está sendo convidado a participar da pesquisa Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental. Seus pais permitiram que você participe.

Nesta pesquisa, queremos saber sobre como ocorre a percepção da fala. Para isso realizaremos dois testes:

Teste 1:

No computador serão apresentadas algumas figuras e você precisará falar os nomes.

Exemplo:



Teste 2:

Neste teste, será apresentado alguns sons de fala e você precisará indicar o que escutou.

Exemplo:



- Você só precisa participar da pesquisa se quiser;
- Caso decida participar, não terá nenhum problema se desistir;
- Ninguém saberá que você está participando da pesquisa; não falaremos a outras pessoas, nem daremos a estranhos as informações que você nos der;

Aceito participar: () SIM

() NÃO

CONSENTIMENTO PÓS-INFORMADO

Eu _____ aceito participar da pesquisa Integração de pistas audiovisuais na percepção de fala: análise desenvolvimental. Entendi que posso dizer "sim" e participar, mas que, a qualquer momento, posso dizer "não" e desistir e que ninguém vai ficar com raiva de mim. Os pesquisadores tiraram minhas dúvidas e conversaram com os meus responsáveis. Recebi uma cópia deste termo de assentimento e li e concordo em participar da pesquisa.

Data: __/__/__

Assinatura do menor

Assinatura do pesquisador