



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Lídia Maurício da Silva

**Integração de pistas audiovisuais em tarefa de percepção de fala
em falantes do Português Brasileiro**

São José do Rio Preto
2024

Lídia Maurício da Silva

**Integração de pistas audiovisuais em tarefa de percepção de
fala em falantes do Português Brasileiro**

Tese apresentada como parte dos requisitos para obtenção do título de Doutora em Estudos Linguísticos, junto ao Programa de Pós-Graduação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP - Câmpus de São José do Rio Preto (SP).

Orientadora: Prof^a. Dr^a. Larissa Cristina Berti

São José do Rio Preto
2024

S586i	Silva, Lídia Maurício da Integração de pistas audiovisuais em tarefa de percepção de fala em falantes do português brasileiro / Lídia Maurício da Silva. -- São José do Rio Preto, 2024 123 p. : il., tabs. Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto Orientadora: Larissa Cristina Berti 1. Linguística. 2. Fonoaudiologia. 3. Percepção da Fala. I. Título.
-------	---

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

IMPACTO CIENTÍFICO E SOCIAL

Com a presente tese esperamos gerar novo conhecimento quanto a percepção multimodal da fala em falantes do português brasileiro de forma a servir de base teórica para as áreas da linguística e fonoaudiologia. Especificamente, esperamos que os achados contribuam para que teorias/ modelos multimodais sejam adotados, em especial, na intervenção fonoaudiológica. Os achados tem potencial para contribuir com os objetivos da ODS, de modo, a propiciar serviços de saúde essenciais de qualidade.

Lídia Maurício da Silva

**Integração de pistas audiovisuais em tarefa de percepção de
fala em falantes do Português Brasileiro**

Tese apresentada como parte dos requisitos para obtenção do título de Doutora em Estudos Linguísticos, junto ao Programa de Pós-Graduação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP - Câmpus de São José do Rio Preto (SP).

Comissão Examinadora

Prof^a. Dr^a. Larissa Cristina Berti
UNESP – São José do Rio Preto
Orientadora

Profa. Dra. Aline Mara de Oliveira
UFSC- Florianópolis

Profa. Dra. Ana Luiza Gomes Pinto Navas
FCMSCSP- São Paulo

Prof. Dr. Lourenço Chacon Jurado Filho
UNESP –São José do Rio Preto

Profa. Dra. Luciani Ester Tenani
UNESP –São José do Rio Preto

São José do Rio Preto
04 de Maio de 2023

Aos meus pais **Janira** e **Ilberto**.

AGRADECIMENTOS

Agradeço aos meus pais, **Janira** e **Ilberto**, por me ensinarem a importância do estudo e oferecerem suporte para que eu pudesse trilhar nesse caminho.

À minha orientadora, **Prof^a. Dr^a. Larissa Cristina Berti**, sem sua assistência e tempo dedicado em todas as etapas desse projeto, esta pesquisa nunca teria sido realizada. Agradeço por compartilhar das suas ideias inovadoras e conhecimento comigo, tornando desta, uma experiência inspiradora para mim.

Ao **Dr. Denis Burnham** e **Dra. Monique Tardif**, pelas sugestões e colaborações para o aperfeiçoamento deste trabalho.

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

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

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

Aos professores membros da banca examinadora de qualificação e de defesa, **Aline Mara de Oliveira**, **Ana Luiza Gomes Pinto Navas**, **Lourenço Chacon Jurado Filho**, **Luciani Ester Tenani**, pelas sugestões e por colaborar para o desenvolvimento dessa pesquisa.

Aos **membros do Laboratório de Análise Acústica e Articulatória**, especialmente à **Mayara de Assis**.

Aos **amigos do doutorado**, em especial, **Larissa Paschoal** e **Cássio Esperandino** e aos **amigos e colegas de trabalho**, **Heidi Ferreira**, **Fernanda Siqueira**, **Letícia Alvieri** e **Maria Helena Rastelli**, pelo suporte oferecido, seja como, ouvintes nos momentos em que precisei de escuta ou como mentes colaboradoras. Obrigada pelos momentos compartilhados em ambiente de trabalho ou fora dele.

Aos meus **pacientes e aos seus responsáveis** por serem pacientes com meu processo de desenvolvimento profissional e por serem minha principal motivação por fazer ciência.

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

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

RESUMO

Introdução: o estudo inovador de McGurk e Macdonald (1976) atribuiu à integração audiovisual um importante papel no fenômeno de percepção da fala. Ao longo dos anos, diferentes estudos buscaram replicar o estudo original, buscando descrever o desempenho em tarefa de percepção multimodal, primordialmente, para língua inglesa e japonesa (SEKIYAMA e TOHKURA, 1991; SEKIYAMA e TOHKURA, 1993; SEKIYAMA e BURNHAM, 2004). O presente estudo assume o fenômeno de percepção da fala como sendo de natureza multimodal (MCGURK e MACDONALD, 1976) e se justifica pela escassez de estudos que investigam a percepção multimodal de fala em falantes do português brasileiro. **Objetivo:** investigar como se dá a integração de pistas audiovisuais em função do padrão silábico da língua considerando três grupos de participantes: adultos, crianças típicas e crianças com distúrbio fonológico. Dessa forma, as hipóteses deste trabalho assim se definiram: H1) haveria integração de pistas audiovisuais na percepção de fala nos falantes do PB, marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV⁻¹ comparativamente à condição AV⁺²; H2) a percepção de fala sofreria influência do padrão silábico, ou seja, haveria uma diferença do grau de influência visual em função da complexidade silábica. Espera-se maior grau de influência visual e maior tempo de reação no padrão silábico complexo; e H3) haveria diferenças no desempenho perceptual entre os grupos. Espera-se que os adultos sofram uma maior influência visual comparado às crianças. E, ainda, que as crianças típicas sofram uma maior influência visual comparadas às crianças com distúrbio fonológico. A influência visual seria marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+. **Método:** os participantes selecionados para a presente pesquisa foram 11 adultos típicos, entre 20 e 30 anos; 10 crianças típicas, entre 6 e 8 anos, que apresentaram desenvolvimento típico de linguagem e sistema fonológico efetivamente adquirido; e 10 crianças com Distúrbio

¹ AV- refere-se aos estímulos apresentados de forma incongruente, ou seja, quando a informação visual não corresponde à informação auditiva;

² AV+ refere-se aos estímulos apresentados de forma congruente, ou seja, quando a informação visual corresponde à informação auditiva.

Fonológico (DF) entre 6 e 8 anos, que apresentaram processos de substituição de líquida não lateral por lateral em padrão silábica complexo ou redução de encontro consonantal. Os estímulos auditivos e visuais foram constituídos por sílabas simples e complexas: /pa/, /ta/, /ka/, /pra/, /tra/ e /kra/. O procedimento experimental de percepção constituiu de um teste de identificação composto por quatro etapas distintas: apresentação dos estímulos por (1) via auditiva; (2) via visual; (3) integração audiovisual congruente e (4) integração audiovisual incongruente. Na análise, foram consideradas as porcentagens de respostas corretas (acurácia) e o tempo de reação nas diferentes condições de apresentação dos estímulos. **Resultados:** os objetivos definidos foram explorados em três estudos. No primeiro estudo, o teste GzLM (*generalized linear models*) demonstrou que os adultos apresentaram maior porcentagem de acerto quando os estímulos foram apresentados na condição audiovisual congruente comparado a condição audiovisual incongruente, entretanto, não houve diferença significativa no tempo de reação na comparação entre as duas condições. No segundo estudo, considerando a investigação de crianças típicas, o teste GzLM não demonstrou diferenças significantes de acurácia ou tempo de reação, na comparação entre as condições de apresentação audiovisual congruente e audiovisual incongruente. No entanto, considerando a investigação de crianças com distúrbio fonológico, o teste GzLM demonstrou que as crianças apresentaram maior porcentagem de acertos quando os estímulos foram apresentados na condição audiovisual congruente comparado a condição audiovisual incongruente, mas não houve diferença significativa no tempo de reação na comparação entre as duas condições. Por fim, no terceiro estudo, considerando os dados de acurácia, o teste GzLM demonstrou uma diferença significativa para a condição de apresentação do estímulo somente visual. Para esta condição de apresentação, o desempenho dos adultos se diferenciou do desempenho das crianças típicas e crianças com DF, na medida em que os adultos apresentaram maior porcentagem de erro comparativamente às crianças. Considerando os dados de tempo de reação, o teste GzLM demonstrou uma diferença significativa para as condições de apresentação somente auditiva e somente visual. Para a condição somente auditiva, o desempenho de crianças típicas se diferenciou do desempenho de crianças com DF, na medida em que crianças com DF demoraram mais para eliciar uma resposta; enquanto, para a

condição somente visual, o desempenho dos adultos se diferenciou do desempenho de crianças com DF, na medida em que crianças com DF demoraram mais para eliciar uma resposta. Além disso, perante análise inferencial dos dados foi possível verificar que o grau de influência visual não varia em função do padrão silábico independente do grupo investigado.

Conclusão: o desempenho perceptual para os falantes do PB parece sofrer influência da pista visual, verificada no desempenho da acurácia e no tempo de reação da resposta. Dessa forma, foi possível verificar a integração de pistas audiovisuais na percepção multimodal da fala para os participantes investigados, falantes do Português Brasileiro. Ainda, o desempenho perceptivo de adultos, crianças típicas e crianças com DF, não é dependente do padrão silábico (simples ou complexo) investigado na presente pesquisa. Além disso, há diferença no desempenho perceptual entre os grupos investigados e essa diferença ocorre a depender da condição de apresentação do estímulo.

Palavras-chave: Percepção Multimodal. Integração audiovisual. Integração sensorial. Distúrbio Fonológico.

ABSTRACT

Introduction: the innovative study by McGurk and Macdonald (1976) emphasized audiovisual integration in the phenomenon of speech perception. Several studies over the years have attempted to replicate the original study, investigating performance in a multimodal perception task for English and Japanese languages (SEKIYAMA and TOHKURA, 1991; SEKIYAMA and TOHKURA, 1993; SEKIYAMA and BURNHAM, 2004). The present study builds upon the phenomenon of speech perception as multimodal in nature (MCGURK and MACDONALD, 1976) and is justified and motivated by the scarcity of research investigating multimodal speech perception in Brazilian Portuguese speakers. **Objective:** to investigate how the integration of audiovisual cues takes place as a function of the syllabic pattern of the language, considering three groups of participants: adults, typical children and children with phonological disorders. Thus, the hypotheses of this investigation were defined as follows: H1) there would be integration of audiovisual cues in the speech perception of BP speakers, marked by the lower percentage of correct answers and longer reaction time in the AV-³ condition compared to the AV+⁴ condition; H2) speech perception would be influenced by the syllabic pattern, that is, there would be a difference in the degree of visual influence depending on syllabic complexity. A greater degree of visual influence and a longer reaction time are expected in the complex syllabic pattern; and H3) there would be differences in perceptual performance between groups. Adults are expected to suffer a greater visual influence compared to children. In addition, that typical children suffer a greater visual influence compared to children with phonological disorders. The visual influence would be marked by the lower percentage of correct answers and longer reaction time in the AV- condition compared to the AV+ condition. **Method:** the participants selected for this research were 11 typical adults, between 20 and 30 years old; 10 typical children, between 6 and 8 years old, who presented typical language development and effectively acquired phonological system; and 10 children with

³ AV- refers to stimuli presented incongruently, that is, when the visual information does not correspond to the auditory information;

⁴ AV+ refers to stimuli presented congruently, that is, when visual information corresponds to auditory information.

Phonological Disorder (PD) between 6 and 8 years old, who presented processes of substitution of non-lateral liquid for lateral in a complex syllabic pattern or reduction of consonant clusters. The auditory and visual stimuli consisted of simple and complex syllables: /pa/, /ta/, /ka/, /pra/, /tra/ and /kra/. The experimental procedure of perception consisted of an identification test composed of four distinct stages: presentation of stimuli by (1) auditory pathway; (2) visual pathway; (3) congruent audiovisual integration; and (4) incongruent audiovisual integration. In the analysis, the percentages of correct answers (accuracy) and the reaction time in the different stimulus presentation conditions were considered. **Results:** the defined objectives were explored in three studies. In the first study, the GzLM test (generalized linear models) showed that adults had a higher percentage of correct answers when the stimuli were presented in the congruent audiovisual condition compared to the incongruent audiovisual condition, however, there was no significant difference in reaction time when comparing the two conditions. In the second study, considering the investigation of typical children, the GzLM test showed no significant differences in accuracy or reaction time, in the comparison between the conditions of congruent audiovisual presentation and incongruent audiovisual presentation. However, considering the investigation of children with phonological disorders, the GzLM test showed that children had a higher percentage of correct answers when the stimuli were presented in the congruent audiovisual condition compared to the incongruous audiovisual condition, but there was no significant difference in the reaction time comparing the two conditions. Finally, in the third study, considering the accuracy data, the GzLM test showed a significant difference for the visual-alone stimulus presentation condition. For this presentation condition, the performance of adults differed from the performance of typical children and children with PD, as adults presented a higher percentage of wrong answers compared to children. Considering the reaction time data, the GzLM test demonstrated a significant difference for auditory-alone and visual-alone presentation conditions. For the auditory-alone condition, the performance of typical children differed from the performance of children with PD, as children with PD took longer to elicit a response; while, for the visual-alone condition, the performance of adults differed from the performance of children with PD, as children with PD took longer to elicit a response. Furthermore, in view of

inferential analysis of the data, it was verified that the degree of visual influence does not vary according to the syllabic pattern, regardless of the investigated group. **Conclusion:** the perceptual performance for BP speakers seems to be influenced by visual cues, verified in the accuracy performance and response reaction time. Thus, it was possible to verify the integration of audiovisual cues in the multimodal perception of the investigated participants' speech, speakers of Brazilian Portuguese. Moreover, the perceptive performance of adults, typical children and children with PD is not dependent on the syllabic pattern (simple or complex) investigated in this research. In addition, there is a difference in perceptual performance between the investigated groups, and this difference occurs depending on the stimulus presentation condition.

Keywords: Multimodal Perception. Audio-visual integration. Sensory integration. Phonological Disorder.

LISTA DE FIGURAS

Figura 1: Opções de resposta após apresentação de estímulo de padrão silábico simples.	66
Figura 2: Vídeo com apresentação do movimento articulatorio do fonema /pa/, sem qualquer informação auditiva associada.....	66
Figura 3: Personagens lúdicos e estímulos auditivos correspondentes para padrão silábico simples.	76
Figura 4- Personagens lúdicos e estímulos auditivos correspondentes para padrão silábico complexo.....	76
Figura 5: Estrutura interna da sílaba	102

LISTA DE QUADROS

Quadro 1: Condições de apresentação x estímulos apresentados	65
--	----

LISTA DE GRÁFICOS

Gráfico 1: Resultado do teste *Generalized Linear Models* comparando valores de acurácia em função dos grupos investigados (adultos, crianças típicas e crianças com distúrbio fonológico (atípicas). Barras que não se encostam diferem-se a 95% pelo teste post post-hoc de Holm. 89

Gráfico 2: Resultado do teste *Generalized Linear Models* comparando valores de tempo de reação em função dos grupos investigados (adultos, crianças típicas e crianças com distúrbio fonológico (atípicas). Barras que não se encostam diferem-se a 95% pelo teste post post-hoc de Holm. 90

LISTA DE TABELAS

Tabela 1: Porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico.....	69
Tabela 2: Tempo de reação em tarefa de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico.	70
Tabela 3: Porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico para crianças típicas.	80
Tabela 4: Tempo de reação em tarefa de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico para crianças típicas.	81
Tabela 5: Porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico para crianças com Distúrbio Fonológico.....	83
Tabela 6: Tempo de reação em tarefa de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico para crianças com Distúrbio Fonológico.....	84
Tabela 7: Média e desvio padrão da porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico por grupo.	89
Tabela 8: Média e desvio padrão do tempo de reação dos estímulos em função das condições de apresentação e do padrão silábico por grupo.....	90
Tabela 9: Comparação de médias acurácia obtidas nos testes para cada grupo utilizando o método post-hoc de Holm.	91
Tabela 10: Comparação de médias tempo de reação obtidas nos testes para cada grupo utilizando o método post-hoc de Holm.	93

LISTA DE ABREVIATURAS E SIGLAS

AAA- Articulate Assistant Advanced
AV+- Audiovisual Congruente
AV- - Audiovisual Incongruente
AVC- Acidente Vascular Cerebral
CAAE- Certificado de Apresentação de Apreciação Ética
CCV- Padrão Silábico Complexo
CEES- Centro de Estudos da Educação e Saúde
CV- Padrão Silábico Simples
dB- Decibel
DF- Distúrbio Fonológico
DP- Desvio Padrão
FAPESP- Fundação de Amparo à Pesquisa do Estado de São Paulo
FFC- Faculdade de Filosofia e Ciências
FLMP- Modelo Lógico da Percepção Difusa
G- Grupo
GPEL- Grupo de Pesquisa Estudos sobre a Linguagem
GzLM- Generalized Linear Models
H- Hipótese
IAFAC- Instrumento de Avaliação de Fala para Análise Acústica
LAAc- Laboratório de Análise Acústica e Articulatória
M- Média
P- Participantes
PB- Português Brasileiro
RM- Ressonância Magnética
SNC- Sistema Nervoso Central
SNR- Sinal-Ruído
STS- Sulco Temporal Superior
TCC- Trabalho de Conclusão de Curso
TCLE- Termo de Consentimento Livre e Esclarecido
TDL- Transtorno do Desenvolvimento da Linguagem
TEA- Transtorno do Espectro do Autismo
TM- Teoria Motora da Fala

TRD- Teoria do Realismo Direto

ULTRA- CATS- The Ultrasonographic Contour Analyzer for Tongue Surfaces

UNESP- Universidade Estadual Paulista

SUMÁRIO

INTRODUÇÃO	22
1 TEORIAS UNIMODAIS DE PERCEPÇÃO DA FALA.....	30
2 TEORIA MULTIMODAL DE PERCEPÇÃO DA FALA	36
2.1 PERCEPÇÃO MULTIMODAL: INTRODUÇÃO AO TEMA	37
2.2 A LÍNGUA COMO FATOR INFLUENTE NA PERCEPÇÃO MULTIMODAL	40
2.3 A IDADE COMO FATOR INFLUENTE NA PERCEPÇÃO MULTIMODAL	47
2.4 OS DISTÚRBIOS DA COMUNICAÇÃO COMO FATOR INFLUENTE NA PERCEPÇÃO MULTIMODAL	52
3 OBJETIVOS E HIPÓTESES.....	60
4 ESTUDOS EXPERIMENTAIS.....	62
4.1 ASPECTOS ÉTICOS DOS TRÊS ESTUDOS.....	62
4.2 ESTUDO 1: Desempenho de percepção multimodal de fala em adultos típicos	63
4.2.1 Participantes.....	63
4.2.2 Experimento de percepção.....	64
4.2.3 Forma de análise	67
4.2.4 Resultados.....	68
4.3 ESTUDO 2: DESEMPENHO DE PERCEPÇÃO MULTIMODAL DE FALA EM CRIANÇAS TÍPICAS E COM DÍSTURBIO FONOLÓGICO.....	73
4.3.1 Participantes.....	73
4.3.2 Experimento de percepção.....	75
4.3.3 Forma de análise	78
4.3.4 Resultados: crianças típicas	79
4.3.5 Resultados: crianças com distúrbio fonológico.....	82
4.4 ESTUDO 3: COMPARAÇÃO DE DESEMPENHO DE PERCEPÇÃO MULTIMODAL ENTRE ADULTOS E CRIANÇAS	86
4.4.1 Participantes.....	86
4.4.2 Experimento de percepção.....	86
3.4.3 Forma de análise	87
4.4.3 Resultados.....	88
5 DISCUSSÃO.....	95
5.1 Integração de pistas audiovisuais na percepção multimodal da fala para falantes do PB	96

5.2	Padrão silábico (simples e complexo) e a percepção multimodal da fala	100
5.3	Diferença no desempenho perceptual entre adultos, crianças e crianças com distúrbio fonológico	104
6	CONSIDERAÇÕES FINAIS.....	110
	REFERÊNCIAS.....	112
	PATER, J.; CHRISTINE, S.; WERKER, J.F. THE PERCEPTUAL ACQUISITION OF PHONOLOGICAL CONTRASTS. LANGUAGE, V. 80, N. 3, P. 384-402, 2004.....	117
	ANEXO A - DOCUMENTO DE APROVAÇÃO DO CEP DA FACULDADE DE FILOSOFIA E CIÊNCIA – UNESP/MARÍLIA	120
	ANEXO B - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)	123

INTRODUÇÃO

Os estudos de Aquisição Fônica buscam investigar a fala, fundamentalmente, pelo aspecto de produção (JESUS e SHADLE, 2002; BONATTO, 2007; RINALDI, 2010; CRISTOFOLINI e SEARA, 2012; SILVA, 2017).

Os pesquisadores que buscam estudar a percepção da fala, em sua maioria, investigam como se dá o fenômeno perceptivo em crianças deficientes auditivas, usuárias de implante coclear e/ou aparelho de amplificação sonora (BARZAGHI, BARBOSA e MALT, 2007; ANGELO, BEVILACQUA, MORET, 2010). Portanto, são relativamente poucos os estudos na literatura nacional que objetivam descrever o desenvolvimento típico e desviante da linguagem, no tocante às habilidades de identificação e discriminação auditiva.

A literatura internacional, diferentemente, é mais robusta quanto às produções considerando o modo como percebemos auditivamente os sons da fala (HEARNshaw, BAKER e MUNRO, 2019; HUNTER et al, 2020; JOHNSON e SJERPS, 2021)

Em suma, as pesquisas que investigam o desempenho perceptivo-auditivo em adultos e crianças com desenvolvimento típico de linguagem têm apontado que, de forma geral, as crianças apresentam um desempenho perceptivo-auditivo significativamente inferior (menos acurado) em relação aos adultos no tocante à habilidade de identificar e discriminar contraste fonológicos da língua, principalmente na percepção do detalhe fonético. Adicionalmente, a literatura aponta que a habilidade das crianças em discriminar contrastes não nativos diminui rapidamente na primeira infância, em função de sua experiência linguística, concomitante ao aumento da acurácia na discriminação dos contrastes fônicos de sua língua. Esse declínio de desempenho perceptivo-auditivo parece ocorrer de forma gradual e ocorre em uma ordem sistemática (BROWN, 2000; PATER, STAGER e WERKER, 2004).

No Brasil, em 2017, a Profa. Dra. Larissa Cristina Berti, pesquisadora do Laboratório de Análise Acústica e Articulatória- LAAC, elaborou o instrumento de Avaliação da Percepção de Fala (PERCEFAL), cuja proposta é avaliar o desempenho perceptivo-auditivo, com base em tarefa de identificação, envolvendo os contrastes fonológicos do Português Brasileiro (PB). Este

instrumento preocupou-se em selecionar palavras que cumprissem os seguintes critérios: contrastar os fonemas do PB de modo a comporem pares mínimos de palavras; serem passíveis de representação por meio de gravuras; pertencerem ao vocabulário infantil; e pertencerem ao rol de palavras elencadas em um estudo prévio.

Sob supervisão da Profa. Dra. Larissa Berti estudos vêm sendo realizados, buscando descrever mais detalhadamente como acontece a aquisição perceptivo-auditiva dos contrastes fonológicos. Especificamente, quais contrastes fonológicos são primeiramente percebidos, quais classes de sons podem favorecer ou não essa percepção, e, ainda, no interior de uma mesma classe de sons, quais os contrastes entre os segmentos que podem ou não favorecer sua percepção.

Berti (2017) propôs verificar se uma determinada classe fônica apresenta maior ou menor dificuldade de identificação por crianças com desenvolvimento de típico de linguagem. A autora verificou que o desempenho perceptivo-auditivo de crianças ocorreu de modo gradual e dependente da classe fônica. A maior acurácia na identificação dos contrastes fônicos parece obedecer a seguinte ordem: vogais, sonorantes, oclusivas e fricativas.

Ademais, ao desempenho perceptual de crianças típicas, estudiosos sugerem que o desempenho perceptual exerce influência na maneira em que os sons da fala são produzidos, dessa forma, crianças com Distúrbio Fonológico (DF) poderiam apresentar dificuldade no plano perceptual (NIJLAND, 2009; CABBAGE et al, 2016; HEARNshaw et al, 2018).

Em seu estudo, Berti et al (2020) propôs investigar a relação entre produção e percepção de fala, em crianças com DF. Os autores ressaltam em seu trabalho que os erros de produção não espelham os erros de percepção de fala. A análise inferencial dos dados demonstrou uma correlação positiva entre erros de produção e erros de percepção de fala dependente da classe fonológica. Sendo que, a classe das fricativas, particularmente, foi a única classe que apresentou correlação positiva entre os erros de produção e percepção de fala, em crianças com DF.

Outros estudos (FREITAS, MEZZOMO e VIDOR, 2015; HEARNshaw et al, 2018; ASSIS et al, 2021) buscaram descrever e/ou comparar o desempenho

perceptivo-auditivo de crianças com desenvolvimento típico de linguagem e crianças com distúrbio fonológico.

A título de exemplificação, em Assis et al (2021), desenvolvemos a descrição de desempenho perceptivo-auditivo em crianças com e sem DF na classe das oclusivas. O estudo teve por objetivo comparar o desempenho perceptivo-auditivo de crianças com e sem distúrbio fonológico em tarefa de identificação dos contrastes entre os fonemas oclusivos. O instrumento PERCEFAL foi utilizado para coleta de identificação de estímulos oclusivos do Português Brasileiro. Os resultados demonstraram que as crianças com distúrbio fonológico apresentaram pior acurácia em relação às crianças sem alteração e, ainda, um maior tempo de reação para os acertos.

Entretanto, ainda são poucos os estudos nacionais (SILVA, PACHECO e CAGLIARI, 2017; ASSIS, 2021 e a presente tese) que buscam investigar a percepção da fala considerando mais de uma via de entrada da informação, por exemplo, auditiva e visual.

Dentre os estudos existentes, majoritariamente as descrições quanto ao uso de pistas audiovisuais na percepção de fala são realizadas internacionalmente (HOCKLEY e POLKA, 1994; DESJARDINS e WERKER, 1996; SEKIYAMA e BURNHAM, 2004; e CHEN, 2017). Considerando a lacuna existente, atualmente, os primeiros estudos que propuseram uma descrição para a língua portuguesa vêm sendo realizados sob orientação da Profa. Dra. Larissa Berti no LAAc (ASSIS, 2021 e a presente tese).

Em Assis (2021), a autora comparou o desempenho de crianças com e sem DF na tarefa de percepção multimodal envolvendo pistas auditivas e visuais em fricativas desvozeadas. Em seu estudo, foi possível verificar que crianças com DF apresentaram pior desempenho e maior tempo de reação na tarefa de percepção multimodal, sugerindo dificuldades na integração de pistas advindas de fontes bimodais.

Meu desenvolvimento na ciência, naturalmente, acompanhou a trajetória e os avanços científicos de um grupo de pesquisadores, como membro do Grupo de Pesquisa Estudos sobre a Linguagem (GPEL/ CNPq) e do subgrupo de pesquisa Aquisição Fonológica. Iniciei minhas produções científicas na investigação dos aspectos de produção de fala, utilizando diferentes recursos tecnológicos de investigação (análise acústica e ultrassonografia) até estreitar

minha relação entre os dados de produção de fala e os dados de percepção de fala, ao realizar atividades como bolsista técnica. Na presente tese, ao assumir a teoria de percepção multimodal como referencial teórico assumo também uma relação indissociável entre aspectos de produção e percepção de fala.

Minha trajetória na ciência se iniciou ainda na graduação em Fonoaudiologia na Faculdade de Filosofia e Ciências (FFC) da Universidade Estadual Paulista (UNESP) – câmpus de Marília, quando sob supervisão da profa. Dra. Larissa Berti, realizei minha primeira pesquisa científica como membro de iniciação científica.

A referida pesquisa realizada em 2014, resultou no trabalho de conclusão de curso (TCC) intitulado: *“Aquisição fonológica das oclusivas: em busca de pistas fonético-acústicas”* e teve como propósito identificar quais as pistas acústicas relevantes para marcar o contraste entre as oclusivas. Para isso, investigamos gravações de 20 crianças, na faixa etária de 5 a 7 anos de idade, com desenvolvimento típico de linguagem e aquisição efetiva dos sons oclusivos. Como recurso de investigação, utilizamos o *software* PRAAT, para extração de sete parâmetros: centroide, variância, assimetria, curtose, VOT, duração da oclusão e duração do estouro. Grosso modo, os resultados demonstraram que os parâmetros acústicos utilizados para diferenciar as oclusivas são dependentes do contexto vocálico e possibilitaram inferir que as crianças da faixa etária estudada não marcaram o contraste entre as oclusivas utilizando todos os parâmetros acústicos.

No mestrado, desenvolvido no Programa de Pós-Graduação em Fonoaudiologia da FFC/UNESP, na área de atuação: Distúrbios da Comunicação Humana, mantivemos a parceria e produzimos, em 2017, a dissertação intitulada: *“Parâmetros ultrassonográficos do movimento da língua em 14 fonemas consonantais do Português Brasileiro”*. A proposta do referido trabalho foi propor parâmetros ultrassonográficos quantitativos que pudessem apreender diferenças de modo e ponto de articulação entre os 14 fonemas linguais do PB, especificamente, buscamos estimar valores de referência e um padrão de variabilidade quanto ao contorno de superfície de língua na produção de fala típica. Para isso, foram investigadas produções de falas de 20 participantes, de 20 a 30 anos de idade, sem alterações de linguagem. Os dados de produção de fala foram coletados e analisados com o uso do ultrassom e dos

softwares AAA (*Articulate Assistant Advanced*) e Ultra- CATS (*The Ultrasonographic Contour Analyzer for Tongue Surfaces*). Foram analisados três parâmetros: índice de anterioridade, média global e índice de anterioridade normalizado. Em suma, os resultados demonstraram que não houve um único índice que pudesse diferenciar todos os pontos e modos de articulação simultaneamente, portanto, a aplicação clínica para análise dos distúrbios fonológicos necessitaria do uso de pelo menos dois índices ultrassonográficos: um para distinguir o ponto e outro para distinguir o modo de articulação.

Após a conclusão do mestrado, fui contemplada com uma Bolsa de Treinamento Técnico- nível III, concedida pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). As atividades correspondentes ao projeto de pesquisa, intitulado: “*Viabilização dos experimentos de percepção da fala em crianças*” foram desenvolvidas na instituição FFC/UNESP, no período de junho à agosto de 2018, sob orientação da profa. Larissa Berti. Como membro do Grupo de Pesquisa Estudos sobre a Linguagem (GPEL) e do subgrupo de pesquisa Aquisição Fonológica (coordenado pela orientadora da presente pesquisa), participei ativamente de reuniões e atividades em grupos compostos por alunos de graduação e pós graduação do curso de Fonoaudiologia e por bolsista de treinamento técnico. Cada membro do grupo desenvolve estudos individuais e conjuntos quanto aos aspectos de produção e/ou percepção da fala. As reuniões visaram discussões dos trabalhos em desenvolvimento e as atividades corresponderam à execução de tarefas necessárias para elaboração dos estudos, como por exemplo, levantamento bibliográfico em bases de dados e coleta e tabulação de dados. A participação nas discussões destes trabalhos possibilitou um melhor entendimento acerca das teorias e instrumentos existentes para análise da percepção da fala, enquanto, a participação na elaboração dos estudos possibilitou expansão e organização de um banco de dados. O banco de dados é composto por crianças com distúrbio fonológico advindas do Estágio Supervisionado de Fonologia Clínica (UNESP- CEES/ Marília- SP). Cabia a bolsista realizar avaliações audiológicas e investigação do movimento da língua com o uso de um ultrassom. Ao final do período estipulado, além da execução dos objetivos propostos, foi realizada a coleta e tabulação de dados referentes à forma como produzimos e percebemos a fala a depender da classe de sons para o PB.

Atualmente, como doutoranda do programa de pós-graduação em Estudos Linguísticos do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista– UNESP - Câmpus de São José do Rio Preto, sob orientação da profa. Dra. Larissa Berti, buscamos investigar o fenômeno da percepção da fala pelo viés teórico da multimodalidade.

A proposta da presente tese, portanto, visa investigar como se dá a integração de pistas audiovisuais em função da complexidade silábica. Para a execução da proposta, elaboramos um teste de percepção multimodal, com base na literatura internacional existente.

Uma grande referência para os estudos dessa natureza, é o Prof. Dr. Denis Burnham, da Western Sydney University. O pesquisador realizou visita ao LAAc- UNESP, em 2018. Sua visita rendeu discussões quanto aos desafios envolvidos no estudo. Como resultado, a elaboração do instrumento e análise dos dados ficou mais criteriosa, a partir de reflexões metodológicas realizadas no *workshop* de pesquisa ocorrido durante a visita do Dr. Burnham. O pesquisador contribuiu, especificamente, na elaboração do procedimento experimental das crianças, na medida em que salientou a influência da alfabetização na percepção da fala. Optamos, portanto, em não utilizar a representação gráfica dos estímulos e sim figuras correspondentes aos estímulos apresentados. A correspondência entre os estímulos e às figuras foi realizada de forma lúdica e será descrita na seção 4.3.2, p. 71. Para melhor compreensão do procedimento experimental realizado com os adultos e com as crianças (típicas e com distúrbio fonológico), didaticamente, os aspectos metodológicos serão apresentados na presente tese em capítulos distintos. Além disso, o pesquisador contribuiu sugerindo o uso do parâmetro de investigação *grau de influência visual* (SEKIYAMA e BURNHAM, 2008). A definição e o cálculo do referido parâmetro será descrita nos aspectos metodológicos da pesquisa (seção 4.2.3. p.64).

A justificativa da proposta, portanto, se dá pela escassez de estudos que investigam a percepção multimodal de fala em falantes do português brasileiro. Acredita-se que a proposta poderá levar a ganhos científicos, como: (1) fornecer informações inéditas sobre a integração de pistas audiovisuais na percepção no desenvolvimento típico para o PB; (2) verificar como crianças com distúrbio fonológico percebem informações de estímulos audiovisuais integrados; e (3)

servir como benefício para base teórica nas áreas da linguística e da fonoaudiologia. Especificamente, quanto às áreas da linguística, os resultados contribuem para sustentar a natureza multimodal da percepção de fala, de modo que teorias/ modelos multimodais sejam adotados. A investigação da percepção da fala sob viés da teoria multimodal, pode contribuir para o conhecimento quanto ao desenvolvimento das representações dos aspectos de produção e percepção da fala, na medida em que possibilita investigar a estrutura linguística considerando uma nova entrada de investigação. Adicionalmente, os achados poderão ser utilizados como referencial teórico quanto ao desenvolvimento típico considerando o desempenho de adultos e crianças em tarefa de percepção multimodal de fala e oferecer auxílio à clínica fonoaudiológica, na análise do desempenho da percepção multimodal de participantes com distúrbio fonológico.

Buscando apresentar os passos do desenvolvimento da pesquisa, até a elaboração do presente texto, esta tese organiza-se em seis capítulos.

O primeiro capítulo, **teorias unimodais de percepção da fala**, apresenta os principais estudos que contribuíram quanto ao conhecimento da percepção da fala, considerando o aspecto auditivo como objeto da percepção.

O segundo capítulo, **teoria multimodal de percepção da fala**, apresenta o referencial teórico adotado para a execução do presente estudo. Inicialmente, apresentamos os principais estudos que contribuíram quanto ao conhecimento da percepção da fala, atribuindo ao gesto articulatório um importante papel para o fenômeno da percepção da fala. Em sequência, apresentamos o estado da arte e revisão teórica da literatura que buscou investigar o fenômeno de percepção da fala considerando a integração de pistas audiovisuais. Para melhor apresentação das informações literárias, esse capítulo é didaticamente dividido em: (1) percepção multimodal: introdução ao tema; (2) a língua como fator influente na percepção multimodal; (3) a idade como fator influente na percepção multimodal; e (3) os distúrbios da comunicação como fator influente na percepção multimodal.

O terceiro capítulo, **objetivos e hipóteses**, apresenta os objetivos e hipóteses do trabalho e informa ao leitor que a presente tese, será norteadada pela execução de três estudos.

O quarto capítulo, **estudos experimentais**, é dividido em quatro seções: (1) aspectos éticos dos três estudos; (2) estudo 1: desempenho de percepção

multimodal de fala em adultos típicos; (3) estudo 2: desempenho de percepção multimodal de fala em crianças típicas e com distúrbio fonológico; e (4) comparação de desempenho de percepção multimodal entre adultos e crianças. Dessa forma, o capítulo apresenta os aspectos éticos, procedimentos metodológicos envolvidos e resultados encontrados em cada um dos três estudos. Na apresentação do procedimento metodológico apresentaremos a descrição dos participantes, descrição do procedimento experimental e a forma de análise dos dados. A apresentação dos resultados contempla os achados quanto às análises descritivas e inferenciais referentes aos três estudos realizados, que serão apresentados em valores absolutos e percentuais, em gráficos e tabelas.

O quinto capítulo, **discussão**, apresentaremos uma discussão geral, com base nos achados referentes aos três estudos realizados. Dialogaremos com a literatura que deu base para nossas hipóteses norteadoras e elaboração do experimento e levantaremos hipóteses explicativas para os principais achados verificados.

Finalmente, no sexto capítulo, **considerações finais**, retomaremos as hipóteses e os objetivos do presente estudo e discutiremos em que medida as hipóteses e os objetivos foram, ou não, confirmados. Ao final do capítulo apresentaremos as limitações do presente trabalho e sugestões para que novos trabalhos sejam realizados buscando contribuir com o conhecimento quanto ao tema.

1 TEORIAS UNIMODAIS DE PERCEPÇÃO DA FALA

Inicialmente, perceber a fala era visto como sinônimo de capacidade auditiva, portanto, os primeiros estudos que buscaram investigar a percepção da fala foram conduzidos considerando primordialmente a audição e seus aspectos acústicos (LIBERMAN et al., 1967; FUJISAKI e KAWASHIMA, 1969; PISONI, 1973, 1974; COLE e SCOT, 1974). Desse modo, as teorias que buscaram investigar a percepção da fala nessa vertente são denominadas teorias auditivas de percepção da fala (BERTI, 2008).

Em teorias auditivas de percepção da fala informações de natureza acústica são essenciais e suficientes para que um ouvinte perceba a diferença entre sons da fala. Para esta concepção o aspecto de produção dos sons da fala, ou seja, os movimentos articulatórios são considerados secundários, sendo utilizados somente em circunstâncias difíceis (BERTI, 2008).

A seguir, serão apresentadas as principais ideias e estudos que buscaram investigar a percepção da fala considerando seu aspecto auditivo.

Dentro dessa perspectiva, um grupo de estudiosos buscou investigar a relação entre percepção de fala e memória. (FUJISAKI e KAWASHIMA, 1969; PISONI, 1973, 1974; COLE e SCOT, 1974). O interesse em investigar a referida relação partiu da premissa de que haveria uma diferença na forma em que percebemos os sons. Essa diferença é verificada na medida em que o sistema auditivo é capaz de diferenciar sons ambientais de sons da fala, bem como perceber diferenças entre os diferentes sons da fala, dessa forma, agrupando e categorizando os sons percebidos como equivalentes (LIBERMAN et al., 1957, 1967).

Segundo ideias apresentadas em Fujisaki e Kawashima (1969) a capacidade de diferenciar dois sons de categorias fonéticas distintas ocorre quando comparamos as suas propriedades auditivas/acústicas armazenadas em nossa memória de curto prazo. Entretanto, haveria uma diferença no funcionamento da memória de curto prazo, a depender do grupo de sons investigado como, por exemplo, as pistas acústicas para a identificação de consoantes são muito breves, e, portanto, não seriam armazenadas na memória de curto prazo, enquanto as pistas acústicas para a identificação de vogais são

mais longas, e, portanto, seriam armazenadas na memória de curto prazo, dessa forma, a memória de curto prazo seria utilizada na percepção de vogais.

Pisoni (1973) aprofundou o conhecimento quanto ao tema quando propôs investigar 16 participantes adultos, falantes da língua inglesa. Para isso, elaborou um experimento que contemplava dois testes: 1) teste de identificação absoluta; e 2) teste de reconhecimento de memória por comparação atrasada. O experimento consistiu na apresentação de dois estímulos, que seriam apresentados com intervalos distintos. Os estímulos contemplavam 2 classes de vogais (vogais longas e curtas, representadas por /i/ e /I/) e 2 classes de consoantes (/ba/ e /da/ e /ba/ e /pa/) e os atrasos de intervalos eram: 0,0; 0,25; 0,50; 1,0 e 2,0 segundos. As vogais apresentavam estado estacionário longo e curto. No contexto longo, as vogais tinham 300 ms de duração e, no curto, 50 ms. As consoantes /ba/ e /da/ foram sintetizadas seguindo os seguintes parâmetros: os valores de F1, F2 e F3 da vogal eram 743 Hz, 1620 Hz e 2862 Hz, respectivamente; com vozeamento de 40 ms em 150 Hz. Enquanto as consoantes /ba/ e /pa/ foram sintetizadas variando o VOT, que partia de 0 até 60ms. A diferença de cada estímulo era de 10 ms. Primeiramente, os participantes deveriam decidir entre as categorias apresentadas: vogais longas, vogais curtas, oclusivas vozeadas e oclusivas bilabiais (teste de identificação absoluta) e, posteriormente, se eram estímulos iguais ou diferentes (teste de reconhecimento de memória por comparação atrasada). Como tendência geral, os resultados demonstraram uma diferença no uso da memória de curto prazo para identificar sons de consoantes e vogais. Entre categorias, a memória de curto prazo parece facilitar a discriminação de vogais e contribuir pouco para discriminar as consoantes; enquanto, dentro das categorias, a memória de curto prazo parece não interferir, pois a identificação é alta tanto para vogais quanto para consoantes e não é afetada pelo aumento de tempo. Ou seja, há diferenças no uso de pistas de vogais e consoantes pela memória auditiva de curto prazo de modo que ela é melhor para as pistas acústicas de vogais do que de consoantes.

Portanto, o experimento de Pisoni (1973) permite inferir que um ouvinte utiliza da memória de curto prazo para perceber categorias fonéticas distintas, mas não utiliza da memória de curto prazo para discriminar os sons de uma mesma categoria acústica.

A diferença no funcionamento da memória de curto prazo para percepção dos sons da fala intrigou vários estudiosos (LIBERMAN et al, 1967; LIBERMAN, 1970; e PISONI, 1973) que levantaram hipóteses a respeito da existência do que eles chamaram de um codificador especial. Para estes autores, a percepção de alguns sons não ocorre utilizando a memória de curto prazo e sim utilizando uma representação fonética ou um código fonético.

Com o propósito de aprofundar a investigação sobre como os ouvintes respondem as diferenças acústicas entre os sons da fala, Pisoni (1974) elabora um novo experimento. Este novo experimento contemplou as oclusivas /pa/ e /ba/ apresentadas sem variação do VOT e sob cinco diferentes variações de VOT totalizando sete estímulos. O experimento testou duas condições de discriminação: entre categorias e dentro de categorias. Nove participantes adultos, falantes do inglês, deveriam dizer se os estímulos apresentados eram iguais ou diferentes. Como medida de análise utilizou-se o tempo de reação de resposta. Os resultados demonstraram quando os estímulos foram percebidos como iguais, os informantes apresentaram tempo de reação menor para estímulos idênticos (AA) do que estímulos parcialmente iguais (Aa). A variação do VOT nos estímulos apresentados possibilitou verificar como os participantes usaram o chamado código fonético.

Cole e Scot (1974), por sua vez, reforçam a ideia de que perceber as diferenças entre os sons da fala não remete somente a habilidades envolvendo aspectos auditivos, portanto, haveria estruturas representacionais responsáveis pela percepção dos sons da fala. Em seu estudo, os autores sugerem a existência de três tipos pistas envolvidas na percepção da fala: pistas invariantes; pistas condicionadas pelo contexto e pistas promovidas pelo envelope da forma de onda. De modo geral, as pistas invariantes referem-se às pistas acústicas de um fonema independentemente do seu contexto vocálico, ou seja, as pistas invariantes seriam as responsáveis por fazer um ouvinte perceber os sons individualmente, entretanto, elas dão oferecem suporte para perceber o ordenamento dos sons. Atribui-se às pistas condicionadas pelo contexto a função de possibilitar a percepção dos sons em um ordenamento temporal da fala. O envelope da forma da onda, por sua vez, possibilitaria perceber os aspectos prosódicos da fala, por exemplo, as alterações de amplitude do som.

Considerando os estudos existentes na época e os avanços considerando os chamados codificadores especiais para a percepção da fala, Kuhl & Miller (1975) questionam se a percepção da fala seria uma habilidade específica dos seres humanos. A fim de testar a habilidade de perceber sons da fala por animais, os autores propõem investigar se chinchilas seriam capazes de perceber a diferença entre /t/ e /d/ em sílabas CV. Os estímulos foram apresentados em diferentes condições: sílabas /t/ e /d/ sem alteração de padrões acústicos; /t/ e /d/ produzidos por diferentes falantes; e /t/ e /d/ em contextos vocálicos e com alteração do VOT. O experimento foi realizado com chinchilas treinadas, chinchilas não treinadas e com humanos. O desempenho apresentado por todos os animais e pelos humanos foram similares. Os resultados desses experimentos demonstram que a distinção de sonoridade em consoantes em posição inicial não é uma habilidade exclusiva de humanos. Esses resultados apontam para o fato de que tais distinções realizadas pela percepção da fala não é um “privilégio” de humanos (NISHIDA, 2012).

Internacionalmente, nas décadas de 1970 e 1980 houve um grande impulso quanto aos estudos (EIMAS et al, 1971; EDWARDS, 1974; STREETER, 1976; TREHUB, 1976; BARTON, 1980; EILERS, GAVIN, OLLER, 1982; MEHLER, 1985) que se propusessem a descrever o desenvolvimento das habilidades de discriminação e identificação de contrastes fônicos de determinada língua.

Tarefa de discriminação refere-se à habilidade para perceber uma diferença entre dois sons, em que não é exigida do ouvinte a categorização dos sons em questão. Por outro lado, tarefa de identificação refere-se à habilidade de categorização dos sons (HAZAN e BARRET, 2000).

Considerando o desenvolvimento da habilidade de discriminação, a literatura internacional aponta que bebês são capazes de discriminar não somente contraste fônicos de sua língua nativa, como também contrastes fônicos de outras línguas (EIMAS et al, 1971; STREETER, 1976; TREHUB, 1976; EILERS, GAVIN, OLLER, 1982; MEHLER, 1985). Além disso, a capacidade em discriminar contrastes não nativos diminui rapidamente na primeira infância, considerando sua experiência linguística, ao passo que ocorre o aumento da acurácia na discriminação dos contrastes fônicos de sua língua (WERKER e POLKA, 1993).

Considerando o desenvolvimento da habilidade de identificação, pesquisadores têm apontado que a habilidade de crianças para categorizar contraste fonológicos de sua língua se desenvolve não somente de maneira gradual, como também obedece a uma ordem sistemática de aquisição (EDWARDS, 1974; BARTON, 1980; BROWN e MATTHEWS, 1997).

São relativamente poucos os estudos na literatura nacional que objetivam descrever o desenvolvimento típico e desviante da linguagem, no tocante às habilidades de discriminação e identificação auditiva.

Berti (2017) propôs investigar o desempenho perceptivo-auditivo de crianças em tarefa de identificação de contrastes fônicos do PB. Especificamente, buscou verificar se uma determinada classe fônica apresenta maior ou menor dificuldade de identificação. Para isso, foram investigadas 66 crianças (de ambos os gêneros), entre 4- 5 anos de idade. Quatro tarefas de identificação foram testadas usando o instrumento PerceFAL (BERTI, 2011). O PerceFAL é composto por um subconjunto de quatro experimentos: (a) PerceVog (que avalia a identificação entre vogais tônicas); (b) PerceOcl (que avalia a identificação entre oclusivas); (c) PerceFric (que avalia a identificação entre fricativas); (d) PerceSon (que avalia a identificação entre sonorantes). Nesse instrumento estão envolvidas, preferencialmente, palavras dissilábicas paroxítonas, familiares às crianças, contendo os 19 fonemas consonantais do PB na posição acentuada. A tarefa consistiu na apresentação do estímulo acústico, e posterior, escolha da gravura correspondente à palavra apresentada, dentre duas possibilidades de gravuras. Como critério de análise considerou-se: 1) a acurácia perceptivo-auditiva; 2) o tempo de reação dos erros e acertos; e 3) o reconhecimento de padrões com o uso do escalonamento multidimensional e matriz de similaridade. Os resultados demonstraram que o desempenho perceptivo-auditivo de crianças ocorreu de modo gradual e é dependente da classe fônica. A maior acurácia na identificação dos contrastes fônicos parece obedecer a seguinte ordem: vogais, sonorantes, oclusivas e fricativas. O tempo de reação das respostas corretas foi menor do que das respostas incorretas (exceto para a classe das vogais). No interior da classe das vogais, o parâmetro ântero-posterior parece exercer um importante papel na saliência perceptual. Para a classe das obstruintes e sonorantes (nasais e líquidas), as pistas acústicas que marcam o vozeamento (no caso das obstruintes) e o modo de

articulação (no caso das sonorantes) são mais robustas perceptualmente do que as pistas de ponto de articulação.

Os estudos acima resenhados fornecem informações importantes quanto à percepção da fala considerando seu aspecto auditivo. Apesar dos avanços científicos, estudiosos (FOWLER, 1996; ODEN e MASSARO, 1978; LIBERMAN, MATTINGLY, 1985) fomentaram novas questões quanto à forma que percebemos os sons, apontando a necessidade de outras pistas, senão as acústicas, para a percepção dos sons.

Dessa forma, investigações considerando o papel dos articuladores na percepção da fala começaram a ser conduzidas (FOWLER, 1996; ODEN e MASSARO, 1978; e LIBERMAN, MATTINGLY, 1985), dando início à proposta que seria conhecida como “solução articulatória para o problema da percepção da fala” (NISHIDA, 2012). As investigações propostas nesta vertente são compreendidas pela Teoria Multimodal da Fala. As ideias centrais da Teoria Multimodal da fala e os principais achados de estudos que se propuseram a investigar a percepção da fala seguindo esse viés teórico serão apresentados no próximo capítulo.

2 TEORIA MULTIMODAL DE PERCEPÇÃO DA FALA

É bem esclarecida na literatura, que a fala envolve uma alta complexidade sensório- motora como, por exemplo, a interação de diferentes estruturas do trato vocal de forma coordenada, organizada e planejada (BARBARENA et al, 2014). Dessa forma, durante décadas, a percepção da fala foi considerada um fenômeno que ocorre separadamente na comunicação (NISHIDA, 2012).

Entretanto, estudos que assumem a linguagem como um processo dinâmico (ALBANO, 2001; BECKNER, 2009), atribuem à percepção de fala um papel ativo na comunicação. Esses estudos assumem que o movimento articulatório tem um papel na compreensão da linguagem, ou seja, oferece suporte para a percepção da fala. Diante dessa perspectiva, a produção e a percepção de fala são habilidades inseparáveis (PEROZZO, KUPSKE, 2021).

Perspectivas dinâmicas contemplam a gramática e o uso da língua, envolvendo tanto o aspecto de produção quanto o aspecto de percepção da fala. Por intermédio da gramática torna-se possível compreender e descrever aspectos da estrutura de determinada língua, a saber: aspectos fonéticos, fonológicos, morfológicos, sintáticos, semânticos e estilísticos (ELLIS, 2008).

Investigar o funcionamento do aspecto perceptivo, articulatório e simbólico como integrado contribui para o conhecimento sobre como ocorrem as representações considerando os aspectos de produção e percepção da fala (PEROZZO, KUPSKE, 2021).

O aspecto fonológico de uma língua contempla uma base biológica, que corresponde às habilidades motoras envolvidas na fala; e uma base cognitivo-linguística, que corresponde ao aprendizado dos sons em funcionamento na língua (STOEL-GAMMON, 2011). Dessa forma, a percepção da fala, possibilita ao ouvinte desenvolver o aspecto fonológico que, por sua vez, conduz ao desenvolvimento de outros aspectos da estrutura da língua como, por exemplo, o aspecto lexical.

Dessa forma, investigar a percepção da fala sob viés da teoria multimodal, que considera o fenômeno como resultado de uma integração de pistas de diferentes vias sensoriais, pode contribuir para o conhecimento quanto ao desenvolvimento das representações dos aspectos de produção e percepção da

fala, na medida em que possibilita investigar a estrutura linguística considerando uma nova entrada de investigação.

A seguir, será apresentado o referencial teórico adotado para a execução da presente pesquisa, a saber: a percepção multimodal. A referida apresentação abrange breve introdução ao tema, seguida de três seções: na primeira seção (2.2) “*a língua como fator influente na percepção multimodal*”, apresentaremos estado da arte quanto aos estudos que consideraram a língua, especificamente, a complexidade silábica como fator influente neste processo; na segunda seção (2.3) “*a idade como fator influente na percepção multimodal*”, abordaremos estado da arte quanto aos estudos que consideraram o desenvolvimento dessa habilidade frente a diferentes idades; e na terceira seção (2.4) “*os distúrbios da comunicação como fator influente na percepção multimodal*”, abordaremos os estudos que investigaram a percepção em crianças com distúrbio fonológico.

2.1 PERCEPÇÃO MULTIMODAL: INTRODUÇÃO AO TEMA

A descoberta do efeito McGurk (MCGURK e MACDONALD, 1976) inaugurou uma nova perspectiva para o entendimento sobre a percepção da fala. Ao invés de conceber a percepção como sendo de natureza unimodal, ou seja, com o uso de uma única via sensorial, descobriu-se a importância da informação visual na percepção da fala. O efeito McGurk refere-se a uma ilusão perceptual como o produto da integração de informações auditiva e visual, demonstrando que a informação visual é usada na percepção da fala mesmo quando a informação auditiva é clara e íntegra (SEKIYAMA e BURNHAM, 2004).

O efeito McGurk consiste na apresentação simultânea de informações audiovisuais. São chamados de estímulos congruentes, aqueles em que a informação visual corresponde à informação auditiva (AV+), enquanto, são chamados de estímulos incongruentes, aqueles em que a informação visual não corresponde à informação auditiva (AV-). No efeito McGurk são apresentados estímulos incongruentes, por exemplo, o movimento articulatório da sílaba /ga/ é apresentado de forma sincronizada ao estímulo auditivo correspondente à produção da sílaba /ba/. Essa integração de estímulos divergentes resulta na percepção da sílaba /da/, ou seja, o resultado do que é percebido é o relato de

uma ilusão de fusão das informações apresentadas (MARQUES e BOGGIO, 2011).

A partir dos achados inovadores apresentados por McGurk, diferentes teorias e modelos foram propostos e argumentam quanto à existência de outras pistas, senão acústicas, necessárias para a percepção dos sons da fala. As investigações que consideram o aspecto acústico insuficiente para descrever a forma como percebemos os sons e assumem a necessidade do aspecto de produção dos sons da fala são compreendidas pela Teoria Multimodal da Fala (BOVO et al, 2009). A percepção multimodal refere-se à integração de diferentes informações sensoriais que serão percebidas como uma única informação. Os principais sentidos envolvidos no fenômeno de percepção de fala são audição e visão, que, em sincronia, proporcionam maior compreensão do conteúdo apresentado (BRANCAZIO et al, 2003).

As ideias centrais da Teoria Motora da Fala (TM), Teoria do Realismo Direto (TRD), Modelo Lógico da Percepção Difusa (FLMP) - modelo *Fuzzy* e os principais achados de estudos que se propuseram a investigar a percepção da fala seguindo esse viés teórico serão apresentados a seguir.

Diferentemente, das teorias auditivas de percepção de fala, a teoria motora da fala atribui ao movimento articulatorio um importante papel na percepção. Dentro dessa perspectiva, o primitivo da percepção da fala seriam os gestos fonéticos pretendidos, que seriam representados no cérebro como comandos motores invariáveis. Os comandos motores invariáveis, por sua vez, referem-se tanto aos elementos físicos de um som da fala, quanto a um grupo de traços distintivos. (LIBERMAN e MATTINGLY, 1985).

Considerando essa teoria, para perceber os gestos pretendidos haveria um elemento biológico especializado (módulo especializado) que permitiria aos ouvintes relacionar o sinal acústico com os processos motores subjacentes. Esse elemento seria, portanto, especial, por ocorrer apenas na fala. Diante dessa teoria, o ouvinte percebe os gestos pretendidos e não os gestos realizados. “Dessa forma, perceber uma sentença seria perceber um padrão específico de gestos pretendidos” (LIBERMAN e MATTINGLY, 1985).

Cabe ressaltar duas considerações apresentadas dentro dessa teoria. A primeira diz respeito à diferença do tempo entre os movimentos dos gestos envolvidos em determinado símbolo, ou seja, consoantes e vogais sucessivas

se sobrepõem umas às outras no tempo. A segunda refere-se ao efeito da coarticulação. Ao considerar o efeito da coarticulação, hipotetiza-se que um mesmo gesto pode possuir mais de um correspondente acústico.

Em suma, dentro dessa perspectiva a fala é considerada como especial e os objetos da percepção estariam relacionados a processos motores subjacentes ao aspecto auditivo.

Uma segunda teoria que também buscava investigar a percepção de fala considerando o seu aspecto articulatório é a Teoria do Realismo Direto (TRD) proposta por Fowler (1996). Apesar de trazer o componente articulatório na investigação da percepção de fala, apresenta ideias diferentes da TM.

A primeira diferença refere-se ao primitivo da percepção de fala sendo gesto articulatório e não o movimento articulatório. Nesta vertente, o gesto articulatório não é sinônimo de movimento articulatório, dessa forma, além de assumir um aspecto fonético assume também um aspecto fonológico. Especificamente quanto ao aspecto fonológico, os gestos dão origem à estrutura do sinal acústico da fala, ou seja, o ouvinte desprenderia da fala os padrões gestuais ou os modos de coordenação gestuais, revelando as generalizações fonológicas (FOWLER, 1996).

A segunda diferença refere-se a questões relacionadas ao indivíduo. Essa perspectiva concebe que os sistemas perceptuais têm uma função universal, ou seja, considera-se que o indivíduo é capaz de perceber a fala, assim como percebe o mundo em geral, de maneira real e direta (GIBSON, 1966). A partir das ideias de Gibson (1966), Fowler (1996) apresenta em sua teoria o conhecimento quanto à percepção multissensorial, ou seja, como a percepção da fala estaria atrelada a outros sentidos.

Assim como a TM, a TRD não assume a percepção de fala como um fenômeno passivo e paralelo a produção de fala, ambas, apresentam ideias que aproximam a percepção da produção da fala. Apesar de a TRD considerar a percepção de fala como uma habilidade multissensorial, a produção e a percepção tornam-se indissociáveis (NISHIDA, 2012).

Considerando o aspecto multimodal no fenômeno de percepção de fala, Oden e Massaro (1978) apresentam o Modelo Lógico da Percepção Difusa (FLMP) - modelo *Fuzzy* da percepção da fala, que busca explicar e formalizar a integração de pistas visuais e auditivas na percepção de fala. Dentro desta

perspectiva, a percepção da fala ocorreria de acordo com um algoritmo geral e esse processo se daria a partir de três operações: a avaliação, integração e decisão.

Especificamente, o sistema sensorial transformaria um evento físico em traços que seriam avaliados integrados e comparados na memória. As pistas visuais e auditivas seriam avaliadas independentemente e posteriormente integradas, a fim de fornecer suporte para a identificação e a interpretação da informação recebida (SILVA, PACHECO e CAGLIARI, 2017).

Em suma, dentro dessa perspectiva, para que uma informação seja percebida, diferentes fontes de informações oferecem suporte para a sua identificação e a interpretação. O referido modelo é defendido, por teóricos que o descrevem como uma forma de investigar as fontes de informação realmente usadas na percepção da fala (SILVA, PACHECO e CAGLIARI, 2017), porém, é criticada, na medida em que é descrita como um modelo muito geral (TRISTÃO, FEITOSA, 2003).

Estudos recentes que assumem o fenômeno da percepção de fala como sendo de natureza multimodal (BURNHAM e DODD, 1998; CHEN e HAZAN, 2007; MARQUES e BOGGIO, 2011; MAGNOTTI e BEAUCHAMP, 2017; MASAPOLLO, POLKA e MÉNARD, 2017; ASSIS, 2021) têm apontado para a existência de fatores que interferem na integração de pistas audiovisuais, tais como: a língua, a idade e os distúrbios da comunicação. A seguir, serão descritos o estado da arte quanto aos estudos que buscaram investigar os três fatores descritos.

2.2 A LÍNGUA COMO FATOR INFLUENTE NA PERCEPÇÃO MULTIMODAL

Diversos estudos (SEKIYAMA e TOHKURA, 1991; HAYASHI e SEKIYAMA, 1998; CHEN e HAZAN, 2007) buscaram replicar o efeito McGurk a fim de verificar como se dá o fenômeno da percepção multimodal em diferentes línguas.

Sekiyama e Tohkura (1991) investigaram a percepção de 10 japonesas, de idade entre 22 e 27 anos, 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/) produzidos por uma falante japonesa. As gravações

de áudio e vídeo foram realizadas separadamente e, posteriormente, sincronizadas, o que resultou em 100 estímulos audiovisuais. Os estímulos foram apresentados para cada participante, individualmente. Os participantes foram solicitados a escrever o que eles escutaram e não o que eles viram, após a apresentação de cada estímulo. Durante o estudo piloto, os autores verificaram poucas ocorrências do efeito *McGurk*, ou seja, os participantes avaliados não foram fortemente influenciados pelo estímulo visual e não perceberam os estímulos como resultantes de uma fusão de informações audiovisuais, na condição de apresentação incongruente. Posteriormente, duas condições de fala foram adicionadas: (1) estímulo auditivo com adição de ruído; e (2) estímulo auditivo sem adição de ruído. Mais uma vez, foram verificadas poucas ocorrências do efeito McGurk na condição de estímulo sem adição de ruído. Diferentemente, na condição com adição de ruído, foram verificadas maior ocorrência do efeito, demonstrando uma percepção da fusão de informações visual e auditiva dos estímulos apresentados. Os autores concluíram que falantes nativos do japonês usaram a informação visual na percepção da fala apenas quando o componente auditivo do estímulo não estava perfeitamente inteligível.

Em estudo posterior, Sekiyama e Tohkura (1993), replicaram a metodologia usada em Sekiyama e Tohkura (1991) a fim de investigar 20 participantes da língua japonesa e, adicionalmente, 20 participantes da língua inglesa. Os participantes tinham idade entre 20 e 30 anos. Cabe ressaltar que os participantes da língua japonesa tinham conhecimento da língua inglesa, bem como os participantes da língua inglesa tinham conhecimento da língua japonesa. Dessa forma, os estímulos apresentados foram 10 sílabas japonesas (/ba/; /da/; /ga/; /pa/; /ta/; /ka/; /ma/; /na/; /ra/ e /wa/) e seus representantes para a língua inglesa. Os estímulos gravados com sílabas japonesas e inglesas foram apresentados para todos os participantes. Os participantes foram solicitados a escrever o que eles escutaram e não o que eles viram após apresentação de cada estímulo audiovisual. Considerando os estímulos gravados com sílabas da língua japonesa, os resultados mostraram uma diferença significativa no desempenho perceptual entre os participantes das duas línguas investigadas, sendo que, os japoneses apresentaram menor influência da via visual quando comparados aos americanos. Considerando os estímulos gravados com sílabas

da língua inglesa, não foram verificadas diferenças significantes no desempenho perceptual entre os falantes da língua japonesa, comparados aos falantes da língua inglesa. Ou seja, os resultados demonstraram que os falantes americanos automaticamente integraram a pista visual com a pista auditiva em sua língua nativa, enquanto, os japoneses, incorporam a pista visual muito menos do que os americanos para perceber sua língua nativa.

Majoritariamente, as investigações quanto à percepção multimodal da fala são investigadas considerando o desempenho de falantes da língua inglesa e japonesa (SEKIYAMA e TOHKURA, 1991; HAYASHI e SEKIYAMA, 1998; CHEN e HAZAN, 2007). Entretanto, alguns estudos se propuseram a investigar o desempenho em tarefa de percepção multimodal por falante do espanhol (FUSTER-DURAN, 1996), alemão (FUSTER-DURAN, 1996), chinês (SEKIYAMA, 1997), finlandês (SAMS, et al, 1998), italiano (BOVO et al, 2009), hebraico (TAITELBAUM-SWED e FORTICK, 2016) (que será resenhado na seção 2.3, p. 47) e, mais recentemente, do português brasileiro (SILVA, PACHECO e CAGLIARI, 2017; ASSIS, 2021).

Fuster- Duran, (1996) propôs investigar a integração audiovisual da fala entre línguas, especificamente entre o espanhol e o alemão. Para isso selecionou 120 participantes entre 20- 30 anos de idade. Os participantes foram divididos em quatro grupos: 1) composto por 30 falantes nativos do alemão que seriam testados com estímulos em língua nativa; 2) composto por 30 falantes nativos do alemão que seriam testados com estímulos em língua não nativa; 3) 30 participantes falantes do espanhol que seriam testados com estímulos em língua nativa; e 4) 30 falantes do espanhol que seriam testados com estímulos em língua não nativa. Os estímulos do experimento contemplaram sílabas em padrão simples (CV): /ba/, /ma/, /da/, /na/, /ga/, /Na/, /la/, /ra/, /rra/ e /Ra/), sílabas em padrão complexo (CCV): (/bra/, /dra/, /gra/, /bla/, /dla/ e /bla/) e palavras de ambas as línguas que contemplavam as sílabas selecionadas. As sílabas /ra/ e /rra/ foram testadas apenas no espanhol, enquanto /Ra/ foi testado apenas no alemão, além disso, no alemão /d/ é normalmente articulado como uma alveolar, e /R/ como uma fricativa velar, enquanto, para o espanhol /d/ é frequentemente articulado como uma dental, /r/ como um tepe alveolar e /rr/ como uma alveolar vibrante. Os estímulos foram gravados por uma falante nativa de cada língua e apresentados considerando quatro condições de apresentação: somente

auditiva, somente visual, audiovisual congruente e audiovisual incongruente. Os resultados demonstraram que houve o efeito McGurk para as duas línguas investigadas, sendo que a magnitude da influência visual foi maior para estímulos não nativos do que para estímulos nativos. Os participantes se ancoraram mais fortemente em pistas visuais, quando foram apresentadas palavras em língua não nativa. Cabe ressaltar, que foram verificadas diferenças no desempenho perceptual considerando as diferenças fonéticas e fonológicas de cada língua, por exemplo, o tepe /ra/ não ocorre em posição de início de palavra no espanhol, dessa forma, os nativos do espanhol identificaram o estímulo como /gra/ em 50% das vezes, na condição apenas auditiva, enquanto, quando o /ra/ foi sincronizado com /ba/, /ga/ ou /da/, os nativos do espanhol perceberam como /bra/ 80% das vezes, /dra/ 70% e /gra/ 67%.

Sekiyama (1997) propôs investigar o desempenho perceptual de falantes da língua chinesa. Para isso, investigou 14 participantes de língua nativa chinesa, de 19 a 30 anos de idade, que viviam no Japão de quatro meses a seis anos. Todos os participantes tinham conhecimento da língua japonesa. Os estímulos selecionados foram 10 sílabas (/ba/, /pa/, /ma/, /da/, /ta/, /na/, /ga/, /ka/ e /ra/ que foram produzidas por dois falantes, uma mulher falante da língua japonesa e um homem falante do inglês. Os estímulos foram gravados e editados de modo a compor quatro formas de apresentação: somente visual, somente auditivo, audiovisual congruente e audiovisual incongruente. Os participantes foram investigados considerando às produções das duas línguas (inglês e japonês). Os achados da presente pesquisa demonstraram que assim como os falantes da língua japonesa, o desempenho de percepção multimodal por falantes do chinês apresentou menor influência visual para estímulos gravados por falantes da língua japonesa comparados aos estímulos gravados por falantes da língua inglesa. Verificou-se que chineses apresentaram menor efeito de ilusão, ou seja, um efeito McGurk mais fraco para os estímulos da língua japonesa comparados aos estímulos da língua inglesa. Apesar de os achados terem demonstrado o desempenho de chineses em tarefa de percepção multimodal, os estímulos apresentados não referem-se à produções de falantes da língua chinesa.

Sams et al (1998) propuseram investigar 65 mulheres entre 19 e 45 anos de idade, falantes do finlandês em tarefa de percepção multimodal da fala. Os

estímulos selecionados foram (1) sílabas isoladas /pa/ e /ka/; (2) palavras que contemplam os fonemas /p/ e /k/ em diferentes contextos vocálicos; e (3) logatomas contendo os fonemas /p/ e /k/ em diferentes contextos vocálicos. Os estímulos foram apresentados nas condições: somente visual, somente auditivo, audiovisual congruente e audiovisual incongruente. Considerando as condições somente auditiva e somente visuais as participantes apresentaram alta porcentagem de acerto dos estímulos; enquanto para a condição audiovisual, as participantes foram influenciadas pela pista visual. O efeito McGurk foi verificado para sílabas isoladas, palavras e logatomas.

Bovo et al. (2009) buscaram investigar o desempenho de falantes da língua italiana em tarefa de percepção multimodal. Para isso, selecionaram 10 participantes entre 18 e 27 anos de idade (quatro homens e seis mulheres). Os estímulos selecionados foram oito sílabas: /ba/, /da/, /ga/, /pa/, /ta/, /ka/, /ma/ e /na/, produzidas por uma falante italiana. Os estímulos foram editados e apresentados na condição audiovisual incongruente. Os participantes foram orientados a anotar o que percebiam do estímulo apresentado. As respostas que não correspondiam ao estímulo auditivo foram consideradas como “influenciadas visualmente”. Os resultados permitiram verificar a ocorrência do efeito McGurk para a língua italiana, na medida em que houve um valor significativo de respostas que correspondem a uma ilusão perceptual.

A investigação da percepção multimodal da fala no Brasil encontra-se em estágios iniciais. Silva, Pacheco e Cagliari (2017) propuseram investigar a percepção audiovisual da fala considerando, especificamente, a forma como falantes do PB percebem os sons fricativos. Para isso, selecionaram 13 participantes, sendo 8 mulheres e 5 homens com idade entre 24 e 45 anos. Os estímulos selecionados foram palavras dissílabas de estrutura silábica CVCV. A primeira consoante correspondia a uma das seis fricativas (/f/, /v/, /s/, /z/, /ʃ/ ou /ʒ/), a segunda consoante correspondia a uma das quatro oclusivas selecionadas (/k/, /g/, /b/ ou /p/) e a primeira e segunda vogal correspondiam a uma das três vogais selecionadas (/a/, /i/ ou /u/). Os estímulos constituíram pares mínimos e pares análogos, na qual a oposição foi dada considerando o vozeamento e o ponto de articulação das fricativas. As palavras foram inseridas na frase-veículo “digo (palavra alvo) baixinho” e foram produzidas por um falante do PB, de 20 anos de idade. Os estímulos foram apresentados aos participantes

sem manipulação do sinal acústico e com manipulação do sinal acústico. A manipulação do sinal acústico ocorreu de duas formas: 1) foram reduzidos três taxas de ruído dos sons fricativos: 25%, 50% e 75%; e 2) foram aplicadas três taxas de manipulação da frequência: valor mínimo, médio e máximo para cada fricativa. Primeiramente, os estímulos sem manipulação foram apresentados nas condições: somente auditiva, somente visual e audiovisual. Posteriormente, os estímulos foram apresentados nas condições somente auditivo e audiovisual incongruente, sendo que a informação auditiva correspondia ao estímulo auditivo manipulado e a informação visual correspondia à gravação original. Os resultados demonstraram que quando as fricativas foram apresentadas sem manipulação do sinal acústico o desempenho perceptivo para as condições somente auditiva e audiovisual foram parecidas. O desempenho na condição somente visual, por sua vez, apresentou uma menor porcentagem de acerto comparativamente às demais condições de apresentação. Entretanto, quando as fricativas foram apresentadas diante da manipulação do sinal acústico, foi verificado um aumento no desempenho perceptual das fricativas na condição de apresentação audiovisual. De maneira geral, a porcentagem de identificação das fricativas foi maior quando os estímulos foram apresentados na condição de apresentação audiovisual.

Ademais ao estudo de Silva, Pacheco e Cagliari (2017) os primeiros estudos vêm sendo realizados dentro do subgrupo de pesquisa Aquisição Fonológica, no Laboratório de Análise Acústica e Articulatória (LAAc) de Marília, com o trabalho de Assis (2021) (que será resenhado na seção 2.4, p. 47) e a presente tese.

Os resultados referentes aos estudos apresentados nesta seção demonstram que a língua é um fator influente na integração de pistas durante a percepção da fala. Uma possível explicação para a influência da língua sobre a percepção multimodal seria a própria estrutura linguística da língua em questão, especificamente, a complexidade silábica da língua analisada seria uma possibilidade explicativa para a diferença no desempenho perceptivo, conforme aponta (SEKIYAMA e BURNHAM, 2004).

Os movimentos articulatorios necessários para a produção de sílabas da língua japonesa exercem pouca função distintiva quando comparados aos movimentos necessários para a produção das sílabas do inglês, ou seja,

fonologicamente, não há consoantes visualmente distintas, como /v/ ou /d/ no japonês (SEKIYAMA e BURNHAM, 2004). Ademais, a língua japonesa pode ser considerada uma língua de baixa complexidade silábica, pois apresenta em seu inventário apenas cinco vogais e sílabas de estrutura simples, não contemplando encontros consonantais. Portanto, japoneses podem ser menos influenciados por informações visuais, tendo em vista a menor complexidade fonético-fonológica da língua (SEKIYAMA e BURNHAM, 2004).

Sekiyama e Burnham (2008) ressaltam em seu estudo que quanto menor a complexidade silábica e maior a distintividade visual dos fonemas de determinada língua, menor a necessidade do uso de pistas visuais para percepção da fala. Em seu estudo, os autores propuseram investigar o desempenho de adultos e crianças em tarefa de percepção multimodal da fala. Para isso, conduziram dois experimentos. No experimento 1: foram investigados 48 participantes, entre 18 e 29 anos de idade, 24 falantes da língua inglesa e 24 falantes da língua japonesa. Os estímulos contemplaram as sílabas /ba/, /da/ e /ga/ que foram produzidas por quatro falantes, sendo um homem e uma mulher para cada língua investigada. Os falantes foram testados em estudo piloto, de modo que a inteligibilidade média da fala auditiva e visual fossem aproximadamente equivalentes entre os dois idiomas. Os estímulos foram apresentados sob quatro formas de apresentação: somente auditiva, somente visual, audiovisual congruente e audiovisual incongruente. Como forma de análise considerou-se a porcentagem acerto e o tempo de reação de respostas. Para os estímulos audiovisuais incongruentes foram consideradas corretas às respostas que se ancoraram na pista auditiva, enquanto, às demais opções de respostas foram consideradas como influências visuais. Ademais, foi calculado o grau de influência visual (pela diferença entre a porcentagem correta da condição audiovisual congruente da porcentagem correta da condição audiovisual incongruente). Os resultados foram apresentados considerando os dois grupos de falantes e considerando se os estímulos apresentados foram produzidos por falantes da língua nativa ou não nativa e demonstraram que a influência visual para os participantes de língua inglesa foi maior do que para os participantes de língua japonesa, embora, ambos os grupos apresentaram maior influência visual para estímulos não nativos. No experimento 2: foram investigados 32 participantes, 16 crianças falantes da língua inglesa e 16 falantes

da língua japonesa para cada uma das três faixas etárias investigadas (6, 8 e 11 anos). Os estímulos investigados e forma de análise foram semelhantes aos do experimento 1. Os resultados demonstraram que aos seis anos o grau de influência visual foi relativamente menor comparado as demais faixas etárias, embora, equivalente para crianças de língua japonesa e inglesa. Para os participantes de língua inglesa, a influência visual aumentou entre seis e oito anos, enquanto para os participantes da língua japonesa, permaneceu quase constante ao longo da idade. Na comparação entre crianças e adultos, verificou-se que o grau de influência visual aumentou entre seis e oito anos, mas apenas para as crianças da língua inglesa. Ademais, houve um aumento da influência visual para estímulos não nativos, a vantagem da influência visual não aumentou entre os 11 anos e a idade adulta. Os resultados sugerem que crianças falantes do inglês entre seis e oito anos podem necessitar de maior número de pistas sensoriais para resgatar informações de uma língua de maior complexidade silábica e visualmente distintivas. Dessa forma, a complexidade silábica justificaria o aumento acentuado no uso das informações visuais para percepção de fala, por crianças dessa faixa etária na língua inglesa.

Os resultados dos estudos resenhados anteriormente demonstram que a língua é um fator influente na integração de pistas durante a percepção de fala. Assim como a língua, a idade tem sido apontada como um fator influente no modo como combinamos pistas auditivas e visuais na percepção da fala, na medida em que alguns estudos apontam diferenças significantes no uso de pistas visuais para a percepção de fala na comparação entre bebês, crianças e adultos (BURNHAM et al, 1991; TAITELBAUM-SWED; LEAH FOSTICK, 2016; CHEN e HAZAN, 2007). Na próxima seção, serão apresentados os estudos dentro da perspectiva multimodal que consideram a idade como um fator influente na percepção multimodal.

2.3 A IDADE COMO FATOR INFLUENTE NA PERCEPÇÃO MULTIMODAL

Similarmente à influência da língua na integração audiovisual na percepção da fala, a idade também tem sido apontada como um fator influente na integração de pistas durante essa percepção.

Alguns estudos apontam diferenças no uso de pistas visuais para percepção de fala na comparação entre bebês, crianças e adultos (DESJARDINS e WERKER, 1996; ROSENBLUM, SCHMUCKLER e JOHNSON, 1997; BURNHAM e DODD, 1998). O fenômeno da percepção multimodal da fala ao nascer ainda não é bem esclarecido. Entretanto, estes estudos demonstram que bebês são capazes de integrar pistas visuais e auditivas para perceber sons da fala. Além disso, pesquisas apontam uma mudança no padrão de atenção visual durante a percepção da fala ao longo do primeiro ano de vida (SALINAS-MARCHANT e MACLEOD, 2021).

Bebês de 2 a 4 meses de idade apresentam maior fixação do olhar para os olhos do falante e aumentam a fixação do olhar para o rosto do falante quando o movimento da boca é congruente com o estímulo auditivo. Kuhl e Meltzoff (1984) propuseram investigar as habilidades de percepção audiovisual de 32 bebês de 18 a 20 semanas de idade. Os participantes assistiram a dois vídeos apresentados lado-a-lado, que correspondiam aos movimentos articulatórios das vogais /a/ e /i/, de forma sincronizada um estímulo auditivo era apresentado, correspondente a uma das vogais. A fixação visual dos bebês para os dois vídeos foi gravada em vídeo. Os resultados mostraram que o estímulo auditivo influenciou a fixação visual dos bebês. Os bebês olharam por mais tempo para o rosto que combinava com o som apresentado.

Aos cinco meses, os bebês apresentam capacidade para discriminar estímulos congruentes de incongruente. Em seu estudo, Rosenblum, Schmuckler e Johnson (1997) buscaram investigar o efeito McGurk em 20 bebês de cinco meses de idade. Todos os participantes tinham o inglês como língua materna. Os participantes foram investigados individualmente, em uma sala tratada acusticamente, foram posicionados nos colos de seus pais. Uma câmera foi posicionada de forma a registrar os movimentos faciais dos participantes e a partir dos registros, foi verificada a fixação visual dos bebês quando os estímulos foram apresentados. Anteriormente ao teste, os bebês foram “adaptados” ao procedimento experimental. O procedimento de adaptação consistiu na apresentação de um estímulo congruente (áudio /va/ e visual /va/) que foi apresentado repetidamente até que o bebê demonstrasse desinteresse, ou seja, o tempo de fixação do olhar fosse menor do que a metade do tempo de apresentação do estímulo, durante duas vezes consecutivas. Após atingirem o

referido “nível de desinteresse pelo estímulo” foram apresentados os estímulos incongruentes (áudio /ba/ e visual /va/ e áudio /da/ e visual /va/). O tempo de fixação do olhar durante fase de adaptação foi utilizado como parâmetro na investigação do presente estudo. Quando os estímulos do teste eram apresentados, os bebês demonstravam maior atenção, que foi verificada na comparação do tempo de fixação do olhar. O objetivo principal foi verificar se os bebês discriminavam um estímulo congruente de estímulos incongruentes. Os resultados demonstram que os bebês foram capazes de discriminar estímulos congruentes de incongruente, uma vez, que houve um aumento significativo do tempo de fixação de olhar para o estímulo áudio /da/ combinado ao visual /va/ comparado ao estímulo áudio /va/ combinado ao visual /va/. O estímulo áudio /ba/ combinado ao visual /va/, foi percebido como similar ao estímulo congruente.

A partir dos seis meses de idade, os bebês mudam gradualmente sua atenção dos olhos para a boca do falante. Tomalski et al. (2013) buscaram investigar 32 bebês ingleses. Os participantes foram divididos em dois grupos. O primeiro contendo 16 bebês de 6-7 meses e o segundo contendo 16 bebês de 8-9 meses. Os estímulos selecionados foram as sílabas /ba/ e /ga/ produzidas por uma falante do inglês e foram apresentados de forma congruente e incongruente. Os participantes foram posicionados no colo dos pais em frente a um monitor com rastreador ocular. Os movimentos dos olhos foram analisados por áreas de fixação: boca, olhos e toda a face. A soma do tempo total de fixação do olhar foi calculada para cada participante. Os resultados verificados mostraram que bebês de 6 a 9 meses aumentaram sua fixação do olhar para a boca do falante somente para apresentação de estímulos audiovisual incongruentes.

Considerando estímulos de fala nativa, aos 12 meses de idade, os bebês diminuem sua fixação do olhar para a boca do falante. Enquanto, considerando estímulos apresentados em língua não nativa, os bebês mantêm uma alta fixação do olhar para a boca do falante. Pons, Bosch e Lewkowicz (2015) buscaram investigar e comparar o desempenho de bebês monolíngues e bilíngues em tarefa de percepção multimodal da fala. Para isso, selecionaram 60 bebês. Os participantes foram agrupados considerando suas faixas etárias: 4, 8 e 12 meses. Os bebês foram posicionados em um assento infantil em frente a um monitor de computador. Os movimentos oculares foram registrados por um

rastreador ocular. Os estímulos consistiram de um monólogo recitado por duas atrizes, uma falante do catalão- espanhol e outra falante do inglês americano. Cada participante assistiu a dois vídeos, um em sua língua nativa e outro em língua não-nativa. Considerando o desempenho de bebês monolíngues, os resultados demonstraram que os bebês fixaram mais o olhar para os olhos do falante aos quatro meses e mais para a boca do falante aos oito meses considerando estímulos de língua nativa e não nativa. Aos 12 meses, fixaram mais o olhar para a boca do falante somente para os estímulos apresentados de língua não nativa. Enquanto, considerando o desempenho de bebês bilíngues, os resultados demonstraram que os bebês fixaram o olhar para os olhos e para a boca igualmente aos quatro meses e fixaram mais o olhara para a boca do falante aos oito meses. Aos 12 meses os bebês mantiveram uma alta fixação de olhar para a boca do falante para estímulos de língua nativa e não nativa.

Ademais ao desempenho perceptivo audiovisual em bebês, a literatura aponta um possível desenvolvimento no uso da integração de pistas, conforme aumento da idade. (HOCKLEY e POLKA, 1994; SEKIYAMA, BUMHAM e ERDENER, 2003).

Hockley e Polka (1994) investigaram o uso da informação visual durante a percepção de fala em diferentes faixas etárias. Foram investigados 46 participantes que 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 demonstram que a influência da informação auditiva diminui com o aumento da idade, enquanto a influência da informação visual e a integração das pistas audiovisuais aumentaram conforme aumento da idade.

Em um segundo estudo, Sekiyama, Bumham e Erdener (2003) buscaram investigar o desempenho da percepção multimodal em falantes das línguas japonesa e inglesa. Foram analisados 48 adultos, na faixa etária entre 18 e 29 anos e 16 crianças, na faixa etária de 6 anos. Os adultos e crianças foram divididos em grupos a depender da sua língua nativa, dessa forma, o estudo consistiu de dois grupos de adultos e dois grupos de crianças. Os estímulos

investigados foram /ba/, /da/ e /ga/ produzidas por quatro falantes, um homem e uma mulher falantes do inglês e um homem e uma mulher falantes do japonês. Os estímulos foram editados de forma a apresentar somente pista visual, somente pista auditiva e pistas audiovisuais: congruentes e incongruentes. Os estímulos foram apresentados sob diferentes níveis de inteligibilidade, para isso, testou-se estímulos com e sem adição de ruído. Todos os participantes foram instruídos a assistir e ouvir os estímulos e, posteriormente, determinar o que haviam percebidos dos estímulos apresentados. A análise inferencial buscou verificar a interação entre as idades, as línguas e as condições de apresentação investigadas. Os resultados demonstraram que estímulos apresentados com adição de ruído exigiram maior grau de influência visual, calculada pela diferença entre a porcentagem correta da condição AV+ da porcentagem correta da condição AV-, de falantes da língua inglesa, do que falantes da língua japonesa, independente da faixa etária investigada. Adicionalmente, os adultos falantes da língua inglesa demonstraram maior grau de influência visual, comparados aos adultos falantes da língua japonesa, para estímulos sem adição de ruído; enquanto, para o grupo de crianças, essa diferença não foi verificada. Dessa forma, os autores concluem que existe uma diferença perceptual entre as línguas investigadas, entretanto, essa diferença é marcada somente com o aumento da idade, sendo que falantes do inglês são mais influenciados pela pista visual comparados aos falantes da língua japonesa.

Taitelbaum-Swed e Fortick (2016), por sua vez, partiram da premissa que a percepção da fala muda ao longo da vida, desenvolvendo-se e amadurecendo desde a infância até a idade adulta e buscaram investigar o desenvolvimento da percepção audiovisual de fala dos quatro aos 80 anos de idade. Participaram do estudo 77 falantes do hebraico que foram subdivididos em cinco faixas etárias: 4-5, 8-9, 20-30, 40-55 e 65-80. Para evitar os efeitos da experiência linguística relacionada à idade, os estímulos selecionados contemplaram palavras monossilábicas e logatomas. Os estímulos foram apresentados em condição visual, auditivo e audiovisual. Para os logatomas, a acurácia visual e auditiva foi menor em comparação as palavras. Segundo os autores, esse resultado era esperado, já que, para as palavras, o léxico também está disponível como fonte informativa para a percepção de fala, além das pistas auditivas e visuais. Especificamente, para as palavras, os autores apresentam 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. Perante os dados observados, os autores levantam a hipótese explicativa de que o sistema auditivo e o processamento visual se desenvolvem e deterioram no decorrer do desenvolvimento. Ou seja, o sistema auditivo amadurece mais cedo do que o visual e se deteriora mais rápido com o avanço da idade.

Os estudos mencionados anteriormente buscaram investigar a percepção de fala por falantes típicos em diferentes faixas etárias, sugerindo que utilizar pistas auditivas e visuais de forma integrada para perceber sons da fala é uma tarefa que pode ser observada muito cedo na infância e é afetada pela experiência com a língua nativa. Entretanto, a área dos estudos sobre as patologias é um terreno fértil para eliciar questões de natureza teórica. 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 distúrbio fonológico.

2.4 OS DISTÚRBIOS DA COMUNICAÇÃO COMO FATOR INFLUENTE NA PERCEPÇÃO MULTIMODAL

Nesta seção, explicaremos a influência dos distúrbios da comunicação humana na percepção multimodal da fala, perpassando por caracterização do desempenho perceptual em diferentes grupos clínicos, na afasia (BAUM et al, 2013), no transtorno do espectro autista (TEA) (JACOB, et al 2019), no transtorno do desenvolvimento da linguagem (TDL) (AULI et al, 2013) e no distúrbio fonológico (DESJARDINS, ROGERS e WERKER, 1997; DODD et a., 2008).

Jakobson (1975) apresenta a ideia de que a afasia seria o espelhamento, ao reverso, do processo de aquisição da linguagem. Goldstein (1948) defende a

ideia de que a filosofia da linguagem e a investigação da afasia podem favorecer-se mutuamente, pois a reflexão sobre os distúrbios da comunicação necessitam “tirar proveito da filosofia da linguagem”. (apud FONSECA, 2002). Os dois autores mencionados demonstram o estudo da patologia como enriquecedor para o conhecimento a respeito do processo dito normal na linguagem.

A literatura sugere que o sulco temporal superior esquerdo (STS) é um local crítico para a integração multimodal de informações auditivas e visuais durante a percepção da fala. Baum et al (2012) descreveram achados de uma paciente, mulher de 63 anos, que sofreu um acidente vascular cerebral (AVC) que lesionou a área tempo-parietal esquerdo, resultando em afasia anômica leve. A ressonância magnética (RM) estrutural mostrou destruição completa do STS médio e posterior esquerdo, além de danos em áreas adjacentes nos lobos temporal e parietal. A paciente participou de tarefa de percepção multimodal. Os estímulos selecionados correspondiam às oclusivas /b/, /g/, /p/, /k/ e /t/ produzidos por um falante nativo do inglês. Os referidos estímulos foram apresentados sob quatro formas de apresentação: somente visual, somente auditivo, audiovisual congruente e audiovisual incongruente. Surpreendentemente, a paciente demonstrou integração multimodal preservada. Uma possível hipótese explicativa levantada pelos autores deve-se ao fato de haver uma reorganização das redes de linguagem cortical da paciente.

Diferentemente do estudo anterior, um segundo grupo de pesquisadores buscou verificar como se dá a percepção da fala na aquisição atípica da linguagem, especificamente, propuseram descrever o desempenho de crianças com transtorno da linguagem em tarefa de integração audiovisual.

Transtorno da linguagem consiste em dificuldades persistentes na aquisição e no uso da linguagem em suas diversas modalidades (falada, escrita, linguagem de sinais ou outra) devido a déficits na compreensão ou na produção de vocabulário, na estrutura das frases e no discurso. (American Psychiatric Association, 2014, p. 42).

O transtorno da linguagem está fortemente associado a outros transtornos do neurodesenvolvimento, como por exemplo, transtorno de déficit de atenção/hiperatividade e transtorno do espectro autista. Na ausência de comorbidades o transtorno da linguagem recebe o nome de transtorno do desenvolvimento da linguagem (American Psychiatric Association, 2014).

Auli et al (2013) buscaram verificar o efeito da relação do ruído na percepção audiovisual da fala em crianças típicas e crianças com transtorno do desenvolvimento da linguagem. Participaram da pesquisa, 55 crianças de 8 anos. 25 crianças compuseram o grupo de participantes com TDL e 30 crianças típicas sem alteração no desenvolvimento da linguagem, compuseram o grupo de participantes típicos. Os estímulos contemplaram as sílabas /apa/, /ata/ e /aka/, que foram produzidas por uma falante do finlandês. Posteriormente, as gravações foram editadas e sincronizadas de forma a compor três diferentes condições de apresentação: somente visual, somente auditiva e audiovisual incongruente. Para os estímulos audiovisuais incongruentes, três intensidades de som (24, 36 e 48 dB) e três níveis de ruído (-12, 0 e 6 dB) foram adicionados aos estímulos acústicos. Considerando o fator intensidade sonora, para a condição de apresentação somente auditiva, verificou-se que crianças típicas sofreram influência da intensidade, ou seja, quando a intensidade do som ficava menor, maior era a ocorrência de efeito McGurk; enquanto as crianças com TDL foram menos influenciadas pela manipulação da intensidade do som. Para a condição de apresentação audiovisual incongruente, verificou-se que crianças típicas sofreram influência da intensidade, ou seja, quando a intensidade do som ficava menor, havia uma maior porcentagem de respostas ancoradas na pista visual. Enquanto, as crianças com TDL não sofreram influência da intensidade. Considerando o fator ruído as crianças típicas sofreram maior influência do ruído comparadas as crianças com TDL, ou seja, as crianças típicas apresentaram maior porcentagem de respostas ancoradas na pista visual quando os estímulos foram apresentados com adição de ruído.

Como mencionado anteriormente, o transtorno de linguagem pode apresentar diferentes comorbidades como, por exemplo, o transtorno do espectro autista. Crianças com diagnóstico de TEA têm demonstrado um pior desempenho perceptual em tarefa de integração audiovisual (JACOB, et al 2019). O transtorno do espectro autista, por sua vez, consiste em alteração comportamental que afeta a habilidade de responder apropriadamente aos estímulos sensoriais e a capacidade de comunicação. Especificamente, quanto aos estímulos sensoriais, sabe-se que as informações sensoriais, advindas de diferentes vias, chegam ao cérebro de forma desorganizada resultando em dificuldades no processamento sensorial (GOULARTE et al, 2020).

Especificamente, considerando a comunicação, um dos sinais característicos do TEA é a dificuldade de direcionar o olhar para a face dos falantes, ou manter o contato visual (IRWIN, BRANCAZIO e WHALEN, 2011). Sabe-se que a pista visual, ou seja, os movimentos articulatórios dão suporte para que representações da fala sejam feitas no cérebro, dessa forma, contribui para o desenvolvimento da linguagem (IRWIN, BRANCAZIO e WHALEN, 2011).

A seguir serão apresentados dois estudos que propuseram investigar o desempenho em tarefa de percepção multimodal por crianças com TEA. O primeiro considerou como fator influente os aspectos sensoriais envolvidos no transtorno e o segundo considerou como fator influente a dificuldade de olhar para a face/ articuladores do falante.

Jacob et. al. (2019) buscou investigar se as diferenças na capacidade de resposta sensorial estariam relacionadas com o desempenho perceptual da fala em crianças no espectro do autismo. Para isso, foram investigadas 36 crianças de 8 a 17 anos, sendo que 18 contemplaram o grupo controle (crianças sem alteração no desenvolvimento da criança) e 18 o grupo de crianças com TEA. Os estímulos selecionados no estudo contemplam as oclusivas /b/, /d/, /g/ e /ta/ produzidas por uma falante da língua inglesa. Os estímulos foram editados, sincronizados e apresentado sob quatro condições de apresentação: apenas auditivo, apenas visual, audiovisual congruente e audiovisual incongruente. Em sua introdução à teoria de base, os autores, descrevem três padrões atípicos de responsividade sensorial: hiporresponsividade (respostas comportamentais reduzidas ou ausentes a estímulos sensoriais), hiperresponsividade (respostas comportamentais exageradas a estímulos sensoriais) e busca sensorial (fascínio incomum ou desejo por sensações sensoriais). Perfil sensorial refere-se ao padrão de respostas adaptativas aos estímulos sensorial de determinado indivíduo. Para o estudo, o perfil sensorial foi mensurado através de dois questionários específicos: questionário de experiências sensoriais (*Sensory Experiences Questionnaire*) (BARANEK et al. 2006) e o questionário do cuidador- perfil sensorial (*Sensory Profile- Caregiver Questionnaire*) (DUNN 1999). Os resultados inferenciais indicam que a precisão de identificação de estímulos apresentados na condição somente visual e audiovisual reduzida e acuidade temporal mais pobre para fala audiovisual estão associadas a respostas sensoriais mais atípicas. Considerando o perfil sensorial, foi verificada

diferença entre os grupos, ou seja, crianças com TEA apresentaram *score* mais alto para o padrão de hiporresponsividade, hiperresponsividade e busca sensorial comparadas às crianças típicas. Na relação entre o perfil sensorial e desempenho perceptual, o padrão de hiporresponsividade foi correlacionado com a menor porcentagem de acerto nas condições somente auditiva, somente visual e audiovisual congruente e incongruente. De modo geral, a menor porcentagem de acerto, à depender da condição de apresentação foi associada à padrões atípicos de adaptações aos estímulos sensoriais.

Irwin, Brancazio e Whalen (2011), por sua vez, propôs investigar a relação entre a fixação do olhar na face/articuladores do falante com o desempenho em tarefa de percepção de fala por crianças com desenvolvimento típico de linguagem e crianças no TEA. Para isso, foram investigadas 26 crianças, de idade entre 5 e 15 anos, falantes da língua inglesa, 13 diagnosticadas no TEA e 13 com desenvolvimento típico da linguagem. O critério de inclusão para a pesquisa garantiu que as crianças conseguiram compreender e executar a tarefa solicitada. Os estímulos consistiram das sílabas /ma/, /na/, /ga/ e /bada/ que foram gravadas por um falante da língua inglesa. Posteriormente, foram editados e sincronizados de forma a contemplar as condições de apresentação: somente visual, somente auditiva, audiovisuais congruentes e audiovisuais incongruentes. Os estímulos acústicos foram manipulados com adição de ruído (5,0, -5, -10, -15 e - 20 dB). Para coleta dos dados, utilizou-se rastreamento visual. Este modelo utiliza de uma unidade que rastreia a posição de um pequeno sensor magnético preso acima do olho esquerdo do participante. Os resultados demonstraram que para os estímulos apresentados com adição de ruído não foram verificadas diferenças significativas entre os grupos, indicando que as crianças com TEA e as crianças típicas foram capazes de identificar as sílabas no contexto de ruído. Em geral, os participantes com TEA apresentaram desempenho pior do que as crianças típicas. Cabe ressaltar, que o recurso metodológico utilizado na presente pesquisa foi sensível para demonstrar que mesmo quando as crianças com TEA direcionam e mantém o olhar nos movimentos articulatórios, elas parecem não se beneficiar dessas informações tanto quanto às crianças típicas.

No tocante à percepção multimodal, tem sido apresentado para a língua inglesa, que crianças com distúrbio fonológico são menos influenciadas pela pista visual (DESJARDINS, ROGERS e WERKER, 1997).

Desjardins, Rogers e Werker (1997) investigaram a percepção multimodal de adultos e crianças com e distúrbio 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. O grupo composto por participantes adultos, por sua vez, foi o que sofreu maior influência da pista visual durante o teste. Verificou-se que os resultados entre os três grupos não foram estatisticamente diferenciados quando o estímulo foi apresentado apenas por via auditiva. Os autores concluíram que a produção correta das consoantes refletiu na representação visual da fala. Dados como esses, segundo os autores, reforçaram a hipótese de que as crianças com distúrbio fonológico apresentariam dificuldades na integração de pistas audiovisuais na percepção de fala.

Por outro lado, Dodd et al (2008) investigaram a percepção multimodal em crianças com desenvolvimento típico de fala e crianças com distúrbio fonológico. Os estímulos selecionados foram os fonemas oclusivos /b/; /d/; /g/ em contexto de /a/. Os referidos estímulos foram apresentados nas seguintes condições de apresentação: somente por via auditiva, somente por via visual, apresentação simultânea dos estímulos auditivo e visual de modo congruente e apresentação simultânea dos estímulos auditivo e visual de modo incongruente. Os resultados apresentados não demonstraram diferença significativa entre grupos. A análise inferencial dos dados também possibilitou verificar que ambos os grupos apresentaram uma pior acurácia para a condição de apresentação somente por via visual e melhor acurácia para a condição de apresentação somente por via auditiva. Na análise do estímulo audiovisual incongruente, os resultados demonstraram que não houve um relato perceptual de um terceiro estímulo (inferida na literatura como ilusão perceptual) para ambos os grupos, indicando que a atenção dos participantes do estudo foi direcionada para apenas um aspecto do estímulo visual ou ao auditivo. Segundo os autores, os resultados encontrados não corroboram os estudos prévios (DESJARDINS e WERKER, 1996; DESJARDINS et al, 1997), na medida em que, a literatura aponta que

crianças com distúrbio fonológico seriam menos influenciadas pela pista visual, refletindo, assim, uma possível dificuldade de integração de pistas. Os resultados apresentados em Dood, et al (2008), por sua vez, não demonstraram diferença significativa entre os grupos investigados (crianças com desenvolvimento típico e crianças com distúrbio fonológico).

Mais recentemente, Assis (2021) assumindo que o distúrbio fonológico poderia estar associado a uma capacidade prejudicada de integrar a informação fonológica de duas fontes de modalidade, propôs comparar o desempenho de crianças com e sem distúrbio fonológico na tarefa de percepção multimodal envolvendo pistas auditivas e visuais em fricativas desvozeadas do Português Brasileiro. Dessa forma, os estímulos selecionados foram: /f/, /s/ e /ʃ/ no contexto da vogal /a/, que foram apresentados nas seguintes condições de apresentação: apenas por via auditiva, apenas por via visual, simultânea dos estímulos auditivos e visuais de forma congruente e simultânea dos estímulos auditivos e visuais de forma incongruente. Como forma de análise considerou-se as porcentagens de repostas corretas dos estímulos, enquanto para os estímulos incongruentes foi considerada a porcentagem de preferência perceptiva, bem como foram extraídos os tempos de resposta para execução da tarefa. Os resultados apresentados considerando a acurácia de respostas demonstraram que houve uma diferença significativa entre os grupos e entre as condições de apresentação. Foi possível verificar uma pior acurácia para a condição de apresentação somente por via visual, para o grupo de crianças típicas. Na análise do estímulo audiovisual incongruente, foi possível verificar que as crianças típicas apresentaram menor porcentagem de respostas para condição auditiva e maiores porcentagens para a condição visual e ilusão perceptual. Em relação ao tempo de reação, os resultados apresentados demonstraram efeito significativo para grupo e condição. Foi possível verificar que as crianças com distúrbio fonológico apresentaram maior tempo de reação em relação às crianças típicas. Especificamente, maior tempo de reação para condição visual em ambos os grupos. Os resultados, portanto, demonstraram que crianças com Distúrbio Fonológico apresentaram pior desempenho e maior tempo de reação na tarefa de percepção multimodal. Dessa forma, os dados apresentados por Assis (2021) corroboram com os achados descritos por Desjardins, Rogers e Werker (1997).

Os achados da literatura apresentados sugerem, portanto, que os distúrbios da comunicação humana parecem influenciar a forma como percebemos a fala considerando a integração de pistas audiovisuais.

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

3 OBJETIVOS E HIPÓTESES

Objetivo Geral

O objetivo geral do presente estudo é investigar como se dá a integração de pistas audiovisuais no Português Brasileiro em função do padrão silábico da língua considerando três grupos de sujeitos: adultos, crianças típicas e crianças com distúrbio fonológico.

Objetivos específicos

- 1) Verificar se há integração de pistas audiovisuais em tarefa de percepção de fala, ou seja, se o desempenho perceptual de fala sofre influência da pista visual no PB;
- 2) Verificar se o desempenho audiovisual de percepção de fala sofre influência do padrão silábico;
- 3) Comparar o desempenho perceptual de fala, especificamente, a integração de pistas audiovisuais em três grupos de sujeitos: adultos, crianças típicas e com distúrbio fonológico.

Dessa forma, as hipóteses deste trabalho assim se definiram:

H1- haveria integração de pistas audiovisuais na percepção de fala nos falantes do PB, marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV-⁵ comparativamente à condição AV+⁶;

H2- a percepção de fala sofreria influência do padrão silábico, ou seja, haveria uma diferença do grau de influência visual em função da complexidade silábica. Espera-se maior grau de influência visual e maior tempo de reação no padrão silábico complexo;

H3- haveria diferenças no desempenho perceptual entre os grupos. Espera-se que os adultos sofram uma maior influência visual comparado às crianças. E, ainda, que as crianças típicas sofram uma maior influência visual

⁵ AV- refere-se aos estímulos apresentados de forma incongruente, ou seja, quando a informação visual não corresponde à informação auditiva;

⁶ AV+ refere-se aos estímulos apresentados de forma congruente, ou seja, quando a informação visual corresponde à informação auditiva.

comparadas às crianças com distúrbio fonológico. A influência visual seria marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+.

Na próxima seção, descreveremos com maiores detalhes os três estudos delineados para responder aos objetivos definidos.

4 ESTUDOS EXPERIMENTAIS

Como exposto no capítulo anterior, os objetivos definidos serão explorados em três estudos. Na seção 3.1 apresentaremos os aspectos éticos dos três estudos; na seção 3.2, descreveremos o estudo 1- *“Desempenho de percepção multimodal de fala em adultos típicos”*-, analisando a porcentagem de acertos, influência visual e tempo de reação, utilizadas por adultos durante tarefa de identificação a depender do padrão silábico (simples e complexo); na seção 3.3, abordaremos o estudo 2- *“Desempenho de percepção multimodal de fala em crianças típicas e com Distúrbio Fonológico”*- analisando a porcentagem de acertos, influência visual e tempo de reação, utilizadas por crianças durante tarefa de identificação a depender do padrão silábico (simples e complexo), além de verificar se haveria diferença no desempenho de crianças com e distúrbio fonológico; por fim, na seção 3.4, apresentaremos o estudo 3- *“Comparação de desempenho de percepção multimodal entre adultos e crianças”*- que compara o desempenho em tarefa de percepção multimodal da fala por participantes adultos e crianças, com base nos resultados obtidos nos estudos 1 e 2.

4.1 ASPECTOS ÉTICOS DOS TRÊS ESTUDOS

Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Faculdade de Filosofia e Ciências – UNESP/Marília (Faculdade de Filosofia e Ciências- UNESP/Marília), sob o registro de nº 3.201.991, CAAE: 05262918.1.0000.5406 (Anexo A). Os responsáveis pelas crianças assinaram o Termo de Consentimento Livre e Esclarecido- TCLE (Anexo B). Este estudo baseou-se em diretrizes e normas regulamentadoras para pesquisas envolvendo seres humanos, determinados pelo Conselho Nacional de Saúde em suas Resoluções 466/12 e 510/16, respectivamente.

Os três estudos caracterizam-se como descritivo, observacional, quantitativo e prospectivo, na medida em que iremos descrever tendências encontradas em tarefa de percepção multimodal, de dados primários, ou seja, dados coletados pelo próprio investigador da presente tese. Os dados serão tabulados numericamente e analisados de forma descritiva e inferencial, por meio da análise estatística.

4.2 ESTUDO 1: Desempenho de percepção multimodal de fala em adultos típicos

Conforme já exposto, o estudo 1 objetivou verificar se há integração de pistas audiovisuais em tarefa de percepção de fala em adultos e, ainda, verificar a existência, ou não, do efeito do padrão silábico.

As hipóteses a serem testadas são as seguintes:

H1- haveria integração de pistas audiovisuais na percepção de fala nos falantes do PB, marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+;

H2- a percepção de fala sofreria influência do padrão silábico, ou seja, haveria uma diferença do grau de influência visual em função da complexidade silábica. Espera-se maior grau de influência visual e maior tempo de reação no padrão silábico complexo;

4.2.1 Participantes

Participaram deste estudo 11 adultos, sendo cinco do sexo feminino e seis do sexo masculino. Os participantes eram residentes da cidade de Marília/ SP e foram selecionados por conveniência.

Os seguintes critérios de inclusão foram satisfeitos para a seleção destes participantes: ter idade entre 20 e 30 anos; ausência de alterações na produção dos sons da fala; ausência de alterações neurológicas de linguagem, visuais e/ou auditivas, uma vez que alterações desta natureza poderiam interferir no desempenho da percepção audiovisual. Os participantes não passaram por avaliações específicas: neurológicas, audiológicas e oftamológicas, a exclusão foi feita a partir da presença de queixa neurológica, auditiva ou visual.

Para seleção da amostra, as produções de fala dos participantes foram submetidas ao julgamento perceptivo auditivo da avaliadora especialista. A avaliadora conduziu individualmente uma conversa semi-dirigida, a fim de julgar a produção de palavras espontâneas de cada participante, no qual todas as produções foram julgadas como corretas perante análise de oitiva.

4.2.2 Experimento de percepção

Procedimento experimental

O procedimento experimental de percepção consistiu em um teste de identificação, também designado tarefa de escolha forçada. Em testes de identificação, um único estímulo é apresentado e o informante deve identificá-lo, escolhendo uma das opções expostas para escolha (KLUGE *et al.*, 2013). Para elaboração do teste, foi utilizado o *software OpenSesame*. Trata-se de um *software* livre que possibilita ao usuário elaborar *scripts* pertinentes a tarefas simples e/ou complexas. Dentre suas atribuições, o *software* suporta a reprodução de áudio e vídeo. (MATHÔT, SCHREIJ e THEEUWES, 2012). Dessa forma, o referido programa pôde ser utilizado como instrumento para a elaboração de testes de percepção.

O teste de identificação é composto por quatro etapas distintas: (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, ou seja, quando os estímulos auditivos e visuais são correspondentes (exemplo: áudio [pa] e vídeo correspondente à produção do [pa]); e (4) apresentação simultânea dos estímulos auditivo e visual de modo incongruente, isto é, quando não há correspondência entre os estímulos auditivos e visuais (exemplo: áudio [pa] e vídeo correspondente à produção de [ka]).

Dessa forma, os estímulos foram combinados conforme síntese contida no Quadro 1 abaixo:

Quadro 1: Condições de apresentação x estímulos apresentados

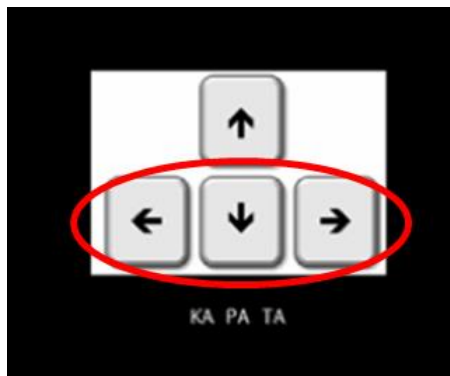
-	Somente auditiva	Somente visual	Integração audiovisual	
			Congruentes (AV+)	Incongruentes (AV-)
Padrão silábico simples	[pa]; [ta] e [ka]	[pa]; [ta] e [ka]	Audio pa _Visual pa Audio ta _Visual ta Audio ka _Visual ka	Audio pa _Visual ta Audio pa _Visual ka Audio ta _Visual pa Audio ta _Visual ka Audio ka _Visual pa Audio ka _Visual ta
Padrão silábico complexo	[pra]; [tra] e [kra]	[pra]; [tra] e [kra]	Audio pra _Visual pra Audio tra _Visual tra Audio kra _Visual kra	Audio pra _Visual tra Audio pra _Visual kra Audio tra _Visual pra Audio tra _Visual kra Audio kra _Visual pra Audio kra _Visual tra

Fonte: Elaborado pela autora

Para a realização do teste de identificação, os participantes foram dispostos confortavelmente em frente à tela do computador (contendo o *software OpenSesame*), com fones acoplados aos ouvidos, em uma sala tratada acusticamente.

Durante a primeira etapa do teste, os participantes escutaram (com apresentação binaural em uma intensidade de 50 dB) um estímulo auditivo e, em seguida, decidiram e indicaram, com a tecla correspondente no computador, qual a sílaba, representada graficamente, correspondente ao estímulo apresentado auditivamente dentre três possibilidades (conforme ilustrado na Figura 1). Cada estímulo auditivo foi apresentado duas vezes.

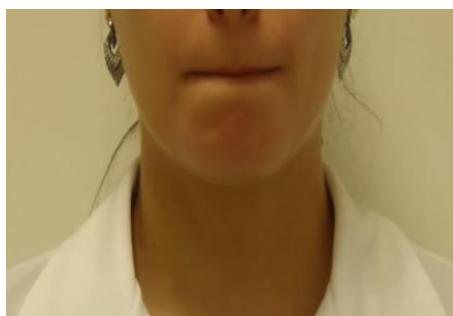
Figura 1: Opções de resposta após apresentação de estímulo de padrão silábico simples.



Fonte: Elaborado pela autora

Para a segunda parte do teste, um estímulo foi apresentado apenas visualmente. Tratou-se de 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, os participantes decidiram e indicaram, dentre três opções dispostas da tela do computador, qual a sílaba correspondente ao estímulo visual apresentado (conforme ilustrado na Figura 2). Cada estímulo visual foi apresentado duas vezes.

Figura 2: Vídeo com apresentação do movimento articulatorio do fonema /pa/, sem qualquer informação auditiva associada.



Fonte: Elaborado pela autora

Para a terceira e quarta etapa do teste, foi apresentado um estímulo contendo informações auditiva e visual: uma gravação de vídeo mostrando movimento articulatorio de boca foi apresentada simultaneamente a uma

informação auditiva. Estas etapas foram constituídas de estímulos congruentes (em que a informação visual corresponde à informação auditiva) e incongruentes (em que a informação visual não corresponde à informação auditiva), apresentados de forma aleatória. A tarefa foi realizada pelo participante, conforme instruções das etapas 1 e 2.

Para as condições de apresentação de estímulo somente auditiva, somente visual e audiovisual congruente, os padrões de resposta considerados foram: acerto (quando o participante identifica o estímulo corretamente) e erro (se não o identifica corretamente). Adicionalmente, o tempo de reação para cada resposta apresentada foi extraído automaticamente pelo programa. Cabe destacar, que todos os estímulos foram apresentados em ótimas condições de comunicação, ou seja, em silêncio e sem a presença de ruído.

Estímulos

Os estímulos apresentados foram constituídos por sílabas em padrão silábico simples e complexo. Em padrão silábico simples, o ataque foi preenchido pelas oclusivas: /p/, /t/ e /k/, enquanto no padrão silábico complexo, o ataque foi preenchido pelas mesmas consoantes seguidas da líquida não lateral /r/.

Estabelecidos os estímulos que compuseram o teste, foram realizadas gravações de vídeo e áudio simultaneamente correspondentes aos estímulos. As gravações foram editadas em um computador, com o uso do *software OpenShop* (WRZESINSKI at. al, 2021), de modo que cada estímulo foi salvo individualmente em formato wav e em formato mp4.

Posteriormente, foi elaborado um *script* para o experimento de identificação a ser executado pelo *software OpenSesame*.

4.2.3 Forma de análise

Para análise dos dados, foram utilizadas as medidas de tempo de reação e porcentagem de acerto dos estímulos de acordo com as seguintes condições de apresentação: somente via auditiva; somente via visual, audiovisual congruente (AV+) e audiovisual incongruente (AV-). Para análise da condição incongruente, consideramos a porcentagem de acerto das respostas baseadas

no estímulo auditivo. Posteriormente, a medida de grau de influência visual foi calculada pela subtração $(AV+) - (AV-)$ (SEKIYAMA e BURNHAM, 2008).

Análise Estatística

Um tratamento estatístico descritivo e inferencial dos dados foi realizado com o uso do *software Jamovi* (versão 2.2.5.) Inicialmente realizou-se a análise de média e desvio padrão das variáveis acurácia, tempo de reação e influência visual. Posteriormente, para a análise inferencial foi adotado o teste GzLM (generalized linear models) adotando distribuição Gamma com função de ligação identidade, a fim de verificar o efeito das condições de apresentação dos estímulos sobre os resultados de acurácia e tempo de reação. Para a comparação de médias, utilizou-se o teste Post Hoc de Holm. Adotou-se $p < 0,05$ como valor mínimo de significância estatística.

4.2.4 Resultados

A Tabela 1, a seguir, apresenta a porcentagem de identificação correta dos estímulos de acordo com as condições de apresentação. Na condição audiovisual incongruente, consideramos a porcentagem de acerto do estímulo auditivo apresentado.

Tabela 1: Porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico.

P	Somente Via Auditiva		Somente Via Visual		Audiovisual Congruente		Audiovisual Incongruente	
	CV	CCV	CV	CCV	CV	CCV	CV	CCV
P1	100	100	66,6	66,6	100	100	66,6	66,6
P2	100	100	83,3	66,6	100	100	100	100
P3	100	100	66,6	66,6	100	100	100	100
P4	100	100	100	66,6	100	100	83,3	100
P5	83,3	83,3	50	50	100	83,3	83,3	83,3
P6	66,6	83,3	83,3	66,6	100	83,3	66,6	66,6
P7	83,3	100	66,6	16,6	66,6	100	83,3	100
P8	100	100	100	100	100	100	66,6	100
P9	100	83,3	83,3	100	100	100	0	0
P10	100	100	100	83,3	100	100	66,6	66,6
P11	100	100	83,3	66,6	100	100	83,3	100
Média	95,1	95,9	83,5	70,9	97,7	97,3	77,1	82
DP	10,7	7,6	16,9	23,8	9,4	6,5	27	29,8

Fonte: Elaborado pela autora. Legenda: (P): participantes; (CV): padrão silábico simples; (CCV): padrão silábico complexo; e (DP) desvio padrão.

A Tabela 2, apresentada a seguir, apresenta o tempo de reação em relação aos acertos dos adultos para as quatro diferentes condições de apresentação dos estímulos.

Tabela 2: Tempo de reação em tarefa de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico.

P	Somente Via Auditiva		Somente Via Visual		Audiovisual Congruente		Audiovisual Incongruente	
	CV	CCV	CV	CCV	CV	CCV	CV	CCV
P1	1221,8	1541,1	1800	1669,7	845,3	654	1376,2	3784,2
P2	1741	1089,3	2705,4	1171,2	741,5	674,1	869,3	677,5
P3	1262	1038,5	1205,5	1195,2	654,3	739	834,1	855,8
P4	668,3	868,8	1057,5	807,5	694,1	747,5	365,6	578,3
P5	1264,4	1237,6	1802,6	1712	610	619	481,2	462,6
P6	889,2	2681,6	986	1056,7	388,6	606,6	310,2	564,7
P7	1743,8	686,8	1743,8	686,8	760,5	356,1	247,8	770,3
P8	1962,5	547,6	1962,5	547,6	1218,6	226,5	814	325
P9	1349	2034,8	1349	2034,8	392,6	245,6	0	0
P10	646	577,5	646	577,5	538,3	371,8	1083,7	1904,5
P11	2108,3	2401,6	2108,3	2401,6	890,5	936,5	1036,8	3935,3
Média	1079,3	1127,7	1288,4	1050,9	587,2	485,6	548,8	911,3
DP	612,4	733,8	773,5	582,7	314,4	230,5	438,2	1076,1

Fonte: Elaborado pela autora. Legenda: (P): participantes; (CV): complexidade simples; (CCV): complexidade complexa; e (DP) desvio padrão. *tempo de reação calculado em milissegundos.

Para melhor exposição, os resultados inferenciais serão dispostos de acordo com as hipóteses norteadoras do presente estudo.

H1- haveria integração de pistas audiovisuais na percepção de fala nos falantes do PB, marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+.

Um primeiro teste estatístico foi conduzido de modo a verificar se os adultos sofreriam influência da pista visual durante tarefa de percepção multimodal e, ainda, verificar se haveria uma diferença significativa no tempo de reação na execução da tarefa considerando as condições de apresentação AV- comparativamente à condição AV+.

Na comparação entre as condições de apresentação (audiovisual congruente e audiovisual incongruente) o teste GzLM demonstrou efeito significativo ($T(4,64)=208, p<0,01$), mostrando maior acurácia para a condição de apresentação audiovisual congruente. Ou seja, os adultos apresentaram maior porcentagem de acertos quando os estímulos foram apresentados na condição audiovisual congruente comparado a condição audiovisual incongruente.

De modo a comparar o tempo de reação em tarefa de identificação de estímulos apresentados nas condições audiovisual congruente e incongruente o teste GzLM não demonstrou efeito significativo ($T(-0,96)=208, p=1$).

H2- a percepção de fala sofreria influência do padrão silábico, ou seja, haveria uma diferença do grau de influência visual em função da complexidade silábica. Espera-se maior grau de influência visual e maior tempo de reação no padrão silábico complexo.

Um segundo teste estatístico foi conduzido, de modo a verificar se o grau de influência visual seria dependente do padrão silábico (simples e complexo). O grau de influência visual foi calculado pela subtração (AV+) – (AV-). Na comparação entre os padrões silábicos simples e complexo o teste GzLM não mostrou diferença significativa ($Z(1,74)= 416, p=1$).

De modo a comparar o tempo de reação em tarefa de identificação de estímulos apresentados nas quatro condições de apresentação à depender do padrão silábico (simples e complexo), o teste GzLM não demonstrou efeito significativo ($Z(0,0662)=326, p=1$).

Perante análise inferencial dos dados foi possível verificar que o grau de influência visual não varia em função do padrão silábico em adultos.

4.3 ESTUDO 2: DESEMPENHO DE PERCEPÇÃO MULTIMODAL DE FALA EM CRIANÇAS TÍPICAS E COM DÍSTURBIO FONOLÓGICO

O estudo 2 objetivou verificar se há integração de pistas audiovisuais em tarefa de percepção de fala em crianças com e sem distúrbio fonológico e, ainda, verificar a existência, ou não, do efeito do padrão silábico. O desempenho perceptual multimodal foi testado separadamente considerando o desempenho para crianças típicas e crianças com distúrbio fonológico.

As hipóteses testadas foram as seguintes:

H1- haveria integração de pistas audiovisuais na percepção de fala nos falantes do PB, marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+;

H2- a percepção de fala sofreria influência do padrão silábico, ou seja, haveria uma diferença do grau de influência visual em função da complexidade silábica. Espera-se maior grau de influência visual e maior tempo de reação no padrão silábico complexo;

4.3.1 Participantes

Para o desenvolvimento deste estudo, fizeram parte dois grupos de participantes. Foram incluídas no grupo 1 (G1): crianças entre 6 e 8 anos que apresentaram desenvolvimento típico de linguagem e sistema fonológico efetivamente adquirido; e no grupo 2 (G2): crianças entre 6 e 8 anos que apresentaram processos de substituição de líquida não lateral por lateral em padrão silábica complexo ou redução de encontro consonantal.

Para seleção dos participantes do G1, foram selecionadas 10 crianças (5 do sexo feminino e 5 do sexo masculino), recrutadas de duas escolas municipais da cidade de Marília e de uma instituição de caridade destinada à crianças e adolescentes.

Para seleção dos participantes do G2, 10 crianças (5 do sexo feminino e 5 do sexo masculino) com diagnóstico de distúrbio fonológico, foram recrutadas da lista de espera do Estágio Supervisionado de Terapia Fonoaudiológica: Fonologia Clínica do Centro de Estudos da Educação e Saúde (CEES), UNESP Marília.

Os participantes do G1 e G2 passaram por avaliação fonológica. A avaliação fonológica foi investigada pela análise oitiva da produção de fala dos participantes em tarefa de nomeação espontânea, sendo utilizadas as palavras que compõem o Instrumento de Avaliação de Fala para Análise Acústica (IAFAC). Este instrumento contempla 96 palavras-alvo representadas por meio de gravuras, contemplando pelo menos três ocorrências de todos os fonemas consonantais do PB acompanhados das vogais [i], [a] e [u] na posição acentuada. Para a presente pesquisa, privilegiou-se a produção de fala das palavras em contexto da vogal [a].

Perante julgamento perceptivo-auditivo, as crianças do G1 deveriam apresentar todas as produções julgadas como correta, enquanto as crianças pertencentes ao G2 deveriam apresentar os processos de substituição de líquida não lateral por líquida lateral em estrutura silábica complexa ou redução de encontro consonantal em porcentagem menor que 85%, indicando que a criança não apresenta efetivamente em seu sistema fonológico a representação dos sons líquidos em estrutura silábica complexa. As crianças deveriam apresentar porcentagem de acerto maior que 85% para as demais classes consonantais, a saber: oclusivas, fricativas e nasais, a fim de garantir aquisição efetiva e representação dos demais sons do PB (YAVAS, HERNANDORENA e LAMPRECHT, 2001).

Por serem recrutados de um Centro de Estudos da Educação da Saúde, os participantes do G2 puderam passar por avaliação audiológica, possibilitando verificar presença ou ausência de alterações otológicas e/ou audiológicas. As avaliações audiológicas foram realizadas pelo serviço de audiologia do CEES – sob responsabilidade da Prof^{fa}. Dr^a. Ana Cláudia Vieira Cardoso.

De 26 crianças, recrutadas da lista de espera do Estágio Supervisionado de Terapia Fonoaudiológica Clínica do Centro de Estudos da Educação da Saúde, UNESP Marília, excluímos do estudo 16 crianças, uma vez que oito crianças apresentaram alteração de linguagem, duas apresentaram disfluências típicas da gagueira, seis não apresentaram trocas fonológicas envolvendo a líquida não lateral, em padrão silábico simples (CV) e padrão silábico complexo (CCV).

Como critério de exclusão da amostra, adotamos parâmetros semelhantes para ambos os grupos, sendo excluídas as crianças com Distúrbio

Fonológico associado a outras alterações de linguagem, alterações neurológicas, bem como criança cujos responsáveis relatassem problemas de acuidade auditiva e/ou visual, uma vez que alterações dessa natureza poderiam interferir no desempenho da percepção audiovisual. Os participantes não passaram por avaliações específicas: neurológicas, audiológicas (com exceção do G2) e oftalmológicas, a exclusão foi pautada na presença de queixa auditiva e visual.

4.3.2 Experimento de percepção

Procedimento experimental

O procedimento experimental deste estudo foi elaborado segundo descrição do estudo 1. Sendo assim, o procedimento experimental de percepção deste estudo consistiu em um teste de identificação elaborado com o uso do *software OpenSesame*.

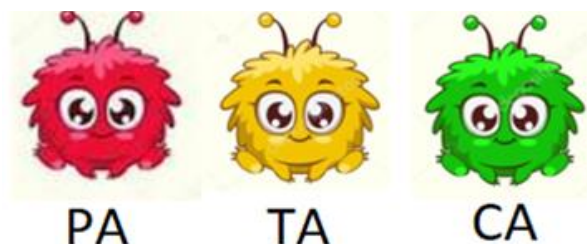
O teste de identificação foi composto por quatro etapas distintas: (1) apresentação de estímulos somente por via auditiva; (2) apresentação de estímulos somente por via visual; (3) apresentação simultânea dos estímulos auditivo e visual de modo congruente; e (4) apresentação simultânea dos estímulos auditivo e visual de modo incongruente.

Ademais, a fim de garantir a compreensão da tarefa a ser realizada e confiabilidade de respostas das crianças, para o presente estudo, optamos por inserir uma etapa de familiarização dos estímulos auditivos.

A fase de familiarização foi elaborada contemplando dois arquivos de *Power Point*, um referente à apresentação de estímulos de padrão silábico simples e outro referente à apresentação de estímulos de padrão silábico complexo.

A familiarização consistiu em duas etapas: (1) associação dos estímulos auditivos com personagens lúdicos diferenciados por cores, vermelho, verde e amarelo, realizando referência ao nome do personagem e (2) fase treino. As Figuras 3 e 4, a seguir, ilustram os personagens lúdicos selecionados e suas respectivas correspondências aos estímulos auditivos.

Figura 3: Personagens lúdicos e estímulos auditivos correspondentes para padrão silábico simples.



Fonte: Imagens coletadas no jogo *Sort Fluffy Balls* (referência) via google.com. Editadas pela autora para elaboração de tarefa experimental.

Figura 4- Personagens lúdicos e estímulos auditivos correspondentes para padrão silábico complexo.



Fonte: Imagens coletadas no site *Seekpng* via google.com. Editadas pela autora para elaboração de tarefa experimental.

Tarefa Experimental

Para a realização do teste, os participantes foram dispostos confortavelmente em frente à tela do computador, em uma sala tratada acusticamente.

A coleta seguiu a seguinte ordem: (1) fase de familiarização, fase treino e teste de identificação de estímulos de padrão silábico simples e (2) fase de familiarização, fase treino e teste de identificação de estímulos de padrão silábico complexo.

A fase treino da familiarização consistiu na apresentação de dois estímulos em suas três condições de apresentação. Os participantes foram orientados a indicar a resposta, a fim de garantir a compreensão da tarefa a ser cumprida.

Ainda posicionados em frente à tela do computador (contendo o *software OpenSesame*), fones de ouvidos foram acoplados nos participantes.

Durante a primeira etapa do teste, os participantes escutaram (com apresentação binaural em uma intensidade de 50 dB) um estímulo auditivo e, em seguida, decidiram e indicaram, qual o personagem, correspondente ao estímulo apresentando auditivamente dentre três possibilidades. Os participantes 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 participantes menores. Cada estímulo auditivo foi apresentado duas vezes.

Para a segunda etapa do teste, um estímulo foi apresentado apenas visualmente. Tratou-se de 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 indicou, dentre três opções distintas na tela do computador, qual o personagem correspondente ao estímulo visual apresentado. Cada estímulo foi apresentado duas vezes.

Para as terceira e quarta etapas do teste, foram apresentados estímulos contendo informações auditiva e visual: uma gravação de vídeo mostrando movimento articulatorio de boca foi apresentada simultaneamente a uma informação auditiva. Estas etapas foram constituídas de estímulos congruentes (em que a informação visual corresponde à informação auditiva) e incongruentes (em que a informação visual não corresponde à informação auditiva), apresentados de forma aleatória. A tarefa foi realizada pelo participante, conforme instruções das etapas 1 e 2.

Para as condições de apresentação de estímulo somente auditiva, somente visual e audiovisual congruente, os padrões de resposta considerados foram: acerto (quando o participante identifica o estímulo corretamente) e erro (se não o identifica corretamente). Adicionalmente, o tempo de reação para cada resposta apresentada foi extraído automaticamente pelo programa. Cabe destacar, que todos os estímulos foram apresentados em ótimas condições de comunicação, ou seja, em silêncio e sem a presença de ruído.

Estímulos

Os estímulos apresentados foram constituídos por sílabas em padrão silábico simples e complexo. Em padrão silábico simples, o ataque foi preenchido pelas oclusivas: /p/, /t/ e /k/, enquanto no padrão silábico complexo, o ataque foi preenchido pelas mesmas consoantes seguidas da líquida não lateral /r/.

Estabelecidos os estímulos que compuseram o teste, foram realizadas gravações de vídeo e áudio 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 e em formato mp4. O Quadro 1 (p.36, seção 3.2.2) exemplifica a referida distribuição dos estímulos selecionados.

Posteriormente, foi elaborado um *script* para o experimento de identificação a ser executado pelo *software OpenSesame*.

4.3.3 Forma de análise

Para análise dos dados, foram utilizadas as medidas de tempo de reação e porcentagem de acerto dos estímulos de acordo com as seguintes condições de apresentação: somente via auditiva; somente via visual, audiovisual congruente (AV+) e audiovisual incongruente (AV-). Para análise da condição incongruente, consideramos a porcentagem de acerto das respostas baseadas no estímulo auditivo. Posteriormente, a medida de grau de influência visual foi calculada pela subtração (AV+) – (AV-) (SEKIYAMA e BURNHAM, 2008).

Análise Estatística

Um tratamento estatístico descritivo e inferencial dos dados foi realizado com o uso do *software Jamovi* (versão 2.2.5.) Inicialmente realizou-se a análise de média e desvio padrão das variáveis acurácia, tempo de reação e influência visual. Posteriormente, para a análise inferencial foi adotado o teste GzLM (generalized linear models) adotando distribuição Gamma com função de ligação identidade, a fim de verificar o efeito das condições de apresentação dos estímulos sobre os resultados de acurácia e tempo de reação. Para a

comparação de médias, utilizou-se o teste Post Hoc de Holm. Adotou-se $p < 0,05$ como valor mínimo de significância estatística.

A seguir, serão apresentados os resultados em função do grupo de participantes investigados; dessa forma, na seção 3.3.4 serão apresentados desempenhos em tarefa de percepção multimodal de fala apresentados por crianças típicas e, na seção 3.3.5, serão apresentados desempenhos em tarefa de percepção multimodal de fala apresentados por crianças com distúrbio fonológico.

4.3.4 Resultados: crianças típicas

A Tabela 3, a seguir, apresenta a porcentagem de identificação correta dos estímulos de acordo com as condições de apresentação. Na condição audiovisual incongruente, consideramos a porcentagem de acerto do estímulo auditivo apresentado.

Tabela 3: Porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico para crianças típicas.

P	Somente Via Auditiva		Somente Via Visual		Audiovisual Congruente		Audiovisual Incongruente	
	CV	CCV	CV	CCV	CV	CCV	CV	CCV
P1	100	100	100	100	100	100	83,3	100
P2	100	100	100	100	100	100	83,3	83,3
P3	100	100	100	100	100	100	83,3	33,3
P4	100	100	100	100	83,3	100	83,3	83,3
P5	100	100	100	100	100	100	83,3	100
P6	100	100	100	100	100	100	83,3	83,3
P7	100	100	100	100	100	100	100	83,3
P8	100	100	100	100	100	100	100	100
P9	100	100	100	100	100	100	100	83,3
P10	100	100	100	100	100	100	83,3	100
Média	100	100	100	100	98,9	100	90,5	86,4
DP	0	0	0	0	4,9	0,3	8,9	19,5

Fonte: Elaborado pela autora. Legenda: (P): participantes; (CV): padrão silábico simples; (CCV): padrão silábico complexo; e (DP) desvio padrão.

A Tabela 4, apresentada a seguir, apresenta o tempo de reação em relação aos acertos das crianças típicas para as quatro diferentes condições de apresentação dos estímulos.

Tabela 4: Tempo de reação em tarefa de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico para crianças típicas.

P	Somente Via Auditiva		Somente Via Visual		Audiovisual Congruente		Audiovisual Incongruente	
	CV	CCV	CV	CCV	CV	CCV	CV	CCV
P1	982,5	1417,6	1503,6	1692,5	1189,3	972,1	2362,4	1049,8
P2	507,1	286,6	2934,8	580,8	675,5	315,8	462,8	250,5
P3	393,1	408,5	3593	2837,1	413,3	317,3	767,4	2303
P4	343,1	185,5	2121,3	323,6	209	279,6	212	272,4
P5	1176,8	658	970,5	1250,1	855,6	684,8	768,4	587,8
P6	209,3	607,3	344,6	1898	362	608,8	300	241,2
P7	1889,5	804,3	2337,8	1199,5	550,1	624,1	772,1	1406,4
P8	1430,8	2074,5	3878,6	5932,3	1312,8	1253,1	1377,3	1211,6
P9	265,6	281	3596	3586,5	554,5	325,8	583	324,2
P10	265,6	961,8	3596	1625,3	554,5	970,3	583	1244,8
Média	638,9	707,8	2089,9	1911,5	573,3	586,6	699,3	817,5
DP	588,1	595,7	1456,1	1687	389,2	360	635	686

Fonte: Elaborado pela autora. Legenda: (P): participantes; (CV): complexidade simples; (CCV): complexidade complexa; e (DP) desvio padrão. *tempo de reação calculado em milissegundos.

Para melhor exposição, os resultados inferenciais serão dispostos de acordo com as hipóteses norteadoras do presente estudo.

H1- haveria integração de pistas audiovisuais na percepção de fala nos falantes do PB, marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+.

Um primeiro teste estatístico foi conduzido de modo a verificar se as crianças típicas sofreriam influência da pista visual durante tarefa de percepção multimodal e, ainda, verificar se haveria uma diferença significativa no tempo de reação na execução da tarefa considerando as condições de apresentação AV- comparativamente à condição AV+.

Na comparação entre as condições de apresentação audiovisual congruente e audiovisual incongruente o teste GzLM não demonstrou efeito significativo ($T(2,70)=208, p>0,286$).

De modo a comparar o tempo de reação em tarefa de identificação de estímulos apresentados nas condições audiovisual congruente e incongruente o teste GzLM não demonstrou efeito significativo ($T(-0,57)=208, p=1$).

H2- a percepção de fala sofreria influência do padrão silábico, ou seja, haveria uma diferença do grau de influência visual em função da complexidade silábica. Espera-se maior grau de influência visual e maior tempo de reação no padrão silábico complexo.

Um segundo teste foi conduzido, de modo a verificar se o grau de influência visual seria dependente do padrão silábico (simples e complexo). O grau de influência visual foi calculado pela subtração (AV+) – (AV-). Na comparação entre os padrões silábicos simples e complexo o teste GzLM não demonstrou diferença significativa ($Z(0,23)= 436, p=1$).

De modo a comparar o tempo de reação em tarefa de identificação de estímulos apresentados nas condições audiovisual congruente e incongruente o teste GzLM não demonstrou efeito significativo ($Z(-0,245)=342, p=1$).

Perante análise inferencial dos dados foi possível verificar que crianças típicas não parecem ser influenciadas pela pista visual e, ainda, o grau de influência visual não varia em função do padrão silábico.

4.3.5 Resultados: crianças com distúrbio fonológico

A Tabela 5, a seguir, apresenta a porcentagem de identificação correta dos estímulos de acordo com as condições de apresentação. Na condição

audiovisual incongruente, consideramos a porcentagem de acerto do estímulo auditivo apresentado.

Tabela 5: Porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico para crianças com Distúrbio Fonológico.

P	Somente Via Auditiva		Somente Via Visual		Audiovisual Congruente		Audiovisual Incongruente	
	CV	CCV	CV	CCV	CV	CCV	CV	CCV
P1	100	100	100	100	100	100	33,3	66,6
P2	100	100	100	100	100	100	83,3	83,3
P3	100	100	100	100	100	100	100	66,6
P4	100	100	100	100	100	100	100	50
P5	100	100	100	100	100	100	66,6	66,6
P6	100	100	100	100	100	100	83,3	33,3
P7	100	100	100	100	100	100	100	33,3
P8	100	100	100	100	100	100	100	100
P9	100	100	100	100	100	100	83,3	100
P10	100	100	50	100	16,6	100	100	83,3
Média	100	100	96,1	100	93,3	100	87,7	71,3
DP	0	0	14,5	0,3	24,2	0,3	20,5	25

Fonte: Elaborado pela autora. Legenda: (P): participantes; (CV): padrão silábico simples; (CCV): padrão silábico complexo; e (DP) desvio padrão.

A Tabela 6, a seguir, apresenta o tempo de reação em relação aos acertos das crianças com distúrbio fonológico para as quatro diferentes condições de apresentação dos estímulos.

Tabela 6: Tempo de reação em tarefa de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico para crianças com Distúrbio Fonológico.

P	Somente Via Auditiva		Somente Via Visual		Audiovisual Congruente		Audiovisual Incongruente	
	CV	CCV	CV	CCV	CV	CCV	CV	CCV
P1	7891,8	8274,6	12099,5	8335,8	4703,5	6397,5	2902,5	3050,2
P2	7615,3	1299,8	5503,1	2706,8	2215	1457,5	1153,2	698,8
P3	671,3	513,1	2693,1	852,3	2231,1	319,5	3400,8	663,7
P4	656,5	9226,3	2032	3166,6	596,8	5637,3	325,6	4779
P5	1645,8	3453,5	1941,3	877,3	2385,8	1832,8	1376,5	1820,7
P6	931,3	1892	1473	1025,3	1586	749,5	2962,2	2994,5
P7	641,1	3153,1	2901,8	4731,8	208,5	4383,6	386,6	2741,5
P8	650,8	582	4134,8	3211,8	623	851,8	765,5	1001,5
P9	2191,8	2275,8	2033,3	2760	1156,5	1739,3	1052	2500,3
P10	1257,3	1035	862	1860,5	659	815,3	1894,3	709,6
Média	2029,7	2891,5	2989,7	2693,6	1380,7	2207,7	1368,5	1914,6
DP	2739,2	3092,4	3263,5	2310,3	1345,7	2210,9	1170,6	1425,3

Fonte: Elaborado pela autora. Legenda: (P): participantes; (CV): complexidade simples; (CCV): complexidade complexa; e (DP) desvio padrão. *tempo de reação calculado em milissegundos.

Para melhor exposição, os resultados inferenciais serão dispostos de acordo com as hipóteses norteadoras do presente estudo.

H1- haveria integração de pistas audiovisuais na percepção de fala nos falantes do PB, marcada pela menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+.

Um segundo teste foi conduzido, de modo a verificar se as crianças com distúrbio fonológico sofreriam influência da pista visual durante tarefa de percepção multimodal e, ainda, verificar se haveria uma diferença significativa no tempo de reação na execução da tarefa considerando as condições de apresentação AV- comparativamente à condição AV+.

Na comparação entre as condições de apresentação (audiovisual congruente e audiovisual incongruente) o teste GzLM demonstrou efeito significativo ($T(4,15)=208, p<0,00$), que mostrou maior acurácia para a condição de apresentação audiovisual congruente. Ou seja, as crianças apresentaram maior porcentagem de acertos quando os estímulos foram apresentados na condição audiovisual congruente comparado a condição audiovisual incongruente.

De modo a comparar o tempo de reação em tarefa de identificação de estímulos apresentados nas condições audiovisual congruente e incongruente o teste GzLM não demonstrou efeito significativo ($T(0,46)=208, p=1$).

H2- a percepção de fala sofreria influência do padrão silábico, ou seja, haveria uma diferença do grau de influência visual em função da complexidade silábica. Espera-se maior grau de influência visual e maior tempo de reação no padrão silábico complexo.

Um segundo teste estatístico foi conduzido, de modo a verificar se o grau de influência visual seria dependente do padrão silábico (simples e complexo). O grau de influência visual foi calculado pela subtração (AV+) – (AV-). Na comparação entre os padrões silábicos simples e complexo o teste GzLM não demonstrou diferença significativa ($Z(-0,70)=436, p=1$).

De modo a comparar o tempo de reação em tarefa de identificação de estímulos apresentados nas condições audiovisual congruente e incongruente o teste GzLM não demonstrou efeito significativo ($Z(1,0215)=342, p=1$).

Perante análise inferencial dos dados foi possível verificar que o grau de influência visual não varia em função do padrão silábico.

4.4 ESTUDO 3: COMPARAÇÃO DE DESEMPENHO DE PERCEPÇÃO MULTIMODAL ENTRE ADULTOS E CRIANÇAS

Diferentemente dos estudos anteriores, o presente estudo objetivou comparar a integração de pistas visuais nos três grupos de participantes investigados: adultos, crianças típicas e com distúrbio fonológico, em tarefa de percepção de fala, com base nos resultados obtidos nos estudos 1 e 2.

A hipótese a ser testada foi a seguinte:

H4- haveria diferenças no desempenho perceptual entre os grupos. Espera-se que os adultos sofram uma maior influência visual comparado às crianças. E ainda, que as crianças típicas sofram uma maior influência visual comparadas às crianças com distúrbio fonológico. A influência visual seria marcada por menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+.

4.4.1 Participantes

Para o desenvolvimento deste estudo, utilizamos os dados coletados e analisados nos estudos 1 e 2. Dessa forma, fizeram parte desse estudo, três grupos de participantes. Foram incluídos no grupo 1 (G1): 11 adultos entre 20-30 anos que apresentaram ausência de alterações na produção dos sons da fala, sendo 5 do sexo feminino e 6 do sexo masculino; no grupo 2 (G2): 10 crianças entre 6 e 8 anos que apresentaram desenvolvimento típico de linguagem, sendo cinco do sexo feminino e cinco do sexo masculino; e no grupo 3 (G3): 10 crianças entre 6 e 8 anos que apresentaram processos de substituição de líquida não lateral em estrutura silábica complexa ou redução de encontro consonantal, sendo cinco do sexo feminino e cinco do sexo masculino. Uma melhor descrição da seleção da amostra foi descrita nas sessões 3.2.1 e 3.3.1.

4.4.2 Experimento de percepção

Procedimento experimental

O procedimento experimental deste estudo foi elaborado segundo descrição do estudo 1. Sendo assim, o procedimento experimental de percepção deste estudo consistiu em um teste de identificação elaborado com o uso do *software OpenSesame*.

O teste de identificação foi composto por quatro etapas distintas: (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.

Estímulos

Os estímulos apresentados foram constituídos por sílabas em padrão silábico simples e complexo. Em padrão silábico simples, o ataque foi preenchido pelas oclusivas: /p/, /t/ e /k/, enquanto em padrão silábico complexo, o ataque foi preenchido pelas mesmas consoantes seguidas da líquida não lateral /r/.

Estabelecidos os estímulos que compuseram o teste, foram realizadas gravações de vídeo e áudio 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 e em formato mp4. O Quadro 1 (p.36, seção 3.2.2) exemplifica a referida distribuição dos estímulos selecionados.

Posteriormente, foi elaborado um *script* para o experimento de identificação a ser executado pelo *software OpenSesame*.

3.4.3 Forma de análise

Para análise dos dados, foram utilizadas as medidas de tempo de reação e porcentagem de acerto dos estímulos de acordo com as seguintes condições de apresentação: somente via auditiva; somente via visual, audiovisual congruente (AV+) e audiovisual incongruente (AV-). Para análise da condição incongruente, consideramos a porcentagem de acerto das respostas baseadas no estímulo auditivo. Posteriormente, a medida de grau de influência visual foi calculada pela subtração (AV+) – (AV-) (SEKIYAMA e BURNHAM, 2008).

O desempenho perceptivo em tarefa de identificação de fala foi verificado perante respostas dos participantes aos testes elaborados e descritos nos estudos 1 e 2.

Análise Estatística

Um tratamento estatístico descritivo e inferencial dos dados foi realizado com o uso do *software Jamovi* (versão 2.2.5.) Inicialmente realizou-se a análise de média e desvio padrão das variáveis acurácia e tempo de reação. Posteriormente, para a análise inferencial foi adotado o teste GzLM (generalized linear models) adotando distribuição Gamma com função de ligação identidade, a fim de verificar o efeito das condições de apresentação dos estímulos sobre os resultados de acurácia e tempo de reação. Comparando o desempenho para os três grupos investigados: (adultos, crianças típicas e crianças com distúrbio fonológico). Para a comparação de médias, utilizou-se o teste Post Hoc de Holm. Adotou-se $p < 0,05$ como valor mínimo de significância estatística.

4.4.3 Resultados

A tabela 7, a seguir, apresenta a média e desvio padrão da porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico, para os adultos, crianças típicas e crianças com distúrbio fonológico.

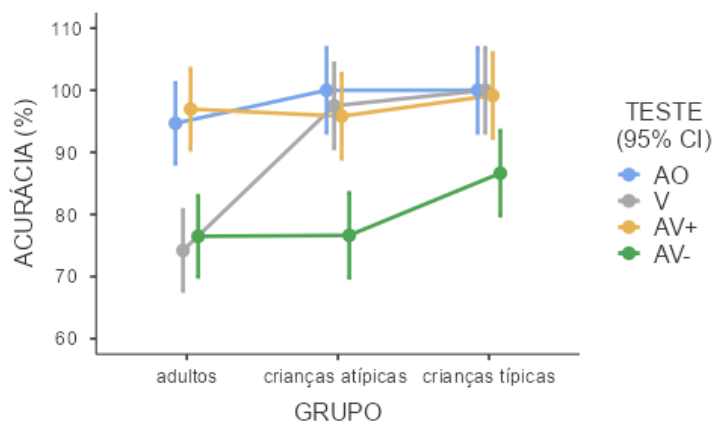
Tabela 7: Média e desvio padrão da porcentagem de identificação correta dos estímulos em função das condições de apresentação e do padrão silábico por grupo.

Grupo	Somente Via Auditiva		Somente Via Visual		Audiovisual Congruente		Audiovisual Incongruente	
	CV	CCV	CV	CCV	CV	CCV	CV	CCV
G1 (M)	95,1	95,9	83,5	70,9	97,7	97,3	77,1	82
G1 (DP)	10,7	7,6	16,9	23,8	9,4	6,5	27	29,8
G2 (M)	100	100	100	100	98,9	100	90,5	86,4
G2 (DP)	0,6	0,3	0,6	0,3	4,9	0,3	8,9	19,5
G3 (M)	100	100	96,1	100	93,3	100	87,7	71,3
G3 (DP)	0,6	0,3	14,5	0,3	24,2	0,3	20,5	25

Fonte: Elaborado pela autora. Legenda: (G1) adultos; (G2) crianças típicas; (G3) crianças com distúrbio fonológico; (CV): padrão silábico simples; (CCV): padrão silábico complexo; (M) média e (DP) desvio padrão.

Uma melhor visualização de desempenho dos três grupos, considerando a acurácia de resposta, é apresentada no Gráfico 1, a seguir:

Gráfico 1: Resultado do teste *Generalized Linear Models* comparando valores de acurácia em função dos grupos investigados (adultos, crianças típicas e crianças com distúrbio fonológico (atípicas)). Barras que não se encostam diferem-se a 95% pelo teste post post-hoc de Holm.



Legenda: Elaborado pela autora. (AO) somente auditivo; (V) somente visual; (AV+) audiovisual congruente; (AV-) audiovisual incongruente.

A tabela 8, a seguir, apresenta a média e desvio padrão do tempo de reação em relação aos acertos para os adultos, crianças típicas e crianças com distúrbio fonológico.

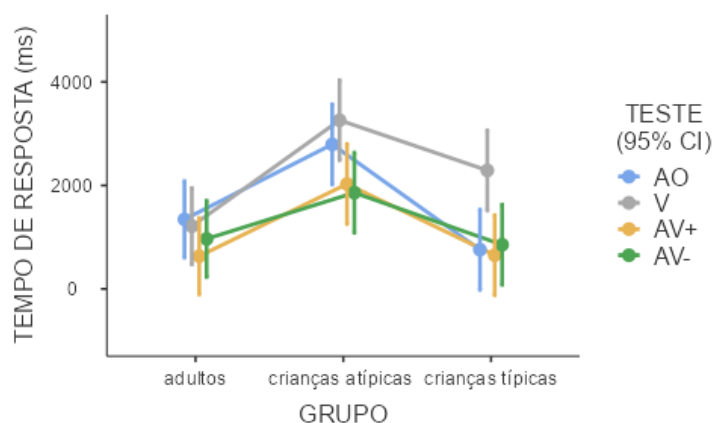
Tabela 8: Média e desvio padrão do tempo de reação dos estímulos em função das condições de apresentação e do padrão silábico por grupo.

Grupo	Somente Via Auditiva		Somente Via Visual		Audiovisual Congruente		Audiovisual Incongruente	
	CV	CCV	CV	CCV	CV	CCV	CV	CCV
G1 (M)	1079,3	1127,7	1288,4	1050,9	587,2	485,6	548,8	911,3
G1 (DP)	612,4	733,8	773,5	582,7	314,4	230,5	438,2	1076,1
G2 (M)	638,9	707,8	2089,9	1911,5	573,3	586,6	699,3	817,5
G2 (DP)	588,1	595,7	1456,1	1687	389,2	360	635	686
G3 (M)	2029,7	2891,5	2989,7	2693,6	1380,7	2207,7	1368,5	1914,6
G3 (DP)	2739,2	3092,4	3263,5	2310,3	1345,7	2210,9	1170,6	1425,3

Fonte: Elaborado pela autora. Legenda: (G1) adultos; (G2) crianças típicas; (G3) crianças com distúrbio fonológico; (CV): padrão silábico simples; (CCV): padrão silábico complexo; (M) média e (DP) desvio padrão. *tempo de reação calculado em milissegundos.

Uma melhor visualização de desempenho dos três grupos, considerando o tempo de reação para resposta, é apresentada no gráfico 2, a seguir:

Gráfico 2: Resultado do teste *Generalized Linear Models* comparando valores de tempo de reação em função dos grupos investigados (adultos, crianças típicas e crianças com distúrbio fonológico (atípicas)). Barras que não se encostam diferem-se a 95% pelo teste post post-hoc de Holm.



Legenda: Elaborado pela autora. (AO) somente auditivo; (V) somente visual; (AV+) audiovisual congruente; (AV-) audiovisual incongruente.

Para melhor exposição, os resultados inferenciais serão dispostos de acordo com a hipótese norteadora do presente estudo.

H3- haveria diferenças no desempenho perceptual entre os grupos. Espera-se que os adultos sofram uma maior influência visual comparado às crianças. E, ainda, que as crianças típicas sofram uma maior influência visual comparadas às crianças com distúrbio fonológico. A influência visual seria marcada por menor porcentagem de acerto e maior tempo de reação na condição AV- comparativamente à condição AV+.

Um primeiro teste estatístico foi conduzido de modo a verificar se haveria uma diferença entre a acurácia em tarefa de percepção multimodal de fala entre os três grupos investigados (adultos, crianças típicas e crianças com distúrbio fonológico), considerando as quatro condições de apresentação do estímulo: somente auditiva, somente visual, audiovisual congruente e audiovisual incongruente. A tabela 9, a seguir, sintetiza os achados do referido teste.

Tabela 9: Comparação de médias acurácia obtidas nos testes para cada grupo utilizando o método post-hoc de Holm.

Teste	Comparação Grupos	Diferencia da média	Desvio Padrão	T	GL	P _{holm}
AO	G1 X G3	-5.314	4.79	-1.1096	164	1
	G1 X G2	-5.314	4.79	-1.1096	164	1
	G3 X G2	1.10E-13	4.9	2.24E-14	164	1
V	G1 X G3	-23.295	4.79	-4.8646	164	< .001*
	G1 X G2	-25.795	4.79	-5.3867	164	< .001*
	G3 X G2	-2.5	4.9	-0.5101	164	1
AV +	G1 X G3	1.134	4.79	0.2367	164	1
	G1 X G2	-2.201	4.79	-0.4597	164	1
	G3 X G2	-3.335	4.9	-0.6804	164	1
AV-	G1 X G3	-10.159	4.79	-2.1214	164	1
	G1 X G2	-10.159	4.79	-2.1214	164	1
	G3 X G2	-10.005	4.9	-2.0412	164	1

Fonte: Elaborado pela autora. Legenda: (G1) adultos; (G2) crianças típicas; (G3) crianças com distúrbio fonológico; (CV): padrão silábico simples; (CCV): padrão silábico complexo;

*Significância estatística $p < 0,05$

Na comparação entre as médias de acurácia para as quatro condições de apresentação dos estímulos o teste GzLM demonstrou diferença significativa para grupo ($F(8)=28, p<0,00$), para condição ($F(20,65)=208, p<001$) e interação de grupo e condição ($F(4,21)=208, p<.001$). Considerando o grupo, os adultos apresentaram um desempenho estatisticamente menor para a condição visual comparado ao desempenho das crianças típicas ($T(-3,93)=28, p<0,002$) e com DF ($T(-2,5)=28, p<0,037$). Considerando a condição de apresentação, o desempenho foi pior para a condição AV- comparado às condições AO ($T(6,96)=208, p<.001$), V ($T(-4,04)=208, p<.001$) E AV+ ($T(6,61)=208, p<.001$). Por fim, considerando a interação entre os grupos e as condições de apresentação, o teste Post Hoc de Holm demonstrou que, a diferença significativa se deu na condição de apresentação somente visual. Para esta condição, o grupo dos adultos se diferenciou do grupo de crianças típicas ($T-25.79)= 4.79, p<.001$), bem como do grupo de crianças com distúrbio fonológico ($T-4.86)= 4.79, p<.001$). Ou seja, os adultos apresentaram menor porcentagem de acerto na condição somente visual comparado às crianças típicas e com DF.

Um segundo teste foi conduzido, de modo a verificar se haveria uma diferença entre o tempo de reação em tarefa de percepção multimodal de fala entre os três grupos investigados (adultos, crianças típicas e crianças com distúrbio fonológico), considerando as quatro condições de apresentação do estímulo: somente auditiva, somente visual, audiovisual congruente e audiovisual incongruente. A tabela 10, a seguir, sintetiza os achados do referido teste.

Tabela 10: Comparação de médias tempo de reação obtidas nos testes para cada grupo utilizando o método post-hoc de Holm.

Teste	Comparação Grupos	Diferencia da média	Desvio Padrão	T	GL	P _{holm}
AO	G1 X G3	-1449.2	535	-2.7084	60.5	0.413
	G1 X G2	586.3	535	1.0957	60.5	1
	G3 X G2	2.04E+03	548	3.72E+00	60.5	0.024*
V	G1 X G3	-2048.7	535	-3.8289	60.5	0.017*
	G1 X G2	-1078.7	535	-2.016	60.5	1
	G3 X G2	970	548	1.7712	60.5	1
AV +	G1 X G3	-1395.1	535	-2.6074	60.5	0.516
	G1 X G2	-19.1	535	-0.0357	60.5	1
	G3 X G2	1376	548	2.5125	60.5	0.645
AV-	G1 X G3	-891.8	535	-1.6667	60.5	1
	G1 X G2	113.1	535	0.2115	60.5	1
	G3 X G2	1005	548	1.835	60.5	1

Fonte: Elaborado pela autora. Legenda: (G1) adultos; (G2) crianças típicas; (G3) crianças com distúrbio fonológico; (CV): padrão silábico simples; (CCV): padrão silábico complexo;

*Significância estatística $p < 0,05$

De modo a comparar o tempo de reação em tarefa de identificação de estímulos apresentados nas quatro condições de apresentação do estímulos o teste GzLM demonstrou diferença significativa para grupo ($F(6,63)=28, p < 0,004$), para condição ($F(12,56)=208, p < .001$) e interação de grupo e condição ($F(2,52)=208, p < 0,02$). Considerando o grupo, as crianças com DF apresentaram um tempo de reação estatisticamente maior comparado ao tempo de reação de adultos ($T(-3,304)=28, p < 0,00$) e crianças típicas ($T(3,005)=28, p < 0,01$). Considerando a condição de apresentação, o tempo de reação depreendido na condição visual foi maior quando comparado ao tempo de reação depreendido nas condições AO ($T(-3,01)=208, p < .0,01$), AV+ ($T(-5,56)=208, p < .001$), e AV- ($T(-4,96)=208, p < .001$). Por fim, considerando a interação entre os grupos e as condições de apresentação, o teste Post Hoc de Holm demonstrou que, a diferença significativa se deu na condição de apresentação somente auditiva. Para esta, o grupo de crianças típicas se diferenciou do grupo de crianças com distúrbio fonológico ($T(3.72)=548, p < 0.02$). Ou seja, crianças com distúrbio

fonológico depreenderam maior tempo de reação para eliciar uma resposta quando o estímulo foi apresentado na condição de apresentação somente auditiva, comparativamente às crianças típicas. Uma segunda diferença significativa se deu na condição de apresentação somente visual. Para esta, o grupo dos adultos se diferenciou do grupo de crianças com distúrbio fonológico ($T(-3.82) = 535, p < 0.01$). Ou seja, crianças com distúrbio fonológico depreenderam maior tempo de reação para eliciar uma resposta quando o estímulo foi apresentado na condição de apresentação somente visual, comparativamente aos adultos.

Portanto, perante análise inferencial dos dados, considerando a acurácia, foi possível verificar uma diferença significativa para a condição de apresentação do estímulo somente visual. Para esta condição de apresentação, o desempenho dos adultos se diferenciou do desempenho das crianças típicas e crianças com DF, na medida em que os adultos apresentaram maior porcentagem de erro comparativamente às crianças. Considerando o tempo de reação, foi possível verificar uma diferença significativa para as condições de apresentação somente auditiva e somente visual. Para a condição somente auditiva, o desempenho de crianças típicas se diferenciou do desempenho de crianças com DF, na medida em que crianças com DF demoraram mais para eliciar uma resposta; enquanto, para a condição somente visual, o desempenho dos adultos se diferenciou do desempenho de crianças com DF, na medida em que crianças com DF demoraram mais para eliciar uma resposta.

Cabe ressaltar que quando os estímulos foram apresentados em condições unimodais (somente auditiva e somente visual) o tempo de reação para eliciar uma resposta foi maior, como tendência numérica, comparado ao tempo de reação depreendido para identificar os estímulos nas condições de apresentação multimodais (audiovisual congruente e audiovisual incongruente) para todos os grupos investigados.

5 DISCUSSÃO

O presente estudo teve por objetivo investigar como se dá a integração de pistas audiovisuais no PB e, ainda, verificar se haveria um efeito em função do padrão silábico da língua.

Especificamente, procuramos verificar o desempenho em tarefa de percepção multimodal por adultos e crianças. Dentre o grupo infantil, procuramos verificar se o distúrbio fonológico poderia influenciar na percepção de estímulos apresentados por vias auditiva, visual e audiovisual.

A investigação da percepção da fala considerando a integração audiovisual no PB é escassa (SILVA, PACHECO e CAGLIARI, 2017; ASSIS, 2021). A literatura internacional aponta três fatores como influentes na tarefa de percepção multimodal: a idade (DESJARDINS e WERKER, 1996; ROSENBLUM, SCHMUCKLER e JOHNSON, 1997; BURNHAM e DODD, 1998), a condição clínica (DESJARDINS, ROGERS e WERKER, 1997; DODD et al., 2008) e a língua (SEKIYAMA e TOHKURA, 1991; HAYASHI e SEKIYAMA, 1998; CHEN e HAZAN, 2007).

A casuística da presente tese foi composta por adultos, crianças com desenvolvimento típico de linguagem e crianças com distúrbio fonológico.

A escolha de um grupo composto por participantes adultos se deu uma vez que este grupo poderia trazer informações quanto ao desempenho típico em tarefa de percepção multimodal; enquanto a escolha de investigar o desempenho de crianças se deu considerando a existência de uma diferença no desempenho perceptual conforme aumento da idade (HOCKLEY e POLKA, 1994). Entretanto, alguns autores apontam que crianças com distúrbio fonológico apresentariam representações perceptivas menos robustas em comparação aos seus pares típicos (DESJARDINS, ROGERS e WERKER, 1997). Dessa forma, também optamos por dividir o grupo de crianças entre: típicas (sem alteração no desenvolvimento da linguagem) e com distúrbio fonológico.

A tarefa experimental foi elaborada conforme o conhecimento descrito para outras línguas. Diferentes estudos (SEKIYAMA e TOHKURA, 1991; SEKIYAMA e BURNHAM, 2004; e SEKIYAMA e BURNHAM, 2008) verificaram uma menor influência da via visual por falantes da língua japonesa quando comparados à língua inglesa. Autores apontam que a complexidade silábica da

língua analisada seria uma possibilidade explicativa para essa diferença no desempenho perceptivo (SEKIYAMA e BURNHAM, 2004; e SEKIYAMA e BURNHAM, 2008).

Tendo em vista os objetivos e metodologia necessária para execução da presente tese, os objetivos foram explorados em três estudos, (1) *Desempenho de percepção multimodal de fala em adultos típicos*, (2) *Desempenho de percepção multimodal de fala em crianças típicas e com Distúrbio Fonológico*; e (3) *Comparação de desempenho de percepção multimodal entre adultos e crianças*.

Neste capítulo, discutiremos os resultados encontrados nos três estudos desenvolvidos e suas possíveis hipóteses explicativas.

Para melhor exposição às hipóteses explicativas serão dispostas em função das três hipóteses centrais da presente tese: (1) haveria integração de pistas audiovisuais na percepção multimodal da fala para falantes do PB; (2) a percepção multimodal da fala sofreria influência do padrão silábico (simples e complexo); e (3) haveria diferença no desempenho perceptual entre adultos, crianças e crianças com distúrbio fonológico.

5.1 Integração de pistas audiovisuais na percepção multimodal da fala para falantes do PB

A análise de acurácia (porcentagem de acerto dos estímulos) e do tempo de reação (tempo depreendido para eliciar uma resposta) possibilitou verificar de que forma adultos, crianças típicas e crianças com distúrbio fonológico manipulam as pistas visuais e auditivas apresentadas de forma combinada, tanto na condição de apresentação dos estímulos congruentes (AV+), quanto na apresentação dos estímulos incongruentes (AV-).

Na análise individual dos grupos, os resultados da acurácia demonstraram uma integração de pistas audiovisuais para o grupo dos adultos e crianças atípicas, marcada pela menor porcentagem de acerto em estímulos apresentados na condição audiovisual incongruente. Diferentemente, as crianças típicas não demonstraram diferenças significantes entre as duas condições de apresentação: AV+ vs AV-. Além disso, no terceiro estudo, os resultados da acurácia demonstraram que a condição AV- se diferenciou das

demais condições de apresentação (somente visual, somente auditiva, audiovisual congruente e incongruente), ou seja, os três grupos investigados, apresentaram pior desempenho em tarefa de identificação dos estímulos quando eles foram apresentados na condição audiovisual incongruente.

Na análise individual dos grupos, considerando o tempo de reação depreendido para execução da tarefa, os resultados não demonstraram diferenças significativas para nenhum dos três grupos (adultos, crianças típicas e crianças com DF). Além disso, no terceiro estudo, observou-se que quando os estímulos foram apresentados em condições unimodais (somente auditiva e somente visual) o tempo de reação para eliciar uma resposta foi maior, como tendência numérica, comparado ao tempo de reação depreendido para identificar os estímulos nas condições de apresentação multimodais (audiovisual congruente e audiovisual incongruente) para todos os grupos investigados.

Os resultados observados, portanto, demonstraram que há influência da pista visual na percepção da fala, verificada no desempenho da acurácia e no tempo de reação da resposta.

Ao assumir a teoria multimodal como base teórica do presente estudo, consideramos que o fenômeno da percepção da fala ocorre de forma integrada, ou seja, as pistas sensoriais envolvidas no processo de percepção de fala não poderiam ser dissociadas.

Dessa forma, foi possível verificar a integração de pistas audiovisuais na percepção multimodal da fala para os participantes investigados, falantes do Português Brasileiro, confirmando, portanto, nossa primeira hipótese.

Nossos resultados corroboram os achados apresentados internacionalmente, na medida, em que são verificadas influências da pista visual no processo perceptivo, para o inglês, japonês (SEKIYAMA e TOHKURA, 1991; SEKIYAMA e TOHKURA, 1993; SEKIYAMA e BURNHAM, 2004), espanhol (FUSTER-DURAN, 1996), alemão (FUSTER-DURAN, 1996), chinês (SEKIYAMA, 1997), finlandês (SAMS, et al, 1998), italiano (BOVO et al, 2009) e hebraico (TAITELBAUM-SWED e FORTICK, 2016). A influência visual nesses estudos é descrita pela presença da ilusão perceptual (efeito McGurk) e/ou pela menor porcentagem de identificação correta de estímulos apresentados na condição de apresentação audiovisual incongruente.

Majoritariamente, as investigações da percepção multimodal da fala são realizadas considerando o desempenho de falantes da língua inglesa e japonesa (SEKIYAMA e TOHKURA, 1991; HAYASHI e SEKIYAMA, 1998; e CHEN e HAZAN, 2007).

A título de exemplificação, Sekiyama e Tohkura (1991) investigaram a percepção de 10 japonesas, de idade entre 22 e 27 anos, durante a apresentação de estímulos audiovisuais. Os autores testaram duas condições de apresentação de fala: (1) estímulo auditivo com adição de ruído; e (2) estímulo auditivo sem adição de ruído. Foram verificadas poucas ocorrências do efeito McGurk na condição de estímulo sem adição de ruído. Diferentemente, na condição com adição de ruído foram verificadas maior ocorrência do efeito, demonstrando uma percepção da fusão de informações visual e auditiva dos estímulos apresentados. Os autores concluíram que falantes nativos do japonês usaram a informação visual na percepção da fala apenas quando o componente auditivo do estímulo não estava perfeitamente inteligível.

Encontra-se na literatura, em menor número de publicações, estudos que buscaram replicar o efeito McGurk e descrever como se dá o fenômeno de percepção multimodal para outras línguas. A seguir, será apresentada breve síntese quanto ao desempenho, descrito na literatura, em tarefa de percepção multimodal por falantes do alemão, espanhol, chinês, finlandês, italiano e hebraico.

Os autores se propuseram a investigar o desempenho em tarefa de percepção audiovisual da fala. Para isso, em cada estudo, foram selecionados estímulos que foram apresentados sob diferentes condições. Em sua maioria, os estudos propõe a investigação de quatro condições de apresentação: somente auditivo, somente visual, audiovisual congruente e audiovisual incongruente. Em geral, os estudos investigaram a condição de apresentação audiovisual incongruente.

Achados considerando o espanhol e o alemão foram verificados em Fuster- Duran (1996). O autor propôs investigar o desempenho de falantes nativos do espanhol e do alemão em tarefa de percepção de estímulos de fala produzidos por falantes da sua língua nativa e não nativa. Foi verificado o efeito McGurk para as duas línguas investigadas, sendo que os participantes se ancoraram mais fortemente em pistas visuais, quando foram apresentadas

palavras em língua não nativa. Sekiyama (1997), por sua vez, ao investigar o desempenho de falantes da língua chinesa em tarefa de percepção de estímulos de fala, produzidos por falantes do inglês e do japonês, verificou que chineses apresentaram menor efeito de ilusão, ou seja, um efeito McGurk mais fraco para os estímulos da língua japonesa comparados aos estímulos da língua inglesa. Considerando o desempenho de percepção audiovisual de falantes do finlandês, Sams, et al (1998) verificaram que as participantes investigadas foram influenciadas pela pista visual, na medida, em que foi verificado o efeito McGurk nos diferentes estímulos apresentados (sílabas isoladas, palavras e logatomas).

Bovo et al. (2009), por sua vez, também verificam um valor significativo de respostas que correspondem a uma ilusão perceptual, no desempenho perceptual de falantes do italiano. Ou seja, o efeito McGurk também foi verificado em tarefa de percepção audiovisual para o italiano. Por fim, considerando o hebraico, Taitelbaum-Swed e Fortick (2016) os autores que propuseram investigar o desempenho da percepção de fala ao longo da idade, verificaram, que o desempenho perceptual tende a apresentar um pior desempenho nas faixas etárias de 4-5 anos e 65-80 anos em comparação às faixas etárias de 8-9 anos, 20-30 anos e 40-55 anos. Entretanto, todos os grupos demonstraram ser influenciados pela pista visual em tarefa de percepção em maior ou menor grau.

Da mesma forma, nossos resultados corroboram os achados apresentados por Silva, Pacheco e Cagliari (2017) na investigação do desempenho de percepção audiovisual em falantes adultos do PB. Em seu estudo, os autores, propuseram investigar a percepção multimodal de sons fricativos. Os estímulos foram apresentados aos participantes sem manipulação do sinal acústico e com manipulação do sinal acústico. Os resultados demonstraram que quando as fricativas foram apresentadas sem manipulação do sinal acústico o desempenho perceptivo para as condições somente auditiva e audiovisual foram parecidas. O desempenho na condição somente visual, por sua vez, apresentou uma menor porcentagem de acerto comparativamente às demais condições de apresentação. Entretanto, quando as fricativas foram apresentadas diante da manipulação do sinal acústico, foi verificado um aumento no desempenho perceptual das fricativas na condição de apresentação audiovisual. De maneira geral, a porcentagem de identificação das fricativas foi

maior quando os estímulos foram apresentados na condição de apresentação audiovisual.

A literatura resenhada, portanto, descreve a existência da integração de pistas audiovisuais da fala considerando as línguas investigadas. Embora o grau de influência visual varie de língua para a língua, parece que a integração sensorial é algo constitutiva do fenômeno da percepção.

Autores apontam a estrutura silábica da língua como hipótese explicativa para as diferenças de desempenho entre as línguas (SEKIYAMA e BURNHAM, 2004). Considerando os estudos apresentados, o japonês e o chinês demonstraram menor influência da pista visual, ou seja, se beneficiam menos da integração audiovisual na percepção da fala, enquanto as demais línguas parecem se beneficiar mais da integração de pistas. Dentre as línguas investigadas (japonês, chinês, inglês, alemão, italiano, espanhol, finlandês e hebraico) o chinês e o japonês são as únicas línguas que não apresentam padrão complexo de estrutura silábica. Nossos achados demonstraram há existência da integração audiovisual em tarefa de percepção de fala para falantes do PB, assim como verificado em Silva, Pacheco e Cagliari (2017), considerando que o PB contempla padrão simples e complexo, podemos inferir que nossos achados corroboram a hipótese explicativa levantada pela literatura internacional.

Cabe ressaltar, a segunda hipótese norteadora da presente tese: “a *percepção multimodal da fala sofreria influência do padrão silábico (simples e complexo)*”. Para essa hipótese, nossos achados demonstraram que o desempenho perceptivo de adultos, crianças típicas e crianças com DF, não foi dependente do padrão silábico (simples ou complexo), entretanto, hipóteses explicativas para a não diferenciação de desempenho considerando os padrões silábicos serão apresentadas na próxima seção.

5.2 Padrão silábico (simples e complexo) e a percepção multimodal da fala

Considerando os achados de acurácia, o parâmetro *grau de influência visual* foi calculado pela subtração da porcentagem de acerto de AV- da porcentagem de acerto de AV+. A fim de investigar se o padrão silábico (simples e complexo) influenciaria no desempenho perceptual de adultos, crianças típicas

e crianças com DF, considerou-se o grau de influência visual e o tempo de reação para eliciar uma resposta.

Os resultados apresentados não demonstraram diferença significativa considerando o grau de influência visual ou o tempo de reação a depender do padrão silábico para nenhum dos grupos investigados, ou seja, o desempenho perceptivo de adultos, crianças típicas e crianças com DF, não foi dependente do padrão silábico (simples ou complexo).

Em relação ao padrão silábico os resultados apresentados não confirmam, portanto, nossa segunda hipótese.

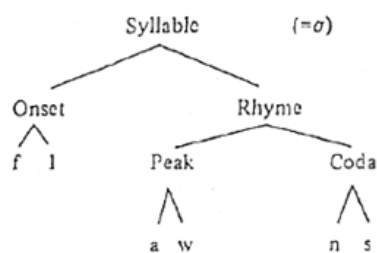
Não há na literatura estudos que investigam o desempenho perceptivo multimodal em função dos diferentes padrões silábicos. A literatura existente somente realiza a comparação de desempenho perceptivo entre falantes de diferentes línguas.

A título de exemplificação, SEKIYAMA e BURNHAM (2008) buscaram comparar o desempenho perceptivo de falantes da língua japonesa, que apresenta exclusivamente o padrão silábico simples (CV) com o desempenho perceptivo de falantes da língua inglesa, que apresenta padrão silábico simples (CV) e complexo (CCV). Os autores apontam a complexidade do sistema fonético-fonológico como um possível fator explicativo para as diferenças no fenômeno perceptual encontradas. Segundo os autores, os falantes da língua inglesa apresentaram maior uso de integração de pistas, pois a maior complexidade do sistema fonético-fonológico exigiria maior uso da pista visual para perceber o estímulo apresentado.

Considerando a inexistência de uma descrição pormenorizada quanto ao desempenho perceptual investigando a influência do padrão silábico para o PB, ao definirmos a segunda hipótese da presente pesquisa, consideramos os achados da literatura internacional existente.

A noção de sílaba é comum a todas as línguas, mas cada língua organiza os fonemas na sílaba de maneira diferente (CAGLIARI, 1981). Selkirk (1982), por exemplo, descreve uma estrutura interna na sílaba que é hierárquica e universal. A figura 5, a seguir, ilustra essa estrutura interna da sílaba.

Figura 5: Estrutura interna da sílaba



Fonte: (SELKIRK, 1982, p. 341)

Diante do modelo não-linear proposto por Selkirk, com exceção do núcleo (tradução para a estrutura *Peak*, representada na figura 5), qualquer constituinte da sílaba pode sofrer ramificação ou não ser preenchido. As regras específicas de determinada língua irão definir quanto ao preenchimento dessas estruturas. Ademais, nem todas as posições das sílabas estão no mesmo plano estrutural, de modo que pode haver relação de dependência e previsibilidade entre os elementos (CAMPOS, TENANI, BERTI, 2011).

O PB, por exemplo, apresenta nove padrões silábicos que vão de V a CCVCC, enquanto, o inglês, apresenta vinte e quatro padrões silábicos que vão de V a CCCVCCCC (BACK e MATTOS, 1972).

Apesar de ambas as línguas apresentarem o padrão CCV, no PB, a segunda consoante da sílaba, pode ser preenchida, apenas, por dois segmentos consonantais, são eles: /l/ e /r/; enquanto, no inglês, a segunda consoante da sílaba, pode ser preenchida por pelo menos 10 segmentos consonantais, por exemplo: /l/, /r/, /n/, /w/, /f/, /t/, /d/, /k/, /g/ e /p/. Além de apresentar um número menor de possíveis segmentos na posição silábica, os fonemas /l/ e /r/ não são tão visualmente distintivos quanto um som bilabial, por exemplo: /p/.

Cabe ressaltar que todos os estímulos testados na presente tese foram apresentados em condição de silêncio, ou seja, não foram manipulados com inserção de ruído. Estudos (SEKIYAMA e TOHKURA, 1991; SEKIYAMA, BUMHAM e ERDENER, 2003) demonstram que estímulos apresentados com adição de ruído tendem a exigir um maior grau de influência visual, dessa forma, o fato dos estímulos terem sido apresentados em boas condições de apresentação (sem adição de ruído) pode ter contribuído para que os

participantes percebessem igualmente estímulos de padrão silábico simples e padrão silábico complexo.

A título de exemplificação Sekiyama, Bumham e Erdener (2003) buscaram investigar o desempenho da percepção multimodal em falantes das línguas japonesa e inglesa. Foram analisados 48 adultos, na faixa etária entre 18 e 29 anos e 16 crianças, na faixa etária de 6 anos. Os adultos e crianças foram divididos em grupos à depender da sua língua nativa, dessa forma, o estudo consistiu de dois grupos de adultos e dois grupos de crianças. Os estímulos investigados foram /ba/, /da/ e /ga/ produzidas por quatro falantes, um homem e uma mulher falantes do inglês e um homem e uma mulher falantes do japonês. Os estímulos foram editados de forma a apresentar somente pista visual, somente pista auditiva e pistas audiovisuais: congruentes e incongruentes. Os estímulos foram apresentados sob diferentes níveis de inteligibilidade, para isso, testou-se estímulos com e sem adição de ruído. Os resultados demonstraram que estímulos apresentados com adição de ruído exigiram maior grau de influência visual, calculada pela diferença entre a porcentagem correta da condição AV+ da porcentagem correta da condição AV-, de falantes da língua inglesa, do que falantes da língua japonesa, independente da faixa etária investigada.

Com isso, levantamos três hipóteses explicativas para que os três grupos investigados não apresentassem diferenças significativas de desempenho perceptivo considerando o padrão silábico. A primeira hipótese deve-se ao fato de a língua portuguesa apresentar apenas dois possíveis segmentos combinatórios para constituir sílabas de padrão complexo (CCV), sendo esses dois segmentos menos distintivos visualmente comparados a segmentos labiais. A segunda hipótese consiste no conhecimento quanto à relação de dependência e previsibilidade entre os elementos. E por fim, a terceira hipótese foi levantada considerando a apresentação dos estímulos em condição de silêncio.

Ademais, destacam-se os resultados apresentados na presente tese, no qual, nenhum dos grupos sofreu influência do padrão silábico durante tarefa de percepção multimodal da fala. Por intermédio da investigação perceptual da fala, torna-se possível fazer inferências quanto ao conhecimento que uma criança, na faixa etária investigada, tem da língua.

5.3 Diferença no desempenho perceptual entre adultos, crianças e crianças com distúrbio fonológico

Para a análise de comparação entre os grupos, foram consideradas as médias da porcentagem de acerto (acurácia) e tempo de reação entre as quatro condições de apresentação (somente auditiva, somente visual, audiovisual incongruente e audiovisual congruente), considerando os três grupos investigados (adultos, crianças típicas e crianças com DF).

Os resultados da acurácia demonstraram uma diferença significativa para a condição de apresentação do estímulo somente visual. Para esta condição de apresentação, o desempenho dos adultos se diferenciou do desempenho das crianças típicas e crianças com DF, na medida em que os adultos apresentaram maior porcentagem de erro comparativamente às crianças.

Diferentemente, os resultados do tempo de reação demonstraram uma diferença significativa para as condições de apresentação somente auditiva e somente visual. Para a condição somente auditiva, o desempenho de crianças típicas se diferenciou do desempenho de crianças com DF, na medida em que crianças com DF demoraram mais para eliciar uma resposta; enquanto, para a condição somente visual, o desempenho dos adultos se diferenciou do desempenho de crianças com DF, na medida em que crianças com DF demoraram mais para eliciar uma resposta.

Perante dados apresentados foi possível verificar diferenças estatísticas entre os grupos investigados, marcadas por uma maior porcentagem de acerto na condição somente visual por crianças típicas e crianças com DF comparativamente aos adultos, bem como, marcadas pela diferença no tempo de reação, quando os estímulos foram apresentados na condição somente auditiva, as crianças com DF apresentaram maior tempo de reação comparativamente às crianças típicas. Quando os estímulos foram apresentados na condição somente visual, as crianças com DF apresentaram maior tempo de reação comparativamente aos adultos.

Dessa forma, em relação à comparação de desempenho de percepção multimodal entre os grupos investigados, os resultados demonstram que há diferença no desempenho perceptual entre adultos, crianças típicas e crianças com DF, sendo que as diferenças verificadas são dependentes da condição de apresentação.

Especificamente, quanto ao desempenho das crianças com DF, os estudos, majoritariamente, demonstram que as crianças com DF são menos susceptíveis à influência visual na percepção multimodal quando comparadas às crianças com desenvolvimento típico.

A título de exemplificação internacional, em seu estudo, Desjardins e Werker (1996), investigaram a integração de pistas audiovisuais em crianças com distúrbio 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. Ou seja, os autores relatam que os erros de fala poderiam estar associados à menor influência da pista visual.

Na literatura nacional, Assis (2021), investigou a integração de pistas audiovisual, comparando o desempenho de crianças com e sem distúrbio fonológico, entre quatro e oito anos. Segundo resultados apresentados no estudo, a autora verificou que a pista visual mostrou-se mais difícil em tarefa de percepção multimodal para as crianças com e sem DF. Além disso, crianças com DF tiveram menor número de acertos comparativamente às crianças típicas, portanto, pior desempenho na tarefa perceptual multimodal, ou seja, o grupo de crianças com DF foi menos influenciado pela pista visual comparativamente ao grupo controle.

Por outro lado, embora, Dodd et al. (2008) também investigaram a percepção multimodal em crianças com desenvolvimento típico e crianças com DF, os autores apontam que perante análise inferencial dos dados, não foram verificadas diferenças estatísticas entre os grupos. Ou seja, em seu estudo, ambos os grupos apresentaram uma pior acurácia para a condição de apresentação somente via visual e melhor acurácia para a condição de apresentação somente por via auditiva. Adicionalmente, na análise do estímulo audiovisual incongruente, os resultados demonstraram que 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 apresentados em Dodd, et al (2008), por sua vez, não demonstraram diferença significativa entre os grupos investigados (crianças com desenvolvimento típico e crianças com distúrbio fonológico).

Dentre os estudos levantados e resenhados ao longo da presente tese, que buscaram investigar a influência do distúrbio fonológico na percepção multimodal da fala demonstram que não há um consenso quanto à integração de vias sensoriais na percepção da fala para o referido grupo clínico.

Considerando os resultados apresentados no presente estudo, nossos achados corroboram com a descrição do fenômeno perceptual de integração de vias descrita em Dood, et al (2008).

Uma possível hipótese explicativa para o fato de crianças típicas e crianças com distúrbio fonológico apresentarem melhor desempenho perceptivo na condição de apresentação somente visual comparativamente com o desempenho perceptivo apresentado por adultos, deve-se ao fato, de que as crianças investigadas se encontram em faixa etária de alfabetização.

Segundo a literatura (BURNHAM e ERDENER, 2013), crianças em fase de alfabetização melhoram o desempenho visual, na medida em que buscam por pistas extras de informação que contribuam na correspondência grafema-fonema, bem como, a pista visual torna-se mais privilegiado em sala de aula.

Sabe-se que a definição de fala e escrita pode ser compreendida por diferentes referenciais teóricos. Dentro da concepção de fala e escrita como atos enunciativos, deve-se considerar a inserção do falante/ escrevente em determinadas práticas de oralidade/letramento como práticas discursivas (CHACON, 2021).

Ainda dentro dessa concepção, fala e escrita correspondem aos atos enunciativos concretos, ou seja, enunciados de determinada prática de oralidade e letramento. Oralidade e letramento, por sua vez, correspondem à diversidade de práticas discursivas que regulam a produção e a circulação dos enunciados falados e escritos, respectivamente (CHACON, 2021).

Segundo Tfouni (1995), a constituição de um sujeito linguístico, como falante e como escrevente ocorre quando práticas de oralidade e letramento são atravessadas simultaneamente. Dessa forma, a percepção de fala poderia ser vista enquanto prática enunciativa e, portanto, estaria atravessada por práticas de letramento.

Isso significa dizer que a fala, enquanto ato enunciativo e fato linguístico resulta do atravessamento (e constituição) do sujeito não apenas por práticas de oralidade, mas também por práticas de

letramento. O mesmo se pode dizer da escrita: enquanto ato enunciativo e fato linguístico, ela resulta do atravessamento (e constituição) do sujeito não apenas por práticas de letramento, mas também por práticas de oralidade. Fala e escrita seriam, pois, constitutivamente heterogêneas, já que se mostram atravessadas, ou constituídas, simultaneamente por características de enunciados produzidos em múltiplas práticas de oralidade e de letramento, que se interpenetram tanto nos atos de fala quanto nos de escrita (CHACON, 2021).

Destaca-se, a contribuição científica que pesquisadores do Grupo de Pesquisa Estudos sobre a Linguagem (GPEL) têm oferecido quanto ao conhecimento sobre a relação entre a percepção-auditiva e a aquisição da escrita (SCHIER, BERTI e CHACON, 2013; e CHACON e VAZ, 2013; SCHIER, 2017).

A título de exemplificação, em Chacon e Vaz (2013) os autores buscaram verificar a relação entre a aquisição da percepção-auditiva e a aquisição da ortografia de consoantes soantes. Para isso, selecionaram e investigaram o desempenho de 53 escolares do 1º ano do Ensino Fundamental e 48 escolares do 2º ano do Ensino Fundamental. Para investigação do aspecto perceptual-auditivo, utilizou-se o instrumento de Avaliação da Percepção de Fala-PERCEFAL (Berti, 2011). 30 palavras foram apresentadas no experimento perceptual-auditivo, e posteriormente, escritas por cada participante individualmente. Para análise dos dados considerou-se à acurácia perceptual-auditiva (porcentagem de erros, acertos e não respostas) e a acurácia ortográfica (porcentagem de erros, acertos, não respostas e não interpretáveis). Os resultados mostraram que alunos do 1º ano do Ensino Fundamental apresentaram maior índice de acertos no desempenho perceptual-auditivo do que no desempenho ortográfico, e maior índice de erros no desempenho ortográfico do que no desempenho perceptual-auditivo. Enquanto, os alunos do 2º ano do Ensino Fundamental apresentaram maior índice de acertos no desempenho ortográfico do que no desempenho perceptual-auditivo, e maior índice de erros no desempenho perceptual-auditivo do que no desempenho ortográfico. Os autores sugerem que a inserção em práticas de letramento afeta a aquisição da percepção- auditiva, pois, embora os desempenhos nas tarefas

de percepção e de ortografia pareçam se estabilizar de maneira gradual, essa estabilização ocorre de maneira mais lenta no desempenho perceptual- auditivo.

Burnham e Erdener (2013) apontam que na busca por realizar a correspondência fonema-grafema, as crianças podem buscar qualquer informação extra para estabelecer essa correspondência, sendo a integração de pistas auditivas e visuais uma das fontes extras de informação. Os autores também ressaltam que em salas de aula, os professores usam um estilo de fala hiperarticulado que presumivelmente destaca os movimentos da face e dos lábios, ou seja, informações visuais da fala são privilegiadas.

A análise de comparação entre os grupos investigados considerando o tempo de reação em tarefa de identificação demonstrou que os desempenhos apresentados para as condições de apresentação do estímulo somente visual e somente auditiva apresentaram diferenças entre os grupos. Quando os estímulos foram apresentados na condição somente auditiva, as crianças com DF apresentaram maior tempo de reação comparativamente às crianças típicas; enquanto quando os estímulos foram apresentados na condição somente visual, as crianças com DF apresentaram maior tempo de reação comparativamente aos adultos.

Considerando os achados da presente pesquisa é possível inferir uma possível gradação no desempenho perceptivo de fala considerando o tempo de reação. Crianças com DF parecem apresentar um tempo de reação mais laborioso, seguidas das crianças típicas e adultos.

A literatura existente que se propõe investigar a integração de pistas visual e auditiva na percepção de fala, majoritariamente descreve um pior desempenho de crianças com DF, comparativamente às crianças com desenvolvimento típico de linguagem e adultos.

A título de exemplificação, em um estudo nacional (ASSIS, 2021), a autora buscou investigar a percepção multimodal de crianças típicas e crianças com DF e verificou que as crianças com DF apresentaram um tempo de reação maior quando comparados às crianças típicas, na identificação dos estímulos apresentados.

Uma possível hipótese explicativa para o fato de crianças apresentarem um tempo de reação superior ao tempo demandado por adultos, consiste na inferência de que o fenômeno perceptual da fala inicia-se por estímulos auditivos

e, posteriormente, desenvolve-se habilidades de identificação visual. Nossos achados corroboram com a literatura. Taitelbaum-Swead (2016) apresenta que, 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. Da mesma forma, Hazan (2009) também verificou uma gradação no desenvolvimento do uso da pista visual para a percepção da fala.

Podemos concluir com nossos achados, que há diferença no desempenho perceptual entre adultos, crianças típicas e crianças com DF, sendo que as diferenças verificadas são dependentes da condição de apresentação. Essa diferença entre os grupos pode ser verificada considerando a porcentagem de acerto dos estímulos apresentados (acurácia) e/ou considerando o tempo de reação apreendido para eliciar uma resposta.

Como limitações do presente estudo, alertamos para o fato do número amostral da presente pesquisa ser limitado. Sugerem-se novas investigações de desempenho considerando diferentes faixas etárias, incluindo por exemplos, um grupo de adolescentes. Dessa forma, seria possível acompanhar a transição perceptual da infância a vida adulta, considerando as diferentes fases de maturação de processamento auditivo e visual. Ademais, para os próximos trabalhos que objetivem investigar o desempenho de crianças com DF, sugere-se que a seleção da amostra considere não somente os processos fonológicos de estrutura da sílaba, como também os processos fonológicos de substituições. Cabe ressaltar, como uma limitação, que na presente tese foi investigada somente uma possibilidade de preencher a estrutura CCV. Apesar, desta não ter se mostrado relevante, sugere-se que novas investigações sejam realizadas considerando as duas possibilidades de preenchimento.

6 CONSIDERAÇÕES FINAIS

Ao apresentar os objetivos propostos na presente tese, esperávamos que os resultados pudessem contribuir teoricamente para o aprofundamento da compreensão quanto ao desempenho em tarefa de percepção multimodal da fala por falantes do PB.

Considerando os estudos nacionais e internacionais existentes que se propuseram a investigar a percepção de fala considerando a integração de pistas sensoriais, é possível concluir que os achados do presente estudo, apesar de iniciais, são inovadores e abrem espaço para que novas investigações sejam realizadas.

Apesar de o número de indivíduos do estudo ter sido limitado, a amostra foi capaz de descrever o desempenho de falantes do PB em tarefa de percepção multimodal e demonstrar a influência da pista visual na percepção de fala.

Em suma, os resultados obtidos, sugerem que indivíduos parecem manipular pistas audiovisuais diferentemente em função da idade, ou seja, haveria um desenvolvimento gradual no uso da integração de estímulos visuais e auditivos. Ainda, as alterações nos sons da fala parecem prejudicar o desempenho de crianças em tarefa de identificação de estímulos de fala. Além disso, a percepção de estímulos apresentados de forma isolada ou integradas não parece ser dependente do padrão silábico para a língua portuguesa.

Em termos de contribuições, não somente, acreditamos que os achados deste estudo podem servir como benefício para base teórica na área da linguística, como para a clínica fonoaudiológica. Especificamente, quanto às áreas da linguística, os resultados contribuem para sustentar a natureza multimodal da percepção de fala, de modo que teorias/ modelos multimodais sejam adotados. A investigação da percepção da fala sob viés da teoria multimodal, pode contribuir para o conhecimento quanto ao desenvolvimento das representações dos aspectos de produção e percepção da fala, na medida em que possibilita investigar a estrutura linguística considerando uma nova entrada de investigação. Adicionalmente, os achados poderão ser utilizados como referencial teórico quanto ao desenvolvimento típico considerando o desempenho de adultos e crianças em tarefa de percepção multimodal de fala e oferecer auxílio à clínica fonoaudiológica, na análise do desempenho da

percepção multimodal de participantes com distúrbio fonológico. Desse modo, inserir no cenário clínico, planos terapêuticos que visem a avaliação e o desenvolvimento da percepção multimodal, poderiam implicar em resultados terapêuticos satisfatório.

REFERÊNCIAS

- ALBANO, E. C. O gesto e suas bordas. esboço de fonologia acústico-articulatória do português brasileiro. Mercado de Letras/ Fapesp, DE LETRAS/FAPESP, 2001.
- American Psychiatric Association, Manual diagnóstico e estatístico de transtornos mentais: DSM-5. 5ª ed.; M.I.C. Nascimento, Trad. Porto Alegre, RS: Artmed. 2014
- ANGELO T.C.S.; BEVILACQUA M.C.; MORET A.L.M. Percepção da fala em deficientes auditivos pré-linguais usuários de implante coclear. **Pró- Fono**, v. 22, n. 3, 2010.
- ASSIS, M.F et al. Desempenho perceptivo-auditivo em crianças com e sem transtorno fonológico na classe das oclusivas. **CODAS**, v. 33, 2021.
- ASSIS, M.F. Integração de pistas em tarefa de percepção de fala de fricativas desvozeadas em crianças com e sem transtorno fonológico. Dissertação de Mestrado. Programa de pós graduação em Fonoaudiologia. Universidade Estadual Paulista. Marília, 2021.
- AULI, M.A.B. et al. Audiovisual speech perception in children with developmental language disorder in degraded listenig conditions. **Journal of Speech, Language and Hearing Research**, v. 56, p. 211- 221, 2013.
- BACK, E.; MATTOS, G. Gramática constritural da lingua portuguesa. São Paulo: FTD, 1972.
- BARANEK, G. T. et al. Sensory Experiences Questionnaire: Discriminating sensory features in young children with autism, developmental delays, and typical development. **Journal of Child Psychology and Psychiatry**, v. 47, p. 591–601, 2006.
- BARBARENA, L.S; SOARES, M. K.; BERTI, L.C. Descrição dos gestos articulatórios envolvidos na produção dos sons /r/ e //l/. **Revista Audiology Communication Research**, v. 19, p. 338-344, 2014.
- BARTON, D. Phonemic perception in children. In: YENI-KOMSHIAN, G.; KAVANAGH, J.; FERGUSON, C. Ferguson (Ed.). **Child Phonology**, v.2. New York: Academic Press, p.97-116, 1980.
- BARZAGHI. L.; BARBOSA, K.; MALT, S.M.E. Deficiência de audição e contraste de vozeamento em oclusivas do português brasileiro: analise acústica e perceptiva. **Distúrbios da Comunicação** v. 19, n. 3, 2007
- BAUM, S. H. et al. Multisensory speech perception without the left superior temporal sulcus. **Neuroimage**, v. 62, n. 3, p. 1825- 1832, 2012.

BECKNER, C. et al. Language is a Complex Adaptive System: Position Paper. **Language Learning**, v. 59, n. 1, p. 1-26, 2009.

BERTI, L.C. Relação entre produção e percepção de fala: coerência com o parâmetro fonético-acústico. **Cadernos de Estudos Linguísticos**, v. 50, n.1, p. 47-67, 2008.

BERTI, L.C. et al. Desempenho perceptivo-auditivo de crianças na identificação de contrastes fonológicos entre as oclusivas. **Jornal Sociedade Brasileira de Fonoaudiologia**, v.24, n.4, p. 348-354, 2012.

BERTI, L.C. PERCEFAL: instrumento de avaliação da identificação de contrastes fonológicos. **Audiology Communication Research**, v. 22, 2017.

BERTI, L.C. Desempenho perceptivo-auditivo de crianças na identificação de contrastes fônicos. **Alfa Revista de Linguística**, v.61, n.1, p. 81-103, 2017.

BERTI, L.C. et al. Relationship between speech production and perception in children with speech sound disorders. **Journal of Portuguese Linguistics**, v. 19, n. 13, p. 1-13, 2020.

BERTI, L.C.; PAGLIUSO, A.; LACAVA, F. Instrumento de avaliação de fala para análise acústica (IAFAC), baseado em critérios linguísticos. **Revista Sociedade Brasileira de Fonoaudiologia**, São Paulo, v. 14, n. 2, p. 305-314, 2009.

BONATTO, M.T.R.L. A produção de plosivas por crianças de três anos falantes do português brasileiro. **Revista CEFAC**, v.9, n.2, p. 199-206, 2007.

BOVO, R. et al. A. The McGurk phenomenon in italian listeners. **ACTA Otorhinolaryngologica Italica**, v. 29, p.203- 208, 2009.

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.

BROWN, C. The Interrelation between speech perception and phonological acquisition from infant to adult, **Linguistics**, 2000

BROWN, C.; MATTHEWS, J. The role of feature geometry in the development of phonemic contrasts. In: HANNAHS, S. J.; YOUNG-SCHOLTEN, M. (Ed.). **Generative Studies in the Acquisition of Phonology**. Amsterdam: John Benjamins Publishing Company, p.67-112, 1997.

BURNHAM, D.; DOOD. B. Familiarity and novelty in infant cross-language studies: Factors, problems, and a possible solution. **Advances in Infancy Research**, v.12, p. 170-187, 1998.

BURNHAM, D.; ERDENER, D. The relationship between auditory- visual speech perception and language- specific speech perception at the onset of

reading instruction in English- speaking children. **Journal of Experimental Child Psychology**, v.116, p. 120- 138, 2013.

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

CAGLIARI, L.C. Elementos da fonética do português brasileiro. Tese (Livre docência), Universidade Estadual de Campinas, Campinas, 1981.

CAMPOS, P.B.B.; TENANI, L.E.; BERTI, L.C. Os registros não-convencionais da coda nasal em dados de EJA. **Estudos Linguísticos**, v.40, n.2, p. 524-538, 2011.

CHACON, L. A relação fala/ escrita em dados não- convencionais de escrita infantil. **Cadernos de Linguística**, v.2, n.1, 2021.

CHACON, L; VAZ, S. Relações entre aquisição da percepção auditiva e aprendizagem da ortografia: consoantes soantes em questão. **Linguagem em Dis(curso)**, v.13, n.3, p. 695-719, 2013.

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

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.

COLE, R.A.; SCOT, B. Toward a theory of speech perception. **Psychological Review**, v.81, p. 348-374,1974.

CRISTOFOLINI, C.; SEARA, I.C. Características acústicas de consoantes plosivas e fricativas produzidas por crianças de 6 a 12 anos: período de refinamento articulatório? **Revista Verba Volant**, v.3, n.1, p.55-71, 2012.

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

DESJARDINS, R.N.; WERKER, J.F. 4-month-old female infants are influenced by visible speech. **International Conference of Infant Studies**, Providence RI, 1997.

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

DUNN, W. The sensory profile: User's manual. San Antonio: **Psychological Corporation**, 1999.

EDWARDS, M. L. Perception and production in child phonology: The testing of four hypotheses. **Journal of Child Language**, v.1, n. 2, p. 205–219, 1974.

ELLIS, N. The dynamics of language use, language change, and first and second language acquisition. **Modern Language Journal**, v. 41, n. 3, p. 232-249, 2008

FONSECA, S.C. O afásico na clínica de linguagem. Tese (Doutorado em linguística aplicada e estudos da linguagem). Pontifícia Universidade Católica de São Paulo, São Paulo/SP, 2002.

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

FREITAS, C.; MEZZOMO, C.; VIDOR, D. Phonemic discrimination and the relationship with other linguistic levels in children with typical phonological development and phonological disorder. **CoDAS**, v. 27, n. 3, p. 236-41, 2015.

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

GIBSON, J.J. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.

GOLDSTEIN, K. *Language and language disturbances; aphasic symptom complexes and their significance for medicine and theory of language*. **Grune & Stratton**.1948

GOULARTE, L. M. et al. Transtorno do Espectro Autista (TEA) e hipersensibilidade alimentar: perfil nutricional e de sintomas gastrointestinais. **Revista Associação Brasileira de Nutrição**, v.11, n. 1. p. 48- 58, 2020.

HAYASHI, Y. e SEKIYAMA, K. Native- foreign langage effect in the McGurk effect: a test with Chinese and japonese. **Auditory-visual speech processing**, v. 4, p. 61-66, 1998.

HAZAN & BARRET, The development of phonemic categorization in children aged 6+12. **Journal of Phonetics**, v. 28, p. 377- 396, 2000.

HEARNSHAW, S. et al. The speech perception skills of children with and without speech sound disorder. **Journal of communication disorders**, v. 71, p. 61-71, 2018.

HEARNSHAW, S.; BAKER, E.; MUNRO, N. Speech perception skills of children with speech sound disorders: a systematic review and meta-analysis. **Journal of Speech, Language, and Hearing Research**, v. 62, 2019.

HOCKLEY, N.S.; POLKA, L. A developmental study of audiovisual speech perception using the McGurk paradigm. **The Journal of the Acoustical Society of America**, v. 96, p. 3309, 1994.

HUNTER, L. L et al. Extended high frequency hearing and speech perception implication in adults and children. **Hearing research**. v. 397, 2020.

IRWIN, J.R, BRANCAZIO, L.; WHALEN, D.H. Can children with autism spectrum disorders “hear” a speaking face? **Child Development**, v. 82, p.1397-1403, 2011.

JACOB, L. F. et al. Brief report: differences in multisensory integration covary with sensory responsiveness in children with and without autism spectrum disorder. **Journal of Autism and Developmental Disorders**, v. 49, p. 397-403, 2019.

JAKOBSON, R. Dois aspectos da linguagem em dois tipos de afasia. **Linguística e Comunicação**, Ed. Cultrix, São Paulo, p. 35- 62, 1975.

JESUS, L.M.T.; SHADLE, C.H. A parametric study of the spectral characteristics of European Portuguese fricative. **Journal of Phonetics**, v. 30, p. 437-464, 2002.

JOHNSON, K.; SJERPS, M. J. Speaker normalization in speech perception. **The Handbook of Speech Perception**, p. 145-176, 2021

KUHL, P.K.; MILLER, J.D. Speech Perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. **Science**, 1975.

LIBERMAN, A.M. et al. Perception of the speech code. **Psychological Review**, v. 74, p. 431-461, 1967.

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

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

MARQUES, L.M.; BOGGIO, P.S. Modulação do efeito McGurk com o uso de estimulação transcraniana por corrente contínua. **VII Jornada de Iniciação Científica**, Universidade Presbiteriana Mackenzie, 2011.

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

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

NISHIDA, G. Sobre teorias de percepção da fala. Tese- Universidade Federal do Paraná. Programa de pós- graduação em letras. Curitiba, 2012.

NULAND, L. Speech perception in children with speech output disorders. **Clinical linguistics e phonetics**. V.23, n. 3, p. 222-239, 2009

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

PATER, J.; CHRISTINE, S.; WERKER, J.F. The Perceptual Acquisition of Phonological Contrasts. **Language**, v. 80, n. 3, p. 384-402, 2004.

PEROZZO, R.V.; KUPSKE, F.F. Speech perception and production as constructs of action: implications for models of L2 development. **Journal of Speech Language and Hearing Research**, v. 16, n. 5, p. 1231- 1257, 2021.

PISONI, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. In: **Perception and Psychophysics**, 13, p.253-260, 1973.

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

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

RINALDI, L.M. Procedimentos para a análise das vogais e obstruentes na fala infantil do português brasileiro. Dissertação (Mestrado em linguística). IEL/Unicamp, Campinas/SP, 2010

Related Wallpaper- monstinhos coloridos. Seekpng. Disponível em https://www.seekpng.com/ipng/u2w7o0o0e6o0r5u2_quadrado-colorido-png-parallel. Acesso em: 11 de abril de 2018.

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

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

SALINAS- MARCHANT, C; MACLEOD, A.A.N. Audiovisual speech perception in children: a scoping review. **Speech, Language and Hearing**, v.25. p. 433-449, 2021.

SCHIER, A.C. Relações entre percepção de fala e ortografia de crianças em ensino fundamental: características fonológicas. Tese de Doutorado apresentada ao Programa de Pós-graduação em Estudos Linguístico do instituto de Biociências, Letras e Ciências Exatas UNESP- São José do Rio Preto (SP), 2017.

SCHIER, A.C., BERTI, L.C., CHACON, L. Desempenho perceptual-auditivo e ortográfico de consoantes fricativas na aquisição da escrita. **CODAS**, v. 25, n.1, p.45-51, 2013.

SEKIYAMA, K. Cultural and linguistic factors in audiovisual speech processing: the McGurk effect in Chinese subjects. **Perception & Psychophysics**. v. 59. n. 1, p. 73-80, 1997.

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

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

SEKIYAMA, K.; BURNHAM, D.; TAM, H.; ERDENER, D. Development of auditory-visual speech perception in Japanese and English speakers. **Speech Processing**, St. Jorioz, France, p. 43-46, 2003.

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

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

SELKIRK, E. O. The syllable. In: HULST, F. V; SMITH, N. **The structure of phonological representations**. Dordrecht: Foris, p. 337-379, 1982.

SILVA, L.M. Parâmetros ultrassonográficos do movimento da língua em 14 fonemas consonantais do Português Brasileiro. Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Fonoaudiologia da Faculdade de Filosofia e Ciências “Júlio de Mesquita Filho” UNESP – Marília (SP), 2017.

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

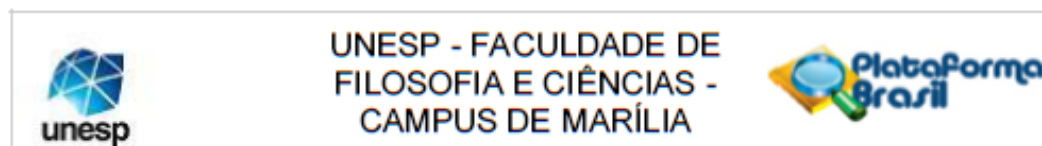
TAITELBAUM-SWEAD, R.; FOSTICK, L. Auditory and visual information in speech perception: A developmental perspective. **Clinical Linguistics & Phonetics**, v. 30, n. 7, p. 531-545, 2016.

TFOUNI, L. V. Letramento e Alfabetização. São Paulo: Cortez. 1995
TRISTÃO, R.M.; FEITOSA, M.A.G. Percepção da fala em bebês no primeiro ano de vida. **Estudos de Psicologia**, v. 8, n.3, 2003.

WRZESINSKI, A. P.; CORREA, M. J. S.; BARRETO, M. M.; FERREIRA, I. F. Software OpenShot: Roteiro para produção, gravação e edição de vídeos. Universidade Federal de Santa Maria. Santa Maria, RS, 2021.

YAVAS, M.; HERNANDORENA, C. L. M.; LAMPRECHT, R. R. Avaliação fonológica da criança: Reeducação e Terapia. Porto Alegre, RS. Editora Artmed, 2001.

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: Efeito da complexidade da língua na integração audiovisual em tarefa de percepção de fala

Pesquisador: Lídia Maurício da Silva

Área Temática:

Versão: 1

CAAE: 05262918.1.0000.5406

Instituição Proponente: Faculdade de Filosofia e Ciências/ UNESP - Campus de Marília

Patrocinador Principal: Faculdade de Filosofia e Ciências/ UNESP - Campus de Marília

DADOS DO PARECER

Número do Parecer: 3.201.991

Apresentação do Projeto:

O presente estudo propõe investigar como se dá a integração de pistas audiovisuais em crianças com alteração de fala e, ainda, verificar se a complexidade da língua interfere nessa integração. Para o desenvolvimento do trabalho, participarão três grupos de indivíduos. Serão incluídas no grupo 1: crianças entre 6 e 8 anos com desenvolvimento típico de linguagem; no grupo 2: crianças entre 6 e 8 anos que apresentam processos de substituição de líquida ou redução de encontro consonantal; e no grupo 3: adultos falantes típicos entre 20-30 anos. Os estímulos auditivos e visuais serão constituídos por sílabas simples e complexas: /pa/, /ta/, /ka/, /pra/, /tra/ e /kra/. O procedimento experimental de percepção consistirá de um teste de identificação composto por três etapas distintas: (1) apresentação de estímulos por via auditiva; (2) via visual; (3a) informações de integração audiovisual congruentes e (3b) incongruentes. Supõe-se que a complexidade da língua seria um fator que exigiria maior grau de integração de pistas para todos os grupos. Espera-se que adultos apresentem melhor desempenho de integração de vias comparado aos demais grupos. Para estímulos apresentados em condição incongruente, espera-se que adultos e crianças típicas apresentem preferência por informação visual, enquanto crianças com alteração de fala priorizem informação auditiva. Esperam-se contribuições para a discussão em torno das teorias de percepção de fala uni vs multimodais.

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: 3.201.991

Objetivo da Pesquisa:

O objetivo do presente estudo é investigar como se dá a integração de pistas audiovisuais em função da complexidade da língua.

Especificamente, procuraremos verificar o desempenho em tarefa de percepção multimodal por indivíduos adultos e crianças. Dentre o grupo infantil, procuraremos verificar se alterações nos sons da fala poderiam influenciar na percepção de estímulos apresentados por vias auditiva, visual e audiovisual.

Avaliação dos Riscos e Benefícios:

Não se aplica.

Comentários e Considerações sobre a Pesquisa:

A pesquisa encontra-se dentro dos critérios éticos.

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:

Aprovado.

Considerações Finais a critério do CEP:

O CEP da FFC da UNESP de MARÍLIA, em reunião ordinária de 27/02/2019, 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 Efeito da complexidade da língua na integração audiovisual em tarefa de percepção de fala

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1208850.pdf	21/12/2018 10:20:57		Aceito
Cronograma	PLANEJAMENTO.pdf	21/12/2018 10:19:41	Lidia Maurício da Silva	Aceito
Projeto Detalhado / Brochura Investigador	PROCESSO_PROJETO.pdf	27/09/2018 10:48:36	Lidia Maurício da Silva	Aceito

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: 3.201.991

TCLE / Termos de Assentimento / Justificativa de Ausência	PROCESSO_TERM0.pdf	27/09/2018 10:47:56	Lidia Maurício da Silva	Aceito
Declaração de Instituição e Infraestrutura	PROCESSO_AUTORIZACAO.pdf	27/09/2018 10:47:29	Lidia Maurício da Silva	Aceito
Folha de Rosto	PROCESSO_FOLHAROSTO.pdf	27/09/2018 10:44:07	Lidia Maurício da Silva	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

MARILIA, 15 de Março de 2019

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

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

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



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Estamos desenvolvendo uma pesquisa no Centro de Estudos da Educação e da Saúde – CEES – UNESP – Marília, intitulado “Efeito da complexidade da língua na integração audiovisual em tarefa de percepção de fala” e gostaríamos que participasse da mesma. O objetivo desta é investigar como se dá a integração de pistas audiovisuais em tarefa de percepção de fala em adultos e crianças com e sem alteração de fala e, verificar se a complexidade da língua interfere nessa integração. Participar desta pesquisa é uma opção e 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 que soubessem que:

- 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 _____ autorizo a participação da pesquisa intitulada “Efeito da complexidade da língua na integração audiovisual em tarefa de percepção de fala” 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
Lidia Maurício da Silva Fone: (11) 98140-4890 Endereço: Rua Lima Verde, 88. Vila dos Andrades. São Paulo/SP – CEP: 02610-060 lidia.mauricio@hotmail.com