



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”

CAMPUS DE MARÍLIA

**PROGRAMA DE PÓS-GRADUAÇÃO EM FONOAUDIOLOGIA NÍVEL DE
MESTRADO**

MAYARA FERREIRA DE ASSIS

**INTEGRAÇÃO DE PISTAS EM TAREFA DE PERCEPÇÃO DE FALA DE
FRICATIVAS DESVOZEADAS EM CRIANÇAS COM E SEM TRANSTORNO
FONOLÓGICO**

MARÍLIA

2021

MAYARA FERREIRA DE ASSIS

**INTEGRAÇÃO DE PISTAS EM TAREFA DE PERCEPÇÃO DE FALA DE
FRICATIVAS DESVOZEADAS EM CRIANÇAS COM E SEM TRANSTORNO
FONOLÓGICO**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Fonoaudiologia da Faculdade de Filosofia e Ciências “Júlio de Mesquita Filho” UNESP –Campus de Marília (SP) – para obtenção do título de mestre em Fonoaudiologia.

Área de concentração: Distúrbios da Comunicação Humana.

Orientadora: Prof^{fa}. Dr^a. Larissa Cristina Berti.

Marília

2021

A848i

Assis, Mayara Ferreira de

Integração de pistas em tarefa de percepção de fala de fricativas desvozeadas em crianças com e sem transtorno fonológico / Mayara Ferreira de Assis. -- Marília, 2021
67 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília
Orientadora: Larissa Cristina Berti

1. Percepção da fala. 2. Percepção visual. 3. Percepção auditiva. 4. Distúrbios fonológicos. 5. Fonologia. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

MAYARA FERREIRA DE ASSIS

**INTEGRAÇÃO DE PISTAS EM TAREFA DE PERCEPÇÃO DE FALA DE
FRICATIVAS DESVOZEADAS EM CRIANÇAS COM E SEM TRANSTORNO
FONOLÓGICO**

Dissertação apresentada ao Programa de Pós-Graduação em Fonoaudiologia, Área de Concentração Distúrbios da Comunicação Humana, da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, para obtenção do título de mestre em Fonoaudiologia.

BANCA EXAMINADORA

Orientadora: _____

Profa Dra Larissa Cristina Berti – Universidade Estadual Paulista – Unesp – Marília – SP.

2º Examinador: _____

Dra Geovana Carina Neris Soncin Santos – Universidade Estadual Paulista – Unesp – Marília – SP.

3º Examinador: _____

Dra Márcia Keske-Soares – Universidade Federal de Santa Maria – UFSM – Santa Maria – RS.

Marília, 06 de maio de 2021

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

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

AGRADECIMENTOS

Agradecimento especial aos meus pais, Rose e Adilson, por me apoiarem em cada momento. Agradeço imensamente ao incentivo e dedicação. Gratidão e amor eterno!

À minha orientadora, **Prof^a. Dr^a. Larissa Cristina Berti**, pela orientação e incontáveis momentos de aprendizagem. Obrigada por contribuir em meu crescimento pessoal e profissional. Foi uma honra tê-la como orientadora.

Aos **membros do Laboratório de Análise Acústica e Articulatória**, especialmente à Lídia da Silva e Elissa Cremasco, pelo apoio na elaboração do instrumento de avaliação e coletas deste estudo. Aos queridíssimos colegas de laboratório, Cássio, à Jhulya, à Luiza, à Thalia, à Grazielly. Sou grata a todos pelos enriquecedores momentos compartilhados.

Aos membros do **Grupo de Pesquisa Estudos sobre a Linguagem**.

Ao **Programa de Pós-Graduação em Fonoaudiologia da UNESP** (Campus de Marília), pela atenção e prontidão durante as dúvidas e dificuldades ao longo do mestrado.

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

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

À **Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP)**, pelo **financiamento concedido (Processo FAPESP nº 2019/12749-2)**, por ter **acreditado no nosso trabalho e por ter possibilitado dedicação exclusiva para execução do mesmo**.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

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

“Ninguém caminha sem aprender a caminhar, sem aprender a fazer o caminho caminhando, refazendo e retocando o sonho pelo qual se pôs a caminhar.”

(Paulo Freire)

RESUMO

Introdução: McGurk e McDonald inauguraram uma nova perspectiva para o entendimento sobre a percepção de fala, reportando a importância da informação visual associada às pistas de ordem acústica nesse processo. Desde então, estudos apontam para a natureza multimodal no processo de percepção de fala. Embora se verifique a presença de estudos voltados para identificação de integração de pistas no processo de percepção de fala, ainda são poucos os estudos que avaliam a integração de pistas audiovisuais em crianças com Transtorno Fonológico (doravante TF). **Objetivo:** O objetivo do presente estudo foi comparar o desempenho de crianças com e sem TF na tarefa de percepção multimodal envolvendo pistas auditivas e visuais em fricativas desvozeadas. Assumindo que o transtorno fonológico poderia estar associado a uma capacidade prejudicada de integrar a informação fonológica de duas fontes de modalidade, supõe-se que crianças com TF apresentem pior desempenho e maior tempo de reação na tarefa de integração de pistas auditivas e visuais em relação às crianças sem TF. **Método:** Este estudo foi aprovado pelo Comitê de Ética em Pesquisa sob o Protocolo nº15416819.6.0000.5406. Participaram do estudo 30 crianças, subdivididas em dois grupos: Grupo experimental (GE), composto por 15 crianças diagnosticadas com transtorno fonológico com acometimento na classe das fricativas; e Grupo controle (GC), composto por 15 crianças com desenvolvimento típico de fala. Os estímulos apresentados consistiram em gravações em áudio e vídeo da produção de sílabas com padrão silábico simples: /f/, /s/ e /ʃ/ no contexto da vogal /a/. Foi utilizado o software Opensesame⁽⁵⁾ para a elaboração da tarefa de percepção multimodal, envolvendo a apresentação de estímulos apenas por via auditiva; (2) apresentação dos estímulos apenas por via visual; e (3) apresentação simultânea dos estímulos auditivos e visuais de forma congruente; (4) apresentação simultânea dos estímulos auditivos e visuais de maneira incongruente. Na análise, foram consideradas as porcentagens de respostas corretas e o respectivo tempo de reação nas diferentes condições de apresentação dos estímulos, enquanto para os estímulos incongruentes, foi considerada a porcentagem de preferência perceptiva e o respectivo tempo de reação. **Resultados:** Considerando a porcentagem de respostas corretas, a Anova de Medidas Repetidas mostrou efeito estatisticamente significativo para o grupo, para as condições de apresentação dos estímulos, assim como para interação entre grupo*condição. O teste Post-hoc de Duncan's mostrou pior acurácia para a condição de apresentação visual isolada somente para o (GE). Na análise do estímulo audiovisual incongruente, a Anova de Medidas Repetidas mostrou efeito significativo para as condições e para a interação entre grupo*condições. O teste Post Hoc de Duncan's mostrou, para o (GE), menor porcentagem de respostas para condição auditiva e maiores porcentagens para a condição visual e ilusão perceptual. Em relação ao tempo de reação, a Anova de Medidas Repetidas mostrou efeito significativo para grupo e condição. Crianças com TF apresentaram maior tempo de reação em relação às crianças típicas. O teste Post-Hoc de Duncan's mostrou maior tempo de reação para condição visual em ambos os grupos. **Conclusão:** Crianças com TF apresentaram pior desempenho e maior tempo de reação na tarefa de percepção multimodal, sugerindo dificuldades na integração das informações fonológicas de fontes bimodais. A condição visual mostrou-se como a mais difícil para ambos os grupos.

Palavras-chave: Percepção Multimodal. Transtorno fonológico. Identificação Sonora.

ABSTRACT

Objective: This study aimed to compare the integration of auditory and visual cues in the multimodal perception task of voiceless fricatives in children with and without PD. Assuming that phonological disorder could influence how children perform the integration of auditory-visual cues, the hypothesis is that children with PD have worse performance in the integration of auditory and visual cues since the PD could be associated with an impaired capacity to integrate phonological information from two-mode sources. **Methods:** This study was approved by the Research Ethics Committee (REC) of the São Paulo State University “Júlio de Mesquita Filho” (Unesp), Campus of Marília, under protocol nº 15416819.6.0000.5406. Thirty children were tested, being 15 diagnosed with phonological disorder involving problems to produce fricatives (GE) and 15 with typical speech development (GC). The presented stimuli consisted of audio and video recordings of the production of syllables with simple syllabic pattern: /f/, /s/ and /ʃ/ in the context of the vowel /a/. The multimodal perception task, using the Opensesame software, involved the stimuli presentation only by auditory pathway; (2) the stimuli presentation only by visual pathway; and (3a) simultaneous presentation of the auditory and visual stimuli congruently; (3b) simultaneous presentation of the auditory and visual stimuli incongruently. In the analysis were considered the accuracy (% of correct answers), % of perceptual preference in incongruent audiovisual condition, and the reaction time through different conditions of the stimuli presentation. **Results:** The repeated measures Anova showed a statistically significant for group, for conditions as well as for interaction between group*conditions. Tukey’s Post-hoc test showed worst accuracy for the PD mainly in the only visual pathway compared to typical children. In the analysis of audiovisual stimuli incongruent, the repeated measures Anova showed a significant main effect for conditions and interaction between group*conditions. Tukey’s Post-Hoc showed that both groups based on the auditory pathway in the incongruent condition, but typically children presented the highest percentage of the correct response. Concerning the reaction time, repeated measures Anova showed a statistically significant for group and conditions. Children with PD presented a longer reaction time than typical children. Tukey’s Post-hoc showed a longer reaction time for the visual condition in both groups. **Conclusion:** Children with PD presented worse performance and longer reaction time in the multimodal perception task, suggesting difficulties in integrating phonological information from two-mode sources.

Keywords: Multimodal Perception. Phonological disorder. Sound identification.

LISTA DE FIGURAS

Figura 1 – Curva acústica dos sons da fala.....	20
Figura 2- Script :Percepção multimodal fricativas.....	43
Figura 3- Apresentação auditiva do fonema /fa/ - com imagem da região orofacial congelada.....	44
Figura 4- Opções de escolha após apresentação da condição auditiva experimental.....	44
Figura 5- Vídeo com apresentação do movimento articulatorio do fonema /fa/ sem qualquer informação auditiva associada.....	45
Figura 6- Opções de escolha após apresentação da condição experimental visual.....	45

LISTA DE GRÁFICOS

Gráfico 1- Comparação entre a acurácia perceptual geral do grupo experimental (GE) e grupo controle (GC)	47
Gráfico 2- Comparação entre acurácia perceptual das condições de apresentação dos estímulos do grupo experimental (GE) e grupo controle (GC).....	48
Gráfico 3- Comparação entre o tempo de reação do grupo experimental (GE) e grupo controle (GC).....	49
Gráfico 4- Estímulo audiovisual incongruente: Comparação relativa à preferência perceptual do grupo experimental (GE) e grupo controle (GC).....	50
Gráfico 5- Comparação entre o tempo de reação, para estímulo incongruente, do grupo experimental (GE) e grupo controle (GC).....	51

LISTA DE QUADROS

Quadro 1- Caracterização dos Sujeitos do Grupo Experimental (GE).....	41
Quadro 2- Caracterização dos Sujeitos do Grupo Controle (GC).....	41
Quadro 3- Condições de apresentação x Estímulos apresentados.....	43

LISTA DE TABELAS

Tabela 1 – Porcentagem de acerto relativa ao grupo experimental (GE) e grupo controle (GC).....	46
Tabela 2 - Tempo de reação referente ao grupo experimental (GE) e grupo controle (GC).....	47
Tabela 3 - Estímulo audiovisual incongruente: preferência perceptual relativa ao grupo experimental (GE) e grupo controle (GC).....	48
Tabela 4 - Tempo de reação relativo ao estímulo audiovisual incongruente.....	50

SUMÁRIO

APRESENTAÇÃO	14
1 INTRODUÇÃO.....	17
1.1 FONEMAS FRICATIVOS.....	17
1.2 Características articulatórias.....	17
1.3 Características acústicas.....	18
1.4 Características auditivas.....	19
1.5 PERCEPÇÃO DE FALA COMO UM EVENTO MULTIMODAL.....	21
1.5.1 Influência da língua no processo de integração de pistas audiovisuais na percepção da fala.....	25
1.5.2 Influência da idade no processo de integração de pistas audiovisuais na percepção da fala.....	29
1.5.3 Influência cultural no processo de integração de pistas audiovisuais na percepção da fala.....	31
1.6. INTEGRAÇÃO DE PISTAS NA PERCEPÇÃO DE FALA EM CRIANÇAS COM TRANSTORNO FONOLÓGICO.....	32
1.6.1 Caracterização do Transtorno Fonológico.....	32
1.6.2 Percepção Multimodal em crianças com Transtorno Fonológico.....	34
2 OBJETIVO E HIPÓTESES.....	37
3 METODOLOGIA.....	38
3.1 Aspectos Éticos.....	38
3.2 Amostra.....	38
3.3 Estímulos.....	41
3.4 Procedimento Experimental.....	41
3.5 Análise Estatística	45

4 RESULTADOS.....	46
5 DISCUSSÃO.....	51
6 CONCLUSÃO.....	55
REFERÊNCIAS BIBLIOGRÁFICAS	57
ANEXO A - DOCUMENTO DE APROVAÇÃO DO CEP DA FACULDADE DE FILOSOFIA E CIÊNCIA – UNESP/MARÍLIA.....	62
ANEXO B - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCL).....	65
ANEXO C- LISTA DE PALAVRAS UTILIZADAS, DURANTE A AVALIAÇÃO DA PRODUÇÃO DE FALA, PARA SELEÇÃO DOS SUJEITOS QUE COMPÔS O (GC) E (GE).....	66

APRESENTAÇÃO

Esta dissertação é derivada dos resultados e reflexões que vimos desenvolvendo desde a iniciação científica (finalizado em 2018). O projeto desenvolvido neste período, com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (Processo 2017/04968-0), intitulado “Desempenho Perceptivo-auditivo em crianças com e sem transtorno fonológico na classe das oclusivas”, foi o início do meu direcionamento para a investigação voltada à percepção de fala, mais especificamente na investigação da relação entre percepção de fala e transtorno fonológico.

Em junho de 2017, demos início ao desenvolvimento do projeto de iniciação científica. Naquela ocasião, (a) comparamos o desempenho perceptivo-auditivo de crianças com e sem transtorno fonológico em tarefa de identificação sonora, envolvendo os fonemas oclusivos do português brasileiro e; (b) 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 transtorno fonológico apresentam pior acurácia 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 uma evidência de que crianças com transtorno fonológico também poderiam apresentar dificuldades no domínio perceptivo-auditivo (CABBAGE, 2016).

Os resultados obtidos apontaram para a necessidade de aprofundarmos a investigação sobre a dificuldade no domínio perceptivo de crianças com TF, a fim de analisar se a natureza de tal dificuldade perceptual estaria restrita ao desempenho perceptivo-auditivo ou se a dificuldade perceptual estaria relacionada à integração de pistas auditivas e visuais, uma vez que tal achado auxiliaria na clínica fonoaudiológica em termos de técnicas e estratégias no processo de reabilitação fonológica.

Quanto à integração de pistas audiovisuais, embora se verifique, em âmbito internacional, a presença de estudos voltados para a identificação de integração de pistas no processo de percepção de fala, não há, na literatura nacional, até onde o levantamento bibliográfico realizado possibilitou chegar, estudos que discorram

sobre percepção multimodal em crianças típicas e/ou com TF, registrando e/ou comparando a existência da integração de informações audiovisuais para percepção de fala no português brasileiro (PB). Essa lacuna propiciou o desenvolvimento de estudos voltados para a investigação do desempenho em tarefa de percepção multimodal de adultos e crianças com e sem alteração da fala, que vêm sendo desenvolvidos pelo Grupo de Pesquisa e Estudos sobre Linguagem, sob supervisão da Prof^a. Dr^a. Larissa Cristina Berti. Atualmente, a doutoranda Lídia Maurício da Silva está investigando a integração de pistas audiovisuais em adultos e crianças com e sem TF, envolvendo a classe das oclusivas.

Nesta dissertação, desejamos comparar o desempenho de crianças com e sem TF em tarefa de percepção multimodal envolvendo pistas auditivas e visuais em fricativas desvozeadas. A investigação envolvendo esta classe fônica justifica-se pelas características acústicas caracterizadas por apresentar altos valores de frequência, que variam entre 1.200 Hz e 8.000 Hz (RUSSO e BEHALU, 1994), e por apresentarem os menores níveis de intensidade no PB (HAUPT, 2008). Tais conjuntos de características acústicas impedem que esses sons sejam beneficiados pelo ouvido humano podendo afetar sua percepção, o que demandaria a ancoragem em pistas de outra ordem, como, por exemplo, as pistas visuais.

Supõe-se que crianças com TF apresentariam pior desempenho na integração de pistas auditivas e visuais, refletida por menor porcentagem de acertos, menor porcentagem de ilusão audiovisual e maior tempo de reação na tarefa perceptual em relação às crianças sem TF.

Esta dissertação foi estruturada do seguinte modo: no primeiro capítulo serão apresentados os aspectos teóricos, abrangendo as seguintes seções: na primeira seção (1.1), os fonemas fricativos de português brasileiro; na segunda seção (1.2), suas características articulatórias; na terceira seção (1.3), suas características acústicas; na quarta seção (1.4), suas características auditivas.

Nas seções seguintes, trataremos do referencial teórico adotado, a saber: (1.5) a percepção de fala como um evento multimodal, com subdivisões: (1.5.1) as influências da língua na integração de pistas audiovisuais na percepção de fala; (1.5.2) a influência da idade no processo de integração de pistas audiovisuais na percepção de fala; (1.5.3) a influência cultural na integração de pistas audiovisuais na percepção de fala.

A partir da seção (1.6), apresentaremos a integração de pistas na percepção da fala em crianças com transtorno fonológico subdividida em duas seções: (1.6.1) caracterização do transtorno Fonológico e (1.6.2) percepção multimodal em crianças com transtorno fonológico.

No segundo capítulo, exporemos a proposta e o objetivo que nortearam a execução do estudo. No terceiro capítulo, abordaremos os aspectos metodológicos, com o delineamento e análise estatística do estudo. Nos capítulos seguintes, quarto e quinto, apresentaremos as informações referentes aos resultados e discussão, respectivamente. Por fim, no sexto capítulo, a conclusão do estudo.

1 INTRODUÇÃO

1.1 FONEMAS FRICATIVOS

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/.

Quando um fonema é fisicamente produzido na fala, sua produção envolve um grande número de características fonéticas, e estudos desta linha (CAGLIARI, 1981; KENT e READ, 2002; LAWRENCE et al., 2007) segmentam tais características a fim de proporcionar uma análise aprofundada de cada aspecto dos sons utilizados na linguagem humana. Essas características são descritas em três tipos: articulatórias, acústicas e auditivas. A seguir, apresentaremos de modo específico as características fonéticas dos sons fricativos do PB.

1.2 Características articulatórias

As características motoras dizem respeito aos movimentos envolvidos nas produções dos fones, avaliados por meio de três parâmetros: modo de articulação, lugar de articulação e vozeamento (CAGLIARI, 1981).

Em relação ao modo de articulação, as fricativas apresentam, no momento de sua produção, obstrução parcial entre os articuladores durante a passagem do ar, que permite sua saída contínua pela cavidade oral. Ao passar por esse espaço estreito entre os articuladores, a corrente aérea sofrerá grande turbulência, que resultará no ruído de fricção. Durante a produção das fricativas, o véu palatino encontra-se elevado (fechado), permitindo que a saída de ar se dê, exclusivamente, pela cavidade oral (CAGLIARI, 1981).

Quanto ao lugar de articulação, no PB, as fricativas são classificadas em três pares: labiodentais [f] e [v] - constrição estreita realizada entre os dentes incisivos superiores e o lábio inferior; alveolares [s] e [z] - constrição estreita realizada entre ponta de língua e região dos alvéolos; e palatoalveolares [ʃ] e [ʒ] - constrição estreita realizada entre a parte anterior da língua e a região anterior do palato. O elemento

contrastivo entre os pares apresentados é a ausência e presença, respectivamente, de vozeamento (surdo/sonoro) (CAGLIARI, 1981; KENT e READ, 2015) .

1.3 Características acústicas

Este tópico refere-se à maneira como os fones se organizam baseados em correlatos físicos, em termos de duração (intervalo de tempo entre um ciclo e outro), frequência (número de ciclos realizados por unidade de tempo, mensurado em Hertz) e intensidade (qualidade relacionada à amplitude, pressão e energia) (RUSSO,1999; KENT e READ, 2015).

Os traços articulatórios essenciais de uma fricativa são: formação de constrição estreita; desenvolvimento de um fluxo aéreo turbulento (movimento do fluxo aéreo se torna altamente complexo); geração de um ruído de turbulência. O ruído resultante apresenta segmentos ruidosos relativamente mais longos em relação às outras classes consonantais, e é esse intervalo extenso de energia que distingue as fricativas como uma classe de sons (KENT e READ; 2015).

No interior desta classe, a duração variará conforme alguns fatores, como local de constrição. Por exemplo, as fricativas palatoalveolares são mais longas que as alveolares, que, por sua vez, são mais longas que as labiodentais. Segundo Haupt (2008), quanto mais posteriorizado for a constrição articulatória, maior será a extensão da corrente de ar, se prolongando, consequentemente, ao longo do tempo.

Em relação à frequência, os valores estão na faixa de 1.200Hz podendo chegar até a 8.000 Hz, sendo estes os sons mais agudos do PB (HAUPT, 2008; RUSSO e BEHLAU;1994). Especificamente, os fones labiodentais [f,v] apresentam uma faixa de frequência ampla, indo de 1.200 Hz a 7.000 Hz ; os fones alveolares [s,z] possuem faixa de frequência entre 4.500 a 8.000 Hz; e os fones palatoalveolares [ʃ,ʒ] apresentam faixa de frequência de 2.500 a 6.000 Hz(RUSSO e BEHLAU, 1994).

No que se a refere à intensidade, esta classe de sons é a mais fraca do PB. Considerando este parâmetro, as fricativas podem ser classificadas como estridentes e não estridentes. O grupo das fricativas estridentes envolve os fones: [s, z,] e [ʃ, ʒ], enquanto o subgrupo não estridente [f,v] são os sons de menor intensidade (KENT e READ, 2015). Ressalta-se que o segundo elemento do par (vozeado, desvozeado) é mais intenso, dado que, de acordo com Haupt, (2008), “a

É seguro dizer, portanto, que o filtro (trato supralaríngeo) atua como uma caixa de ressonância acentuando e atenuando determinadas frequências a depender da configuração do trato (FANT, 1960; LAAD, 2014). O sinal acústico resultante será detectado pela orelha e pode beneficiar ou dificultar a percepção de determinado fone.

A identificação das fricativas ocorre a partir da percepção do componente aperiódico de duração relativamente longa no sinal acústico (LAWRENCE et al., 2007), caracterizado, como já exposto anteriormente, pelos altos valores de frequência, entre 1.200 Hz a 8.000 Hz, e por apresentar os menores níveis de intensidade dentre as classes consonantais do PB (HAUPT, 2008).

Esse conjunto de características acústicas impede que o ouvido humano seja beneficiado pelo som, o que poderia afetar sua percepção. Sabe-se que o ouvido humano não é um sistema de alta fidelidade, tendo em vista que suas próprias características anatomofisiológicas promovem alteração do sinal acústico (atenuando e/ou incrementando intensidade) desde a orelha externa até a orelha interna (LAWRENCE et al., 2007).

Analisando o impacto do sinal acústico na percepção auditiva das fricativas, nota-se que a faixa de frequência em que as fricativas se encontram (1.200Hz a 8.000 Hz) não é favorável à percepção, já que frequências acima de 5.000Hz não recebem qualquer incremento pelo sistema auditivo (JOHSON,1997). Além disso, o ruído de turbulência característico desta classe promove um mascaramento desfavorável à sua percepção (LAWRENCE et al., 2007). Portanto, essa classe sonora é menos saliente acusticamente, a julgar pela alta faixa de frequência e baixa intensidade característica das fricativas, desfavorecendo assim sua percepção.

No campo de estudos perceptivo-auditivos, Berti (2017) verificou pior desempenho em crianças na faixa etária entre quatro e cinco anos, em tarefa de identificação de contrastes fônicos envolvendo a classe das fricativas. Adicionalmente, neste mesmo estudo, a autora também constatou que o domínio perceptivo-auditivo em crianças com desenvolvimento típico de linguagem, na identificação de contrastes fonológicos entre as fricativas, ainda não se mostrava efetivo na faixa etária estudada (cinco-seis anos), apresentando uma acurácia de acerto em torno de 75%. Outro resultado descrito no estudo foram os diferentes graus de similaridades entre os contrastes fricativos, em que houve melhor acurácia na percepção das fricativas vozeadas vs desvozeadas comparativamente à acurácia

na percepção entre as fricativas alveolares vs palatais. Essa discrepância foi explicada pela presença de pistas mais robustas, como as pistas acústicas que marcam a distinção do vozeamento, as quais favoreceriam a percepção dos contrastes fonológicos fricativos.

Ademais, do ponto de vista perceptivo-auditivo, a percepção dos contrastes fônicos fricativos é a última a ser estabilizada, de acordo com a seguinte ordem decrescente: vogais, sonorantes, oclusivas e fricativas (BERTI, 2017).

Em virtude das características acústicas e dos diferentes graus de similaridade perceptual entre os contrastes fricativos, podemos inferir que pistas somente de ordem acústica poderiam não ser suficientes para propiciar o resgate de um contraste fonológico no interior desta classe, levantando a hipótese de que outras pistas diferentes das acústicas sejam necessárias para a percepção das fricativas.

O trabalho assume, portanto, enquanto concepção teórica, a perspectiva multimodal da percepção de fala, que será apresentada a seguir.

1.5 PERCEPÇÃO DE FALA COMO UM EVENTO MULTIMODAL

A percepção como um evento multimodal refere-se a uma perspectiva teórica que considera a integração de diferentes informações. Os principais sentidos envolvidos no processo de percepção de fala são audição e visão, que, de modo integrado, proporcionam maior compreensão do conteúdo apresentado (BRANCAZIO et al., 2003).

Essa perspectiva sobre a percepção de fala, como um evento multimodal, foi inaugurada por McGurk e MacDonald (1976), que, ao invés de conceber a percepção como sendo de natureza unimodal, ou seja, com o uso de uma única via sensorial, demonstraram a importância da informação visual para a percepção de fala.

No estudo desenvolvido, McGurk e MacDonald (1976) analisaram a influência da informação visual na percepção de fala. Participaram do estudo crianças (3 a 8 anos) e adultos (18 a 40 anos). Os participantes do estudo foram subdivididos em 3 grupos etários: (G1) crianças de 3 a 5 anos; (G2) crianças de 7 a 8 anos; (G3) adultos de 18 a 40 anos. O desenvolvimento se deu por meio de um teste de ilusão audiovisual, no qual os estímulos foram de natureza auditiva e visual, obtidos a partir

de gravação de áudio e vídeo de sílabas de padrão silábico simples: /ga/;/ba/;/pa/;/ka/.

A partir das gravações, foram elaboradas quatro condições de estímulos: (1) estímulos auditivos, caracterizados por áudio que apresentava a produção vocal dos fonemas-alvo; (2) estímulos visuais, caracterizados pelo gesto articulatório que apresentava a produção dos fonemas-alvo, e nessa condição foi apresentado o vídeo sem áudio associado dos fonemas-alvo; (3) estímulos audiovisuais congruentes, caracterizados pela apresentação simultânea dos estímulos auditivo e visual de modo correspondente (Exemplo: áudio [ga] e vídeo correspondente à produção articulatória do [ga]); e (4) estímulo audiovisual incongruente, caracterizado pela apresentação simultânea incongruente dos estímulos auditivos e visuais, isto é, não ocorreu correspondência entre os estímulos auditivos e visuais (Exemplo: áudio [ga] e vídeo correspondente à produção de [ba]).

Como resultado, os autores observaram que, para as condições auditiva e audiovisual congruente, os participantes relataram as sílabas com precisão. 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 já expresso na condição incongruente. Especificamente, para a combinação [áudio /ga/ - vídeo /ba/], o relato perceptual foi /da/; já para combinação [áudio /pa/ - vídeo /ka/], o relato perceptual foi de /ta/.

O relato de um “terceiro estímulo” diante da condição experimental audiovisual incongruente demonstra a interação entre a audição e visão na percepção de fala, uma vez que, se esta fosse de natureza unimodal (guiada por uma única via perceptual seja auditiva ou visual), a resposta perceptual dos participantes perante a condição experimental audiovisual incongruente seria pautada em apenas um dos componentes (auditivo ou visual) do estímulo incongruente. Os autores inferiram, portanto, que o relato do “terceiro estímulo”, perante condição experimental incongruente, seria fruto de uma ilusão perceptual (mais tarde também denominada como efeito McGurk), decorrente da integração das pistas audiovisuais. A partir destes dados, foi inaugurada a perspectiva teórica que considera a percepção de fala como um evento multimodal.

Diante de tal achado, os autores salientaram a importância da informação visual na percepção de fala, uma vez que a ilusão perceptual gerada na condição de

apresentação incongruente seria produto da integração da informação visual e auditiva durante a percepção de fala. Os autores destacam que houve diferença no desempenho entre os grupos, na medida em que o grupo de crianças de três a cinco anos e de sete a oito anos obtiveram menor influência visual, observado a partir de um menor relato da ilusão audiovisual (efeito McGurk), perante apresentação dos estímulos incongruentes, em relação aos adultos. Os autores atribuíram tal achado a um aumento gradual (desenvolvimento) da influência visual durante a infância até a idade adulta, tal aspecto será abordado na seção (3.2).

Desde então, estudos apontam (MARQUES; BOGGIO, 2011; MAGNOTTI; BEAUCHAMP, 2017; MASAPOLLO et al., 2017) para uma natureza multimodal no processo de percepção de fala, demonstrando que a informação visual é usada na percepção de fala mesmo quando a informação auditiva é clara e íntegra (SEKIYAMA e BURNHAM, 2004).

Atualmente, a integração de pistas audiovisuais vem sendo investigada através 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. Em seu estudo, Ojanen (2005) 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” são uma atividade cortical ativada 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 articulatorio. 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 (percepção do gesto articulatorio) 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 multimodal (*cross-modal*) de fala.

Diante de tais resultados, podemos notar que ambos os estudos apontam a influência conjunta da informação visual e auditiva para a percepção de fala, sugerindo assim a presença de integração de pistas audiovisuais.

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, a idade, a complexidade das estruturas

linguísticas (SEKIYAMA e TOHKURA, 1993) e alterações na produção dos sons da fala (DESJARDINS e WERKER, 1996; DODD, 2008) .

Assim, quando falamos em percepção multimodal, há necessidade de considerar diferentes aspectos, já que cada fator (cultural, linguístico, idade e condição clínica do sujeito) pode interferir no modo de percepção de fala. Na próxima seção, trataremos desses aspectos.

1.5.1 Influência da língua no processo de integração de pistas audiovisuais na percepção da fala

No tocante à língua, Massaro (1993) objetivou em seu estudo investigar a integração audiovisual em diferentes línguas (espanhol, inglês, holandês e japonês) a fim de verificar se a integração de pistas (visuais e auditivas) seria de natureza específica de uma dada língua ou de natureza universal. Apesar das diferenças derivadas das características de cada língua investigada, os autores observaram uma influência perceptual visual entre os falantes, ou seja, houve integração audiovisual durante a percepção de fala para todos os grupos investigados. Os autores concluíram, portanto, que a integração audiovisual seria de natureza universal.

Outros estudos (SEKIYAMA e TOHKURA, 1991; HAYASHI e SEKIYAMA, 1998; CHEN e HAZAN, 2007; SEKIYAMA e BURNHAM, 2008) objetivaram replicar o efeito McGurk, verificando como se dá o processo da percepção multimodal em diferentes línguas, isto é, se ocorreria algum tipo de mudança perceptual a depender da língua. Estes autores questionavam se a complexidade da língua analisada interferiria no desempenho perceptivo.

Sekiyama e Tohkura (1991) investigaram a percepção de japoneses adultos durante a apresentação de estímulos audiovisuais. Os estímulos de fala apresentados foram os seguintes: (/ba/; /da/; /ga/; /pa/; /ta/; /ka/; /ma/; /na/; /ra/ e /wa/). Durante o estudo piloto, os autores verificaram evidências fracas do efeito *McGurk*, ou seja, os participantes avaliados não foram fortemente influenciados pelo estímulo visual e não perceberam os estímulos auditivos como uma fusão de informações. Posteriormente, duas condições de fala foram adicionadas: estímulo auditivo com adição de ruído e estímulo auditivo sem adição de ruído. Foram verificadas fracas evidências do efeito *McGurk* na condição de estímulo sem adição

de ruído; diferentemente, na condição com adição de ruído, foram verificadas fortes evidências do efeito, demonstrando uma percepção da fusão de informações visual e auditiva dos estímulos apresentados nesta última condição. Os autores concluíram que falantes nativos do japonês usaram maior informação visual na percepção da fala apenas quando o componente auditivo do estímulo não estava perfeitamente inteligível. Os autores discutem três aspectos que justificariam tais resultados: (1) a menor influência visual diante de um estímulo auditivo em condição sem ruído é justificada pela própria característica silábica da língua japonesa, que permite apenas padrão silábico simples, exigindo menor integração de pistas audiovisuais para resgatar o conteúdo apresentado; (2) os aspectos culturais da população japonesa justificaria a menor influência visual durante a percepção de fala já que, culturalmente, essa população apresenta menor contato visual, ou seja, a percepção multimodal não é de natureza filogenética, ao contrário, depende das características da língua; (3) a maior influência visual diante apresentação dos estímulos em condição de ruído foi justificada pelo fato de que a pista visual seria fundamental diante da incerteza auditiva .

Hayashi e Sekiyama (1998) analisaram o efeito McGurk para estímulos da língua chinesa e japonesa. Participaram do estudo 40 sujeitos subdivididos em dois grupos: (G1) 17 falantes nativos de chinês; (G2) 23 falantes nativos de japonês. 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 e audiovisual de modo congruente e incongruente, em japonês e em chinês, para ambos os grupos. 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 incongruente, os resultados indicaram que os sujeitos japoneses usaram mais a informação visual na percepção de fala quando os oradores eram estrangeiros (estímulos apresentados em chinês), enquanto os sujeitos chineses não mostraram tal assimetria. Os autores reportaram tal achado pelo fato de o chinês ser uma língua tonal, em que o significado da palavra não é determinado somente pela estrutura silábica, mas também pelo tom. Deste modo, tais características fomentariam uma forte dependência da informação auditiva, resultando menor influência visual.

Por outro lado, Chen e Hazan (2007) analisaram a influência visual na percepção de fala em falantes de língua tonal e não tonal, bem como investigaram o efeito da língua estrangeira sob a influência visual de acordo com os falantes nativos e não nativos, na percepção de fala. A hipótese levantada por esses autores é que os falantes de língua tonal apresentariam melhor desempenho auditivo na percepção de fala e menor influência visual, já que os traços acústicos seriam os principais traços distintivos entre os fonemas destas línguas. O estudo contou com a participação de 22 chineses e 18 ingleses, 10 tailandeses e 10 japoneses, entre 20 e 54 anos. Os estímulos consistiram nas sílabas /ba/, /da/, /ga/ apresentados em diferentes condições: visual isolada, auditiva isolada, audiovisual congruente e incongruente, em situação de silêncio e ruidosa, proferidas por quatro falantes (dois chineses e dois ingleses, um homem e uma mulher em cada língua). Para investigar a influência visual na percepção de fala em falantes de língua tonal e não tonal, definido como primeiro objetivo do estudo, os autores analisaram o desempenho dos participantes chineses – língua tonal (22 participantes) e ingleses – língua não tonal (18 participantes). Os autores investigaram o desempenho dos participantes a partir dos estímulos nativos do seu idioma.

Como resultado, para a condição incongruente, os chineses apresentaram maior grau de influência visual em comparação aos ingleses; já para as condições visuais e auditivas isoladas, ambos os grupos apresentaram desempenhos semelhantes. A hipótese de que há uma menor influência de pistas visuais para as línguas tonais, levantada pelos autores, não foi corroborada a partir destes resultados. Os autores discutem que a informação visual exerceria influência distintiva mesmo em língua tonais.

Para investigar o efeito da língua estrangeira sob a influência visual na percepção de fala, segundo objetivo do estudo, os autores analisaram o desempenho dos participantes tailandeses -língua tonal (dez participantes) e japoneses – língua não tonal (dez participantes), já que para estes grupos todos os estímulos foram apresentados em chinês e inglês, ou seja, nenhum dos estímulos era no idioma nativo. Foi analisado também o desempenho dos chineses e ingleses diante de estímulos não nativos. Como resultado, os autores observaram influência visual forte para todos os grupos, anulando, assim, qualquer viés auditivo que possa advir de falantes de línguas tonais ou não tonais, ou seja, há maior influência visual na fala ao ouvir falantes não nativos independente dos padrões entoacionais da

língua. Dentre os grupos, os ingleses apresentaram maior influência visual, o que pode ser explicado pela possível similaridade entre as línguas chinesa, tailandesa e japonesa, em que os falantes em contato com uma destas línguas precisariam de menos pistas para resgatar o estímulo apresentado, justificado pela própria similaridade entre essas línguas.

Em estudo posterior, Sekiyama e Burnham (2008) compararam a percepção audiovisual na fala entre as línguas japonesa e inglesa. Participaram do estudo 96 sujeitos subdivididos em dois grupos: (G1) 48 estudantes universitários monolíngues (24 falantes nativos do inglês e 24 falantes nativos do japonês), entre 18 e 29 anos; (G2) 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; (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”). Para o grupo 1, os resultados demonstraram influência visual mais fraca para os participantes japoneses; no grupo 2, o grau de influência visual foi baixo e equivalente para os japoneses e ingleses 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. Conclusão: os japoneses apresentaram menor influência da via visual quando comparados aos americanos.

Alguns autores (SEKIYAMA e TOHKURA,1991; HAYASHI e SEKIYAMA,1998; SEKIYAMA e BURNHAM, 2008) apontam que a complexidade da língua analisada seria uma possibilidade explicativa para essa diferença no desempenho perceptivo.

Sekiyama e Burnham (2008) ressaltam em seu estudo que quanto menor a complexidade fonológica e menos visualmente distintiva forem os sons de determinada língua, menor a necessidade do uso de pistas visuais para a percepção de fala.

Os resultados dos estudos resenhados anteriormente, tomados juntos, demonstram que a língua é um fator influente na integração de pistas durante a percepção de fala.

1.5.2 Influência da idade no processo de integração de pistas audiovisuais na percepção da fala

Similarmente à influência da língua na integração audiovisual na percepção de fala, a idade tem sido também reportada como um fator influente na integração de pistas durante a percepção de fala na medida em que alguns estudos reportam diferenças significantes no uso de pistas visuais para a percepção de fala na comparação entre bebês, crianças e adultos (BURNHAM et al., 1991; CHEN, 2017; TAITELBAUM-SWED e LEAH FOSTICK, 2016). Nestes estudos, os autores observaram um aumento gradual (desenvolvimento) da influência visual durante a infância, na percepção de fala.

Hockley e Polka (1994) indicam em seu estudo que a percepção de fala no adulto é fortemente influenciada pelas informações visuais, enquanto a percepção de crianças entre cinco e oito anos mostra uma influência visual baixa e um forte viés auditivo. No entanto, com aumento da idade, a influência auditiva diminui, enquanto a influência da informação visual e a integração audiovisual aumentam entre as faixas etárias. Especificamente, os autores investigaram o uso da informação visual durante a percepção de fala em diferentes faixas etárias. Os participantes foram separados em cinco grupos: quatro grupos de crianças, sendo cada um composto por uma faixa etária (5-7-9 e 11 anos) e um grupo composto por adultos. Os participantes foram testados em quatro diferentes condições: (1) apenas auditiva; (2); apenas visual; (3) audiovisual congruente; e (4) audiovisual incongruente. Os estímulos analisados foram os seguintes: /ba/, /va/, /ta/, /da/, e /ga/. Os resultados apresentados demonstraram uma tendência geral de aumento da acurácia conforme aumento da idade, para as seguintes condições: apenas visual e audiovisual congruente. Tal achado vai de encontro ao estudo de McGurk e MacDonald (1976), no qual crianças de três a cinco anos e de sete a oito anos obtiveram menor influência visual na percepção de fala em relação aos adultos. Os autores atribuíram tal achado a um aumento gradual (desenvolvimento) da influência visual durante a infância até a idade adulta.

Embora não seja bem esclarecido na literatura (ROSENBLUM et al., 1997; DODD, 1998; BURNHAM e DODD, 2004), estudos em bebês demonstram que esta população é capaz de detectar a congruência de pistas visuais e auditivas na percepção de fala.

Dodd (1979) investigou o uso da informação visual durante a percepção de fala em 12 bebês de 10 a 16 semanas, aos quais foram apresentadas canções com sons da fala e 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 da criança para os estímulos apresentados. Os resultados mostraram que a duração da fixação visual foi menor diante apresentação dos estímulos incongruentes em comparação à apresentação congruente. Esse achado indicou à autora que os bebês detectam congruência entre movimentos labiais aos sons da fala.

Apesar de esses estudos reportarem a presença de um desempenho perceptivo audiovisual em bebês, estudos posteriores sugerem um possível desenvolvimento no uso da integração de pistas, conforme o aumento da idade.

Taitelbaum-Swed (2016) investigaram o desenvolvimento da integração de pistas audiovisuais na percepção de fala ao longo da vida, em falantes de hebraico. Participaram do estudo 77 sujeitos entre 4 e 80 anos subdivididos em 5 grupos de acordo com a faixa etária: 4-5, 8-9, 20-30, 40-55, 65-80. Os estímulos apresentados foram palavras monossilábicas com padrão silábico (CVC), e as palavras consistiam em 2 grupos : palavras com valor semântico para a língua (12 listas com 10 palavras em cada lista); palavras sem valor semântico para a língua (12 listas com 10 palavras em cada lista), apresentadas em condição visual, auditiva e audiovisual. Para as palavras sem valor semântico para a língua, a acurácia visual e auditiva foi menor em comparação as palavras com valor semântico para a língua. Segundo os autores, esse resultado era esperado, já que, para as palavras com valor semântico, o léxico também está disponível como fonte informativa para a percepção de fala, além das pistas auditivas e visuais. Especificamente, para as palavras com valor semântico, os autores reportaram os resultados para cada uma das condições de apresentação. Para a condição visual, observou-se associação positiva com a idade, desempenho semelhante nos grupos de 4-5 anos e 8-9 anos e desempenho inferior em comparação ao grupo de 20-30 anos. Já para a condição auditiva, os resultados mostraram um padrão de U invertido, com desempenho mais baixo nas idades 4-5 e 65-80 em comparação aos participantes com idade entre 8-55 anos. O resultado mostrou-se semelhante para a condição audiovisual, menor desempenho para as faixas etárias entre 4-5 anos e 65-80 anos, em comparação aos grupos entre 8-9, 20-30 e 40-55 anos. Com base nesses resultados, os autores concluíram que o

sistema auditivo e o processamento visual se desenvolvem e deterioram no decorrer do desenvolvimento. Especificamente, os autores reportam que o sistema auditivo amadurece mais cedo do que o visual, mas se deteriora mais rápido na idade avançada.

De acordo com os estudos resenhados nesta seção, observa-se o uso da informação visual durante a percepção de fala já em bebês, obtendo maior influência desta pista ao longo do desenvolvimento. No entanto, há uma divergência na literatura quanto à idade em que a maturação da integração audiovisual ocorreria: alguns estudos reportam entre 6 e 8 anos (MASSARO, 1986), enquanto outros indicam que ocorreria por volta dos 14 anos de idade (HOCKLEY e POLKA, 1994; BRANDWEIN et al., 2011; ROSS et al., 2011).

1.5.3 Influência cultural no processo de integração de pistas audiovisuais na percepção da fala

Conforme já mencionado nas seções anteriores, a literatura registra uma possível influência da língua sobre a percepção multimodal considerando a própria estrutura linguística em questão aliada à idade do sujeito como fatores influentes na percepção de fala.

Por sua vez, os aspectos culturais, embora menos explorados pela literatura, também poderiam exercer influência no modo pelo qual nos comunicamos, influenciando, assim, aspectos da percepção de fala. A título de exemplificação, Sekiyama e Burnham (2004) relataram que não é convencional na cultura japonesa direcionar a atenção visual para os movimentos de face e/ou articuladores do falante durante o discurso, ou seja, para o exemplo mencionado, os hábitos culturais da língua poderiam ser influentes no uso da pista visual durante o processo de percepção de fala.

Por outro lado, Chen e Hazan (2007) apontam em seu estudo que o hábito de evitar contato visual, na cultura japonesa, pode estar se tornando menos válido na geração mais nova. Nota-se que a dinâmica da comunicação sofre impacto cultural inerente e natural na sociedade (ISSEI-JAAKOLA, 2006).

Em síntese, neste capítulo foi apresentada a fundamentação da perspectiva da percepção de fala como um evento multimodal, que considera a integração de

diferentes informações. Para esta perspectiva, os principais sentidos envolvidos na percepção de fala são a audição e visão.

Os estudos mencionados até aqui investigaram a percepção multimodal em falantes típicos considerando como variáveis a língua, faixas etárias e culturas. Diante da apresentação destes estudos, nota-se que a percepção multimodal é de caráter universal e estaria presente em todas as línguas.

No entanto, o grau de integração de pistas audiovisuais dependerá de alguns fatores, como: (1) Língua, já que cada idioma apresenta complexidades e características específicas; (2) Faixa etária, tendo em vista que a integração de pistas audiovisuais é de caráter desenvolvimental; (3) Cultura, uma vez que os hábitos culturais poderiam ser influentes no modo pela qual as pistas audiovisuais são utilizadas durante o processo de percepção de fala.

Os estudos resenhados foram desenvolvidos com falantes típicos. Na próxima seção, serão apresentados os estudos dentro da perspectiva multimodal na vertente dos distúrbios da comunicação humana, especialmente os estudos desenvolvidos em crianças com transtorno fonológico.

1.6 INTEGRAÇÃO DE PISTAS NA PERCEPÇÃO DE FALA EM CRIANÇAS COM TRANSTORNO FONOLÓGICO

Nesta seção, explicaremos a influência do transtorno fonológico na integração de pistas audiovisuais, perpassando por caracterização do transtorno fonológico, e desempenho perceptual nessa população.

1.6.1 Caracterização do Transtorno fonológico

As crianças com transtorno fonológico constituem um amplo e heterogêneo grupo, e isso se reflete nas diferentes denominações e descrições de suas manifestações. O diagnóstico dessa população deve ser baseado em um sistema de classificação: Perspectiva Médica (observação clínica); Perspectiva Linguística (padrões de erros); ou Perspectiva Etiológica (fatores etiológicos). Abordaremos a classificação sob a perspectiva médica, adotada neste trabalho. Esta perspectiva é adotada pelo Manual Diagnóstico e Estatístico de Transtornos Mentais, que tem sido

uma das bases de diagnósticos de saúde mental mais usadas no mundo, por psicólogos, médicos e terapeutas.

A perspectiva médica é reconhecida e adotada mundialmente. O diagnóstico é feito através da observação clínica, que classifica as alterações da produção da fala de acordo com o diagnóstico clínico.

Citemos as características diagnósticas do transtorno fonológico advinda do Manual Diagnóstico e Estatístico de Transtornos Mentais 5ª edição - DSM-5 (2014, p.44):

crianças com dificuldades para produzir a fala podem apresentar dificuldades no reconhecimento fonológico dos sons da fala ou na capacidade de coordenar movimentos para falar, nos mais variados graus. O transtorno da fala é desse modo, heterogêneo em seus mecanismos subjacentes, incluindo transtorno fonológico e transtorno da articulação.

O transtorno fonológico é caracterizado por uma inadaptação, desorganização ou anormalidade no sistema de sons da criança em relação ao padrão-alvo adulto, sem qualquer comprometimento orgânico por parte das crianças (KESKE-SOARES et al., 2007), sugerindo um problema de base cognitiva que envolve um déficit na categorização linguística dos sons (DSM-IV-TR, 2002). Deste modo, a produção da fala não ocorre como esperado, de acordo com a idade e estágio de desenvolvimento da criança.

Em suma, crianças com transtorno fonológico apresentam erros de fala não mais esperados para a sua idade (RVACHEW, 2016).

Várias condições (biológica e circunstancial) podem explicar a ocorrência do transtorno fonológico, no entanto, como este transtorno não apresenta uma causa aparente, alguns estudos destacam o fato de que crianças com transtorno fonológico também poderiam apresentar dificuldades no domínio perceptual (RVACHEW e JAMIESON, 1989; BROWN, 2000), uma vez que a dificuldade de compreender os sons da fala poderia estar relacionada à ocorrência de alterações fonológicas durante o período da infância (FREITAS et al., 2015).

Uma vertente de estudos (DESJARDINS e WERKER, 1996; DODD et al., 2008), sob a perspectiva da percepção de fala como um evento multimodal, assume que essa população apresentaria déficit na integração de pistas audiovisuais, interferindo na percepção e representação dos fonemas linguísticos. A seguir, abordaremos a percepção de fala em crianças com transtorno fonológico sob a perspectiva multimodal.

1.6.2 Percepção multimodal em crianças com Transtorno fonológico

No tocante à percepção multimodal, tem sido reportado, para a língua inglesa, que crianças com transtorno fonológico são menos influenciadas pela pista visual, refletindo, assim, uma possível dificuldade de integração de pistas, interferindo na percepção dos fonemas linguísticos.

Desjardins e Werker (1996) investigaram a integração de pistas audiovisuais em crianças com Transtorno Fonológico, entre três e cinco anos de idade. Os autores verificaram que as crianças que realizavam substituições fonológicas apresentaram menor influência visual na percepção audiovisual quando comparadas às crianças que não apresentavam substituições, sugerindo um descompasso na integração de pistas durante a percepção de fala, o que poderia influenciar, conseqüentemente, na representação fonêmica no grupo de crianças com substituições fonológicas.

Posteriormente, em outro estudo, Desjardins et al. (1997) compararam a percepção multimodal da fala em 16 adultos típicos com 2 grupos de pré-escolares, 5 crianças com erros de substituição envolvendo os sons fricativos e 6 crianças com produção de fala típica. Os estímulos, apresentados em condições auditiva isolada, visual isolada, audiovisual (incongruente), foram: /ba/;/va/, /ða/ e /da/. Os autores verificaram que, para a condição auditiva isolada, não houve diferença entre os grupos; no entanto, para a condição visual isolada, as crianças mostraram menor influência visual em comparação aos adultos. Entre o grupo de crianças, o subgrupo com erros de substituição apresentou menor influência visual, tanto para as condições visual isolada quanto para a condição audiovisual incongruente em relação ao subgrupo de crianças com desenvolvimento típico de fala. Dados como esses, segundo os autores, reforçaram a hipótese de que as crianças com TF apresentariam dificuldades na integração de pistas audiovisuais na percepção de fala.

Por outro lado, Dodd et al. (2008) compararam a integração de pistas audiovisuais em 55 crianças com TF e 55 com desenvolvimento típico de fala, levando em consideração as seguintes condições: (1) apresentação de estímulos somente por via auditiva; (2) apresentação de estímulos somente por via visual; (3) apresentação simultânea dos estímulos auditivo e visual de modo congruente; e (4)

apresentação simultânea dos estímulos auditivo e visual de modo incongruente. Os estímulos foram apresentados em nível silábico envolvendo os fonemas oclusivos /b/; /d/; /g/ em contexto de /a/. Na análise, foram consideradas as porcentagens de respostas corretas nas diferentes condições de apresentação dos estímulos; para os estímulos incongruentes, foi considerada a porcentagem da ilusão audiovisual (inferida pelos autores, a partir do relato perceptual referente a um terceiro estímulo, não correspondente nem ao estímulo auditivo nem ao estímulo visual presente na condição audiovisual incongruente). Como resultado, a Anova de Medidas Repetidas não mostrou efeito significativo para o grupo nem para interação entre grupo*condições, evidenciando diferença estatística significativa para as condições de apresentação dos estímulos. O teste Post-hoc de Bonferroni mostrou pior acurácia para a condição de apresentação visual isolada e melhor para a condição auditiva isolada. Na análise do estímulo audiovisual incongruente, a Anova de Medidas Repetidas não mostrou efeito significativo para grupo nem para interação entre grupo*condição. Ou seja, não houve um relato perceptual de um terceiro estímulo (inferida na literatura como ilusão perceptual) para ambos os grupos, indicando que atenção dos participantes do estudo foi direcionada para apenas um aspecto do estímulo visual ou ao auditivo.

Os resultados encontrados neste estudo, segundo os autores, não vai de encontro aos resultados obtidos em estudos prévios (DESJARDINS e WERKER, 1996; DESJARDINS et al., 1997), isto é, crianças com transtorno fonológico seriam menos influenciadas pela pista visual, refletindo, assim, uma possível dificuldade de integração de pistas, na medida em que não houve diferença entre os grupos. Com relação ao estímulo incongruente, de forma mais específica, o fato de ambos os grupos não apresentarem relato perceptual de um terceiro estímulo (inferida na literatura como ilusão perceptual), seria justificado, segundo os autores, pela perturbação gerada na fala para as condições audiovisuais incongruentes, isto é, para a condição experimental audiovisual incongruente, as crianças usariam estratégias de se concentrar em um aspecto do estímulo.

Ainda no mesmo estudo, Dodd et al. (2008) compararam a integração de pistas audiovisuais em 34 crianças com TF, divididas em 2 subgrupos: (G1) 12 indivíduos com TF consistente; e (G2) 22 indivíduos com atraso fonológico. Os estímulos foram apresentados em nível de palavras: cat, gloves, helicopeter, book, sing. Os estímulos incongruentes foram submetidos à análise considerando a

porcentagem de preferência perceptiva. Como resultado, os autores observaram que ambos os grupos não se diferenciaram quanto à porcentagem da ilusão (relato perceptual de um terceiro estímulo), mas diferiram na preferência do componente perceptual em que indivíduos com atraso fonológico foram mais propensos a se ancorarem na informação visual em comparação ao grupo de crianças com TF. Tal achado sugere, portanto, que a integração de pistas audiovisuais poderia variar a depender da especificidade do grupo analisado. Segundo os autores, a estratégia de se apoiar menos na pista visual, adotada pelas crianças com transtorno fonológico consistente, poderia estar associada a uma capacidade prejudicada de integrar informações fonológicas de duas fontes de modalidade.

Encerrada a apresentação da assunção teórica, apresentaremos, no capítulo seguinte, o objetivo e hipóteses de nosso estudo.

2 OBJETIVO E HIPÓTESES

Embora se verifique, em âmbito internacional, a presença de estudos voltada para a identificação de integração de pistas no processo de percepção de fala, na literatura nacional, o levantamento bibliográfico realizado apontou uma lacuna de estudos que discorram sobre percepção multimodal em crianças típicas e/ou com TF, registrando e/ou comparando a existência da integração de informações audiovisuais para a percepção de fala no português brasileiro (PB).

O objetivo do presente estudo, portanto, foi comparar o desempenho de crianças com e sem TF na tarefa de percepção multimodal envolvendo pistas auditivas e visuais em fricativas desvozeadas. Assumindo que o transtorno fonológico poderia estar associado a uma capacidade prejudicada de integrar informação fonológica de duas fontes de modalidade, presumimos que crianças com TF apresentariam pior desempenho na integração de pistas auditivas e visuais, refletida por (a) menor porcentagem de acertos, (b) menor porcentagem de ilusão audiovisual e (c) maior tempo de reação na tarefa perceptual em relação às crianças sem TF.

3 METODOLOGIA

3.1 Aspectos Éticos

Este estudo, transversal prospectivo, foi aprovado pelo Comitê de Ética em Pesquisa (CEP) da Faculdade de Filosofia e Ciências – Unesp Campus de Marília, sob o Protocolo nº 15416819.6.0000.5406. Os responsáveis pelos indivíduos participantes assinaram o Termo de Consentimento Livre e Esclarecido e houve assentimento por parte das crianças participantes. A instituição onde ocorreram as coletas autorizou o desenvolvimento deste estudo.

3.2 Amostra

Participaram deste estudo 30 crianças (8 do sexo feminino e 22 do sexo masculino), entre 4 e 8 anos de idade, sendo 15 diagnosticadas com transtorno fonológico (TF), as quais apresentavam, dentre outros processos fonológicos, aqueles envolvendo a classe das fricativas (GE-grupo experimental), e 15 com desenvolvimento típico de fala (GC-grupo controle).

As crianças do (GE) foram recrutadas da lista de espera do Estágio Supervisionado de Terapia Fonoaudiológica: Fonologia Clínica do Centro de Estudos da Educação da Saúde (CEES), UNESP Marília. Elas passaram por avaliação audiológica, avaliação fonológica e avaliação motora da fala .

A avaliação audiológica foi realizada pelo serviço de audiologia do CEES – sob responsabilidade da Prof^a. Dr^a. Ana Cláudia Vieira Cardoso.

A avaliação fonológica permitiu a análise da produção de fala em tarefa de nomeação, sendo utilizadas as palavras que compõem o *Instrumento de Avaliação de Fala para Análise Acústica em contexto da vogal /a/ IAFAC* (Berti, Pagliuso e Lacava, 2009), bem como o *Percefal – Instrumento de Avaliação da Identificação de contrastes fonológicos para fricativas* (Berti, 2017).

O PERCEFAL foi aplicado somente nas crianças que apresentaram acometimento nas fricativas durante a avaliação da produção da fala com o uso do IAFAC. As listas de palavras aplicadas durante a avaliação de fala encontra-se no Anexo I.

A Avaliação motora da fala permitiu realizar um diagnóstico diferencial entre Transtorno Fonológico (TF) e Apraxia de fala infantil (API), cuja avaliação constituiu de um teste composto por cinco provas (VASSOLER, 2020) : (I) Medidas de produção de fala; (II) Repetição de palavras multissilábicas; (III) Prova da frase lexical; (IV) Avaliação motora da fala; (V) Prova de inconsistência de fala. Este teste possibilitou realizar o diagnóstico diferencial entre TF e API.

Avaliamos 28 crianças, recrutadas da lista de espera do Estágio Supervisionado de Terapia Fonoaudiológica: Fonologia Clínica do Centro de Estudos da Educação da Saúde (CEES), Unesp Marília. Ao final da avaliação fonológica – composta pelo levantamento do inventário fonético, sistema de fones contrastivos, análise dos processos fonológicos e da avaliação motora da fala –, excluímos do estudo 13 crianças as quais quatro apresentaram alteração de linguagem, duas apresentaram disfluências típicas da gagueira, seis não apresentaram trocas fonológica envolvendo a classe das fricativas e uma apresentou dificuldade na acuidade visual (miopia).

Portanto, selecionamos apenas 15 crianças com diagnóstico de TF, que apresentavam processos fonológicos envolvendo a classe das fricativas, sem alteração auditiva e, sem dificuldade motora associada.

As crianças com desenvolvimento típico de fala, que compuseram o (GC), foram selecionadas de duas escolas municipais da cidade de Marília e de uma instituição de caridade destinada a crianças e adolescentes. As produções de fala foram submetidas ao julgamento perceptivo auditivo do avaliador especialista, a partir da análise oitiva da produção de fala em tarefa de nomeação, com o uso do IAFAC em contexto da vogal /a/.

Para composição do grupo experimental, selecionamos as crianças que obtiveram o diagnóstico de Transtorno Fonológico (com acometimento na classe das fricativas) sem qualquer alteração auditiva verificada a partir de uma avaliação audiológica, enquanto as crianças do grupo controle foram selecionadas de modo pareado ao grupo experimental quanto à idade e ao sexo e, ainda, que não tiveram alteração na triagem auditiva (realizada previamente).

Já como critério de exclusão da amostra, adotamos parâmetros semelhantes para ambos os grupos, sendo excluídas as crianças com transtorno fonológico associado a outras alterações de linguagem, alterações neurológica e/ou auditiva, bem como criança cujos responsáveis relataram problemas de acuidade visual.

Quanto à acuidade visual, os participantes não passaram por avaliação oftalmológica, a exclusão foi pautada na presença de queixa visual e/ou uso de óculos, uma vez que alteração dessa natureza pode interferir na percepção audiovisual. A seguir, disposto nos Quadros 1 e 2, segue a relação dos participantes presentes em cada grupo.

Quadro 1: Caracterização dos Sujeitos do Grupo Experimental (GE)

Grupo	Sujeito	Idade	Processos Fonológicos	PCC-R	Grau
GE	1	6	Posteriorização de fricativa	87	leve
GE	2	4	Dessonorização de fricativa	82	levemente-moderado
GE	3	4	Anteriorização de fricativa	80	levemente-moderado
GE	4	6	Plosivização de fricativa	80	levemente-moderado
GE	5	6	Apagamento de coda fricativa	71	levemente-moderado
GE	6	5	Dessonorização de fricativa	85	levemente-moderado
GE	7	8	Apagamento de coda fricativa	76	levemente-moderado
GE	8	6	Posteriorização de fricativa	88	leve
GE	9	8	Anteriorização de fricativa	94	leve
GE	10	4	Posteriorização/Dessonorização de fricativa	84	levemente-moderado
GE	11	6	Posteriorização de fricativa	73	levemente-moderado
GE	12	5	Anteriorização de fricativa	60	moderadamente-severo
GE	13	8	Anteriorização/Dessonorização de fricativa	52	moderadamente-severo
GE	14	7	Posteriorização de fricativa	69	levemente-moderado
GE	15	7	Anteriorização/Dessonorização de fricativa	91	leve

Fonte: Elaborado pela autora

Quadro 2: Caracterização dos Sujeitos do Grupo Controle (GC)

Grupo	Sujeito	Idade	PCC-R
GC	1	6	100
GC	2	4	100
GC	3	4	100
GC	4	6	100
GC	5	6	100
GC	6	7	100
GC	7	8	100
GC	8	6	100
GC	9	8	100
GC	10	4	100
GC	11	6	100
GC	12	5	100
GC	13	8	100
GC	14	5	100
GC	15	7	100

Fonte: Elaborado pela autora

O **PCC-R** (Porcentagem de Consoantes Corretas-revisado) foi o cálculo realizado neste estudo a partir das amostras de fala dos sujeitos avaliados. O valor deste cálculo permite definir o grau do transtorno fonológico conforme a porcentagem de acerto, de acordo com o seguinte critério: Grau leve: acima de 85%; Grau levemente-moderado: entre 65% e 85%; Grau moderadamente-severo: entre 50% e 65%; Grau severo: abaixo de 50% (SHRIBERG e KWIATKOWSKI, 1982). O PCC-R proporciona melhor separação entre as crianças com aquisição típica de fala e as com atraso de fala, e aponta somente os erros de omissão e substituição (WERTZNER, AMARO e TERAMOTO, 2005).

3.3 Estímulos

Os estímulos do teste são de natureza auditiva e visual referente a sílabas de padrão silábico simples, de modo que o ataque foi preenchido pelas fricativas: /f/, /s/ e /ʃ/ em contexto da vogal /a/.

Estabelecidos os estímulos do teste, realizamos gravações de áudio e vídeo simultaneamente correspondentes aos estímulos. As gravações foram editadas em um computador, com o uso do software Openshop, de modo que cada estímulo foi salvo individualmente em formato .wav para os estímulos auditivos, e em formato .mp4 para os estímulos visuais.

A partir das gravações, elaboramos quatro condições de estímulos: (1) estímulos auditivos, que se caracterizaram por áudio que apresentava a produção vocal dos fonemas-alvo; (2) estímulos visuais, que se caracterizaram pelo gesto articulatório que apresentava a produção dos fonemas-alvo, nessa condição era apresentado o vídeo sem áudio associado dos fonemas-alvo; (3) estímulos audiovisuais congruentes, que se caracterizaram pela apresentação simultânea dos estímulos auditivo e visual de modo correspondente (Exemplo: áudio [fa] e vídeo correspondente à produção do [fa]); e (4) estímulo audiovisual incongruente, que se caracterizou pela apresentação simultânea incongruente dos estímulos auditivos e visuais, isto é, não ocorreu correspondência entre os estímulos auditivos e visuais (Exemplo: áudio [fa] e vídeo correspondente à produção de [sa]).

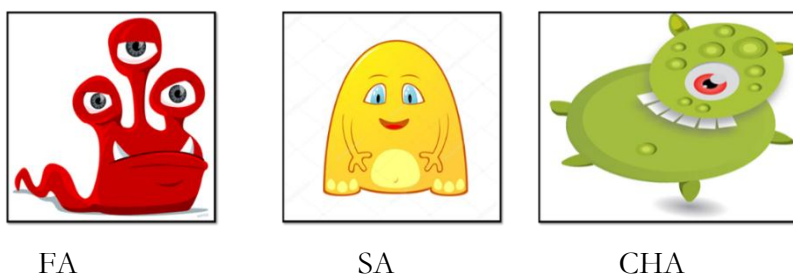
Dessa forma, os estímulos foram combinados conforme síntese contida no Quadro 3.

Quadro 3: Condições de apresentação x estímulos apresentados

Condições de apresentação dos estímulos	Somente auditivo	Somente visual	Integração audiovisual	
			Congruentes	Incongruentes
Estímulos	[fa]; [sa] e [ja]	[fa]; [sa] e [ja]	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 ja _Visual sa Audio sa _Visual fa Audio sa _Visual ja

3.4 Procedimento Experimental

Após a definição dos estímulos, elaboramos um *script* para o experimento de identificação a ser executado pelo *software OpenSesame* (MATHÔT, SCHREIJ e THEEUWES, 2012). O *script* contou com figuras de três personagens, em que cada personagem foi associado a um fonema-alvo e sua representação grafêmica correspondente, conforme ilustra a Figura 2.

Figura 2: Script – Percepção Multimodal fricativas

O teste de percepção multimodal refere-se a um teste de identificação, também designado tarefa de escolha forçada. Em testes de identificação, um estímulo é apresentado, e o informante deve identificá-lo, escolhendo uma das opções oferecidas (KLUGE *et al.*, 2013). Para a elaboração do teste, foi utilizado o *software OpenSesame* (MATHÔT, SCHREIJ e THEEUWES, 2012).

O teste de identificação foi composto por duas etapas distintas. Na primeira etapa – fase treino –, apresentamos aos participantes os estímulos visuais e

auditivos presentes no teste, por meio de um *power-point*, e cada estímulo foi associado a um desenho lúdico (realizando referência ao nome do personagem). Tal processo proporcionou familiaridade da criança com os estímulos.

Na segunda etapa – fase teste –, com uso do *software OpenSesame*, apresentamos, respectivamente, os estímulos: (1) auditivos (2) visuais; (3) audiovisuais congruentes e (4) audiovisuais incongruentes. A ordem de apresentação foi baseada na complexidade da própria natureza do estímulo. Nesta etapa, os indivíduos foram dispostos confortavelmente em frente à tela do computador (contendo o *software OpenSesame*), com fones acoplados aos ouvidos, com nível de intensidade acusticamente confortável.

Durante a primeira parte do teste, os indivíduos escutaram (com apresentação binaural, em um nível de intensidade acusticamente confortável) um estímulo auditivo, com imagem da região orofacial congelada (em posição neutra) na tela do computador e, em seguida, precisaram decidir e indicar qual foi a sílaba correspondente ao estímulo auditivo apresentado (representada por imagem do personagem e pelo seu grafema correspondente) dentre três possibilidades dispostas na tela do computador, conforme ilustra a Figura 3.

Figura 3: apresentação auditiva do fonema /fa/ com imagem da região orofacial congelada

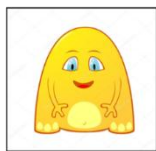


- apresentação auditiva do fone /fa/ - com imagem da região orofacial congelada.

Figura 4: opções de escolha após apresentação do estímulo auditivo



FA



SA



CHA

- opções de escolhas a partir do estímulo auditivo apresentado

Os indivíduos foram orientados a apontar para uma das opções apresentadas, enquanto a avaliadora foi responsável por teclar a resposta

correspondente a fim de facilitar a tarefa a ser executada por indivíduos menores. Cada estímulo auditivo foi apresentado duas vezes.

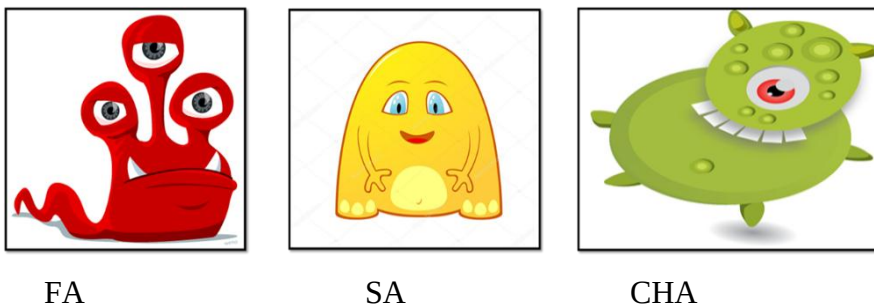
Para a segunda parte do teste, um estímulo foi apresentado apenas visualmente. Tratou-se gravação de vídeo, mostrando movimento articulatorio correspondente à produção dos estímulos, apresentada sem qualquer informação auditiva associada. Similarmente à etapa anterior, após a apresentação do estímulo visual, o indivíduo deveria decidir e indicar, dentre três opções dispostas da tela do computador, qual a sílaba correspondente ao estímulo visual apresentado (conforme Figura 5). Cada estímulo visual foi apresentado duas vezes.

Figura 5: vídeo com apresentação do movimento articulatorio do fonema /fa/, sem qualquer informação auditiva associada.



- vídeo com apresentação do movimento articulatorio do fonema /fa/

Figura 6: opções de escolha após apresentação da condição experimental visual



- opções de escolhas a partir do estímulo visual apresentado

Por fim, para a terceira parte do teste, apresentamos os estímulos audiovisuais, caracterizados por conter tanto a informação auditiva quanto a informação visual, ou seja, uma gravação de vídeo relativa à produção dos fonemas 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 indivíduo, conforme instruções das etapas 1 e 2.

Para as condições de apresentação dos estímulos auditivos, visuais e audiovisuais congruentes, os padrões de resposta considerados foram: acerto (quando o indivíduo identifica o estímulo corretamente) e erro (se não o identifica corretamente). Adicionalmente, consideramos os tempos de reação para apresentação de uma resposta de cada criança.

Para a condição audiovisual incongruente, os padrões de resposta considerados foram: preferência auditiva (quando o indivíduo seleciona a resposta correspondente ao estímulo auditivo), preferência visual (quando o indivíduo seleciona a resposta correspondente ao estímulo visual) e presença de ilusão audiovisual (quando o indivíduo seleciona a resposta que não condiz com o estímulo auditivo ou visual).

4.5 Análise Estatística

Realizamos um tratamento estatístico descritivo e inferencial dos dados com o uso do *software IBM SPSS Statistics* (versão 2.2). Para a análise inferencial, comparamos os tempos de reação, bem como a porcentagem de acerto para as quatro condições investigadas (somente auditiva, somente visual e integração audiovisual *congruente/incongruente*), entre os grupos de crianças (GC e GE), a partir do teste Anova de Medidas Repetidas, considerando-se, como fator intergrupo, a condição clínica das crianças (típicas - GC e com transtorno fonológico - GE), e como fatores intragrupos, as quatro condições de estímulos investigadas (somente auditiva, somente visual e integração audiovisual *congruente/incongruente*). Para análise da condição incongruente, consideramos a porcentagem de acerto das respostas baseadas no estímulo auditivo, uma vez que a literatura indica que a percepção de fala da criança e um primeiro momento seria pautado majoritariamente no componente auditivo (HOCKLEY e POLKA, 1994); e para o tempo de reação, a média do tempo de resposta para as respostas baseadas também no estímulo auditivo. Adotamos um valor de $\alpha \leq 0,05$.

4 RESULTADOS

A Tabela 1 mostra a porcentagem de acertos das crianças dos dois grupos nas quatro diferentes condições do teste experimental. Na condição audiovisual incongruente, consideramos a porcentagem de acerto do estímulo auditivo apresentado.

Tabela 1: Porcentagem de acerto relativa ao grupo experimental (GE) e grupo controle (GC)

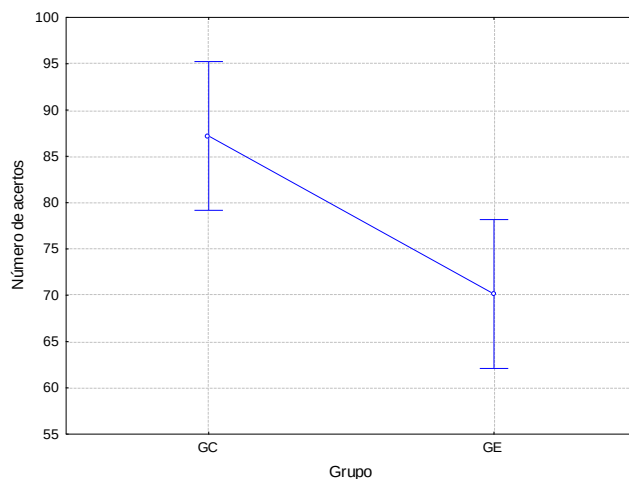
Média (% de acerto)	AO	AV+	AV-	V
	Média (DP)	Média (DP)	Média (DP)	Média (DP)
GE	74,4 (±26,5)	83,2 (±21,8)	75,3 (±20,8)	47,6 (±22,6)
GC	88,8 (±13,5)	86,6 (±16,8)	91,6 (±13,9)	82,2 (±19,3)

Legenda: **GE** - Grupo experimental **GC** - Grupo controle **AO** - Estímulo auditivo isolado **AV+** - Estímulo audiovisual congruente **AV-** - Estímulo audiovisual incongruente **V** – Estímulo visual isolado **DP** – Desvio padrão.

A Anova de Medidas Repetidas mostrou um efeito significativo para o grupo ($F(1, 28) = 9,4673$, $p < 0,05$), para as condições de apresentação dos estímulos ($F(3, 84) = 11,720$, $p < 0,05$) e para a interação entre o grupo e as condições ($F(3, 84) = 5,6855$, $p < 0,05$).

As crianças com transtorno fonológico (GE) apresentaram maior número de erros comparativamente às crianças típicas (GC), portanto, pior acurácia perceptual, conforme observamos no Gráfico 1.

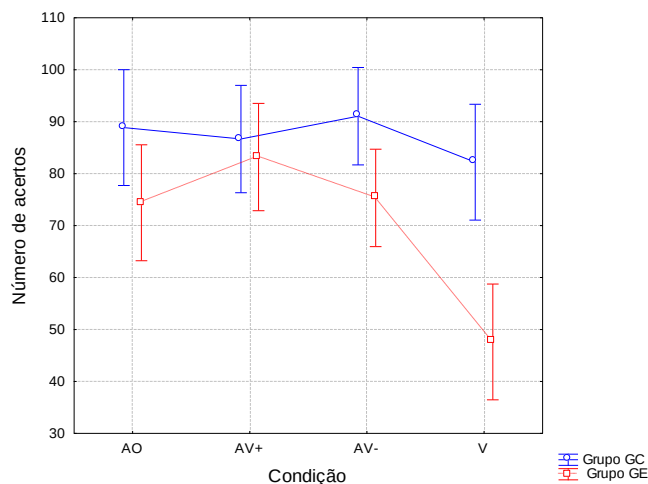
Gráfico 1: Comparação entre a acurácia perceptual geral do grupo experimental (GE) e grupo controle (GC).



Legenda: GE- grupo experimental; GC-Grupo Controle.

Em virtude do efeito significativo na interação entre grupo e condições experimentais, realizamos o teste Post Hoc de Duncan's, que mostrou pior acurácia para a condição de apresentação visual somente para o grupo de crianças com transtorno fonológico $p(p<0,05)$, como ilustra o Gráfico 2.

Gráfico 2: Comparação entre acurácia perceptual das condições de apresentação dos estímulos do grupo experimental (GE) e grupo controle (GC).



Legenda: AO auditivo isolado; AV+ estímulo audiovisual congruente; AV- estímulo audiovisual incongruente; V estímulo visual isolado; GC grupo controle; GE grupo experimental.

De modo a verificar a presença de integração de pistas audiovisuais, analisamos separadamente a condição audiovisual incongruente (AV-). A Tabela 2 mostra a porcentagem para preferência perceptual do estímulo auditivo ou visual, bem como a presença de ilusão perceptual (resultante do relato perceptual de um terceiro estímulo não correspondente nem ao estímulo auditivo nem ao estímulo visual) perante apresentação do estímulo incongruente (AV-).

Tabela 2: Estímulo audiovisual incongruente: Preferência perceptual relativa ao grupo experimental (GE) e grupo controle (GC).

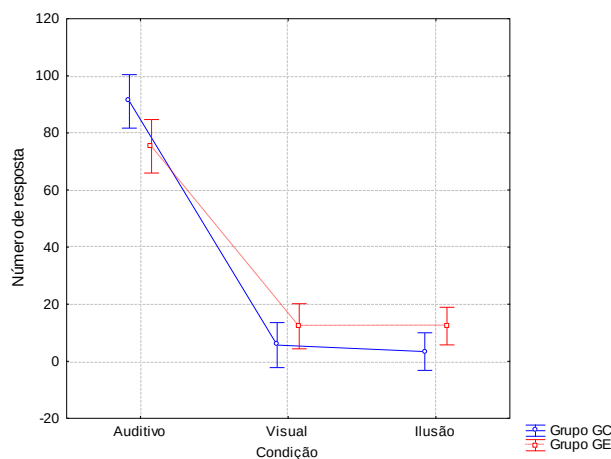
Média (% preferência perceptual)	Auditivo	Visual	Ilusão
	Média (DP)	Média (DP)	Média (DP)
GE	75,3 (±20,8)	12,6 (±19,4)	12,3 (±16,1)
GC	91,1 (±13,9)	5,7 (±8,3)	3,4 (±7,0)

Legenda: GE - Grupo experimental; GC - Grupo controle; DP – Desvio padrão.

A Anova de Medidas Repetidas não mostrou efeito significativo para grupo ($F(1,28) = 3,1500$, $p > 0,05$), mas houve efeito significativo para preferência perceptual

($F(2,56) = 161,66$, $p < 0,005$) e para a interação entre grupo e preferência perceptual ($F(2,56) = 4,0196$, $p < 0,05$). Em virtude do efeito significativo para preferência perceptual e interação entre grupo e preferência perceptual, aplicamos o teste Post Hoc de Duncan's a fim de realizar a comparação por pares. Como resultado, verificamos uma diferença entre os grupos em cada uma das preferências perceptual, de modo que as crianças com TF tiveram menor porcentagem de respostas para a preferência perceptual auditiva e maiores porcentagens para a preferência perceptual visual e ilusão perceptual, conforme ilustra o Gráfico 3.

Gráfico 3: Estímulo audiovisual incongruente: comparação relativa à preferência perceptual do grupo experimental (GE) e grupo controle (GC).



Legenda: GC- Grupo controle; GE- Grupo experimental.

No tocante ao tempo de reação, a Tabela 3 mostra o tempo de reação em relação aos acertos das crianças dos dois grupos para as quatro diferentes condições do teste experimental. Vale destacar que, na condição audiovisual incongruente, consideramos o tempo de reação do estímulo auditivo apresentado.

Tabela 3 : Tempo de reação referente ao grupo experimental (GE) e grupo controle (GC).

TR acertos (ms)	TR_AO	TR_AV+	TR_AV-	TR_V
	Média (DP)	Média (DP)	Média (DP)	Média (DP)
GE	2445 (±1207)	1624 (±894)	1980 (±1495)	5130(±2848)
GC	1189 (±887)	964 (±681)	1052 (±510,2)	3734 (±1849)

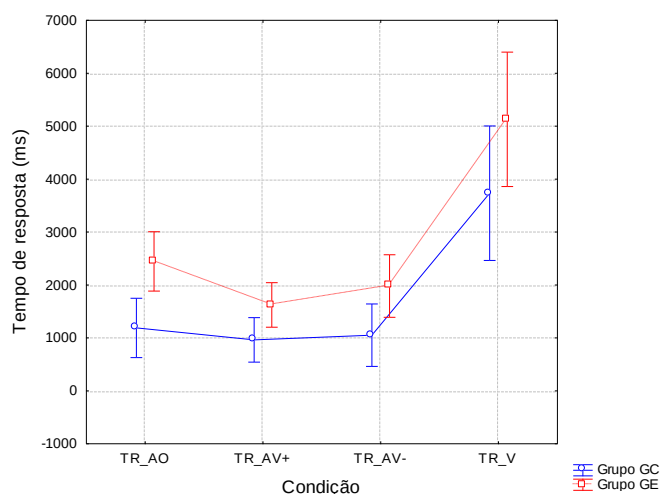
Legenda: GE - Grupo experimental GC - Grupo controle TR_AO - Tempo de reação em relação ao Estímulo auditivo isolado TR_AV+ - Tempo de reação em relação ao Estímulo audiovisual

congruente **TR_AV-** - **Tempo de reação em relação ao** Estímulo audiovisual incongruente **TR_V** – Tempo de reação em relação ao Estímulo visual isolado **DP** – Desvio padrão.

Para os dados referentes ao tempo de reação, a Anova de Medidas Repetidas mostrou efeito significativo para grupo ($F(1,28)=7,1689$, $p < 0,05$) e para as condições de apresentação dos estímulos ($F(3,84)=47,254$, $p < 0,05$), mas não houve efeito significativo para interação entre grupo e condição ($F(3,84)= 0,60662$, $p > 0,05$). Particularmente, as crianças com TF apresentaram maior tempo de reação (diferença de 1060 ms).

Em virtude do efeito significativo para as condições de apresentação dos estímulos, realizamos o teste Post Hoc de Duncan's, que mostrou, para ambos os grupos, maior tempo de reação para a condição de apresentação visual, tal como ilustra o Gráfico 4.

Gráfico 4: Comparação entre o tempo de reação do grupo experimental (GE) e grupo controle (GC).



Legenda: TR-AO Tempo de reação para estímulo auditivo isolado; TR-AV+ Tempo de reação para estímulo audiovisual congruente; TR-AV- Tempo de reação para estímulo audiovisual incongruente; TR-V Tempo de reação para estímulo visual isolado; GC grupo controle; GE grupo experimental, ms milissegundos.

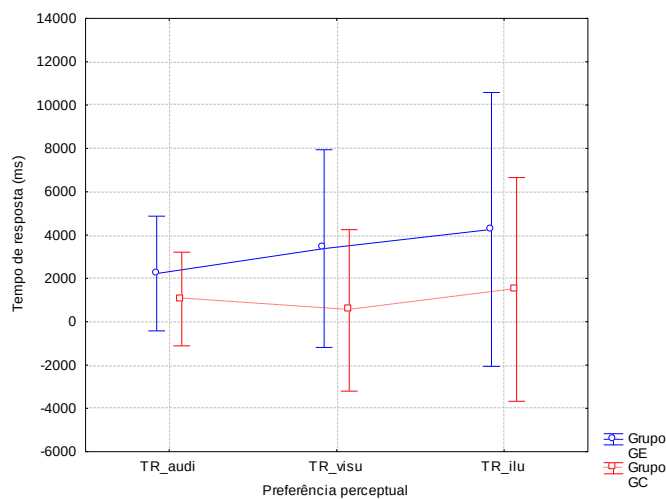
Para a condição audiovisual incongruente, especificamente, no tocante ao tempo de reação, a Tabela 4 mostra o tempo de reação das crianças dos dois grupos para as respostas baseadas na preferência perceptual do estímulo auditivo ou visual, bem como a presença de ilusão perceptual (resultante do relato perceptual de um terceiro estímulo não correspondente nem ao estímulo auditivo nem ao estímulo visual) perante a apresentação do estímulo incongruente (AV-).

Tabela 4: Tempo de reação relativo ao estímulo audiovisual incongruente.

TR (ms)	TR_auditivo	TR_visual	TR_ilusão
	Média (DP)	Média (DP)	Média (DP)
GE	1980 (± 1495)	3463 (± 3276)	2463 (± 23464)
GC	1052 (± 510)	831 (± 700)	1490 (± 880)

Legenda: **GE** - Grupo experimental; **GC** - Grupo controle; **DP** – Desvio padrão.

Para os dados referentes ao tempo de reação para a condição incongruente, a Anova de Medidas Repetidas não mostrou efeito significativo para o grupo ($F(1,3)=1,6183$, $p > 0,05$), para a preferência perceptual ($F(2,6)=2,0177$, $p > 0,05$), nem para a interação entre grupo e preferência perceptual ($F(2,6)= 1,0883$, $p > 0,05$). Especificamente para a condição de apresentação incongruente, os grupos não se diferenciaram quanto ao tempo de resposta.

Gráfico 5 : Comparação entre o tempo de reação, para estímulo incongruente, do grupo experimental (GE) e grupo controle (GC).

Legenda: TR-audi - Tempo de reação para preferência perceptual auditiva; TR-visu - Tempo de reação para preferência perceptual visual; TR-ilu - Tempo de reação para o relato da ilusão; GC grupo controle; GE grupo experimental, ms milissegundos.

5 DISCUSSÃO

Neste capítulo, apresentaremos as tendências encontradas neste estudo e suas possíveis hipóteses explicativas.

Tratando-se do desempenho de crianças com e sem TF em tarefa de percepção multimodal envolvendo pistas auditivas e visuais, presumimos que crianças com TF apresentariam pior desempenho na integração de pistas auditivas e visuais, refletida por (a) uma menor porcentagem de acertos, (b) menor porcentagem de ilusão audiovisual e (c) maior tempo de reação na tarefa perceptual em relação às crianças sem TF.

Apresentaremos as tendências encontradas e suas possíveis hipóteses explicativas na seguinte ordem: porcentagem de acertos, porcentagem de ilusão audiovisual e tempo de reação.

Em relação à porcentagem de acertos, os resultados mostraram que o grupo de crianças com transtorno fonológico (GE) teve menor número de acertos comparativamente às crianças típicas (GC), portanto, pior desempenho na tarefa perceptual. Esses resultados confirmaram integralmente a primeira hipótese e vai de encontro com alguns estudos anteriores (RVACHEW & DONALD, 1989; BROWN, 2000; HEARNSHAW, 2018; NIJALAND, 2009), os quais apontam que crianças com transtorno fonológico apresentariam dificuldades no domínio perceptual da fala. Os autores apontam que crianças com TF apresentariam representações perceptivas, fundamentais para discernir a diferença entre valores fonêmicos e fonéticos, menos robustas em comparação aos seus pares típicos. Edwards et al. (1999) descrevem que a variação na produção da fala resultaria em parte de danos subjacentes nas representações perceptivas.

Tendo em vista o efeito significativo na interação entre grupo e as condições, realizamos o teste estatístico Post Hoc de Duncan's, que mostrou pior acurácia para a condição de apresentação visual isolada somente para o grupo de crianças com transtorno fonológico. A partir deste dado, podemos inferir que o grupo de crianças com TF pode ser menos influenciado pela pista visual comparativamente ao grupo controle.

Estudos existentes para a língua inglesa, sobre percepção multimodal, também têm mostrado que crianças com transtorno fonológico são menos influenciadas pela pista visual, refletindo assim em uma possível dificuldade de

integração de pistas interferindo, portanto, na percepção dos fonemas linguísticos (DESJARDINS e WERKER ,1996; SIVA et al., 1995). Mais especificamente, estes estudos relatam que os erros de fala poderiam estar associados à menor influência da pista visual.

Desta forma, esses resultados podem ser mais uma evidência de que as crianças com transtorno fonológico são menos susceptíveis à influência visual na percepção multimodal quando comparadas às crianças que não apresentam TF, sugerindo um descompasso na integração de pistas durante a percepção de fala, o que poderia influenciar, conseqüentemente, na representação fonêmica no grupo de crianças com transtorno fonológico.

No tocante à integração de pistas audiovisuais, analisamos separadamente a condição audiovisual incongruente, a fim de verificar presença de integração audiovisual, a partir do relato da ilusão audiovisual. Verificamos uma diferença entre os grupos em cada uma das preferências perceptual, de modo que as crianças com TF tiveram menor porcentagem de respostas para a preferência perceptual auditiva e maiores porcentagens para a preferência perceptual visual e ilusão perceptual.

Em relação ao estímulo incongruente, considerando as crianças com TF, pressupomos que a menor porcentagem de respostas para preferência perceptual auditiva seja fruto de uma dificuldade perceptivo-auditiva apresentada por essa população, conforme apontam estudos prévios (RVACHEW & DONALD, 1989; BROWN, 2000; HEARNSHAW,2018; NIJALAND, 2009). Nesse sentido, possivelmente, a dificuldade perceptiva auditiva das crianças com TF implicaria na ancoragem de outras vias sensoriais de modo a compensar sua dificuldade perceptivo-auditiva, o que justificaria maiores porcentagens para preferência perceptual visual e ilusão perceptual quando comparadas às crianças com desenvolvimento típico de fala. Destacamos, porém, que essa compensação seria insuficiente para equiparar o desempenho entre as condições, viabilizando uma integração audiovisual efetiva durante a percepção de fala para equiparar o desempenho aos seus pares com desenvolvimento típico de fala.

No tocante ao tempo de reação, os resultados obtidos demonstram que as crianças com Transtorno fonológico apresentam, para o acerto, um tempo de reação mais laborioso quando comparado ao grupo controle, confirmando a terceira hipótese. A condição visual para ambos os grupos foi a que exigiu maior tempo de reação.

O fato de o grupo de crianças com TF ter apresentado um tempo de reação superior ao grupo de crianças típicas pode evidenciar maior demanda de processamento por parte das crianças com TF, tal como verificada por Attoni (2010). Ao analisar as respostas encontradas na avaliação do processamento auditivo em crianças com desenvolvimento normal de fala e com transtorno fonológico, a partir de testes de escuta diótica, monótica e dicótica, Attoni verificou que as crianças com desenvolvimento normal de fala obtiveram valores considerados normais em todos os testes do processamento auditivo, mas apresentaram valores alterados nos testes de escuta dicótica. Os autores concluíram que crianças com TF apresentam dificuldades no processamento auditivo.

Destacamos, ainda, que ambos os grupos apresentaram maior tempo de reação envolvendo a condição visual isolada. Este achado pode ser mais uma evidência de que a percepção multimodal não é inata ao indivíduo na medida em que este sofre influência da língua, da cultura e da idade. A faixa etária das crianças avaliadas neste estudo varia de quatro a oito anos. Há uma discrepância na literatura com relação à idade em que a habilidade perceptual de fala é estabilizada, variando de 6 a 12 anos (PENIDO, 2013; BERTI, 2017).

A literatura também reporta uma gradação no desenvolvimento do uso da pista visual para percepção de fala (CHEN e HAZAN, 2009). Inicialmente, a criança não se beneficia dos recursos visuais, sua influência na percepção de fala aumenta com a idade. De modo geral, num primeiro momento, a percepção de fala da criança é pautada majoritariamente pela pista auditiva, havendo um aumento gradual do uso da pista visual conforme o aumento da idade (TAITELBAUM-SWEAD, 2016).

Considerando a faixa etária avaliada neste estudo (quatro a oito anos) e o desenvolvimento gradual da pista visual, para a percepção de fala, supomos que a habilidade perceptual de fala dos grupos investigados ainda esteja em desenvolvimento, o que justificaria o maior tempo de reação para a condição experimental visual, destacamos que essa informação precisaria ser validada com o desempenho dos adultos em tarefa de percepção multimodal de fala.

Analisamos separadamente o tempo de reação para a condição audiovisual incongruente. Os resultados obtidos não evidenciaram diferenças entre grupos e nem preferência perceptual. A partir desta não diferenciação, inferimos que a incongruência do estímulo parece ser um fator dificultador para ambos os grupos.

Para o GE, um dado importante de ser ponderado é a gravidade do transtorno fonológico que, para o grupo de crianças analisadas, variou de leve a moderadamente severo. Estudos apontam uma relação entre produção e percepção de fala: embora não seja uma relação de espelhamento, essas duas habilidades não são dissociadas entre si (BERTI, 2020).

Como já mencionamos na seção de metodologia, a gravidade é obtida a partir da análise de produção. No entanto, considerando a relação existente entre produção e percepção de fala, podemos inferir que o índice de gravidade possivelmente não reflita apenas o índice de produção, mas também as habilidades gerais envolvidas no circuito da fala, perpassando, assim, pelas habilidades de percepção e produção.

Diante desses dados, a gravidade torna-se um índice crucial a ser considerado, já que o desempenho perceptual dessas crianças poderia variar a depender da gravidade, implicando diretamente no processo terapêutico, pois crianças com TF de gravidade moderadamente severo e severo apresentariam dificuldades perceptuais mais acentuadas refletidas por um pior desempenho nas condições experimentais (auditiva, visual, audiovisual congruente e incongruente).

Em suma, os resultados, tomados juntos, sugerem que a percepção multimodal parece se desenvolver ao longo do processo de aquisição da linguagem implicando, assim, em um desenvolvimento no uso da integração de pistas. Além disso, crianças com transtorno fonológico apresentaram menor influência da informação visual na percepção audiovisual quando comparadas às crianças típicas, sugerindo um descompasso na integração de pistas durante a percepção de fala, o que poderia influenciar, conseqüentemente, na representação fonêmica no grupo de crianças com transtorno fonológico.

6 CONCLUSÃO

Crianças com TF, para a classe das fricativas, apresentaram pior desempenho e maior tempo de reação na tarefa de percepção multimodal, sugerindo dificuldades na integração das informações fonológicas de fontes bimodais (auditiva e visual). Adicionalmente, crianças com TF tiveram maior dificuldade na identificação do estímulo na condição visual.

Alertamos como limitação desse estudo para o fato de não ser considerada a gravidade do TF, bem como a investigação da percepção multimodal ser restrita à classe das fricativas.

Em termos de gravidade, esse índice pode refletir diretamente no desempenho perceptual, já que a gravidade do TF possivelmente não reflita somente a gravidade manifestada na produção, mas também as diversas habilidades envolvidas no processo de produção e percepção de fala. Desta forma, sugerimos a investigação para estudos futuros da percepção multimodal considerando a gravidade do TF.

Quanto ao estudo ser restrito à classe das fricativas, sugerimos, para estudos futuros, a investigação da percepção multimodal para as demais classes fonológicas do PB, já que, em termos de aquisição perceptual, estudos apontam uma gradação entre as classes fonológicas do PB. Analisar a percepção multimodal, tanto na população típica quanto na população com TF, considerando as demais classes consonantais, acarretará ganhos para o entendimento da percepção multimodal, pouco investigado no panorama nacional.

Em síntese, com base nos resultados obtidos, podemos concluir que a presente dissertação possibilitou o direcionamento sob uma perspectiva referente à percepção multimodal da fala, inédita no PB, permitindo a condução de estudos futuros e discussões com o intuito de aprofundar esta análise, na população infantil e adulta, com e sem alteração de fala.

Outro ganho desse estudo foi a contribuição para a clínica fonoaudiológica, especialmente na reabilitação de crianças com TF já que os resultados até aqui obtidos refletem a importância da pista audiovisual na percepção de fala, deste modo os modelos terapêuticos com enfoque nos aspectos auditivos e visuais do fonema alvo, poderiam implicar em resultados terapêuticos satisfatórios.

Acreditamos que há muito a ser feito (discutido, analisado, criticado), destacando-se a necessidade de replicação de estudos com o uso do instrumento ora proposto para que este seja aperfeiçoado a partir das demandas perante sua aplicação. Destacamos, ainda, a necessidade do desenvolvimento do estudo sobre percepção multimodal em diferentes faixas etárias, tanto na população infantil (com e sem alteração de fala) quanto adulta, a fim de obtermos resultados em nível longitudinal, permitindo análises mais profundas quanto ao desenvolvimento da integração de pistas audiovisuais, bem como possíveis dificuldades na população desviante, para o português brasileiro.

Muitos são os desdobramentos desta dissertação, mas, como Vinicius de Moraes escreveu, “Por mais longa que seja a caminhada, o mais importante é dar o primeiro passo”...

REFERÊNCIAS

1. CABBAGE, K. et.al. Speech perception differences in children with dyslexia and persistent speech delay. **Speech Communication**, v. 82, p. 14-25, 2016.
2. RUSSO, I.; BEHLAU, M. **Percepção da Fala**: análise acústica do Português Brasileiro. São Paulo: Lovise, p.57, 1994.
3. HAUPT, C. As fricativas [s],[z], e do português brasileiro. **Letras & Letras**, Uberlândia (MG), v. 24, n.1, p.59-71, 2008.
4. CAGLIARI LC. **Elementos de fonetica do português brasileiro**. Campinas. Tese [Livre-Docência] - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem; 1981.
5. KENT, R. D.; READ, Charles. The acoustic characteristics of consonants. **The acoustic analysis of speech**. 2nd ed., Canada: Singular Thomson Learning, p. 139-88, 2002.
6. RAPHAEL, Lawrence J.; BORDEN, Gloria J.; HARRIS, Katherine S. **Speech science primer: Physiology, acoustics, and perception of speech**. Lippincott Williams & Wilkins, 2007.
7. Kent, Ray D., and Charles Read. As características acústicas das consoantes. In: Kent, Ray D., and Charles Read. **Análise acústica da fala**. Cortez Editora, p. 263-280, 2015.
8. RUSSO, I. C. P. **Acústica e psicoacústica aplicadas à fonoaudiologia**. São Paulo: Lovise, p. 263, 1999.
9. Fant, C. Gunnar M. **Acoustic theory of speech production with calculations based on x-ray studies of Russian articulations**. The Hague, Mouton, 1960.
10. LADD, D. Robert; COUTO, Tradução de Letícia Rebollo; SEARA, Izabel Christine. **O que é prosódia?**. Working Papers em Linguística, Florianópolis, v. 20, n. 1, p. 8-46, set. 2019.
11. JOHNSON, K. **Acoustic and auditory phonetics**. London: Blackwell, 1997.
12. BERTI, Larissa Cristina. DESEMPENHO PERCEPTIVO-AUDITIVO DE CRIANÇAS NA IDENTIFICAÇÃO DE CONTRASTES FÔNICOS. **Alfa, rev. linguíst. (São José Rio Preto)**, São Paulo , v. 61, n. 1, p. 81-103, 2017.
13. BRANCAZIO, L.; MILLER, J. L.; PARÉ, M. A. Visual influences on the internal structure of phonetic categories. **Perception & Psychophysics**, v. 65, n. 4, p. 591-601, 2003.
14. MCGURK, H. ; MACDONALD, J. Hearing lips and seeing voices. **Nature**, Reino Unido, v. 264, n. 5588, p. 746-748, 1976.

15.MARQUES, L.M. ; BOGGIO, P.S. Modulação do efeito McGurk com o uso de estimulação transcraniana por corrente contínua. In: **Jornada de Iniciação Científica**, 7, 2012, Universidade Presbiteriana Mackenzie.

16.MAGNOTTI, J.F. ; BEAUCHAMP, M.S. A causal Inference model explains perception of the McGurk effect and other incongruent audiovisual speech. **PLoS Comput Biol**, v.13, n. 2, 2017.

17.MASAPOLLO, M.; POLKA, L. ; MÉNARD, L. A universal bias in adult vowel perception- by ear or by eye. **Cognition**, v. 166, p. 358-370, 2017.

18.SEKIYAMA, K. ; BURNHAM, D. Issues in the development of Auditory- visual speech perception: adults, infants, and children .In: International Conference on Spoken Language Processing, 8,2004, International Convention Center Jeju, Jeju Island. **Paper [...]** . Grenoble, França: ISCA Archive , p. 1137-1140,2004.

19.OJANEN, V. et al. **Neurocognitive mechanisms of audiovisual speech perception**. Helsinki University of technology, 2005.

20.SEKIYAMA, K.; BURNHAM, D.; TAM, H., ERDENER. Auditory-visual speech perception development in Japanese and English speakers. In: International Conference on Audio;Visual Speech Processing (AVSP) 2003,St. Jorioz, France. **Paper [...]**. Grenoble, França: ISCA Archive, p. 43-47,2003.

21. Rizzolatti, G., e Arbib, M. A. Language within our grasp. **Trends Neuroscience**, v. 21, p.188-194,1998.

22.Buccino, G., Binkofski, F., e Riggio, L. (2004). The mirror neuron system and action recognition. **Brain and Language**, v. 89, p.370–376, 2004.

23.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.

24.SEKIYAMA, K; TOHKURA, Y. Inter-language differences in the influence of visual cues in speech perception. **Journal of Phonetics**, v.21,n.4 , p.427-444,1993.

25.DESJARDINS, R.N. ; WERKER, J.F. 4-month-old female infants are influenced by visible speech. In: International Conference of Infant Studies, Providence RI, 1996.**Abstracts [...]**. Infant Behavior & Development,p. 421,1996.

26.DODD,B.; McIntosh, B.; Erdener, D; Burnham,D. 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.

27. MASSARO, D. W. et al. Bimodal speech perception: An examination across languages. **Journal of Phonetics**, v. 21, n. 4, p. 445-478, 1993.

28. SEKIYAMA, K. ; TOHKURA, Y. McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. **J. Acoust. Soc. Am.**, v 90, p. 1797-1805, 1991.
29. HAYASHI, Y.; SEKIYAMA, K. Native-foreign language effect in the mcgurk effect: A test with chinese and japanese. In: AVSP'98 International Conference on Auditory-Visual Speech Processing, 1998, Terrigal-Sydney, Australia. **Paper [...]**. Grenoble, França: ISCA Archive, p.61-66, 1998.
30. CHEN, Y.; HAZAN, V. Language effects on the degree of visual influence in audiovisual speech perception. In: **Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrueken, Germany**, p. 6-10, 2007.
31. SEKIYAMA, K.; BURNHAM, D. Impact of language on development of auditory-visual speech perception. **Developmental science**, v. 11, n. 2, p. 306-320, 2008.
32. Burnham, D. K., Earnshaw, 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.
33. CHEN, L.; Lei, J. The development of visual speech perception in Mandarin Chinese-speaking children. **Clinical Linguistics & Phonetics**, v. 31, n. 7-9, p. 514-525, 2017.
34. 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.
35. HOCKLEY, N.S. ;POLKA, L. A developmental study of audiovisual speech perception using the McGurk paradigm. **J. Acoust. Soc. Am.**, v. 96, n.5, p. 3309, 1994.
36. ROSENBLUM, L.D.; SCHMUCKLER, M.A.; JOHNSON, J.A. The McGurk effect in infants. **Perception & Psychophysics**, v. 59, n. 3, p. 347-357, 1997.
37. Burnham, D., & Dodd, B. Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. **Developmental Psychobiology**, v.45, p.204–220, 2004.
38. 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.
39. Massaro, D. W., Thompson, L. A., Barron, B., Laren, E. Developmental changes in visual and auditory contributions to speech perception. **Journal of Experimental Child Psychology**, v.1,p. 93–113, 1986.
40. Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. The development of audiovisual multisensory integration across childhood and early adolescence: A high-density electrical mapping study. **Cerebral Cortex**, v. 21, p. 1042–1055, 2011.
41. Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. The development of multisensory speech perception continues into the

late childhood years. **The European Journal of Neuroscience**, v. 33, p.2329–2337, 2011.

42. ISEI-JAAKKOLA, T. Cognition and Physio-acoustic Correlates—Audio and Audio-visual Effects of a Short English Emotional Statement: On JL2, FL2 and EL1. In: International Conference on Natural Language Processing (in Finland),2006, Heidelberg, Berlin. **Paper**,p. 161-173,2006.

43.ASSOCIAÇÃO AMERICANA DE PSIQUIATRIA. **Manual de diagnóstico e estatística dos transtornos mentais DSM-V**. 5. ed. Porto Alegre (RS): Artmed, 2013.

44. KESKE-SOARES, M. et.al. Estudo sobre os ambientes favoráveis à produção da líquida não-lateral /r/ no tratamento do desvio fonológico. **Rev. Soc. Bras. Fonoaudiol.**, São Paulo; v.12, n.1,p. 48-54, 2007.

45. ASSOCIAÇÃO AMERICANA DE PSIQUIATRIA. **Manual Diagnóstico e estatística de transtornos mentais DSM-IV-TR** .4 ed.. Porto Alegre (RS): Artmed,2002.

46.RVACHEW, S.; BROSSEAU-LAPRÉ, F. **Developmental phonological disorders: foundations of clinical practice**. San Diego, CA: Plural Publishing, 2016.

47.RVACHEW, S; JAMIESON, D. G. Perception of voiceless fricatives by children with a functional articulation disorder. **J Speech Hear Disord**, v. 54, n. 2, p. 193-208, 1989.

48.BROWN, C. The interrelation between speech perception andphonological acquisition from infant to adult. **Second language acquisition and linguistic theory**, v. 1, p. 4-64, 2000.

49.FREITAS, C.R; MEZZOMO, C. L; VIDOR, D.C.G.M. Discriminação fonêmica e a relação com os demais níveis linguísticos em crianças com desenvolvimento fonológico típico e com desvio fonológico evolutivo. **CoDAS**, São Paulo , v. 27, n. 3, p. 236-241, 2015 .

50.DESJARDINS, R.N.; ROGERS, J; WERKER, J.F. An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. **Journal of Experimental Child Psychology**, v. 66, n. 1, p. 85-110, 1997.

51. BERTI, L.C; PAGLIUSO, A.; LACAVA, F. Instrumento de avaliação de fala para análise acústica (IAFAC) baseado em critérios linguísticos. **Rev. Soc. Bras. Fonoaudiol**, São Paulo, v. 14, n. 3, p. 305-314, 2009.

52. BERTI, Larissa Cristina. PERCEFAL: instrumento de avaliação da identificação de contrastes fonológicos. **Audiology-Communication Research**, v. 22, 2017.

53. VASSOLER, A.O ; POLLI, L. ; VESCHI, G. V. ; ESPERANDINO. C. ; BERTI, L. C. . Speech production measures in Brazilian Portuguese children with and without Speech Sound Disorder. In: Elena Babatsouli. (Org.). Speech production measures

in Brazilian Portuguese children with and without Speech Sound Disorder.

1ed.Bristol, Reino Unido: **Communication Disorders Across Languages**, v. 1, p. 380-400,2020.

54. SHRIBERG, Lawrence D.; KWIATKOWSKI, Joan. Phonological disorders I: a diagnostic classification system. **Journal of Speech and Hearing Disorders**, v. 47, n. 3, p. 226-241, 1982.

55. WERTZNER, Haydée Fiszbein; AMARO, Luciana; TERAMOTO, Suzana Sumie. Gravidade do distúrbio fonológico: julgamento perceptivo e porcentagem de consoantes corretas. **Pró-Fono Revista de Atualização Científica**, v. 17, n. 2, p. 185-194, 2005.

56. 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.

57. 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.

58. HEARNshaw, S.; BAKER, E.; MUNRO, N. The speech perception skills of children with and without speech sound disorder. **J Commun Disord**, v. 71, n. 61-71, 2018.

59. NIJLAND, L. Speech perception in children with speech output disorders. **Clin Linguist Phon**, v 23, n. 3, p. 222-239, 2009.

60. EDWARDS, J., FOURAKIS, M., BECKMAN, M. E., FOX, R. A. Characterizing knowledge deficits in phonological disorders. **Journal of Speech, Language, and Hearing Research**, v.42, n1, p.169-186, 1999.

61. SIVA, NITHYA et al. A comparison between cerebral-palsied and normal adults in the perception of auditory-visual illusions. **The Journal of the Acoustical Society of America**, v. 98, n. 5, p. 2983-2983, 1995.

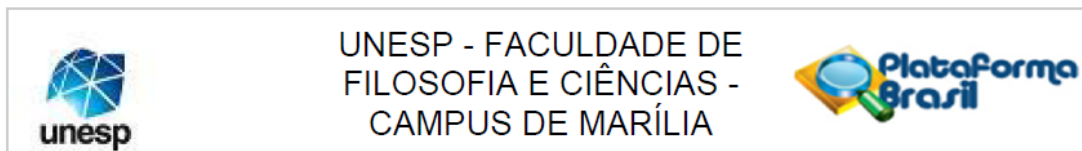
62. ATTONI T, QUINTAS V, MOTA H. Avaliação do processamento auditivo e da discriminação fonêmica em crianças com desenvolvimento fonológico normal e desviante. **Braz. j .otorhinolaryngol**.v.76,n.6, p.762-768, 2010.

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

64. 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.

65. BERTI, L.C, et al. Relationship between speech production and perception in children with Speech Sound Disorders. **Journal of Portuguese Linguistics**. v. 19, n. 1, 2020.

ANEXO A - DOCUMENTO DE APROVAÇÃO DO CEP DA FACULDADE DE FILOSOFIA E CIÊNCIA – UNESP/MARÍLIA



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Percepção Multimodal nas Fricativas Desvozeadas em Crianças com e sem Transtorno Fonológico

Pesquisador: MAYARA FERREIRA DE ASSIS

Área Temática:

Versão: 1

CAAE: 15416819.6.0000.5406

Instituição Proponente: UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.484.427

Apresentação do Projeto:

O objetivo do presente estudo é comparar a integração de pistas auditivas e visuais em tarefa de percepção multimodal de fricativas desvozeadas em crianças com e sem TF. Assumindo que o transtorno fonológico poderia influenciar o modo pelo qual as crianças realizam a integração de pistas audiovisuais, supõe-se que crianças com TF

apresentem pior desempenho na integração de pistas auditivas e visuais em relação a crianças sem TF, já que este poderia estar associado a uma capacidade prejudicada de integrar informações fonológicas de duas fontes de modalidade. Participarão do estudo 30 crianças, sendo 15 diagnosticadas com transtorno fonológico com acometimento na classe das fricativas (G1) e 15 com desenvolvimento típico de fala (G2). Com o uso do software OpenSesame, os estímulos apresentados serão constituídos por gravações de áudio e vídeo da produção de sílabas com padrão silábico simples, de modo que o ataque seja preenchido pelas fricativas: /f/, /s/ e // em contexto da vogal /a/. A tarefa de percepção multimodal envolverá: a apresentação dos estímulos somente por via auditiva; (2) a apresentação dos estímulos somente por via visual; (3a) a apresentação simultânea dos estímulos auditivo e visual de modo congruente; e (3b) a apresentação simultânea dos estímulos auditivo e visual de modo incongruente. Para análise dos dados, serão consideradas: (1) acurácia (% de acertos) e tempo de resposta na tarefa de identificação somente auditiva; (2) acurácia e tempo de resposta em tarefa de identificação somente visual; e (3a) acurácia em tarefa de identificação em

Endereço: Av. Hygino Muzzi Filho, 737
Bairro: Campus Universitário
UF: SP **Município:** MARÍLIA
Telefone: (14)3402-1346 **CEP:** 17.525-900
E-mail: cep.marilia@unesp.br



UNESP - FACULDADE DE
FILOSOFIA E CIÊNCIAS -
CAMPUS DE MARÍLIA



Continuação do Parecer: 3.484.427

integração audiovisual para estímulos congruentes e incongruentes. Uma análise estatística inferencial será feita para verificar se há diferença entre os grupos e as condições de apresentação dos estímulos. Acredita-se que a proposta poderá levar a ganhos científicos, como: (1) fornecer informações inéditas sobre a integração de pistas na percepção dos contrastes entre as fricativas surdas no PB; (2) verificar como crianças com transtorno fonológico percebem informações de estímulos audiovisuais integrados; e (3) auxiliar a clínica fonoaudiológica na análise do desempenho da percepção multimodal de indivíduos com alterações fonoaudiológicas.

Objetivo da Pesquisa:

O objetivo do presente estudo é comparar a integração de pistas auditivas e visuais em tarefa de percepção multimodal de fricativas desvozeadas em crianças com e sem TF.

Avaliação dos Riscos e Benefícios:

Não se aplica.

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:

Foram apresentados e analisados os termos solicitados pelo Comitê de Ética em pesquisa com seres humanos.

Conclusões ou Pendências e Lista de Inadequações:

Diante do exposto o projeto "Percepção Multimodal nas Fricativas Desvozeadas em Crianças com e sem Transtorno Fonológico", CAAE: 15416819.6.0000.5406 será realizados dentro dos critérios exigidos pelo Comitê de Ética em pesquisa com seres humanos.

Sugiro aprovação do projeto.

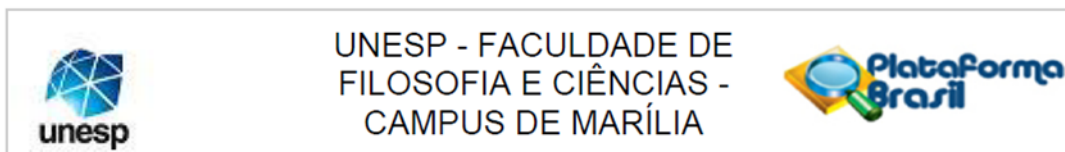
Considerações Finais a critério do CEP:

O CEP da FFC da UNESP de MARÍLIA, em reunião ordinária de 03/07/2019, 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 o projeto de pesquisa Percepção Multimodal nas Fricativas Desvozeadas em Crianças com e sem Transtorno Fonológico

Endereço: Av. Hygino Muzzi Filho, 737
Bairro: Campus Universitário
UF: SP **Município:** MARILIA
Telefone: (14)3402-1346

CEP: 17.525-900

E-mail: cep.marilia@unesp.br



Continuação do Parecer: 3.484.427

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1321833.pdf	10/06/2019 15:03:56		Aceito
Folha de Rosto	FOLHA.pdf	09/06/2019 12:37:38	MAYARA FERREIRA DE ASSIS	Aceito
Projeto Detalhado / Brochura Investigador	Brochura.pdf	06/06/2019 17:18:34	MAYARA FERREIRA DE ASSIS	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TERMO.pdf	06/06/2019 17:15:26	MAYARA FERREIRA DE ASSIS	Aceito
Declaração de Instituição e Infraestrutura	DECLARACAO.pdf	06/06/2019 17:14:10	MAYARA FERREIRA DE ASSIS	Aceito
Cronograma	CRONOGRAMA.pdf	06/06/2019 17:06:23	MAYARA FERREIRA DE ASSIS	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

MARILIA, 04 de Agosto de 2019

Assinado por:
SIMONE APARECIDA CAPELLINI
(Coordenador(a))

Endereço: Av. Hygino Muzzi Filho, 737
Bairro: Campus Universitário
UF: SP **Município:** MARILIA

CEP: 17.525-900

Telefone: (14)3402-1346

E-mail: cep.marilia@unesp.br

ANEXO B - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCL)

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Estamos desenvolvendo uma pesquisa no Centro de Estudos da Educação e da Saúde – CEES – UNESP – Marília, intitulado “Percepção Multimodal nas Fricativas Desvozeadas em Crianças com e sem Transtorno Fonológico” e gostaríamos que participasse da mesma. O objetivo desta é comparar a integração de pistas audiovisuais em tarefa de percepção de fala em crianças com e sem alteração de fala. Participar desta pesquisa é uma opção. No caso de não participarem ou desistirem a qualquer momento, não haverá perda de qualquer benefício no tratamento que estiverem fazendo nesta Universidade.

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

- A) A avaliação é segura, não invasiva, não é dolorosa e não utiliza radiação;
- B) Os senhores estarão contribuindo para a descrição sobre o uso de integração de pistas audiovisuais na percepção de sons da fala;
- C) 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;
- D) 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;
- E) Qualquer indivíduo participante da pesquisa poderá desistir em qualquer momento da pesquisa, sem que haja prejuízo de qualquer natureza.

Eu, _____ portador (a) do RG _____ responsável pelo(a) _____ autorizo a participação da pesquisa intitulada “Percepção Multimodal nas Fricativas Desvozeadas em Crianças com e sem Transtorno Fonológico” a ser realizada no CEES – UNESP – Marília. Declaro ter recebido as devidas explicações sobre a referida pesquisa e concordo que minha desistência poderá ocorrer em qualquer momento da pesquisa sem que haja quaisquer prejuízos físicos, mentais ou no acompanhamento deste serviço. Declaro ainda estar ciente de que a participação é voluntária e que fui devidamente esclarecido (a) quanto aos objetivos e procedimentos desta pesquisa.

Autorizo,

Data: ____/____/____

Assinatura

Responsáveis pela pesquisa:

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

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

ANEXO C: LISTA DE PALAVRAS UTILIZADAS DURANTE A AVALIAÇÃO DA PRODUÇÃO DE FALA, PARA SELEÇÃO DOS SUJEITOS QUE COMPÕS O (GC) E (GE):

IAFAC (Contexto da vogal /a)	Percefal para fricativas (PerceFric)
SAPO	CAÇAR
RABO	CAÇAR
PALHAÇO	CAVAR
LATA	FACA
ZAGA	JACA
NABO	VACA
FACA	FANTA
CHAVE	SANTA
MAPA	FORA
BALA	CHORA
CAPA	FORRO
PAPA	ZORRO
DADO	ROCHA
TACO	ROSA
GATO	SAPA
CUNHADO	CHAPA
VACA	SELO
JACA	GELO
BARATA	VEIA
PRATO	CHEIA
TRAVE	VELA
CRAVO	SELA
PLACA	XIS
CLAVE	GIZ
PASTA	ZANGADA
SANTO	JANGADA
CARTA	-
SALTO	-