



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
FACULDADE DE FILOSOFIA E CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE E COMUNICAÇÃO
HUMANA

LÍVIA MAYRA ROQUE DE OLIVEIRA

**PERCEPÇÃO PROSÓDICA AUDIOVISUAL DE ATOS DE FALA NA AQUISIÇÃO
DO PORTUGUÊS BRASILEIRO**

MARÍLIA

2026

UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
FACULDADE DE FILOSOFIA E CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE E COMUNICAÇÃO
HUMANA

LÍVIA MAYRA ROQUE DE OLIVEIRA

**PERCEPÇÃO PROSÓDICA AUDIOVISUAL DE ATOS DE FALA NA AQUISIÇÃO
DO PORTUGUÊS BRASILEIRO**

Tese de doutorado apresentada ao Programa de Pós-graduação em Ciências da Saúde e Comunicação Humana, como parte das exigências para a obtenção do título de doutora em fonoaudiologia pela Faculdade de Filosofia e Ciências “Júlio de Mesquita Filho” UNESP – Marília (SP)

Área de Concentração: Comunicação Humana
Orientadora: Prof^ª Dra. Geovana Carina Neris Soncin Santos

R786p

Roque de Oliveira, Livia Mayra

Percepção prosódica audiovisual de atos de fala na aquisição do português brasileiro / Livia Mayra Roque de Oliveira. -- , 2026
124 f.

Tese (doutorado) - Universidade Estadual Paulista (UNESP),
Faculdade de Filosofia e Ciências, Marília,
Orientadora: Geovana Carina Neris Soncin Santos

1. Aquisição de linguagem. 2. Análise prosódica (Linguística). 3.
Atos de fala (Linguística). 4. Percepção da fala. I. Título.

Lívia Mayra Roque de Oliveira

**PERCEPÇÃO PROSÓDICA AUDIOVISUAL DE ATOS DE FALA NA AQUISIÇÃO
DO PORTUGUÊS BRASILEIRO**

Tese apresentada ao Programa de Pós-graduação em Ciências da Saúde e Comunicação Humana da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), como requisito parcial para a obtenção do título de doutora em Ciências da Saúde e Comunicação Humana. Área de concentração: Comunicação Humana.

BANCA EXAMINADORA:

Orientador(a): _____
Dra. Geovana Carina Neris Soncin Santos
Universidade Estadual Paulista – UNESP FFC/Marília-SP

2º Titular(a): _____
Dra. Larissa Cristina Berti
Universidade Estadual Paulista – UNESP FFC/Marília-SP

3º Titular(a): _____
Dra. Manuella Carnaval
Universidade Federal do Rio de Janeiro (UFRJ)

4º Titular(a): _____
Dra. Lúcia Regiane Lopes-Damasio
Universidade Estadual Paulista – UNESP IBILCE/S. J. do Rio Preto-SP

1º Suplente: _____
Dra. Viviane Cristina de Castro Marino
Universidade Estadual Paulista – UNESP FFC/Marília-SP

2º Suplente: _____
Dra. Aveliny Mantovan Lima
Universidade de Brasília – Câmpus Universitário Ceilândia

3º Suplente: _____
Dr. Ubiratã Kickhöfel Alves
Universidade Federal do Rio grande do Sul

Dedico esse trabalho aos meus pais, Agostinho Tavares Roque e Maria Jesuína Rodrigues Roque, e ao meu esposo, Lourenço Gabriel de Oliveira.

AGRADECIMENTOS

Durante os anos de elaboração deste trabalho, vivi experiências desafiadoras, mas também momentos de grande alegria e aprendizado.

Agradeço primeiramente a Deus, por me sustentar até aqui, concedendo-me força, sabedoria e todos os recursos necessários para a concretização desta tese. Sou especialmente grata pela cura do meu irmão Mateus, sua vitória sobre o câncer que marcou profundamente este percurso.

Aos meus pais, *Agostinho Tavares Roque* e *Maria Jesuína Rodrigues Roque*, deixo meu reconhecimento pelo exemplo de dedicação, pelo carinho constante e pelo incentivo incondicional.

Ao meu esposo, *Lourenço Gabriel de Oliveira*, agradeço por todo apoio, paciência e presença amorosa nos momentos mais intensos da escrita.

À minha orientadora, *Prof.^a Dr.^a Geovana Soncin*, expresso minha admiração e profunda gratidão pela amizade, pela orientação científica e pelo cuidado humano com que sempre me conduziu, com empatia e gentileza.

Aos professores membros da banca de qualificação e de defesa, agradeço pelas valiosas contribuições e sugestões que enriqueceram este trabalho.

Aos colegas do Laboratório de Análise Articulatória e Acústica (LAAC), especialmente *Manoela das Graças Rodrigues do Nascimento* e *Felipe Ferro Erlo*, agradeço pela colaboração, pelo apoio e pelas trocas de conhecimento ao longo dessa jornada.

À minha querida mentora, *Maristela Kanashiro de Souza*, obrigada por seu amor, cuidado e por ser presença constante e inspiradora em minha vida.

Por fim, à família Lourenço, e em especial à *Helen*, minha melhor amiga, agradeço pela companhia diária, pela amizade sincera e por celebrar comigo cada conquista. Agradeço também aos seus pais, *Ivete* e *Deoclécio*, por me acolherem em sua casa com tanto carinho durante o período de aulas.

“As palavras nunca são meramente palavras: são portadoras de mundos.”

— C. S. Lewis, *Studies in Words* (1960)

RESUMO

Introdução: A prosódia constitui um recurso central para a marcação da força ilocucionária e para a expressão de atitudes e intenções comunicativas, articulando-se tanto à estrutura linguística quanto à dimensão pragmática da fala (Cole, 2015; Fónagy, 1991, 2001). A manifestação prosódica da força ilocucionária ocorre não apenas no sinal acústico, mas na integração multimodal fala-gesto (McNeill, 1985; 1992; 2005; Rusiewicz, 2011; Gilli-Fivela, 2018) a partir da qual pistas auditivas e visuais funcionam como marcadores prosódicos que, na percepção de fala, contribuem para a reconstrução dos sentidos mobilizados pelo falante (Schwartz et al., 2004; Swerts; Krahmer, 2005, 2008, 2010). Embora a percepção da multimodalidade seja considerada consolidada em adultos (Swerts; Krahmer, 2020), sua aquisição pressupõe um processo que ocorre de forma gradual na infância em função do amadurecimento linguístico, cognitivo e sociopragmático dos sujeitos (Esteve-Gibert; Prieto, 2013; Hübscher et al., 2018; Rusiewicz e Esteve-Gibert, 2018). Com base nessas premissas, este trabalho investiga a percepção de atos de fala diferenciados prosodicamente no Português Brasileiro por pistas auditivas e visuais ao longo do processo de aquisição prosódica e busca cotejar a relevância das pistas de natureza visual para a percepção dos atos de fala no processo de aquisição dessa língua, haja vista que tradicionalmente as pistas de natureza auditiva são privilegiadas nos estudos de percepção de fala e, com efeito, às pistas visuais tende-se a atribuir um estatuto acessório ou compensatório em relação às pistas auditivas quando as condições de apresentação dessas últimas é limitada ou problemática (Munhall et al., 1996; Barkhuysen; Krahmer; Swerts, 2010; Pilling; Thomas, 2011). **Objetivos:** 1) comparar a acurácia perceptual de diferentes grupos etários na identificação de atos de fala diferenciados prosodicamente por pistas auditivas e visuais; 2) avaliar o efeito de pistas prosódicas auditivas e visuais na percepção dos diferentes atos de fala pelos grupos etários; e 3) verificar se, em condição de estímulos com ruído, as pistas visuais exerceriam um efeito compensatório na identificação dos diferentes atos de fala pelos grupos etários. O trabalho se justifica pela necessidade de se levantar dados de referência sobre o desempenho perceptual prosódico de crianças brasileiras e, ainda, pela relevância de se mapear mudanças que ocorrem ao longo da aquisição no que diz respeito ao modo como as crianças percebem auditiva e visualmente a força ilocucionária marcada prosodicamente nos atos de fala. **Método:** Participaram da pesquisa 90 crianças, de 4 a 12 anos, organizadas em 4 grupos etários, e um grupo de 20 adultos, totalizando 110 participantes. O grupo de adultos foi utilizado como parâmetro de comparação por representar o modelo linguístico alvo no processo de aquisição da linguagem (Clark, 2009), enquanto a organização das crianças por faixas etárias baseia-se na concepção de que a aquisição ocorre de forma gradual, não linear e heterogênea, especialmente nas competências pragmáticas e discursivas (Tomasello, 2003; Snow, 1995). Por meio do software Open Sesame, aplicou-se experimento de percepção audiovisual de diferentes atos de fala. No experimento, os participantes tinham como tarefa identificar os estímulos como declaração, pergunta, ordem ou pedido. Os estímulos foram apresentados em quatro modalidades: apenas auditiva (AO); apenas visual (VO); audiovisual congruente (AV+) e audiovisual incongruente (AV-). Em termos linguísticos, os estímulos foram compostos pela sentença “Vamos”, que foi enunciada por falante adulta do português brasileiro como declaração, pergunta, ordem e pedido. Na produção da falante, a diferenciação desses atos de fala contou com marcadores prosódicos acústicos e gestuais, os quais foram devidamente descritos para a caracterização dos estímulos. A aplicação do experimento foi dividida em duas etapas. Na primeira etapa, os estímulos foram apresentados sem ruído. Na segunda etapa, os estímulos envolvidos foram os mesmos, porém contavam com ruído sobreposto. Para a análise dos resultados, mensurou-se o percentual de identificação de cada ato de fala nas modalidades AO, VO e AV+; a preferência perceptual (auditiva ou visual) na modalidade AV-; e o índice de efeito visual na identificação de cada ato de fala nas condições com e sem ruído. Os dados obtidos foram analisados por estatística descritiva e inferencial por

meio do teste estatístico ANOVA de medidas repetidas. **Resultados:** Com relação ao objetivo I, comparar a acurácia perceptual de diferentes grupos etários na identificação dos atos de fala, os resultados mostraram que tanto o desempenho perceptual dos grupos quanto o efeito das pistas auditivas e visuais são dependentes do tipo de ato de fala. A ordem foi o ato de fala para o qual todos os grupos apresentaram maior acurácia perceptual, indicando maior saliência perceptual em termos de marcadores auditivos e visuais, enquanto o ato de fala pedido apresentou os menores índices de acurácia, sugerindo maior complexidade prosódica para a percepção. Na percepção de perguntas, maior acurácia foi identificada a partir dos 8 anos, mostrando que, a partir dessa idade, ocorre o refinamento na percepção de contornos entoacionais ascendentes, pistas essas de natureza auditiva identificadas como mais robustas para a identificação desse ato de fala. Já a acurácia perceptual na identificação da declaração variou entre os grupos de crianças e foi identificada como significativamente maior apenas no grupo dos adultos, reforçando que o domínio perceptual dos padrões entoacionais se consolida ao longo da aquisição. Quanto ao objetivo II, sobre o efeito de pistas auditivas e visuais, verificou-se mudança ao longo do processo de aquisição na percepção de declarações e perguntas, pois, para esses atos de fala, as pistas auditivas se mostraram mais predominantes para a percepção dos adultos e de crianças mais velhas em relação a crianças mais novas, porém, na percepção de atos de fala de ordem e pedido, o efeito das pistas visuais se mostrou significativo desde a infância, mantendo-se robusto até a idade adulta. Em relação ao Objetivo III, verificou-se que a presença de ruído não conferiu às pistas visuais um efeito compensatório. Esse resultado indica que a informação visual constitui um componente intrínseco ao próprio ato de fala, participando do processamento perceptivo de maneira integrada, independentemente das condições de degradação acústica. **Conclusão:** No processo de aquisição da prosódia do Português Brasileiro, a percepção da força ilocucionária marcada prosodicamente é adquirida de forma gradual e heterogênea. Observa-se um aumento da acurácia perceptual à medida que, no processo de aquisição, o ouvinte passa a reconhecer a atuação integrada das pistas auditivas e visuais na sinalização dos atos de fala, considerando que a configuração desse conjunto saliente e indissociável se modifica de acordo com as características semântico-pragmáticas de cada ato de fala. Esse reconhecimento acontece em decorrência do amadurecimento linguístico e sociopragmático dos sujeitos, que permite relacionar determinados marcadores prosódicos auditivos e/ou visuais a contrastes linguísticos e a nuances semântico-pragmáticas dos atos de fala.

Palavras-chave: Percepção de Fala; Multimodalidade; Aquisição da Prosódia; Força ilocucionária.

ABSTRACT

Introduction: Prosody constitutes a central resource for marking illocutionary force and expressing communicative attitudes and intentions, articulating with both the linguistic structure and the pragmatic dimension of speech (Cole, 2015; Fónagy, 1991, 2001). The prosodic manifestation of illocutionary force occurs not only in the acoustic signal but also within the multimodal speech-gesture integration (McNeill, 1985; 1992; 2005; Rusiewicz, 2011; Gilli-Fivela, 2018), through which auditory and visual cues function as prosodic markers that, in speech perception, contribute to the reconstruction of the meanings mobilized by the speaker (Schwartz et al., 2004; Swerts; Krahmer, 2005, 2008, 2010). Although the perception of multimodality is considered well-established in adults (Swerts; Krahmer, 2020), its acquisition presupposes a process that occurs gradually during childhood as a function of the subjects' linguistic, cognitive, and socio-pragmatic maturation (Esteve-Gibert; Prieto, 2013; Hübscher et al., 2018; Rusiewicz & Esteve-Gibert, 2018). Based on these premises, this study investigates the perception of prosodically differentiated speech acts in Brazilian Portuguese via auditory and visual cues throughout the prosodic acquisition process. It also seeks to contrast the relevance of visual cues for speech act perception during the acquisition of this language, given that auditory cues are traditionally favored in speech perception studies and, consequently, visual cues tend to be assigned an accessory or compensatory status relative to auditory cues when the presentation conditions of the latter are limited or problematic (Munhall et al., 1996; Barkhuysen; Krahmer; Swerts, 2010; Pilling; Thomas, 2011). **Objectives:** 1) to compare the perceptual accuracy of different age groups in identifying speech acts prosodically differentiated by auditory and visual cues; 2) to evaluate the effect of auditory and visual prosodic cues on the perception of different speech acts across age groups; and 3) to verify whether, under noisy stimulus conditions, visual cues exert a compensatory effect on speech act identification across age groups. This work is justified by the need to gather baseline data on the prosodic perceptual performance of Brazilian children, as well as the relevance of mapping changes that occur throughout acquisition regarding how children auditorily and visually perceive the prosodically marked illocutionary force in speech acts. **Method:** Participants included 90 children, aged 4 to 12 years, organized into 4 age groups, and a group of 20 adults, totaling 110 participants. The adult group served as a benchmark for comparison because it represents the target linguistic model in the language acquisition process (Clark, 2009), while the grouping of children by age is based on the conception that acquisition occurs gradually, non-linearly, and heterogeneously, especially regarding pragmatic and discourse skills (Tomasello, 2003; Snow, 1995). Using Open Sesame software, an audiovisual perception experiment of different speech acts was conducted. In the experiment, the participants' task was to identify the stimuli as a statement, question, command, or request. The stimuli were presented in four modalities: auditory-only (AO); visual-only (VO); congruent audiovisual (AV+); and incongruent audiovisual (AV-). In linguistic terms, the stimuli consisted of the sentence "Vamos" ("Let's go"), which was uttered by a female adult speaker of Brazilian Portuguese as a statement, question, command, and request. In the speaker's production, the differentiation of these speech acts relied on acoustic and gestural prosodic markers, which were duly described for stimulus characterization. The experiment was administered in two stages. In the first stage, stimuli were presented without noise. In the second stage, the same stimuli were used, but with superimposed noise. For data analysis, the study measured the identification percentage of each speech act in the AO, VO, and AV+ modalities; the perceptual preference (auditory or visual) in the AV- modality; and the visual effect index on speech act identification under both noisy and quiet conditions. The data obtained were analyzed using descriptive and inferential statistics via repeated measures ANOVA. **Results:** Regarding Objective I—comparing the perceptual accuracy of different age groups in identifying speech acts—the results showed that both the perceptual performance of the groups and the effect of auditory and visual cues depend on the

type of speech act. Commands were the speech act for which all groups showed the highest perceptual accuracy, indicating greater perceptual salience in terms of auditory and visual markers, whereas requests showed the lowest accuracy rates, suggesting greater prosodic complexity for perception. In the perception of questions, higher accuracy was identified starting at age 8, showing that a refinement in the perception of rising intonation contours occurs from this age onward, with these auditory cues being identified as more robust for the identification of this speech act. Conversely, perceptual accuracy in identifying statements varied among the children's groups and was found to be significantly higher only in the adult group, reinforcing that the perceptual mastery of intonational patterns consolidates throughout acquisition. Regarding Objective II—on the effect of auditory and visual cues—a change was observed throughout the acquisition process in the perception of statements and questions; for these speech acts, auditory cues proved more predominant for adults and older children compared to younger children. However, in the perception of commands and requests, the effect of visual cues was significant from childhood and remained robust into adulthood. Regarding Objective III, it was found that the presence of noise did not grant a compensatory effect to visual cues. This result indicates that visual information constitutes an intrinsic component of the speech act itself, participating in perceptual processing in an integrated manner, regardless of acoustic degradation conditions. **Conclusion:** In the prosodic acquisition process of Brazilian Portuguese, the perception of prosodically marked illocutionary force is acquired gradually and heterogeneously. An increase in perceptual accuracy is observed as the listener, throughout the acquisition process, begins to recognize the integrated role of auditory and visual cues in signaling speech acts, considering that the configuration of this salient and inseparable set changes according to the semantic-pragmatic characteristics of each speech act. This recognition occurs as a result of the subjects' linguistic and socio-pragmatic maturation, which allows them to map specific auditory and/or visual prosodic markers to linguistic contrasts and semantic-pragmatic nuances of speech acts.

Keywords: Speech Perception; Multimodality; Prosody Acquisition; Illocutionary Force.

LISTA DE FIGURAS

Figura 1 – Hierarquia de constituintes prosódicos Nespor e Vogel (1986)

Figura 2 - Modalidades de apresentação

Figura 3 –Alternativas de respostas do experimento de percepção dos atos de fala

Figura 4 –Análise entoacional ato de fala declarativo

Figura 5 –Análise entoacional ato de fala pergunta com surpresa

Figura 6 –Análise entoacional ato de fala ordem

Figura 7 – Análise entoacional ato de fala pedido

Figura 8 – Análise gestual dos tipos de atos de fala

LISTA DE GRÁFICOS

Gráfico 1– Desempenho perceptual na identificação dos tipos de atos de fala por grupo etário

Gráfico 2– Acurácia perceptual dos grupos etários por modalidade na identificação do ato de fala declarativo

Gráfico 3– Acurácia perceptual dos grupos etários por modalidade na identificação do ato de fala pergunta

Gráfico 4– Acurácia perceptual dos grupos etários por modalidade na identificação de ordem

Gráfico 5– Acurácia perceptual dos grupos etários por modalidade na identificação do pedido

Gráfico 6– Preferência perceptual na identificação do ato de fala declarativo por parte dos grupos

Gráfico 7– Preferência perceptual na identificação do ato de fala pergunta por parte dos grupos

Gráfico 8– Preferência perceptual na identificação do ato de fala ordem por parte dos grupos

Gráfico 9– Preferência perceptual na identificação do ato de fala pedido por parte dos grupos

LISTA DE QUADROS

Quadro 1 – Amostra de sujeitos participantes da pesquisa

Quadro 2 – Descrição dos contextos usados na gravação dos estímulos para eliciar os atos de fala

Quadro 3 – Detalhamento dos estímulos

Quadro 4 - Análise sêmica dos atos de fala utilizados como estímulo

Quadro 5 - Resumo dos resultados

LISTA DE TABELAS

Tabela 1 – Análise entoacional dos atos de fala do português brasileiro

Tabela 2 – Média e desvio padrão da acurácia perceptual dos grupos por ato de fala

Tabela 3 – Média e desvio padrão da acurácia perceptual dos grupos por modalidade na identificação do ato de fala declarativo

Tabela 4 – Média e desvio padrão da acurácia perceptual dos grupos por modalidade na identificação do ato de fala pergunta

Tabela 5 – Média e desvio padrão da acurácia perceptual dos grupos por modalidade na identificação do ato de fala ordem

Tabela 6 – Média e desvio padrão da acurácia perceptual dos grupos por modalidade na identificação do ato de fala pedido

Tabela 7 – Efeito compensatório da pista visual por condição de apresentação dos estímulos

SUMÁRIO

APRESENTAÇÃO	16
1 INTRODUÇÃO	19
1.1 AQUISIÇÃO PROSÓDICA: A PERCEPÇÃO EM FOCO	20
1.2 AQUISIÇÃO PROSÓDICA: A PERCEPÇÃO SEGUNDO A PERSPECTIVA MULTIMODAL.....	27
1.3 REFERENCIAIS TEÓRICOS	38
1.3.1 Uma breve definição de pragmática.....	38
1.3.2 Teoria dos atos de fala.....	38
1.3.3 As funções prosódicas.....	45
1.3.3.1 A perspectiva de Fónagy (1991, 2001, 2003)	45
1.3.3.2 A perspectiva de Cole (2015)	46
1.3.4 Fonologia Prosódica	49
1.3.5 Fonologia entoacional.....	52
2 OBJETIVOS E HIPÓTESES	58
2.1 OBJETIVOS	58
2.2 HIPÓTESES.....	58
3 MATERIAL E MÉTODO.....	61
3.1 DESCRIÇÃO METODOLÓGICA.....	61
3.2 ASPECTOS ÉTICOS.....	61
3.3 DESIGN EXPERIMENTAL	61
3.4 AMOSTRA	62
3.5 PROCEDIMENTO EXPERIMENTAL.....	64
3.6 ESTÍMULOS	66
3.7 DESCRIÇÃO DOS ESTÍMULOS.....	69
3.7.1 Descrição semântico-pragmática	69
3.7.2 Descrição acústica.....	73
3.7.3 Descrição gestual	77
3.8 VALIDAÇÃO DOS ESTÍMULOS.....	79
3.9 FORMA DE ANÁLISE DOS RESULTADOS	80
3.10 ANÁLISE ESTATÍSTICA	80
4 RESULTADOS.....	82
4.1 ACURÁCIA PERCEPTUAL GERAL DOS GRUPOS NA IDENTIFICAÇÃO DOS ATOS DE FALA DO PB	82

4.2 ACURÁCIA PERCEPTUAL AUDITIVA E VISUAL DOS GRUPOS POR ATO DE FALA	84
4.2.1 Declaração.....	84
4.2.2 Pergunta	86
4.2.3 Ordem.....	87
4.2.4 Pedido	89
4.3 PREFERÊNCIA PERCEPTUAL NA IDENTIFICAÇÃO DOS ATOS DE FALA	90
4.3.1 Declaração.....	91
4.3.2 Pergunta	92
4.3.3 Ordem.....	93
4.3.4 Pedido	94
4.4 ÍNDICE DE EFEITO VISUAL NA IDENTIFICAÇÃO DOS ATOS DE FALA.....	95
5 DISCUSSÃO.....	96
5.1 COMPARAÇÃO DA ACURÁCIA PERCEPTUAL DOS GRUPOS ETÁRIOS NA IDENTIFICAÇÃO DOS DIFERENTES ATOS DE FALA	96
5.3 EFEITO DAS PISTAS PROSÓDICAS AUDITIVAS E VISUAIS NA PERCEPÇÃO DOS ATOS DE FALA	98
5.5 EFEITO COMPENSATÓRIO DA PISTA VISUAL NA IDENTIFICAÇÃO DOS ATOS DE FALA EM CONDIÇÃO DE RUÍDO	100
6 CONSIDERAÇÕES FINAIS	103
REFERÊNCIAS BIBLIOGRÁFICAS	105
ANEXOS.....	117

APRESENTAÇÃO

Este trabalho apresenta o produto de uma investigação sobre a percepção prosódica audiovisual de atos de fala na aquisição do Português Brasileiro e integra as ações de pesquisa do projeto “Produção e percepção da prosódia em crianças com desenvolvimento fonológico típico e atípico” (Proc. FAPESP 20/10144-3), desenvolvido pelo Núcleo de Estudos em Aquisição da Prosódia do Laboratório de Análise Articulatória e Acústica (LAAc), da UNESP, câmpus de Marília.

No Português Brasileiro, a prosódia é fundamental para marcar a força ilocucionária e expressar atitudes e intenções comunicativas, articulando dimensões linguísticas e pragmáticas (Cole, 2015; Fónagy, 1991, 2001). Sua interpretação envolve a multimodalidade da fala, uma vez que diferentes funções linguísticas, contrastes e nuances de sentido são marcados tanto por características acústicas quanto por gestos (McNeill, 1985; 1992; 2005; Rusiewicz, 2011; Gilli-Fivela, 2018), que, respectivamente, funcionam como pistas auditivas e visuais para a percepção multimodal da fala (Schwartz et al., 2004; Swerts; Krahmer, 2005, 2008, 2010). A percepção da multimodalidade na comunicação é atestada por diferentes trabalhos e considerada estabilizada em adultos (Swerts; Krahmer, 2020), mas pressupõe um processo de aquisição que se caracteriza como gradual na infância em razão do amadurecimento linguístico, cognitivo e sociopragmático dos sujeitos (Esteve-Gibert; Prieto, 2013; Hübscher et al., 2018; Rusiewicz e Esteve-Gibert, 2018). Com base nessas premissas, este estudo investiga como pistas auditivas e visuais contribuem para a identificação de atos de fala do português brasileiro em diferentes grupos etários, comparando adultos e crianças, e insere-se, assim, no campo dos estudos de aquisição de linguagem, mais precisamente com ênfase na aquisição da prosódia.

Para tanto, a investigação foi norteada pelos seguintes objetivos específicos: 1) comparar a acurácia perceptual de diferentes grupos etários na identificação de atos de fala diferenciados prosodicamente por pistas auditivas e visuais no Português Brasileiro; 2) avaliar o efeito de pistas prosódicas auditivas e visuais na percepção dos atos de fala pelos grupos etários; e 3) mensurar a influência visual na identificação dos diferentes atos de fala pelos grupos etários, bem como verificar se essa influência seria maior em condições de estímulos com presença de ruído. As justificativas para o delineamento desses objetivos específicos são apresentadas na Introdução desta tese de forma cientificamente fundamentada.

Embora em âmbito internacional vários estudos investiguem a aquisição prosódica, poucos visam compreender como essa aquisição se dá do ponto de vista da percepção

multimodal, especificamente no português brasileiro, tal como se propõe nesta pesquisa. Destaca-se a importância da investigação, tendo em vista que, como propósito maior, o presente trabalho visa contribuir com a ampliação dos estudos sobre aquisição de linguagem, centrando-se no plano da aquisição da prosódia e, para tanto, busca oferecer contribuições no âmbito científico e no âmbito aplicado. No âmbito científico, a pesquisa busca contribuir para a compreensão da aquisição perceptual de atos de fala que se diferenciam prosodicamente, oferecendo uma descrição linguística sob o olhar da multimodalidade da fala. No âmbito aplicado, a pesquisa busca fornecer subsídios relevantes para a prática fonoaudiológica, seja no âmbito da avaliação, ao disponibilizar dados de referência sobre a identificação perceptual auditiva e visual dos atos de fala por crianças em aquisição típica de linguagem, seja no campo da intervenção, considerando que crianças com processos desviantes de aquisição de linguagem podem apresentar comprometimentos no plano prosódico e/ou na esfera da percepção de fala.

A fim de descrever a construção científica, o conteúdo desta tese está organizado da seguinte maneira: inicialmente, apresenta-se uma introdução aos estudos que se dedicaram à investigação da aquisição da prosódia, centrando-se particularmente naqueles que compreenderam a prosódia como multimodal; em seguida, apresenta-se a fundamentação teórica que norteou o desenvolvimento da pesquisa.

Posteriormente, apresenta-se a proposta deste estudo com os seus respectivos objetivos, bem como são apresentados os aspectos metodológicos, que caracterizam o delineamento experimental do estudo, os critérios de análise dos dados e a forma de análise dos resultados.

Após a seção metodológica, apresentam-se os resultados, organizados conforme os objetivos específicos deste estudo. Por fim, a seção de discussão dedica-se a interpretar os achados frente ao referencial teórico e às hipóteses previamente estabelecidas, encerrando-se com as considerações finais desta pesquisa.

1 INTRODUÇÃO

Na literatura, a aquisição prosódica, compreendida como o processo pelo qual a criança adquire os padrões rítmicos, entoacionais e acentuais característicos da língua materna (Juszyk, 1997; Christophe et al., 2008; Prieto; Esteve-Gibert, 2018), tem sido amplamente abordada sob duas dimensões complementares: a da produção e a da percepção (Frota; Butler, 2018). Enquanto o ponto de vista da produção se concentra em descrever como os padrões prosódicos emergem e se estabilizam na fala infantil, a abordagem perceptiva busca compreender como o ouvinte, no processo de aquisição, é capaz de detectar, discriminar e interpretar as variações melódicas e rítmicas que estruturam os diferentes tipos de enunciado (Filipe et al., 2017; Wells et al., 2004). A literatura tem mostrado que esse processo de aquisição prosódica é multifacetado, pois se, por um lado, se inicia antes mesmo do nascimento por meio da sensibilidade dos bebês de detectar e discriminar padrões prosódicos da fala da mãe e da futura língua materna, por outro, se complexifica e ocorre de forma gradual, tornando-se dependente de outros domínios linguísticos e cognitivos, de modo que o uso do sistema prosódico no que diz respeito à relação entre forma e função conforme o padrão do adulto só é alcançado na adolescência (Chen et al., 2020).

Sob uma perspectiva fonológica, a prosódia é concebida como um componente do sistema fonológico que abrange uma série de fenômenos interligados e interativos; são eles: acento, ritmo, fraseamento e entoação (Arvaniti, 2020). Dentre eles, a entoação – fenômeno prosódico envolvido nos testes de percepção do presente trabalho – é compreendida, sob o olhar da Fonologia Entoacional (Ladd, 1996; 2014), como um fenômeno linguístico estruturado que associa eventos tonais a posições proeminentes e a posições relacionadas a fronteiras prosódicas, possibilitando a codificação de contrastes entoacionais que veiculam funções linguísticas, pragmáticas e discursivas e, dessa forma, caracterizam melodicamente as línguas do mundo. Nos últimos anos, porém, têm ganhado força estudos perceptivos que abordam a entoação sob perspectiva multimodal, segundo os quais a percepção prosódica resulta da integração de pistas auditivas e visuais no processamento da fala (Esteve-Gibert; Prieto, 2018; Swerts; Krahmer, 2005). Desse modo, a aquisição prosódica é compreendida como um processo perceptivo dinâmico e integrado, no qual o desenvolvimento da sensibilidade aos componentes prosódicos, ou seja, relativos aos padrões rítmicos e melódicos da língua, interage com a capacidade de interpretar expressões faciais, movimentos labiais e gestos coocorrentes (McNeill, 2005; 1992; 2005; Rusiewicz, 2011). Essa integração entre modalidades perceptivas não apenas potencializa a compreensão da entoação dos enunciados, mas também explicita a

natureza interacional e multimodal do desenvolvimento prosódico na linguagem infantil. Especificamente, neste trabalho, investiga-se o modo como crianças entre 4 e 12 anos de idade adquirem, do ponto de vista perceptivo audiovisual, os atos de fala do português brasileiro. De modo sucinto, esses atos são compreendidos, à luz de Austin (1962), como ações realizadas por meio da linguagem, nas quais o enunciado não apenas descreve um estado de coisas, mas desempenha uma função comunicativa específica em um determinado contexto interacional.

A partir dessas perspectivas teóricas, a presente seção introdutória está organizada em três partes complementares. A primeira parte dedica-se a discutir a aquisição prosódica com foco na percepção, abordando o modo como a criança passa a reconhecer contrastes entoacionais e a atribuir-lhes valor linguístico e pragmático ao longo do processo de aquisição. A segunda parte volta-se à aquisição prosódica sob a perspectiva da percepção multimodal, apresentando estudos que investigam a integração entre pistas auditivas e visuais na compreensão de enunciados e na identificação de atos de fala. Por último, apresenta-se o arcabouço teórico que fundamenta este trabalho.

1.1 AQUISIÇÃO PROSÓDICA: A PERCEPÇÃO EM FOCO

Nas línguas, a prosódia é caracterizada formalmente por padrões de ritmo e melodia (Ladd, 2014) e desempenha um papel central na organização da estrutura informacional dos enunciados, entendida como a forma pela qual a informação é distribuída e hierarquizada no discurso, envolvendo distinções como figura e fundo, tópico e comentário, bem como informação nova e dada (Halliday, 1967; Prince, 1981; Chafe, 1997). Além disso, a prosódia manifesta significados semântico-pragmáticos fundamentais na interação verbal, ou seja, significados que se definem na relação entre o enunciado dito e a situação de produção do dizer, uma vez que contribui para a marcação de tipos de atos de fala, foco, estados epistêmicos, atitudes e estados afetivos do falante (Cole, 2015; Esteve-Gibert; Prieto, 2018; Frota et al., 2022), que podem ser compreendidos como valores de modalização, no sentido da articulação entre o *dictum* e o *modus*¹ (Castilho, 2010). Ainda e não menos importante, em diversas línguas, a prosódia também codifica contrastes fonológicos no nível lexical por meio do acento de

¹ De acordo com Castilho (2010), o *dictum* diz respeito à informação veiculada por um enunciado linguístico, geralmente manifestado na forma de uma sentença, enquanto o *modus* diz respeito à avaliação que o falante faz do *dictum*, considerando-o real, irreal, possível ou necessário. Ademais, o autor esclarece que cada *dictum* vem associado a um ato de fala e é o *modus* o responsável por explicitar de que ato de fala se trata (p. 437-438). Trataremos disso mais adiante com o devido detalhamento.

palavra, como ocorre no português, ou por meio de marcação tonal, conforme ocorre em línguas tonais como o mandarim (Ladd, 2014; Gussenhoven; Chen, 2020). Essas várias funções se manifestam por meio de um conjunto restrito de fenômenos rítmicos e melódicos, quais sejam: agrupamentos frasais prosódicos (comumente chamados na literatura de fraseamento prosódico), proeminência acentual e modulações entoacionais (Esteve-Gibert; Prieto, 2018; Arvaniti, 2020), que se comportam de modo variável e distinto a depender da função que manifestam e a depender da língua.

É por essa razão que, no processo de aquisição de linguagem, adquirir um sistema prosódico implica a aquisição de diferentes habilidades interconectadas, a saber: (i) manifestar, na produção de fala, propriedades formais desse sistema; tais como padrões entoacionais (variações de frequência fundamental), organização rítmica e acentual, marcação de proeminências, delimitação de fronteiras prosódicas e modulação de parâmetros acústicos como duração e intensidade; (ii) variar essas propriedades formais apropriadamente em conformidade com o contexto comunicativo; (iii) perceber os sentidos marcados na fala por esse sistema; (iv) processar informação prosódica, transformando-a em compreensão linguística (cf. Chen et al., 2020; Esteve-Gibert; Prieto, 2018).

Apesar da complexidade envolvida na aquisição prosódica, a qual demanda a realização de estudos tanto no âmbito da produção quanto no âmbito da percepção de fala e, sempre que possível, considerando aspectos relativos às funções linguísticas que desempenham nas situações de comunicação, os estudos sobre a aquisição da prosódia se voltam, principalmente, para o ponto de vista da produção de fala, isto é, para a investigação no que diz respeito ao domínio acústico e, particularmente, para análises de padrões entoacionais (Swerts; Krahmer, 2005).

No tocante à percepção, os estudos existentes concentram-se sobretudo na aquisição da entoação em diferentes línguas, assim como na investigação da identificação de fronteiras prosódicas (Vigário; Frota, 1992; Esteve-Gibert; Pietro, 2018) e de padrões acentuais e de focalização (Chen et al., 2020; Esteve-Gibert; Guellaï, 2018). Especificamente, com relação aos estudos nacionais, se voltam, de modo geral, para a investigação da percepção de fronteiras prosódicas em adultos (Soncin; Tenani; Berti, 2017; Soncin, 2017; Soncin; Tenani; Berti, 2019; Fontes, 2020; Caldas, 2024) e em bebês de 13 meses (Silva; Name, 2014), para o reconhecimento de contornos entoacionais, padrões de acento e foco em adultos (Silva et al., 2017; Silva; Moraes; Rilliard, 2020; Caenaval, 2014; Carnaval; Moraes; Rilliard, 2022; Lima; Gonçalves; Soncin, 2025) e em crianças de 6 a 12 anos (Borges, 2025) e para reconhecimento de emoções e atitudes veiculadas pela prosódia em crianças de 6 a 8 anos (Nascimento et al.,

2024)². No que tange à aquisição inicial, descrita por alguns autores como o processo pelo qual a criança internaliza os padrões rítmicos, entoacionais e acentuais característicos de sua língua materna (Jusczyk, 1997; Christophe et al., 2008; Prieto; Esteve-Gibert, 2018), a prosódia é reconhecida por diferentes trabalhos da área de aquisição de linguagem como essencial na aquisição lexical e sintática, pois favorece, pela via perceptual, tanto o reconhecimento da segmentação de palavras no sinal de fala quanto o mapeamento semântico das palavras, assim como atua como facilitadora no reconhecimento de informações sintáticas (Splendore et al., 2019; Frota; Name, 2017). Em termos teóricos, o papel central da prosódia na aquisição inicial foi conceitualizado pela proposta do chamado "bootstrapping prosódico"³ (*prosodic bootstrapping*) (cf. Höhle, 2009).

A teoria do *bootstrapping* prosódico, proposta inicialmente por Gleitman e Wanner (1982) e atualizada por Morgan e Demuth (1996), defende que crianças pequenas exploram pistas prosódicas disponíveis no *input* para aprender sobre propriedades lexicais e gramaticais abstratas não disponíveis de forma explícita no mesmo *input*, mas com as quais as pistas prosódicas estão relacionadas. Em outras palavras, a proposta do *bootstrapping* prosódico assume que pistas prosódicas manifestadas na marcação acentual, rítmica e entoacional auxiliam a criança a segmentar o sinal de fala em unidades linguísticas relevantes e a categorizar essas unidades como itens lexicais e sintáticos (Höhle, 2009).

Como exemplo, Cutler e Carter (1987) mostram que, no inglês, nomes e verbos bissilábicos com a mesma cadeia segmental são discriminados pelo acento lexical, uma vez que as crianças reconhecem a seguinte regularidade: nomes tendem a portar acento na sílaba inicial (*REcord*) enquanto verbos tendem a portar acento na sílaba final (*reCORD*). Por sua vez, Snedeker e Yuan (2008) mostram, no que diz respeito à análise sintática, que pistas que demarcam fronteiras prosódicas, como alongamento duracional, pausa e variação entoacional, são recursos que orientam as crianças na interpretação de sentenças cuja estrutura sintática é ambígua, como em “Você consegue tocar o sapo com a pena?” (tradução do inglês: *Can you touch the frog with the feather?*), sentença na qual “com a pena” pode modificar tanto o nome *sapo* quanto o verbo *tocar*. Já os estudos sobre o Português Brasileiro revelam que a prosódia

² Deve-se destacar que os dois últimos trabalhos reportados (Borges, 2025; Nascimento et al., 2024) e que investigaram a percepção de aspectos prosódicos em crianças foram produzidos por pesquisadores que são membros do mesmo grupo de pesquisa e do mesmo laboratório no qual essa tese foi desenvolvida.

³ O conceito de *bootstrapping* foi introduzido por Pinker (1984) no campo da aquisição de linguagem à luz de uma abordagem gerativa como uma metáfora para a hipótese de que a criança interage com o *input* que recebe em seu meio e, a partir dele, por meio de mecanismos de aprendizagem implícitos, extrai informações que a auxiliam a adquirir propriedades estruturais da língua nativa. A partir desse conceito introduzido por Pinker (1984), foram desenvolvidas propostas de diferentes tipos de *bootstrapping*, dentre os quais o *bootstrapping* prosódico é um dos mais reconhecidos e disseminados no campo da aquisição de linguagem (cf. Höhle, 2009).

desempenha um papel computacional ativo na linguagem, atuando como um mecanismo que orienta o *parser* (analisador sintático) e otimiza o processamento ao evitar o gasto de tempo com interpretações incorretas em estruturas ambíguas (Almeida; Santos, 2018). Essa dinâmica mostra que o tamanho dos constituintes é um fator determinante, visto que frases mais longas tendem a gerar um maior número de fronteiras prosódicas, o que altera diretamente a percepção da hierarquia sintática (Serra, 2016). De acordo com os pressupostos da Hipótese da Prosódia Implícita investigados por Almeida (2017) e Silva (2015), essa interface é fundamentalmente bidirecional: enquanto a sintaxe fornece o mapeamento estrutural que sugere os pontos potenciais de pausa, é a configuração prosódica final, seja ela produzida na fala ou projetada mentalmente na leitura, que decide qual interpretação sintática o falante deve assumir, podendo inclusive reverter preferências estruturais de anexação (Maia et al., 2004; Leite, 2011).

Conforme apontado por diferentes estudos, o fato importante e necessário para que a prosódia atue como desencadeadora da aquisição de padrões de língua que se manifestam em outras dimensões, como acontece com a dimensão lexical e com a dimensão sintática, de acordo com a proposta do *bootstrapping* prosódico, é a necessária existência da fala dirigida à criança: o famoso “manhês”. Por sua característica de hiperarticulação, os padrões prosódicos da fala dirigida à criança (ou ao bebê) ganham maior saliência e, assim, facilitam a aquisição da linguagem (Khul et al, 1997; Esteve-Gibert; Prieto, 2018; Vihman, 2019).

Contudo, a aquisição prosódica articula-se a outras dimensões linguísticas e ao desenvolvimento de habilidades sociais e pragmáticas, não se restringindo ao contexto da fala dirigida à criança (“manhês”). Estudos demonstram que a sensibilidade a padrões prosódicos emerge precocemente e contribui para processos como a segmentação da fala, o reconhecimento de intenções comunicativas e a organização da interação social, tanto em contextos de fala dirigida à criança quanto na fala entre adultos (Fernald, 1989; Jusczyk, 1997; Kemler Nelson et al., 1989).

Segundo Esteve-Gibert e Prieto (2018), o desenvolvimento prosódico tanto no plano perceptual quanto no plano expressivo da produção de fala é, portanto, de vital importância para a maturação de habilidades linguísticas mais complexas e de habilidades gerais de comunicação de uma criança que acontece de forma gradual ao longo do processo de aquisição. Como exemplo, em termos de trajetória de aquisição, as autoras afirmam que as habilidades iniciais de discriminação rítmica dão lugar às habilidades relacionadas à segmentação de palavras e ao agrupamento sintático por meio de pistas prosódicas e, posteriormente, às habilidades relacionadas à compreensão de informações epistêmicas na fala de outrem (Esteve-Gibert; Prieto, 2018).

Especificamente, alguns estudos mostram que bebês apresentam sensibilidade a pistas prosódicas e tais estudos têm sido considerados argumentos que sustentam a hipótese do *bootstrapping* prosódico. O estudo⁴ de Mehler et al. (1988) mostrou que bebês ainda no útero reconhecem e preferem sua língua materna em detrimento de outras línguas ritmicamente diferentes. Esse fato é explicado pela experiência pré-natal de exposição ao *input* da fala da mãe, uma vez que o ouvir se inicia no terceiro trimestre de gestação, com a particularidade de que, no útero, o sinal de fala chega com as propriedades prosódicas preservadas, enquanto propriedades segmentais são suprimidas (cf. Chen et al., 2020). Já entre 3 e 5 meses, estudos mostram que os bebês, guiados por pistas prosódicas, discriminam não apenas sua língua nativa, mas também línguas que não lhe são familiares, ou seja, línguas estrangeiras com base em suas propriedades rítmicas e entoacionais se essas forem distintas entre a língua materna e a estrangeira (Nazzi et al., 1998; Nazzi et al., 2000; Chen et al., 2020). Após os 6 meses de idade, porém, enquanto a sensibilidade aos padrões entoacionais nativos é mantida e aprimorada, a sensibilidade aos padrões entoacionais não nativos começa a diminuir. Esse fenômeno, denominado na literatura de “reorganização perceptual”, indica que a criança vivencia um processo necessário de especialização linguística, no qual os padrões de sua língua materna se consolidam como sistema regulador da percepção.

Assim como mostram Chen et al. (2020), a partir de então, as crianças passam a perceber limites frasais prosódicos (como frases fonológicas) e os utilizam para identificar os limites entre as palavras do léxico da sua língua. Alguns meses depois, por volta dos 10 meses de idade, os bebês exploram a estrutura prosódica de um enunciado e acessam a partir deles alguns constituintes sintáticos; esse acesso, por sua vez, os ajudará a descobrir o significado de novas palavras (Chen et al., 2020).

No que diz respeito à entoação, os estudos mostram que os bebês demonstram preferência perceptual por contornos ascendentes, tons agudos e melodias expandidas e mostram ainda que os bebês são sensíveis também a mudanças que possam ocorrer nesses contornos melódicos (Frota; Butler, 2018). No que diz respeito ao processo em termos macro, os autores ponderam que a percepção de um determinado contraste entoacional é um pré-

⁴ No estudo de Mehler et al. (1988), a metodologia baseou-se em experimentos de percepção com recém-nascidos para investigar a discriminação rítmica entre línguas. Os autores utilizaram o paradigma de sucção não nutritiva de alta amplitude (*high-amplitude sucking*), no qual bebês com poucos dias de vida ouviam estímulos de fala enquanto sugavam uma chupeta conectada a um sensor. A taxa de sucção era registrada como indicador de atenção e discriminação auditiva. Inicialmente, os recém-nascidos eram habituados a frases em uma língua; em seguida, testava-se a mudança para frases de outra língua pertencente a uma classe rítmica diferente ou semelhante. O aumento na taxa de sucção diante da mudança indicava que os bebês discriminavam os padrões rítmicos entre as línguas, permitindo inferir sensibilidade precoce às propriedades prosódicas da fala.

requisito para a aquisição da gramática e dos aspectos pragmáticos que a ele estão envolvidos. É por essa razão que, segundo Frota e Butler (2018), a aquisição dos contrastes entoacionais segue trajetórias distintas. A percepção da entoação manifesta-se precocemente e de forma sensível aos contrastes da língua materna, mas a habilidade perceptual não é adquirida de maneira simultânea para todos os tipos de contraste (Frota; Santos, 2023). Entre os diferentes contrastes entoacionais, a diferenciação entre interrogativas do tipo sim-não em relação a declarações, por exemplo, destaca-se por sua maior saliência acústica, sobretudo em termos de altura e direção tonal das primeiras em relação às segundas, sendo, portanto, esse contraste entoacional adquirido perceptualmente em estágios primários (Frota; Butler, 2018). Em contrapartida, contrastes entoacionais baseados em diferenças acústicas mais sutis e fortemente dependentes do contexto pragmático tendem a ser adquiridos mais tardiamente, como apontam Frota e Butler (2018), uma vez que exigem a integração entre aspectos prosódicos e pragmáticos, tais como a identificação da intenção do falante, para a interpretação do enunciado.

Retomando a assunção de que o processo de aquisição da prosódia é gradual e multifacetado, conforme propõem tanto Chen et al (2020) quanto Esteve-Gibert e Prieto (2018), por meio dos estudos previamente reportados, buscou-se explicitar que a aquisição prosódica começa muito cedo com a sintonia perceptual a padrões entoacionais e rítmicos e a pistas de fronteiras prosódicas. No entanto, a maturação perceptual a padrões entoacionais e rítmicos e, especialmente, a percepção funcionalmente orientada para a identificação de padrões formais que marcam diferentes funções comunicativas se estende pela infância e adentra a adolescência (cf. a respeito Esteve-Gibert; Prieto, 2018; Chen et al., 2020). A seguir, passa-se a reportar os estudos que investigaram a percepção de funções prosódicas ao longo da infância.

De modo geral, conforme apresenta a revisão de Chen et al (2020), pesquisas indicam que, entre os 3 e os 11 anos, as crianças desenvolvem, de maneira simultânea, diferentes funções da prosódia, ainda que seja possível identificar uma gradualidade na aquisição das funções.

Utilizando a mesma metodologia para a coleta de dados a partir da aplicação do protocolo *Profiling Elements of Prosody in Speech - Communication (PEPS-C)* (McCann; Peppé, 2003), Wells et al (2004), Filipe et al (2012) e Sousa et al (2020) mostraram, respectivamente, para o inglês, o português europeu e o português brasileiro, que, entre os 6 e 9 anos, ocorre a maturação inicial do sistema prosódico, incluindo as habilidades de diferenciar enunciados, como afirmações, perguntas do tipo sim-não e enunciados imperativos, por meio de pistas entoacionais salientes, como altura e direção tonal. Nesse período, as crianças aprimoram gradualmente a percepção e a produção dos contornos prosódicos mais salientes,

enquanto contrastes mais sutis se consolidam de forma mais lenta. Entre os 10 e 14 anos, observa-se continuidade da maturação prosódica, com avanços na interpretação de enunciados complexos que envolvem funções pragmáticas, como ironia, surpresa ou foco contrastivo. Esse processo irá culminar na puberdade, quando o desempenho prosódico em termos de acurácia se aproxima do desempenho dos adultos, permitindo à criança compreender e produzir enunciados de complexidades variadas de maneira funcional e eficaz nas situações comunicativas.

De mesmo modo, Kalathottukaren e Purdy (2017), centrando-se particularmente na descrição do desempenho perceptual, mostram, para o inglês da Nova Zelândia, que o desenvolvimento das habilidades prosódicas receptivas em crianças ocorre principalmente entre os 7 e 9 anos de idade. Durante esse período, as crianças aprimoram sua capacidade de perceber e interpretar padrões prosódicos, como entonação, ritmo e acentuação, que são essenciais para a compreensão de diferentes tipos de enunciados e para a identificação de informações sintáticas, semânticas e pragmáticas. Os autores destacam que esse desenvolvimento não se limita à percepção isolada de sons, mas envolve a integração de múltiplas funções da prosódia, permitindo que a criança utilize pistas prosódicas para interpretar intenções comunicativas e organizar semanticamente as informações no discurso.

Como se pode observar até aqui, a descrição do processo de aquisição perceptual da prosódia privilegia a percepção do ponto de vista auditivo, ou seja, privilegia, seguindo a tradição dos estudos prosódicos, o reconhecimento de padrões prosódicos de uma língua, considerando o seu funcionamento em termos de pistas auditivas, as quais se manifestam no plano acústico por meio de parâmetros fonéticos. No entanto, nas últimas décadas, inaugurou-se uma perspectiva de análise da aquisição prosódica. Um crescente número de estudos têm mostrado que a prosódia também se expressa por meio de movimentos corporais produzidos com a mão, a cabeça ou o rosto (Esteve-Gibert; Guellaï, 2018). Sob essa perspectiva, contemplar a aquisição prosódica implica considerar como a associação desses movimentos gestuais é adquirida de forma relacionada aos padrões prosódicos sonoros.

Em outras palavras, tendo em vista essa tendência no interior das pesquisas sobre prosódia, a partir da qual se incluem os gestos junto aos padrões de natureza melódica e rítmica da fala, a prosódia tem sido entendida e estudada a partir de uma perspectiva multimodal (McNeill, 1985; 1992; 2005; Rusiewicz, 2011; Gilli-Fivela, 2018). Portanto, a aquisição da prosódia tem sido revisitada a partir desse olhar (Esteve-Gibert; Guellaï, 2018). É sobre os gestos enquanto intrinsecamente associados a aspectos sonoros, rítmicos e melódicos da fala que tratamos a seguir, apresentando a perspectiva multimodal da prosódia, assumida no presente trabalho.

1.2 AQUISIÇÃO PROSÓDICA: A PERCEPÇÃO SEGUNDO A PERSPECTIVA MULTIMODAL

A relevância da multimodalidade na percepção da fala foi primeiramente apresentada por McGurk e MacDonald (1976), ao demonstrarem que pistas visuais, como movimentos labiais, podem influenciar diretamente a identificação de sons da fala. Posteriormente, essa noção, inicialmente restrita ao plano segmental, foi expandida para o plano prosódico.

Gili-Fivela (2018) sistematiza os princípios da abordagem multimodal, relacionando-os de modo mais específico ao plano prosódico. Segundo a autora, sob a ótica da multimodalidade, parâmetros gestuais, perceptíveis pela via visual, ocorrem de forma simultânea às pistas verbais prosódicas percebidas pela via auditiva e tradicionalmente descritas em análise acústica, como duração, intensidade e frequência fundamental. Adicionalmente, a autora reitera que, segundo a perspectiva multimodal, pistas auditivas e visuais atuam de forma conjunta na sinalização das diferentes funções atribuídas à prosódia.

Em termos teóricos, de acordo com McNeill (1985), na perspectiva da multimodalidade, a fala é composta de forma constitutiva por gestos, isto é, por *co-speech gestures* na nomenclatura usada pelo autor. Esses gestos não são considerados acessórios, mas sim integrados ao sistema linguístico, compondo, portanto, um único sistema comunicativo de natureza heterogênea. Para sustentar essa concepção, o autor apresenta distintos argumentos, dos quais destacam-se aqui três: (i) os gestos e as unidades verbais sonoras da fala expressam valores semânticos equivalentes ou correlatos e desempenham funções pragmáticas semelhantes; (ii) há sincronia temporal entre as modalidades auditiva e visual; e (iii) o desenvolvimento gestual na infância está intimamente relacionado à aquisição linguística, uma vez que, em estágios iniciais, os gestos desempenham de modo exclusivo e concreto a função dêitica na interação e, gradualmente, integram-se à fala para constituírem um sistema pragmático funcional.⁵

Em trabalho de relevância notória, Swerts e Krahmer (2005) mostraram que expressões faciais, movimentos de cabeça e gestos funcionam como pistas visuais associadas

⁵ Em relação à equivalência semântica (i), McNeill (1985) destaca que o gesto não repete a fala, mas a complementa; por exemplo, ao dizer que um objeto é "grande", o falante pode indicar com as mãos a "forma circular" desse objeto, integrando dimensão e formato em uma unidade de sentido. No que tange à aquisição (iii), a literatura experimental demonstra que a combinação gesto-palavra (ex: apontar para um pão + dizer "quero") precede e prepara a sintaxe biverbal, evoluindo de uma função puramente indicativa para uma coordenação rítmica onde o gesto marca o foco informativo do enunciado.

às auditivas, auxiliando na interpretação da entoação e das intenções comunicativas. Por sua vez, Kendon (2004) e Wagner et al (2014) exploram o conceito de “*visible action as utterance*”, o qual descreve gestos ou ações corporais que atuam como unidades comunicativas, podendo ser interpretados como equivalentes visuais de enunciados falados. Esses gestos podem constituir um enunciado completo ou complementar a fala, sendo integrados ao discurso por meio de pistas visuais que reforçam ou modulam o significado verbal. Miranda et al (2020) destacam que, quando tais ações visuais ocorrem de forma concomitante à fala, elas podem ser compreendidas como uma manifestação de prosódia visual, desempenhando papel fundamental na comunicação multimodal ao fornecer informações sobre intenção, ênfase e estrutura discursiva.

Em revisão teórica sobre os estudos de aquisição que se fundamentaram na perspectiva multimodal, Rusiewicz e Esteve-Gibert (2018) reiteram que, durante a produção de fala, os gestos apresentam sincronia e alinhamento com elementos proeminentes e com fronteiras prosódicas, funcionando como uma forma de prosódia visual que complementa e reforça a estrutura prosódica e a sinalização da intenção comunicativa. Essas proeminências funcionam como pontos de ancoragem desses movimentos sincronizados, o que é explicado pelos autores como um sincronismo baseado em pulsos motores.

O sincronismo entre fala e corpo fundamenta-se na Dynamic Systems Theory (DST) e na *Manual and Speech Draggged Systems Theory* (TEMSS), que concebem o gesto e a fala como partes de um único sistema comunicativo coordenado. Segundo essa perspectiva, as proeminências prosódicas — ou pulsos motores — caracterizadas pelo aumento do esforço fisiológico e acústico em sílabas tônicas funcionam como pontos de ancoragem. Esses pulsos 'atraem' e sincronizam, no espaço e no tempo, outros movimentos corporais, como gestos manuais, inclinações de tronco e elevação de sobrancelhas (Rusiewicz, 2011).

De acordo com Rusiewicz e Esteve-Gibert (2018), essa coordenação mútua influencia a magnitude e a duração dos movimentos, de modo que os picos de proeminência nos acentos nucleares se alinham sistematicamente aos ápices dos gestos associados. Nessa abordagem, a fala e o gesto não são sistemas independentes, mas trabalham de forma intrinsecamente integrada para expressar nuances de significação, com o alinhamento específico dependendo tanto das sílabas tônicas quanto das fronteiras entre as unidades prosódicas. Embora a teoria não aborde diferenças observáveis no processo, seus princípios também são aplicáveis às crianças (Rusiewicz; Esteve-Gibert, 2018).

Do ponto de vista formal, todos os gestos apresentam características motoras semelhantes e exigem precisão motora comparável. Contudo, eles diferem significativamente

em termos funcionais, de modo que, quanto mais complexa for a função pragmática representada pelo gesto, mais tardiamente ele tende a emergir no processo de aquisição (Esteve-Gibert; Prieto, 2014). Por exemplo, movimentos de cabeça, expressões faciais ou gestos manuais podem indicar incerteza, atenuação, envolvimento do interlocutor ou intenção comunicativa implícita, contribuindo para a interpretação do enunciado para além do conteúdo verbal.

Em essência, os tipos de gestos que os bebês conseguem produzir em cada estágio estão estreitamente relacionados aos padrões prosódicos e de fala aos quais esses gestos estarão alinhados. Assim, o padrão de surgimento dos diferentes tipos de gestos parece depender do desenvolvimento linguístico, comunicativo e cognitivo da criança (Rusiewicz; Esteve-Gibert, 2018).

Segundo Iverson e Thelen (1999) e Iverson e Fagan (2004), bebês de aproximadamente seis meses ainda não produzem gestos comunicativos (como os gestos de apontar) nem iniciaram a fase de palavras isoladas. No entanto, o estreito ajuste entre os sistemas vocal e motor nesse período sugere que gesto e fala se desenvolvem de forma paralela, podendo constituir um sistema integrado desde os estágios iniciais da linguagem e da comunicação. Nessa fase, os bebês produzem movimentos rítmicos orais coordenados com movimentos rítmicos manuais, o que é interpretado como o início das regras de sincronização de pulso que regularão, posteriormente, a coordenação entre as proeminências prosódicas e gestuais.

Parladé e Iverson (2011) apontam que, em determinados estágios do desenvolvimento da linguagem, os bebês tendem a privilegiar uma modalidade em detrimento de outra para cumprir funções linguísticas específicas. Segundo os autores, à medida que o vocabulário oral (palavras faladas) se expande, a comunicação multimodal, caracterizada pela combinação de fala, gestos e toques, diminui. Posteriormente, com a consolidação da fala, essa forma de comunicação multimodal volta a aumentar.

Sequencialmente, os gestos dêiticos são os primeiros a emergir e têm a função de localizar objetos ou eventos no espaço, de forma concreta quando o referente está fisicamente presente, ou de forma abstrata quando se trata de elementos ausentes ou apenas contextualmente inferidos. Normalmente envolvem o uso do dedo indicador, embora outras partes do corpo, como cabeça, queixo ou nariz, também possam ser empregadas. Esses gestos surgem geralmente por volta dos 9 e 10 meses e constituem um dos sinais mais claros de que os bebês desenvolveram comunicação intencional referencial (Iverson; Goldin-Meadow, 2005; Özçalışkan; Goldin-Meadow, 2005). Eles podem assumir formas como gestos de mostrar, dar

ou apontar. Inicialmente, o bebê aponta para usar o adulto como meio de obter um objeto ou evento desejado (apontar protoimperativo) ou para usar o objeto como meio de captar a atenção do adulto (apontar protodeclarativo) (Bates et al., 1975; Volterra et al., 2005).

Por volta dos 12 meses, observam-se gestos com funções mais refinadas: imperativa (solicitar um objeto), expressiva (compartilhar interesse) e informativa. Dessa forma, o desenvolvimento pragmático determina as funções comunicativas que os bebês podem expressar por meio de gestos de apontar, que, por sua vez, parecem impulsionar a aquisição de substantivos na fala (Rusiewicz; Esteve-Gibert, 2018). Essa relação com a aquisição de substantivos explica-se pelo fato de que os gestos de apontar estabelecem um referente claro no ambiente, criando condições privilegiadas para que o adulto associe uma forma linguística a um objeto, pessoa ou evento específico (Bates et al., 1979; Iverson e Goldin-Meadow, 2005). Ao destacar entidades concretas e perceptualmente salientes, o apontar favorece episódios de atenção conjunta nos quais os cuidadores tendem a nomear aquilo que está sendo indicado, oferecendo ao bebê modelos verbais que frequentemente correspondem a substantivos (Carpenter; Nagell; Tomasello, 1998; Tomasello, 2003). Desse modo, tais gestos não apenas expressam intenções comunicativas mais refinadas, mas também funcionam como um suporte para o mapeamento palavra-referente, impulsionando especialmente a aprendizagem de palavras que designam objetos e entidades antes da aquisição mais robusta de outras classes gramaticais, como verbos e adjetivos (Goldin-Meadow, 2007; Tomasello, 2003).

Os gestos seguintes a emergir são os gestos representacionais, que podem ser icônicos (representando objetos ou eventos concretos) ou metafóricos (representando conceitos e ideias abstratas, como por exemplo, a criança pode mover a mão para frente ao falar sobre algo que “vai acontecer”, utilizar as mãos para indicar aumento ou diminuição ao se referir à intensidade de uma situação, ou ainda empurrar a mão para longe ao rejeitar uma proposta ou tema.). As formas iniciais desses gestos surgem por volta dos 15 meses, como o gesto de não, o gesto de tchau, gestos que representam alto e baixo ou ações como beber em um copo. A frequência e a variedade de gestos representacionais aumentam significativamente com o avanço do desenvolvimento lexical e cognitivo, geralmente por volta dos 26 meses. O surgimento desses gestos indica o desenvolvimento de habilidades linguísticas e gramaticais paralelas (Özçalışkan et al., 2013; Özçalışkan; Goldin-Meadow, 2011). No entanto, o padrão difere em relação aos gestos dêiticos: enquanto nos gestos dêiticos o uso para referência precede a fala, nos gestos icônicos o uso para se referir a verbos tende a ocorrer após a fala com a mesma função (Rusiewicz; Esteve-Gibert, 2018).

Observa-se, portanto, uma relação diferenciada entre tipos de gestos representacionais e classes lexicais em aquisição (Iverson; Goldin-Meadow, 2005; Özçalışkan; Goldin-Meadow, 2011). Enquanto os gestos dêiticos, ao estabelecerem referência a entidades do ambiente, estão particularmente associados à aprendizagem de substantivos, os gestos icônicos, por representarem ações e eventos, relacionam-se mais diretamente ao desenvolvimento do léxico verbal (Bates et al., 1979; Tomasello, 2003; Özçalışkan et al., 2013). Contudo, diferentemente do que ocorre com os gestos dêiticos, cujo uso frequentemente precede a produção verbal correspondente, os gestos icônicos tendem a emergir após a criança já empregar formas verbais com função semelhante, refletindo avanços nas capacidades simbólicas e gramaticais necessárias à expressão de relações predicativas (Iverson; Goldin-Meadow, 2005; Rusiewicz; Esteve-Gibert, 2018).

Por fim, os gestos de batimento são os últimos a emergir ontogeneticamente. São movimentos bifásicos, rítmicos e coocorrentes com a fala que destacam itens no discurso, mas não possuem relação semântica direta com o que é dito. As habilidades prosódicas rítmicas típicas de adultos parecem se desenvolver mais tardiamente e acompanham a maturação desses gestos de batimento (Colletta et al., 2015; Mathew et al., 2017).

Na produção de fala, os gestos de batimento (também chamados de *beat gestures* ou bastões) tendem a emergir de forma consistente entre os 5 e 7 anos de idade (Colletta et al., 2015; Igualada et al., 2016; Mathew et al., 2014). Os autores afirmam que esses gestos de batimento não carregam conteúdo semântico, ou seja, não representam objetos, ações ou conceitos, como fazem os gestos icônicos ou metafóricos (McNeill, 1992; Kendon, 2004). Sua função principal é estruturar o discurso, marcando proeminência informacional, foco e organização rítmica da fala (McNeill, 1992; Wagner et al., 2014). Nesse sentido, os gestos de batimento atuam como sinais visuais de ênfase e segmentação, contribuindo para a hierarquização dos constituintes do enunciado (Krahmer; Swerts, 2007). Assim, quando estudos mostram que ouvintes utilizam o alinhamento entre gestos de batimento e a fala para processar frases ambíguas, isso não implica que tais gestos representem significado lexical ou conceitual. O que ocorre é que eles fornecem pistas adicionais sobre a organização prosódica e sintática da sentença, por exemplo, indicando onde uma fronteira sintagmática é mais provável (Rusiewicz; Esteve-Gibert, 2018; Biau; Soto-Faraco, 2013). Essas pistas ajudam o ouvinte a decidir entre diferentes interpretações possíveis, funcionando como marcadores de estrutura, e não como portadores de conteúdo semântico propriamente dito (Prieto et al., 2013). Por exemplo, os falantes podem destacar elementos contrastivamente focalizados em um enunciado não apenas

por meio de pistas prosódicas, mas também através de gestos de batimento realizados com a cabeça (Esteve-Gibert et al., 2017; Kim et al., 2014).

Essa relação estreita entre os pulsos motores gestuais e acústicos na constituição da fala como sistema comunicativo dinâmico, conforme defendido por Rusiewicz (2011), está em consonância teórica com o modelo de percepção multimodal da fala proposto por Rosenblum (2008). Segundo este autor, a percepção da fala possui natureza intrinsecamente multimodal e baseia-se em informações amodais, isto é, o cérebro não trata os estímulos auditivos e visuais como canais independentes que seriam integrados apenas posteriormente. Em vez disso, o sistema perceptivo extrai diretamente do sinal uma assinatura dinâmica e rítmica compartilhada entre as modalidades. Nessa perspectiva, os movimentos corporais e de cabeça mencionados anteriormente não desempenham um papel meramente acessório ou redundante em relação ao som. Ao contrário, eles integram a própria constituição primitiva e unificada da fala, oferecendo pistas estruturais e salientes que são processadas de maneira imediata e automática pelo ouvinte para auxiliar na resolução de ambiguidades e na interpretação da informação linguística.

Como se procurou explicitar com o relato dos trabalhos desenvolvidos à luz da perspectiva multimodal, a aquisição prosódica pode ser descrita considerando a intrínseca relação entre pistas prosódicas gestuais e pistas prosódicas verbais-sonoras. Deve-se destacar, porém, que os trabalhos reportados foram realizados com crianças cuja língua materna não é o português brasileiro. Nessa língua, é escassa a investigação da aquisição prosódica sob a perspectiva multimodal, pois a maioria dos estudos prosódicos desenvolvidos nessa perspectiva se voltam para a investigação da percepção prosódica em sujeitos adultos (cf. Miranda et al., 2020; Miranda et al., 2021; Carnaval, 2021; Carnaval et al., 2022; Carnaval et al., 2023). No entanto, os trabalhos de Cavalcante (2012, 2015) e Cavalcante et al. (2017) investigam a emergência do que a autora denomina "matriz multimodal da linguagem". Seus estudos demonstram que, no Português Brasileiro, a criança utiliza a coordenação entre contornos entoacionais e movimentos corporais rítmicos como um suporte estruturante para a organização sintática, destacando que a integração entre gesto e fala não é apenas funcional, mas constitutiva do processo de aquisição linguística em nossa língua.

No que tange a resultados sobre percepção de diferentes atos de fala, tema sob investigação no presente trabalho no âmbito da aquisição prosódica, Miranda et al (2020), mostram que ouvintes brasileiros adultos integram informações auditivas e visuais para identificar diferentes tipos de atos de fala, a saber: asserção, interrogação (questão-eco e questão parcial) e exclamação. Os resultados de Miranda et al. (2020) evidenciam a natureza integrada desse processo, uma vez que a presença da pista visual, em comparação à condição estritamente

auditiva, elevou significativamente a precisão na identificação dos atos de fala. Corroborando essa visão, Miranda et al. (2021) argumentam que esse processamento multimodal não apenas otimiza o desempenho perceptivo, como é fundamental para a construção da interpretação semântica dos enunciados.

Tem-se, assim, que, em consonância com os resultados da literatura internacional, os resultados existentes obtidos com dados do Português Brasileiro sobre a percepção prosódica multimodal, ainda que privilegiadamente com adultos, mostram que pistas auditivas e visuais atuam de forma integrada, contribuindo para a identificação de sentidos que se definem no plano pragmático, como é o caso da identificação de atos de fala que se diferenciam prosodicamente.

Em continuidade ao panorama apresentado, torna-se relevante examinar como diferentes aspectos da prosódia têm sido investigados sob a ótica da multimodalidade e identificar em que condições experimentais a convergência entre as pistas auditivas e visuais torna-se mais determinante. Busca-se compreender como essa integração indissociável, importante para o processo de aquisição desde o seu início, potencializa a interpretação linguística e em quais contextos essa sinergia perceptiva é mais exigida pelo ouvinte. Embora em termos teóricos, os estudos argumentem em defesa da integração indissociável entre pistas auditivas e visuais, a literatura recente tem priorizado investigações que, experimentalmente, avaliem o efeito das pistas visuais em contextos de escuta adversa. Nessas situações, a degradação do sinal acústico de um conjunto de estímulos dos testes experimentais possibilita avaliar se haveria um efeito compensatório da modalidade visual em relação à auditiva, ou seja, permite mensurar se a informação visual ancoraria a compreensão quando o acesso ao canal sonoro está comprometido. Tal recorte metodológico tem possibilitado avançar no entendimento de como componentes prosódicos, como contornos entoacionais, proeminências e atitudes, são processados de forma robusta, demonstrando que, diante da degradação acústica, o sistema perceptivo recorre à sinergia multimodal para garantir a interpretação dos atos de fala.

Portanto, no que se refere à investigação do efeito compensatório da pista visual na percepção de atos de fala em condições com presença de ruído, os estudos encontrados, tanto internacionais quanto nacionais, foram realizados apenas com grupos de adultos, evidenciando como a informação visual atua na facilitação do processamento linguístico em ambientes acusticamente desafiadores (Van Engen et al., 2022). Os estudos encontrados examinaram a percepção de contornos entoacionais e tipos frasais, incluindo a distinção entre afirmação e interrogação e entre sentenças declarativas e interrogativas totais, bem como a integração audiovisual (Miranda et al., 2024; Barkhuysen; Krahmer; Swerts, 2020; Miranda et al., 2020; Miranda et al., 2021; Carnaval; Miranda; Souza, 2022, Munhall et al., 1996; Pilling; Thomas,

2011). Além disso, os estudos também se voltaram para a investigação da percepção de proeminência prosódica e o efeito compensatório de pistas visuais específicas, como a relevância de diferentes regiões da face (olhos, boca e movimentos de cabeça) para a identificação de tipos frasais no Português Brasileiro (Carnaval, 2021; Souza, 2018; Swerts; Krahmer, 2008, 2010) e, por último, outros estudos abordaram a percepção de aspectos pragmáticos e atitudinais em perspectiva multimodal, como o reconhecimento de atitudes (ironia, sarcasmo, polidez e grosseria) e da intenção comunicativa em diferentes tipos de sentenças (Oliveira, 2019; Gowda et al., 2021).

Nesse contexto, Van Engen et al. (2022) ampliam a compreensão do efeito compensatório da pista visual ao demonstrarem que o benefício da informação visual em ambientes ruidosos se sustenta em dois mecanismos principais: a discriminação fonética e a predição temporal. Segundo as autoras, os movimentos articulatórios faciais oferecem pistas complementares que permitem ao ouvinte identificar e diferenciar contrastes linguísticos que se tornam difíceis de perceber quando o sinal acústico está degradado. Paralelamente, a percepção visual do rosto atua como um marcador temporal da fala, possibilitando a antecipação da organização rítmica do discurso contínuo. Esse processo reduz significativamente a quantidade de competidores lexicais ativados e favorece a segregação do sinal linguístico em meio ao ruído ambiental. Assim, a contribuição da modalidade visual não se limita ao aumento da inteligibilidade da fala, mas também atua na redução do esforço cognitivo exigido para o processamento e a interpretação de atos de fala complexos.

De modo geral, os estudos convergem ao demonstrar que a percepção prosódica, especialmente no Português Brasileiro, é inerentemente multimodal, na qual as modalidades auditiva e visual operam como um sistema unificado. Essa natureza multimodal torna-se particularmente evidente em condições de ruído, situações em que o efeito compensatório da pista visual assume um papel determinante para assegurar a compreensão, suprimindo a degradação do sinal acústico para garantir a decodificação dos sentidos pretendidos. Especificamente, as evidências indicam que, na distinção entre tipos frasais e contornos entoacionais (como afirmações vs. interrogativas), a saliência das pistas é redistribuída conforme a qualidade do sinal: enquanto a modalidade auditiva costuma predominar em condições ideais, ela perde confiabilidade sob mascaramento acústico. Nesse cenário de degradação, o sistema perceptivo aciona o efeito compensatório de pistas faciais e movimentos de cabeça, que passam a desempenhar um papel determinante na identificação da intenção comunicativa (Barkhuysen; Krahmer; Swerts, 2010; Miranda et al., 2020, 2021; Carnaval; Miranda; Souza, 2022). Tais estudos revelam, ainda, que as sentenças interrogativas exibem

uma marcação visual mais estereotipada do que as declarativas, o que potencializa a facilitação visual e torna a identificação de perguntas particularmente robusta em contextos multimodais.

Além disso, na percepção de sentenças completas em inglês (fala encadeada), identificou-se que movimentos de cabeça sincronizados com a fala aumentam a acurácia. Assim, a prosódia visual, e não apenas os movimentos labiais, constitui um componente essencial para favorecer a interpretação do sinal de fala em ambientes ruidosos (Munhall et al., 1996; Pilling; Thomas, 2011).

O estudo de Miranda et al. (2024), ao investigar a distinção entre atos de fala do tipo pergunta-QU e exclamação-QU no português brasileiro, reforça a premissa da integração indissociável ao constatar que o processamento audiovisual, comparado às modalidades isoladas, resultou em maiores taxas de acerto e menores tempos de resposta. Embora a entonação (F0) se mantenha como uma pista de alta saliência para a identificação de perguntas no PB, a convergência multimodal mostrou-se fundamental para a robustez do sistema perceptivo. Os autores destacam que as pistas visuais, como a elevação das sobrancelhas e a abertura dos olhos, foram determinantes para a caracterização das exclamações, evidenciando como a expressividade facial se integra à acústica na construção do sentido. Por fim, o trabalho demonstra que, diante de pistas ambíguas ou da degradação do sinal sonoro, a modalidade visual assume um papel de filtro de desambiguação, exercendo um efeito compensatório que permite ao ouvinte interpretar com segurança a força ilocucionária do enunciado.

Essa integração indissociável estende-se também à percepção de proeminência e foco informacional. Estudos demonstram que o sistema perceptivo recorre sistematicamente a 'batidas visuais', movimentos rítmicos de cabeça e sobrancelhas, para localizar constituintes destacados, especialmente quando as variações de F0 e intensidade são mascaradas (Carnaval, 2021). Conforme propõem Swerts e Krahmer (2008, 2010), essas marcações visuais constituem pistas extremamente robustas de foco, assumindo o protagonismo na identificação da proeminência em cenários de degradação acústica. A sensibilidade dos ouvintes à sincronia entre o ápice da entonação e o pico do movimento reforça a natureza unitária desse processamento. Complementarmente, Souza (2018) evidencia que essa carga informacional prosódica se concentra predominantemente na parte superior da face e nos movimentos de cabeça, diferenciando-se dos movimentos labiais, cuja função é prioritariamente associada à articulação segmental.

No tocante à interpretação pragmática e atitudinal, a integração indissociável entre as modalidades torna-se ainda mais estratégica. Estudos revelam que, em condições de ruído, a identificação de atitudes como ironia, sarcasmo ou polidez depende criticamente da expressão

facial, com ênfase na região dos olhos (Oliveira, 2019). Nesses cenários, a intenção comunicativa é preservada graças à redundância⁶ prosódica entre os canais: o sistema perceptivo recupera a força ilocucionária por meio da convergência audiovisual, garantindo a interpretação mesmo quando o conteúdo lexical encontra-se mascarado (Gowda et al., 2021).

Em conjunto, essas evidências demonstram que a prosódia, ao manifestar-se simultaneamente nas modalidades auditiva e visual, opera como um sistema de integração indissociável, garantindo a robustez comunicativa e a compreensão mesmo em condições de escuta adversa. Se, por um lado, está consolidado que adultos possuem elevada acurácia para processar essa dinâmica multimodal e recuperar os sentidos pragmáticos na interação, por outro, permanece em aberto como esse refinamento perceptual se consolida ao longo da aquisição linguística. Torna-se necessário, portanto, investigar em que medida a integração entre pistas auditivas e visuais contribui para a identificação dos atos de fala pela criança, tanto em contextos ideais de acesso à informação quanto em ambientes ruidosos e desafiadores, como o cenário escolar, onde a função da pista visual pode ser determinante para a compreensão de fala.

Portanto, os resultados da literatura não apenas confirmam a natureza multimodal da percepção prosódica, mas também apontam para a necessidade de investigar sua emergência e consolidação durante o processo aquisitivo, especialmente no que se refere à interpretação de atos de fala que se diferenciam prosodicamente no português brasileiro dada a relevância comunicativa que a prosódia assume ao marcar a força ilocucionária dos enunciados linguísticos (Cole, 2015).

Dessa maneira, visando contribuir com a literatura, esta tese busca preencher essa lacuna ao oferecer, de forma inédita, respostas para as seguintes perguntas de pesquisa:

1) A acurácia perceptual relativa à identificação de atos de fala diferenciados prosodicamente por pistas auditivas e visuais, a saber, declaração, pergunta, ordem e pedido, se transformaria ao longo do processo de aquisição prosódica do português brasileiro?;

2) Qual seria o efeito de pistas prosódicas auditivas e visuais na identificação desses diferentes atos de fala pelos grupos etários?;

⁶ O conceito de redundância prosódica, conforme postulado por Gowda et al.(2021), refere-se à natureza multifacetada da sinalização das intenções comunicativas, em que um único contraste pragmático (como a distinção entre sentenças declarativas e condicionais) é codificado simultaneamente por múltiplas pistas acústicas sobrepostas, incluindo variações de frequência fundamental (F0), intensidade e, crucialmente, pistas temporais (duração e pausas). Essa sobreposição funciona como um mecanismo de segurança perceptiva: em condições de ruído severo, onde componentes segmentais ou variações tonais específicas podem ser mascarados, a permanência das pistas temporais garante a sobrevivência da informação prosódica e a correta identificação da intenção do falante, conferindo ao sistema de reconhecimento da prosódia uma robustez superior ao reconhecimento estritamente verbal das sentenças.

3) A modalidade visual exerceria um efeito compensatório na identificação dos atos de fala no português brasileiro pelos grupos etários quando os estímulos são apresentados em condição de ruído?

A fim de responder essas perguntas e, conseqüentemente, a fim de atender aos objetivos da pesquisa que estão a elas diretamente relacionados, apresentam-se, na próxima seção, conceitos teóricos que fundamentaram a investigação realizada. Inicia-se, abordando o conceito de ato de fala.

1.3 REFERENCIAIS TEÓRICOS

1.3.1 Uma breve definição de pragmática

De acordo com Levinson (1983), a pragmática é definida como o campo de estudo que analisa a relação entre a estrutura de um sistema semiótico, em especial, a linguagem, e seu uso em contextos de interação. Integrando, ao lado da semântica, a teoria geral do significado, a pragmática concentra-se particularmente nos aspectos implícitos da comunicação, envolvendo inferências, pressupostos e conteúdos não ditos. Nesse sentido, investiga de que modo a organização estrutural da linguagem se articula com elementos contextuais e cognitivos, permitindo ao interlocutor construir sentidos que não estão explícitos no conteúdo proposicional expresso.

Segundo Armengaud (2006), a teoria clássica dos atos de fala parte da ideia de que a unidade mínima da comunicação humana não é a frase nem qualquer outra expressão isolada, mas sim a realização, ou seja, o *uso* de um enunciado linguístico, tendo em vista que é o uso que transforma o enunciado em ato de fala, haja vista que o uso situa a produção desse enunciado ao associá-lo a um contexto comunicativo e aos participantes da interação (falante e ouvinte). John L. Austin, precursor dessa perspectiva, elaborou uma extensa lista desses atos, incluindo afirmar, questionar, ordenar, prometer, descrever, agradecer, criticar, pedir, desafiar, autorizar, entre outros, inaugurando a Teoria dos Atos de Fala com sua obra *How to Do Things with Words* (1962).

Armengaud (2006) explica que a ideia subjacente à teoria dos atos de fala é considerar que, ao produzir um enunciado, o falante pode realizar diversas ações por meio da produção verbal. Essas ações dependerão da intenção e do contexto comunicativo em que falante e ouvinte se inserem. Desse modo, a Teoria dos Atos de Fala consiste em uma teoria de natureza pragmática, ao investigar sistematicamente a relação entre signos e intérpretes e a forma como os interlocutores operam e compreendem diferentes tipos de signos em situações comunicativas, tendo como unidade mínima da comunicação não a frase, mas o ato de fala em si.

1.3.2 Teoria dos atos de fala

A teoria dos atos de fala, conforme mencionado anteriormente, foi desenvolvida por John L. Austin (1962, 1990). No interior dessa teoria, o autor define os atos de fala como *ações realizadas por meio da linguagem*. Para ele, falar não é apenas enunciar palavras com

significado literal, mas executar atos que têm força social e comunicativa. Em outras palavras, os atos de fala podem ser definidos como eventos comunicativos que se configuram quando um falante ou escritor (S) produz um enunciado (U), dirigido a um ouvinte ou leitor (H), em um contexto específico (C). Mais do que simples construções linguísticas, esses atos constituem práticas sociais, nas quais a linguagem assume função performativa e interacional. Dessa forma, assim como afirmam Allan, Lamarque e Asher (1997), os atos de fala devem ser compreendidos não apenas em sua estrutura formal ou semântica, mas, sobretudo, enquanto manifestações do comportamento social e como componentes essenciais da interação comunicativa.

Um aspecto central da teoria de Austin (1962/1990) é a crítica à visão descritivista da linguagem, dominante na tradição filosófica anterior, que reduzia o enunciado à função de transmitir verdades ou falsidades. Aristóteles, por exemplo, já havia observado que nem todas as sentenças podem ser consideradas enunciados (*apophantikos*), mas apenas aquelas que comportam valores de verdade (Allan; Lamarque; Asher, 1997, p. 454-466). Para Austin (1962/1990), muitos enunciados não são avaliáveis em termos de condições de verdade, mas sim em termos de felicidade ou infelicidade: promessas, desculpas e ordens só “funcionam” se certas condições forem satisfeitas (como a autoridade do falante, um contexto apropriado e aceitação legítima do interlocutor). Tais condições foram posteriormente sistematizadas como condições de felicidade (*felicity conditions*). Desse modo, a proposta de Austin (1962/1990) surge como uma reação a essa tradição que enfatiza que o enunciado constativo, tradicionalmente valorizado pela filosofia por sua natureza lógica, possui como característica central a possibilidade de ser verdadeiro ou falso (Allan; Lamarque; Asher, 1997, p. 454-466; Santos, 2009). Para tanto, ganham posição central na teorização do autor os enunciados performativos, contrapondo-se aos constativos, privilegiados pela lógica clássica. Austin (1962/1990) explicita que, enquanto os primeiros dizem respeito a enunciados que realizam ações ao serem ditos (como “Eu prometo que estarei aqui”), os últimos realizam relatos ou descrições sobre algo (como “A mosca caiu no copo”).

A posição de Austin (1962/1990) de que os enunciados performativos não possuem valores de verdade foi amplamente contestada desde cedo, revelando-se, em grande medida, problemática. Em termos gerais, o ponto central não está na possibilidade de atribuir verdade ou falsidade a tais enunciados, mas sim na sua dimensão comunicativa, que se manifesta sobretudo naquilo que ele denominou “força ilocucionária” do enunciado. Além disso, o próprio autor reconheceu que atos de fala podem ser realizados mesmo na ausência de verbos explicitamente performativos. Diante dessas dificuldades, o filósofo reformula sua abordagem e passa a caracterizar os atos de fala considerando duas outras categorias (Allan; Lamarque;

Asher, 1997; Santos, 2009).

No nível central da caracterização dos atos de fala está o ato ilocucionário. Nesse nível, o falante, ao produzir um enunciado para um ouvinte em um contexto, realiza uma ação socialmente reconhecida, como afirmar um fato ou dar uma opinião, confirmar ou negar algo, fazer uma previsão, promessa ou pedido, agradecer, convidar, aconselhar, ordenar, autorizar, nomear ou realizar um juramento. Nesse sentido, o enunciado carrega uma força ilocucionária ou ilocução que corresponde ao tipo de ação comunicativa expressa pelo enunciado (Austin, 1962/1990; Allan; Lamarque; Asher).

Na maior parte das interações comunicativas, além de produzir uma locução – ou seja, produzir uma proposição linguística ou, em outros termos, uma sentença com sentido e referência (Austin, 1962/1990) – e expressar uma ilocução, o falante também busca provocar um efeito no interlocutor, o que Austin denomina de ato perlocucionário. Esses efeitos incluem, por exemplo, convencer, persuadir, confortar, assustar, advertir, obter uma resposta ou induzir uma ação. O efeito perlocucionário de um enunciado decorre do reconhecimento, por parte do ouvinte, do sentido locucionário e da força ilocucionária do enunciado (Allan; Lamarque; Asher, 1997). Dessa maneira, um ato de fala, de acordo com Austin (1962/1990) é constituído pelo ato locucionário (conteúdo proposicional), pelo ato ilocucionário (sua força em termos de ação) e pelo ato perlocucionário (o efeito provocado no interlocutor).

Embora fundamentais para a teoria da comunicação, os atos perlocucionários extrapolam os limites da linguística estritamente formal, pois dizem respeito às respostas e reações do interlocutor diante do ato de fala. Ainda assim, os linguistas podem investigar as intenções ilocucionárias do falante, isto é, a intenção de produzir determinados efeitos perlocucionários por meio do uso da linguagem (Allan; Lamarque; Asher, 1997, p. 454-466).

Austin (1962/1990) propôs ainda uma primeira tipologia dos atos ilocucionários, distinguindo-os em veredictivos (atos de julgamento, como avaliar ou diagnosticar), exercitivos (atos de ordenar, permitir, aconselhar), comissivos (atos de se comprometer, como prometer ou jurar), comportamentais (atos de expressar atitudes sociais, como agradecer, felicitar, desculpar-se) e expositivos (atos de esclarecer, explicar ou argumentar). Essa classificação foi posteriormente reformulada e ampliada por Searle (1969, 1981), que estabeleceu cinco categorias fundamentais: assertivos, diretivos, comissivos, expressivos e declarativos.

Searle (1969, 1981), então, dá continuidade aos estudos de Austin, aprofundando o desenvolvimento da teoria dos atos de fala. Assim como Austin, ele parte da premissa de que todo enunciado linguístico funciona como um ato particular, sendo a força ilocucionária o elemento que confere ao enunciado o valor de ato de fala. Para o autor, um mesmo enunciado

pode assumir diferentes forças ilocucionárias, assim como pode ser proferido sem carregar valor ilocucional algum.

Em sua sistematização, Searle (1981) distingue três tipos de atos de fala: atos de enunciação, que consistem na simples produção de palavras; atos proposicionais, que se referem ao ato de designar e predicar; e atos ilocucionários, que dizem respeito à força ou ao propósito comunicativo do enunciado, como afirmar, perguntar ou ordenar.

A esses elementos, o autor incorpora o ato perlocucionário, já proposto por Austin (1962/1990), que se refere aos efeitos do enunciado sobre as ações, pensamentos ou crenças do interlocutor. Na proposta de Searle (1981), para que o valor ilocucionário se converta efetivamente em valor perlocucionário, é necessário que sejam respeitadas certas condições de sucesso. De acordo com Kerbrat-Orecchioni (2005), essas condições incluem fatores como a posição hierárquica dos participantes, especialmente em atos diretivos, como a ordem, e o contexto da interação.

Sobre as condições de sucesso, Levinson (2007) explica que Searle retoma e reorganiza o conceito de condições de felicidade proposto por Austin (1962/1990), já que o autor identifica quatro tipos de condições que ajudam a comparar os atos de fala: o conteúdo proposicional, as condições preparatórias, as condições de sinceridade e a condição essencial.

A respeito da classificação dos atos de fala, a tipologia de Searle (1981), conforme apresentada por Levinson (2007), avança em relação à classificação dos verbos performativos de Austin. Searle (1981) propõe um modelo de cinco tipos básicos de atos que se fundamentam nas condições de verdade. Essa classificação dos atos de fala difere da perspectiva original de J. L. Austin porque não se fundamenta diretamente nas condições de felicidade. Em Austin, essas condições dizem respeito às circunstâncias sociais e institucionais que tornam um ato de fala válido ou bem-sucedido, como a autoridade do falante e a adequação do contexto. Já Searle (1981) organiza sua tipologia a partir de critérios como a intenção do falante, o propósito do ato e a relação entre linguagem e mundo. Assim, embora Searle reformule as condições de felicidade em termos de condições preparatórias, de sinceridade e essenciais, elas não constituem a base principal de sua classificação. Por isso, sua proposta desloca o foco das condições externas de êxito do ato para suas propriedades pragmáticas e funcionais.

Na classificação de Searle (1981), os atos se categorizam em: 1) atos representativos, quando o falante se compromete com a veracidade de uma afirmação (afirmar e concluir); 2) atos diretivos, quando o falante tenta fazer com que o ouvinte realize uma ação (pedir, ordenar e perguntar); 3) atos comissivos, quando o falante se compromete com uma ação futura (prometer, ameaçar e oferecer); 4) atos expressivos, quando o falante expressa um estado

psicológico ou emocional (agradecer e pedir desculpas) e 5) atos declaratórios, que provocam uma mudança imediata na realidade, geralmente dentro de um contexto institucional (excomungar, batizar ou declarar guerra).

Considerando a categorização de Searle (1981), os atos de fala alvos de avaliação perceptual neste trabalho são de dois tipos: representativos e diretivos. Enquanto a declaração - ou afirmação, nas palavras do autor - é um ato de fala representativo, pergunta, ordem e pedido são atos de fala diretivos. Entoalcionalmente, no Português Brasileiro, declaração e pergunta do tipo sim/não se contrastam de maneira categórica pela presença de acentos nucleares distintos (Frota et al, 2015). Trataremos dessa caracterização mais adiante. Por sua vez, ordem e pedido apresentam diferenças que se marcam prosodicamente de maneira mais sutil. A discussão sobre o que permite diferenciar esses atos de fala, no entanto, não se limita ao plano da descrição prosódica.

O próprio Searle (1981), ao analisar pedido e ordem como subtipos de atos diretivos, enfatiza a condição comum que esses atos compartilham enquanto tentativas de levar o ouvinte a agir e reconhece diferenças de força e de contexto entre esses atos. Porém, o autor não desenvolve uma distinção sistemática entre eles, privilegiando os aspectos formais comuns. Com base na premissa teórica de que os atos de fala só são executados com sucesso se seguirem certas regras específicas, tanto para o ato de pedir quanto para o ato de ordenar, Searle (1981) estabelece quatro condições. A primeira, a regra de conteúdo proposicional, exige que a ação pretendida seja um evento a ser realizado no futuro. Em seguida, as regras preparatórias determinam que o ouvinte deve ser capaz de realizar a ação e, ainda, que essa não seria realizada por ele no curso normal dos acontecimentos. Por fim, a regra de sinceridade requer que o falante queira genuinamente que a ação seja executada. Desse modo, as regras definem esses atos de forma comum como tentativas de fazer o ouvinte realizar a ação.

No que diz respeito às diferenças, a distinção entre um pedido e uma ordem é um ponto de debate entre diferentes áreas de estudo da linguagem (Santos, 2009). Esse debate é relevante de ser aqui apresentado uma vez que, nesta tese, ordem e pedido estão entre os atos de fala alvos de identificação nos experimentos de percepção multimodal.

De acordo com Santos (2009), para a fonologia, a distinção entre esses atos de fala se manifesta na entonação, o que os diferencia como entidades linguísticas. Já a pragmática tem uma visão diferente. Sob essa perspectiva, uma ordem é, na verdade, uma forma mais enfática de pedido, sendo que a distinção entre eles não residiria na estrutura da frase, mas sim no contexto em que são usados e, principalmente, no grau de hierarquia entre as pessoas que se comunicam.

De acordo com Castilho (2010), autor cujas obras são referência na descrição linguística dos usos do Português Brasileiro, valores diretivos como ordem e pedido são, na visão da gramática tradicional, associados ao modo imperativo; contudo, no Português Brasileiro, os valores desses atos de fala, destaca o autor, não se restringem a esse modo verbal. Funcionalmente, Castilho (2010) demonstra que o modo indicativo é amplamente utilizado para a expressão de diretividade, incluindo ordens, pedidos e conselhos, sobretudo em contextos interacionais específicos. Assim, embora o imperativo represente a forma morfosintática esperada e ditada pela tradição gramatical para a manifestação de ordens e pedidos, o indicativo é a forma verbal recorrentemente usada pelos falantes para expressar esses atos de fala. Como consequência, haja vista que o indicativo é também uma das formas verbais usadas para atos representativos, como a afirmação, a interpretação do enunciado e a identificação do ato de fala resulta da articulação entre *dictum* e *modus*, sendo fortemente orientada por fatores discursivos e prosódicos.

Castilho (2010), ao centrar-se na descrição semântica e pragmática das sentenças do Português Brasileiro, associa ao conceito de ato de fala a distinção entre *dictum* e *modus* e afirma que é o *modus* o responsável por indicar o ato de fala. Nessa visão, portanto, o que diferencia uma ordem de um pedido, considerando especificamente que esses atos de fala se manifestam por formas de verbo idênticas – seja no imperativo, menos usual, seja no indicativo mais usualmente, ou até no subjuntivo – é o *modus*.

O *dictum* corresponde ao conteúdo proposicional do enunciado, isto é, ao estado de coisas descrito por uma sentença, enquanto o *modus* expressa a atitude do falante em relação a esse conteúdo, englobando valores como declarar, indagar, ordenar, pedir, entre muitos outros. Nas palavras do autor: “Cada *dictum* vem associado a um ato de fala. O *modus* evidencia de que ato de fala se trata” (Castilho, 2010, p. 438). Usando uma categoria gramatical, o autor correlaciona o ato de fala indicado pelo *modus* à modalidade da sentença: declarativa, interrogativa, imperativa, por exemplo.

No português brasileiro, conforme afirmação de Castilho (2010), o *modus* pode ser codificado por diferentes meios linguísticos, incluindo recursos morfosintáticos, lexicais e, de modo central, por recursos prosódicos, especialmente a entoação. Dessa forma, Castilho (2010) enfatiza que a distinção entre modalidades e/ou atos de fala não se reduz à estrutura sintática, mas resulta da articulação entre *modus* e *dictum*, para a qual a entoação tem papel de extrema relevância ainda que não seja o único recurso da língua.

No português brasileiro, essa concepção se manifesta de forma particularmente clara no domínio dos atos diretivos, como ordens e pedidos. Como já destacamos, embora o

imperativo seja considerado pela tradição gramatical o modo verbal característico dos atos diretivos, Castilho (2010) observa que as formas do modo imperativo nos verbos tendem a ser restringidas, sobretudo em situações de interação verbal cotidiana. Em contextos nos quais o falante busca preservar a face do interlocutor e reduzir o grau de imposição do ato diretivo, há uma preferência pelo emprego de formas do indicativo e do subjuntivo, frequentemente em estruturas sintáticas mais complexas.

O indicativo, tradicionalmente associado à asserção, pode assumir valor diretivo de ordem ou pedido quando mobilizado no nível do *modus*, funcionando como estratégia de polidez (*Vamos fazer isso* versus *Vá fazer isso*). Já o subjuntivo, por expressar não facticidade, dependência e projeção de estados desejados, mostra-se particularmente produtivo na formulação de pedidos e ordens atenuadas, sobretudo em sentenças complexas (*Se você pode comer doce de leite, você deve comê-lo todos os dias*). Nesses casos, a força diretiva do enunciado é deslocada do plano da imposição para o da negociação interpessoal (Castilho, 2010).

Assim, no Português Brasileiro, a polidez manifesta-se como atitude que referencia o *modus*, permitindo ao falante modular a força ilocucionária dos enunciados, diferenciando uma ordem de um pedido, por exemplo. Nessa perspectiva, Castilho (2010) afirma que os atos de fala resultam da interação entre forma sintática, prosódia e contexto, o que reforça o papel central do *modus* no funcionamento pragmático da língua. Destaca-se que a prosódia, especialmente a entoação, desempenha papel central nesse processo, orientando a interpretação do enunciado e facilitando a identificação de um ato de fala como ordem ou pedido, ou ainda como outros atos (convite, súplica, incentivo).

Portanto, a partir dos trabalhos aqui reportados, como contribuição principal a teoria dos atos de fala estabelece que, do ponto de vista pragmático, a unidade de análise da linguagem é o ato de fala, e não a sentença isolada. Nesse contexto de análise, a prosódia atua como uma ferramenta crucial capaz de sinalizar a força ilocucionária do ato de fala e a atitude do falante frente ao seu dizer e ao ouvinte, ou seja, contribui para a sinalização do *modus* nas palavras de Castilho (2010).

Assumir que a prosódia marca a força ilocucionária e atua como recursos de diferenciação que permite a identificação de atos de fala implica entrar no âmbito das funções desempenhadas pela prosódia. No interior dos trabalhos realizados por pesquisadores especialistas em prosódia, Cole (2015) aprofunda a abordagem das funções desempenhadas pela prosódia ao demonstrar detalhadamente como as funções marcadas pela prosódia, associada a diferentes domínios estruturais de organização da língua, marcam tanto contrastes

estritamente linguísticos quanto aspectos mais pragmáticos e situacionais, sinalizando a relação entre o falante enquanto participante de uma situação comunicativa e a manifestação sonora de seu dizer. Também Fónagy (1991, 2001, 2003) destaca que os elementos prosódicos constituem recursos fundamentais para a expressão de atitudes, emoções e intenções do falante, contribuindo decisivamente para a interpretação pragmática dos enunciados. É sobre esses trabalhos – de relevância notória para a caracterização das funções prosódicas – que tratamos a seguir.

1.3.3 As funções prosódicas

1.3.3.1 A perspectiva de Fónagy (1991, 2001, 2003)

Na obra *La modulation de la parole* (1991), Fónagy realiza uma análise dos elementos prosódicos, chamados por ele de elementos suprasegmentais da fala, como entonação, ritmo, intensidade e pausas, enfatizando seu papel na codificação de atitudes, emoções e intenções do falante. O autor parte do princípio de que a linguagem não se reduz à proposição semântica das palavras: a forma sonora do enunciado constitui um recurso comunicativo essencial, capaz de orientar a interpretação pragmática do ouvinte. Para sistematizar essa função, Fónagy (1991) distingue três dimensões principais da modulação da fala: a expressiva, relacionada à manifestação de estados afetivos (como surpresa ou raiva); a atitudinal, que sinaliza a postura do falante diante do interlocutor (como dúvida ou ironia); e a ilocucionária, que conecta a forma sonora à função pragmática do ato de fala (como ordenar ou prometer), demonstrando que a prosódia é constitutiva dos sentidos produzidos nas situações de comunicação.

Essa concepção é aprofundada pelo conceito de dupla codificação (Fónagy, 2001), segundo o qual a prosódia opera como um sistema híbrido. De um lado, atua como um código linguístico arbitrário e convencional, sujeito às regras da língua; de outro, manifesta-se como um “gesto vocal” expressivo e natural, que espelha estados psicofísicos do locutor. Nessa perspectiva, a força ilocucionária de um enunciado nasce da tensão dialética entre o código linguístico e o código gestual analógico. Enquanto o primeiro assegura a codificação do conteúdo proposicional, o segundo se realiza por meio de modulações prosódicas que funcionam como índices atitudinais. O gesto vocal configura, assim, uma “linguagem dentro da linguagem”, fornecendo o componente expressivo necessário para que o ouvinte infira a intenção do locutor, preenchendo lacunas pragmáticas que a gramática estritamente convencional não capta isoladamente. Dessa forma, considerando a perspectiva do autor, o

significado de um ato de fala emerge do diálogo contínuo entre a forma linguística e a expressão prosódica, permitindo ao ouvinte avaliar simultaneamente o conteúdo proposicional e a intenção pragmática.

Em Fónagy (2003), numa tentativa de organizar em categorias a atuação da prosódia no funcionamento da linguagem, o autor elenca e caracteriza diferentes funções que ele atribui à entoação. São elas: (i) a função demarcadora, responsável por separar unidades do discurso e por organizar o enunciado linguístico em partes; (ii) a função culminativa ou enfática, responsável por destacar uma parte específica do enunciado; (iii) a função sintática, responsável por atuar na desambiguação de enunciados; (iv) a função modal, que determina a modalidade dos enunciados, diferenciando enunciados declarativos de enunciados interrogativos, por exemplo; (v) a função preditiva, responsável por antecipar o conteúdo do discurso; (vi) a função expressiva, relacionada à sinalização de emoções e atitudes do falante; (vii) e a função indicial, que oferece pistas sobre a identidade dos falantes e dos gêneros discursivos em que a fala se manifesta.

Nesses trabalhos, o autor se concentra em apresentar uma abordagem fonética dos parâmetros prosódicos como gestos vocais que conferem expressividade à linguagem e que permitem organizar um conjunto extenso de funções desempenhadas particularmente pela entoação. Por sua vez, Cole (2015) trata das funções prosódicas e as organiza em três eixos principais a partir de abordagem teórica que relaciona estrutura prosódica e parâmetros fonéticos, como é apresentado a seguir.

1.3.3.2 A perspectiva de Cole (2015)

Cole (2015) analisa as funções da prosódia a partir de uma perspectiva fonológica. A autora organiza essas funções com base na noção de "contexto", que pode se referir desde ao contexto estrutural no âmbito lexical até o contexto situacional mais amplo. Sua proposta de organização das funções desempenhadas pela prosódia se alinha ao modelo de Nespor e Vogel (1986), assumindo que a prosódia tem sua base na estrutura fonológica. Segundo a autora, os efeitos do contexto nos parâmetros prosódicos, como a proeminência de certas unidades e a marcação de fronteiras, são mediadas pelas estruturas prosódicas que se organizam hierarquicamente⁷.

Assumindo que a organização prosódica da língua é hierárquica, Cole (2015)

⁷ A hierarquia prosódica será abordada na próxima seção.

propõe subdivisões específicas para cada função atribuída à prosódia. Embora todas essas funções estejam relacionadas a informações contextuais, a autora enfatiza que elas interagem com diferentes planos linguísticos, o que justifica sua distinção e a atualização da noção de contexto aplicada a cada função. Além disso, essa interação explica por que, em termos de estruturação prosódica, as diversas funções se associam a níveis distintos da hierarquia prosódica, refletindo sua complexidade e multifuncionalidade na organização do discurso.

A primeira função refere-se ao papel da prosódia na sinalização da estrutura linguística. Nesse caso, a prosódia atua demarcando o contexto estrutural constitutivo do enunciado linguístico, seja em nível lexical, sintático ou discursivo. Cole (2015) esclarece que essa demarcação se concretiza por meio da sinalização de fronteiras entre unidades linguísticas e pela indicação dos pontos de maior proeminência acentual no interior dessas unidades, permitindo que pistas prosódicas sinalizem tanto limites de palavras e sintagmas quanto unidades de discurso, como os tópicos discursivos, e, assim, facilitem o processamento linguístico do sinal de fala pelos ouvintes.

A segunda função está relacionada à sinalização do significado discursivo. Aqui, a prosódia assume um papel interpretativo ao refletir informações derivadas do contexto pragmático no qual os interlocutores estão inseridos. Em outras palavras, a forma prosódica do enunciado contribui para a construção de significados ao marcar aspectos como a força ilocucionária de um ato de fala, a presença de foco informacional em determinados constituintes, bem como atitudes e estados afetivos do falante. Nesse âmbito, a prosódia não se restringe a um recurso de organização estrutural, mas assume o papel de marcar sentidos que se definem pragmaticamente, afetando de forma direta a interpretação comunicativa e orientando a compreensão do ouvinte sobre intenções, ênfases e posicionamentos do falante (Cole, 2015).

A terceira função diz respeito à sinalização de informações situacionais, na medida em que, de acordo com a autora, a prosódia também fornece pistas sobre o contexto imediato de produção da fala. Entre os aspectos que podem ser sinalizados pela prosódia nessa função encontram-se características da identidade do falante, como sua origem geográfica ou perfil sociolinguístico, além de informações sobre a dinâmica da produção da fala, por exemplo, se esta é mais cuidadosa, espontânea, lida, formal ou coloquial. Esse conjunto de elementos prosódicos possibilita identificar não apenas nuances estilísticas, mas também um estilo situacional de fala, que se relaciona diretamente às circunstâncias específicas em que o enunciado é produzido.

Cole (2015) ainda enfatiza que a prosódia atua de forma articulada nos níveis lexical, sintático e discursivo, o que permite compreendê-la como uma dimensão capaz de

assumir simultaneamente funções distintas em diferentes planos da estrutura linguística.

Quanto à função de sinalização dos significados pragmáticos, Cole (2015) explica que a prosódia pode produzir diferentes efeitos semânticos ao longo da interação. Esses efeitos se manifestam tanto na proeminência prosódica, que assinala foco e estrutura informacional do enunciado, quanto na marcação da força ilocucionária, responsável por distinguir intenções comunicativas, além da codificação de aspectos afetivos, por meio da qual o falante expressa emoções, atitudes e estados subjetivos.

Especificamente, no que se refere à marcação da força ilocucionária, que interessa particularmente ao presente trabalho, Cole (2015) argumenta que a ilocução de um enunciado é determinada pela articulação de acentos de tom sobre constituintes proeminentes e de fronteiras prosódicas em nível frasal. Essa combinação resulta em um contorno entoacional que confere ao enunciado uma significação semântico-pragmática específica.

Conceitualmente, a força ilocucionária é originária da Teoria dos Atos de Fala de Austin (1962, 1990), apresentada na sessão anterior, a qual postula que a fala não apenas descreve ou representa o mundo, mas constitui, por si só, a realização de uma ação. Como já afirmado, esse princípio eleva o ato de fala à unidade de análise primária, em detrimento da sentença. Relacionando o fato apresentado por Cole (2015) de que a prosódia sinaliza a força ilocucionária de um enunciado com a Teoria dos Atos de Fala, tem-se que a prosódia atua na caracterização formal sonora de um ato de fala e, assim, contribui para que ouvintes de uma língua reconheçam tais atos pelos padrões prosódicos que neles se identificam.

Portanto, a exposição da proposta de Cole (2015) sobre as funções da prosódia demonstra que, para a autora, a prosódia e o “contexto” estão intrinsecamente ligados. O contexto, em suas dimensões estrutural, pragmática e situacional, exerce influência na percepção da fala. Essa influência se manifesta nas variações prosódicas, que produzem efeitos perceptivos e comunicativos responsáveis por modular a interpretação e a significação dos enunciados.

Reitera-se, por fim, que, na perspectiva de Cole (2015), as estruturas prosódicas são mediadoras da manifestação das diferentes funções da prosódia na linguagem. A autora parte do pressuposto de que a prosódia interage com diferentes dimensões linguísticas – estruturais, pragmáticas e situacionais, justamente porque a organização prosódica de uma língua é complexa e hierárquica. Ao assumir a existência dessa hierarquia, Cole (2015) adota os princípios defendidos pela Fonologia Prosódica (Nespor e Vogel, 1986, 2007), ao conceber que os constituintes prosódicos se organizam em níveis estruturados e interdependentes, responsáveis por articular e delimitar as unidades fonológicas, sintáticas e discursivas. Assim,

compreende-se que a prosódia, além de desempenhar funções comunicativas e estruturais, funciona como um ponto de convergência entre os diferentes níveis de organização linguística, conforme possibilita conceber o modelo teórico da Fonologia Prosódica, que será abordado a próxima seção.

Antes de passar a apresentação da Fonologia Prosódica, vale destacar que, comparativamente, embora Cole (2015) apresente uma análise detalhada das funções prosódicas e de sua relação com a estrutura prosódica, sua abordagem não estabelece distinções explícitas entre forças ilocucionárias dos atos de fala. Por essa razão, a perspectiva de Fónagy (1991, 2001) oferece recursos teóricos que favorecem compreender a interrelação entre funcionamento linguístico e modulação prosódica, enfatizando como a modulação vocal codifica simultaneamente nuances expressivas de interação entre atitude e força ilocucionária.

A seguir será apresentado o modelo teórico da Fonologia Prosódica. Nesse quadro teórico, conforme assumido por Cole (2015), a prosódia é concebida como um conjunto de unidades fonológicas organizadas em níveis hierárquicos, como o sintagma fonológico e o sintagma entoacional. Tais unidades estabelecem a arquitetura rítmica responsável pela organização do fluxo da fala.

1.3.4 Fonologia Prosódica

A Fonologia Prosódica, conforme formulada por Nespor e Vogel (1986, 2007), entende a prosódia como um componente do sistema fonológico responsável por organizar a língua em sua manifestação falada. Nessa perspectiva, pressupõe-se a existência de uma hierarquia prosódica própria de cada língua, formada por diferentes constituintes delimitados por fronteiras específicas e caracterizados pela atribuição de proeminências em distintos níveis da estrutura prosódica.

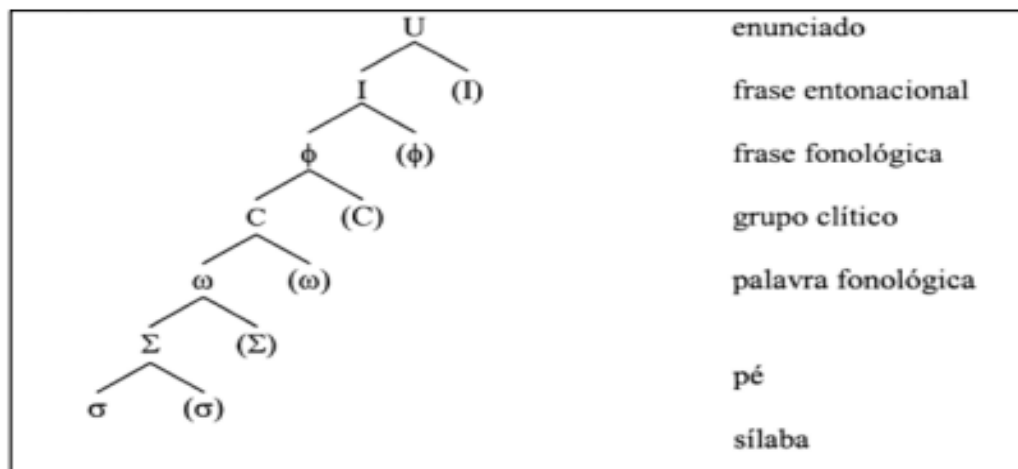
De acordo com Nespor e Vogel (1986), a fala não pode ser descrita apenas em termos de contrastes sonoros segmentais ou de estruturas sintáticas, mas deve incluir níveis intermediários de organização prosódica. Nessa perspectiva, a Fonologia Prosódica considera que os enunciados linguísticos se estruturam em unidades prosódicas hierarquicamente organizadas, que permitem a explicação de diversos padrões fonológicos e rítmicos da fala. Segundo Arvaniti (2020), essas unidades são fundamentais para a organização e a manifestação de fenômenos prosódicos inter-relacionados, tais como acento, ritmo, entoação e fraseamento, que operam conjuntamente para estruturar o fluxo da fala e sinalizar propriedades prosódicas importantes para a interpretação linguística. Processos fonológicos específicos, como a

assimilação e o sândi vocálico, também ocorrem dentro dessa organização hierárquica, mostrando que a estrutura prosódica fornece a base formal necessária para a interação entre aspectos segmentais e suprasegmentais da fala.

Especificamente, essa proposta teórica parte do princípio de que existe uma estrutura hierárquica prosódica subjacente à realização fônica dos fenômenos prosódicos na cadeia da fala. Como explicam Soncin e Tenani (2016), essa estrutura abstrata confere às línguas regularidades e previsibilidades na organização dos sons, permitindo mapear quais regras fonológicas podem ser aplicadas ou bloqueadas no interior e entre limites de certos constituintes prosódicos, assim como mapear em quais posições ocorrem proeminências acentuais e eventos de natureza entoacional, os chamados eventos tonais.

A hierarquia prosódica do modelo de Nespor e Vogel (1986; 2007) é constituída por sete níveis, sendo do menor para o maior: sílaba (σ), pé (Σ), palavra fonológica (ω), grupo clítico (C), frase fonológica (ϕ), frase entoacional (I) e enunciado fonológico (U), conforme ilustra a figura 1 em tradução para o Português feita por Bisol (2005).

Figura 1 - Hierarquia de constituintes prosódicos Nespor e Vogel (1986)



Fonte: Bisol (2005, p. 244)

Cada um desses níveis da hierarquia funciona como domínio específico para a aplicação de regras fonológicas. Como destacam Nespor e Vogel (2007, p. 15): “A fonologia prosódica deve ser concebida como um componente independente da gramática, cuja estrutura não é derivada diretamente da sintaxe, mas estabelecida por meio de regras de mapeamento que correlacionam constituintes sintáticos e prosódicos”. Dessa forma, embora a hierarquia seja considerada pelas autoras como universal, a maneira como cada língua realiza seus constituintes pode variar, refletindo propriedades particulares de cada sistema linguístico (Nespor; Vogel,

1986; Selkirk, 1984). A relevância dos diferentes níveis da hierarquia para a descrição de regras fonológicas e para a caracterização rítmica e entoacional do português brasileiro foi atestada e discutida por diferentes trabalhos, dentre os quais se destacam: Frota e Vigário (2000), Tenani (2002), Bisol (2005), Fernandes (2007) e Serra (2009).

Os constituintes da hierarquia prosódica se diferenciam pelas informações de que necessitam para a sua formação e, de acordo com o modelo, quanto mais alto o nível da hierarquia maior a dependência de informações de outros componentes da gramática (como a morfologia, a sintaxe e a semântica). Porém, deve-se salientar que o modelo tem como princípio central que não há, necessariamente, uma correspondência exata entre um constituinte prosódico e uma estrutura de outro componente gramatical (cf. Silva et al., 2014; Soncin e Tenani, 2016). Por exemplo, não há isomorfia obrigatória entre uma frase entoacional e uma oração no plano sintático, assim como não necessariamente são correspondentes uma palavra fonológica e uma palavra do ponto de vista morfológico. É por essa razão que se chama a Fonologia Prosódica de um modelo de interface entre fonologia e outros componentes gramaticais, com predominância para a interface com a sintaxe. De forma geral, entende-se que os constituintes até o nível da palavra fonológica podem fazer interagir informações fonológicas e morfológicas. Enquanto os constituintes acima do nível da palavra fazem interagir informações fonológicas e sintáticas. O último nível – o enunciado fonológico – considera também informações semânticas e são previstos, ainda, aspectos da natureza da execução do enunciado, ou seja, de natureza pragmática.

Considerado o modelo teórico da Fonologia Prosódica, a caracterização prosódica dos atos de fala – alvos da investigação perceptual realizada no presente estudo – implica a mobilização de dois níveis da hierarquia prosódica: a frase entoacional e o enunciado fonológico, uma vez que os atos de fala se identificam por uma configuração entoacional e, com ela, significam na situação comunicativa.

A frase entoacional é formada por uma ou mais frases fonológicas e, na interação com a estrutura sintática, se configura, em geral, no nível da sentença. Em termos de informação fonológica, o contorno entoacional e a produção de pausas são pistas de fundamental importância para a formação desse constituinte, uma vez que a variação desses elementos pode alterar os limites das frases entoacionais na organização da fala (Soncin; Tenani, 2016). De acordo com Nespor e Vogel (1986), a regra básica para a formação de uma frase entoacional é que os seus limites coincidem com pontos na cadeia de fala onde pausas podem ocorrer e, ainda, onde se observam eventos tonais que delimitam suas fronteiras. Assim, essencialmente, a frase entoacional constitui o nível hierárquico onde se manifestam os padrões melódicos da fala,

como contornos de entoação, pausas e acentuação final.

O enunciado fonológico, por sua vez, diz respeito ao constituinte mais alto da hierarquia prosódica e, como tal, é formado por uma ou mais frases entoacionais, que se encontram no nível imediatamente inferior. Na formação de um enunciado fonológico, além da interação entre informação fonológica e informação sintática, informação de natureza semântica é necessária, uma vez que, segundo Nespor e Vogel (1986, 2007), a completude semântica é fator relevante para a identificação desse constituinte. Especificamente, no caso de enunciados fonológicos que são formados por mais de uma sentença no plano sintático, é crucial que haja uma relação semântica entre essas sentenças, podendo essa relação ser marcada por conectivos (como *e*, *mas*, *porque*, *portanto*, etc.), caso contrário, não se pode considerar a sequência fônica como um enunciado fonológico. Destaca-se, ainda, que as autoras preveem condições pragmáticas que devem ser atendidas para que um enunciado fonológico possa ser formado por mais de uma sentença, são elas: (a) as sentenças devem ser enunciadas por um mesmo falante; e (b) as sentenças devem ser direcionadas a um mesmo interlocutor (cf. Nespor e Vogel, 2007, p. 240).

Compreende-se, assim, nesta tese, que os atos de fala se constituem, em termos de constituição estrutural prosódica, em enunciados fonológicos, pois significam no interior de uma situação comunicativa, ou seja, apresentam completude semântica ao serem produzidos por um falante e direcionados a um (ou mais) interlocutores. A qualidade da significação produzida nesses enunciados fonológicos, que permite identificar o ato de fala em si, ou seja, sua *força ilocucionária*, é definida pela entoação, que, por sua vez, se configura no domínio que compõe fonologicamente o enunciado num nível imediatamente inferior: a frase entoacional.

Uma vez que a frase entoacional está diretamente ligada à configuração melódica dos enunciados, é essencial que sua análise leve em conta a Fonologia Entoacional, cujos princípios de base foram apresentados por Pierrehumbert (1980) e, posteriormente, foram teorizados por Ladd (1996/2008), de forma associada à hierarquia prosódica. A seguir, trata-se dessa proposta teórica de descrição entoacional de modo a explicitar como diferentes enunciados podem ser identificados por sua configuração entoacional.

1.3.5 Fonologia entoacional

A Fonologia Entoacional trata fonologicamente da estrutura da entoação e de sua implementação fonética e a considera como um fenômeno prosódico autônomo e

linguisticamente estruturado (Silva et al., 2014; Arvaniti, 2022). Isso significa que a melodia da fala não é um fenômeno aleatório, mas possui uma organização própria, ou seja, funciona de forma autônoma em relação a outros fenômenos prosódicos e é descrita na análise acústica numa camada independente da camada segmental, embora a ela esteja associada. Nesse sentido, Frota e Vigário (2000) defendem que as características da entoação são uma das pistas mais importantes, dentre outras existentes, para a compreensão da estrutura prosódica de uma língua.

O sistema entoacional é composto por elementos básicos, os tons, unidades contrastivas que formam os contornos melódicos da fala (Silva et al., 2014). De acordo com os princípios teóricos da Fonologia Entoacional (Ladd, 1996/2008), a frequência fundamental (F0), considerada como correlato acústico da entoação, pode ser entendida como uma sequência de eventos fonológicos discretos. Conforme argumenta Gussenhoven (2004), a F0 é o traço mais significativo na determinação do padrão entoacional de um enunciado. As modulações de F0 são percebidas pelos ouvintes como variações de altura melódica, uma propriedade perceptual auditiva chamada *pitch*.

Pierrehumbert (1980) lança os princípios fundamentais dessa proposta teórica com base nos estudos de Liberman (1975) ao apresentar uma análise linguisticamente estruturada da entoação, propondo uma representação fonológica baseada em tons altos (H – High) e baixos (L – Low), responsáveis por gerar contrastes no inglês. Além disso, estabelece regras para o mapeamento das representações fonológicas tonais no nível fonético da entoação, explicando como os movimentos melódicos abstratos se realizam na fala (Miranda, 2019). O trabalho pioneiro de Pierrehumbert (1980) foi central para que, na análise entoacional desenvolvida à luz de uma perspectiva fonológica, se instaurasse a compreensão conceitual da entoação como “modulação de F0 linguisticamente estruturada e pragmaticamente significativa”, conforme explicita Arvaniti (2022, p. 25). Em termos de marco teórico, a teorização da proposta como modelo foi apresentada por Ladd, em 1996, como a publicação do livro *Intonational Phonology*.

No modelo teórico, chamado de Fonologia Entoacional Autossegmental e Métrica ou de Modelo Autossegmental e Métrico da Entoação (Ladd, 1996/2008), a partir dos tons básicos, H e L, são identificados diferentes tipos de eventos tonais que descrevem as variações entoacionais: os acentos tonais, os tons de fronteira e os acentos frasais. Os acentos tonais estão associados a níveis mais baixos da hierarquia prosódica, como a palavra ou a frase fonológica, podendo ser simples — representados por H* ou L* (onde o asterisco indica a sílaba tônica) — ou complexos, formados pela combinação de tons, como H+L*, L+H*, entre outros. Já os tons de fronteira assinalam os limites de constituintes prosódicos mais altos, como a frase entoacional ou o enunciado fonológico, sendo representados por H% ou L%. Por fim, os acentos

frasais (H- ou L-) marcam fronteiras intermediárias dentro do contorno prosódico. Assim, a Fonologia Entoacional permite não apenas caracterizar as unidades tonais que compõem os enunciados, mas também compreender como elas são foneticamente realizadas no fluxo da fala (Silva et al., 2014). No português, os padrões de destaque frasal são predominantemente de tipo românico. Isso significa que, por padrão, a proeminência frasal recai sobre o elemento mais à direita de uma frase entoacional. Assim, a palavra prosódica final é considerada a cabeça (ou seja, a mais proeminente) desse constituinte, e é nela que se localiza o acento nuclear (Frota et al., 2015).

A principal contribuição da Fonologia Entoacional (Pierrehumbert, 1980; Ladd, 1996/2008) consiste em demonstrar que os movimentos melódicos relevantes de um contorno entoacional estão associados a posições métricas fortes, como a sílaba tônica. Nesse sentido, o modelo permite mapear de que forma as variações de F0, ascensões e descidas se realizam em sílabas acentuadas, possibilitando a descrição dos contrastes fonológicos entre diferentes contornos entoacionais (Miranda, 2019). A partir da proposta inicial de Pierrehumbert (1980), desenvolvimentos posteriores a endossaram e ofereceram ferramentas que facilitaram o poder descritivo da Fonologia Entoacional, como é o caso do sistema de notação ToBI (*Tones and Break Indices*), de Silverman et al. (1992), desenvolvido inicialmente como uma aplicação notacional da teoria para a descrição entoacional do inglês, o qual foi posteriormente aplicado a outras línguas, como o Português. Em publicações mais recentes, Arvaniti (2020) e Arvaniti e Fletcher (2018) apresentam as versões mais atualizadas do modelo teórico que, a despeito de manter as mesmas bases teóricas cunhadas por Pierrehumbert (1980) e desenvolvidas por Ladd (1996/2008), contam com refinamentos obtidos a partir de pesquisas experimentais e de análises envolvendo um grande número de línguas.

A partir da perspectiva da Fonologia Entoacional (Pierrehumbert, 1980; Ladd, 1996, 2008), em sua aplicação ao sistema ToBI (*Tones and Break Indices*), Frota et al. (2015) descrevem a entoação do português europeu e do português brasileiro com base em um inventário de acentos tonais, tons de fronteira e padrões de fraseamento prosódico. Essa abordagem permite representar formalmente os contornos entoacionais observados em diferentes tipos de sentenças, assegurando comparabilidade entre variedades e situando o português no quadro tipológico das línguas românicas. O estudo, desenvolvido no âmbito do projeto *Interactive Atlas of the Prosody of Portuguese* (InAPoP), constitui ainda uma contribuição decisiva para a proposta do *Portuguese_ToBI*, primeira tentativa de sistematizar, em um modelo unificado, a anotação entoacional das variedades do português.

Os resultados apresentados por Frota et al. (2015) revelam um panorama rico em

semelhanças e contrastes entre o português europeu e o português brasileiro, permitindo compreender com maior precisão como cada variedade organiza entoacionalmente os diferentes tipos de enunciado. Nas declarações neutras, ambas as variedades convergem na presença de um tom de fronteira baixo, que confere caráter conclusivo ao enunciado. Contudo, divergem quanto à escolha do acento nuclear: no português europeu predomina o padrão H+L* L%, caracterizado por um movimento descendente no núcleo seguido de queda final, enquanto no português brasileiro é recorrente o padrão L+H* L%, em que há ascensão tonal no núcleo antes da queda final. Essa diferença não é apenas formal, mas também perceptual: enquanto no português europeu a descida reforça a assertividade da declaração, no português brasileiro o contorno ascendente confere uma melodia mais marcada e contribui para a impressão de maior dinamismo entoacional (Frota et al., 2015).

Nas interrogativas globais, também chamadas de perguntas sim/não, as diferenças entre as duas variedades tornam-se ainda mais salientes. O português brasileiro apresenta de forma estável o contorno L+H* L%, no qual a subida iniciada no acento nuclear se prolonga até a fronteira final, resultando em uma curva ascendente clara e uniforme, que é interpretada como marca inequívoca da interrogatividade. No português europeu, ao contrário, não há um padrão único e estável: as perguntas sim/não podem ser realizadas com contornos descendentes associados a fronteiras complexas, como HL% ou LH%, que combinam quedas e subidas finais, produzindo efeitos entoacionais mais variáveis e contextualmente dependentes. Essa diferença é um dos traços mais distintivos entre português europeu e português brasileiro, frequentemente percebida pelos próprios falantes na comparação entre as variedades (Frota et al., 2015).

As interrogativas ou perguntas parciais, chamadas no inglês de *wh-questions*, por se definirem pela presença de pronome interrogativo, apresentam maior convergência entre as duas variedades. Tanto no português europeu quanto no português brasileiro, o padrão nuclear típico é H+L* L%, marcado por um movimento descendente no acento nuclear seguido de queda final. A semelhança estrutural indica uma base prosódica comum na codificação desse tipo de interrogativa, embora se observem diferenças de natureza fonética: no português brasileiro, tende a haver maior amplitude melódica e maior extensão do movimento descendente, enquanto no português europeu a realização é geralmente mais contida (Frota et al., 2015).

No caso de atos de fala diretivos, ambos os dialetos recorrem ao contorno H+L* L% em contextos de ordem direta, garantindo a expressividade necessária a enunciados de caráter impositivo. Entretanto, quando se trata de atenuar a força ilocucionária, como no caso de pedidos, as variedades se afastam: no português brasileiro é frequente a ocorrência do

contorno L+H* H%, que aproxima pedidos e sugestões do padrão interrogativo e suaviza a carga impositiva do ato de fala. No português europeu, por sua vez, essa variação é muito menos produtiva, e mesmo em contextos atenuados tende a prevalecer o contorno descendente, revelando um sistema mais conservador e menos flexível na expressão entoacional de nuances pragmáticas (Frota et al., 2015).

Os vocativos constituem o domínio em que se observa maior semelhança entre português europeu e português brasileiro. Em ambas as variedades, a chamada de saudação costuma ser realizada com o contorno (L+)H* L%, em que a subida nuclear é seguida de queda, conferindo efeito de fechamento típico de uma interação breve e amistosa. Já as chamadas insistentes recorrem ao contorno (L+)H* H%, em que a subida é prolongada até a fronteira final, criando a sensação de expectativa aberta e insistência na convocação do interlocutor. Destaca-se que, nos vocativos, o movimento de subida pode não ocorrer em casos em que a configuração do acento tonal iniciar diretamente com tom alto, motivo que justifica a inserção de parênteses no tom L na notação dos acentos nucleares que caracterizam os vocativos (Frota et al., 2015).

Em síntese, a descrição de Frota et al. (2015) demonstra que, embora o português europeu e o português brasileiro compartilhem uma base prosódica comum, essas variedades do português diferem-se quanto à configuração entoacional de alguns atos de fala.

A descrição entoacional desenvolvida por Frota et al. (2015) mostra, de forma detalhada, como o português brasileiro organiza prosodicamente, no nível do enunciado fonológico, diferentes atos de fala e quais recursos entoacionais são mobilizados na codificação de contrastes estruturais e pragmáticos. Esses resultados reforçam a centralidade da prosódia para a percepção da fala, especialmente no que se refere à marcação da força ilocucionária em atos de fala como declarações, perguntas, ordens e pedidos.

No entanto, embora o inventário prosódico do português brasileiro já esteja sistematicamente descrito, permanece uma lacuna importante: ainda não se sabe como esses contornos são processados por crianças em fase de aquisição, nem em que medida a integração indissociável de pistas auditivas e visuais molda essa percepção. Observa-se, portanto, uma ausência de investigações sob a perspectiva multimodal sobre como as crianças percebem a força ilocucionária dos atos de fala. Consequentemente, faltam estudos que avaliem o papel da pista visual ao longo da aquisição prosódica do português brasileiro.

O presente trabalho se justifica, assim, pela necessidade de levantar dados de referência sobre o desempenho perceptual prosódico de crianças brasileiras e pela relevância de mapear as mudanças que ocorrem ao longo da aquisição no que diz respeito à integração de

informações auditivas e visuais na interpretação da prosódia dos atos de fala.

Inserindo-se nesse contexto de investigação e com o propósito de ampliar os estudos sobre aquisição da linguagem centrados na descrição da aquisição da prosódia, o presente trabalho propõe uma investigação experimental voltada a examinar o efeito de pistas auditivas e visuais na percepção de atos de fala diferenciados prosodicamente em crianças falantes do português brasileiro. Como antecipamos na apresentação desta tese, a pesquisa busca contribuir para a compreensão da aquisição prosódica dos diferentes tipos de atos de fala e, ao mesmo tempo, fornecer subsídios relevantes para a prática fonoaudiológica, seja no âmbito da avaliação, ao disponibilizar informações normativas sobre a identificação perceptual audiovisual dos atos de fala, seja no campo da intervenção, considerando que crianças com processos de aquisição de linguagem desviantes podem apresentar comprometimentos também no plano prosódico. Os objetivos específicos desta investigação são apresentados na próxima seção.

2 OBJETIVOS E HIPÓTESES

2.1 OBJETIVOS

Esta tese teve como objetivo geral descrever o desempenho perceptual multimodal de crianças falantes do Português Brasileiro ao longo do processo de aquisição da prosódia na identificação de atos de fala diferenciados prosodicamente: declaração, pergunta, ordem e pedido.

Para atender tal objetivo, foram delineados três objetivos específicos, conforme mencionados na introdução deste trabalho:

- 1) Comparar a acurácia perceptual de diferentes grupos etários na identificação de atos de fala diferenciados prosodicamente por pistas auditivas e visuais;
- 2) Avaliar o efeito de pistas prosódicas auditivas e visuais na percepção dos diferentes atos de fala pelos grupos etários;
- 3) Verificar se, em condição de estímulos com ruído, as pistas visuais exerceriam um efeito compensatório na identificação dos diferentes atos de fala pelos grupos etários.

Os objetivos ora apresentados foram formulados à luz do estado da arte, tanto da literatura internacional quanto da nacional, bem como das lacunas identificadas nos estudos sobre a aquisição prosódica do Português Brasileiro. Em articulação com esses objetivos e em consonância com os resultados discutidos na revisão apresentada, foram estabelecidas as hipóteses que orientam o desenvolvimento desta tese, as quais são explicitadas a seguir.

2.2 HIPÓTESES

Com base nos estudos anteriores (cf. Miranda et al., 2020; Miranda et al., 2021; Carnaval, 2021; Carnaval et al., 2022; Carnaval et al., 2023), realizados com adultos falantes do Português Brasileiro e na assunção de que a aquisição de linguagem é processo gradual, não linear e heterogêneo, particularmente no que diz respeito ao desenvolvimento das competências pragmáticas e discursivas (Tomasello, 2003; Snow, 1995), assumimos as seguintes hipóteses:

- (1) No que diz respeito ao primeiro objetivo específico, relativo à comparação da

acurácia perceptual de diferentes grupos etários⁸ na identificação de atos de fala diferenciados prosodicamente por pistas auditivas e visuais, assumiu-se como hipótese que o desempenho dos grupos na identificação dos atos de fala seria (i) diferente a depender do ato de fala apresentado (Miranda; Moraes e Rilliard, 2020; Carnaval; Miranda e Souza, 2022) e (ii) impactado pela idade. Ou seja, espera-se que grupos de maior faixa etária apresentem desempenho perceptual mais acurado em relação aos grupos de crianças mais novas, ainda que a acurácia perceptual desses grupos possa ser distinta a depender da complexidade envolvida nos diferentes atos de fala (Miranda, Moraes e Rilliard 2020; Miranda et al., 2024; Frota et al., 2016; Sousa et al., 2020).

(2) No que diz respeito ao segundo objetivo específico, o qual busca avaliar o efeito de pistas prosódicas auditivas e visuais na percepção dos diferentes atos de fala pelos grupos etários; assumiu-se como hipóteses que (i) a modalidade audiovisual congruente⁹ favoreceria maior acurácia dos grupos na identificação dos atos de fala em relação às modalidades unimodais (somente auditiva ou somente visual) (Miranda et al., 2020; Miranda et al., 2024) e (ii) que os diferentes grupos etários apresentariam preferências perceptuais distintas na tomada de decisão quanto à identificação dos atos de fala quando expostos a estímulos audiovisuais incongruentes¹⁰ (Miranda et al., 2024).

Além dessas hipóteses gerais, estabeleceram-se expectativas específicas quanto à direção dessas diferenças, fundamentadas na literatura prévia: espera-se que grupos etários mais novos apresentem preferência perceptual por pistas auditivas (Miranda et al., 2024; Souza, 2018), enquanto grupos etários mais velhos apresentem preferência perceptual por pistas visuais (Miranda et al., 2021). Desse modo, enquanto as hipóteses formuladas investigam a existência de efeitos da modalidade e da idade na identificação dos atos de fala, as expectativas explicitam a direção prevista desses efeitos com base nos resultados de estudos anteriores.

(3) No que diz respeito ao terceiro objetivo específico, relativo à verificação se, em condição de estímulos com ruído, as pistas visuais exerceriam um efeito compensatório na identificação dos diferentes atos de fala pelos grupos etários, assumiu-se como hipótese que as

⁸ A inclusão de um grupo de adultos como controle visa estabelecer o padrão do sistema linguístico estabilizado, seguindo protocolos de percepção multimodal no PB (cf. Miranda; Moraes; Rilliard, 2020). Essa abordagem supre a ausência de parâmetros comparativos adultos em pesquisas de aquisição voltadas estritamente ao público infantil (cf. Sousa et al., 2020).

⁹ Modalidade audiovisual congruente é a condição em que as informações auditivas e visuais coincidem e transmitem o mesmo tipo de informação prosódica, de modo que ambas as pistas se reforçam mutuamente e favorecem a interpretação do enunciado pelo ouvinte (cf. Barkhuysen; Krahmer; Swerts, 2020; Miranda et al., 2020).

¹⁰ Condição experimental que permite verificar qual modalidade sensorial — auditiva ou visual — exerce maior peso na interpretação do enunciado (Barkhuysen; Krahmer; Swerts, 2020; Miranda et al., 2020).

pistas visuais poderiam exercer maior efeito em condições de estímulos com ruído, exercendo, assim, um efeito compensatório (Miranda et al., 2024; Sumbly; Pollack, 1954; Miranda et al., 2020; Miranda et al., 2021; Barkhuysen, Krahmer; Swerts, 2010).

3 MATERIAL E MÉTODO

3.1 DESCRIÇÃO METODOLÓGICA

A pesquisa realizada foi delineada como um estudo experimental, transversal, quantitativo e prospectivo. O processo de amostragem foi não aleatório por conveniência de modo a formar amostra de 90 crianças e 20 adultos, com total de 110 participantes de ambos os gêneros, organizados em grupos de diferentes faixas etárias.

A pesquisa foi realizada no Laboratório de Análise Articulatória e Acústica (LAAc) na UNESP, câmpus de Marília, e está vinculada ao desenvolvimento do projeto de pesquisa intitulado “Produção e percepção da prosódia em crianças com desenvolvimento fonológico típico e atípico” (FAPESP 2020/10144-3), coordenado pela orientadora deste trabalho.

3.2 ASPECTOS ÉTICOS

A proposta foi encaminhada para apreciação do Comitê de Ética em Pesquisa (CEP) da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) e aprovada sob parecer nº 6.018.084. Os responsáveis por todas as crianças participantes da pesquisa assinaram o Termo de Consentimento Livre e Esclarecido, assim como as crianças concordaram em participar voluntariamente da pesquisa por meio do Termo de Assentimento. Conforme protocolo ético adotado pelo CEP, os procedimentos experimentais se iniciaram após a aprovação e recomendação do referido comitê.

3.3 DESIGN EXPERIMENTAL

Foi elaborado e desenvolvido um design experimental que consistiu em testar a identificação de atos de fala do PB diferenciados exclusivamente por pistas prosódicas de natureza auditiva e visual. Para tanto, delineou-se um experimento de percepção de fala caracterizado como um teste de identificação de quatro atos de fala. Os quatro tipos de atos de fala do português brasileiro alvos de identificação no experimento foram: declaração, pergunta, ordem e pedido. Esse conjunto de atos de fala foi selecionado por contemplar, de um lado, atos de natureza predominantemente linguística, como a declaração e a pergunta, cuja distinção se manifesta com padrões entoacionais distintivos que implicam gramaticalmente diferentes

modalidades de enunciado, e, de outro, atos de natureza sociopragmática, como a ordem e o pedido, que envolvem maiores nuances de expressividade e dependem de valores semântico-pragmáticos como hierarquia, polidez e negociação, conforme discutido por Castilho (2010) e por Pietroforte e Lopes (2014).

Após a apresentação de cada estímulo, os participantes respondiam se interpretavam o enunciado apresentado como declaração, pergunta, ordem ou pedido, assinalando uma dentre as quatro respostas disponíveis como opção. Os estímulos foram apresentados aleatoriamente em quatro diferentes modalidades¹¹ de apresentação: somente auditivo (AO), somente visual (VO), audiovisual congruente (AV+) e audiovisual incongruente (AV-). A descrição dos estímulos e a caracterização das modalidades de apresentação são feitas mais adiante.

3.4 AMOSTRA

Participaram do estudo falantes nativos do Português Brasileiro com desenvolvimento típico de linguagem, residentes de Ribeirão Preto/SP e de Marília/SP. Os participantes foram organizados em cinco grupos etários, destes cinco grupos etários da pesquisa, quatro deles (G1, G2, G3, G4) foram formados por crianças de diferentes idades, enquanto um grupo (G5) foi formado por adultos, com os quais os grupos de crianças foram comparados, uma vez que, no que diz respeito ao processo de aquisição de linguagem, a fala adulta é considerada o modelo alvo a ser alcançado em termos de padrão linguístico (Clark, 2009).

A divisão das crianças em grupos etários distintos encontra respaldo na literatura que concebe a aquisição da linguagem como um processo gradual, não linear e heterogêneo, particularmente no que diz respeito ao desenvolvimento das competências pragmáticas (Tomasello, 2003; Snow, 1995), assim como se assume neste trabalho. Esses estudos mostram que a compreensão de intenções comunicativas e de atos de fala emerge progressivamente ao longo da infância, o que justifica a comparação entre diferentes faixas etárias infantis, ainda que individualmente crianças no interior de um mesmo grupo possam apresentar ritmos distintos. A divisão dos participantes da pesquisa em grupos etários é apresentada no Quadro 1:

¹¹ O termo *modalidade* no contexto experimental é empregado para designar o(s) canal(is) sensorial(is), por meio dos quais o enunciado é apresentado aos participantes.

Quadro 1 – Amostra de sujeitos participantes da pesquisa

Grupos	Faixa etária	N
G1	4 a 5 anos e 11 meses	21
G2	6 a 7 anos e 11 meses	23
G3	8 a 9 anos e 11 meses	23
G4	10 a 12 anos e 11 meses	23
G5	18 a 25 anos	20
Total de participantes		110

Fonte: Dados da Pesquisa (2025).

Sequencialmente, dentre os participantes, 11 eram do gênero masculino e 10 do gênero feminino para o G1; 10 eram do gênero masculino e 13 do gênero feminino para o G2; 9 eram do gênero masculino e 14 do gênero feminino para o G3; 14 eram do gênero masculino e 9 do gênero feminino para o G4 e, por fim, 3 eram do gênero masculino e 17 do gênero feminino para o G5.

As crianças que formaram os grupos G1, G2, G3 e G4 foram recrutadas nas escolas EMEI Santa Maria Goreti e EMEF Prof. Eduardo Romualdo de Souza, na cidade de Ribeirão Preto/SP. Por sua vez, os adultos que formaram o grupo G5 foram recrutados do conjunto de discentes do curso de graduação em Fonoaudiologia da UNESP - Marília/SP.

Para os grupos de crianças, os experimentos foram aplicados nas escolas em salas disponibilizadas pela coordenação e, para o grupo de adultos, os experimentos foram aplicados na sala de gravação e análise de fala e voz do Laboratório de Análise Acústica - LAAC da UNESP/Marília. Todos os equipamentos necessários para a realização da pesquisa foram fornecidos pelo LAAC e utilizados nos locais de aplicação dos experimentos.

Para todos os grupos, foram adotados como critérios de inclusão a ausência de quaisquer queixas de problemas otológicos e/ou auditivos e de queixas de acuidade visual mediante preenchimento de questionário. Como critérios de exclusão, foram considerados quaisquer problemas neurológicos e/ou de linguagem detectados por triagem fonoaudiológica.

Após entrega dos termos de consentimento livre e esclarecido, as crianças passaram por triagem fonoaudiológica e foram avaliadas pela pesquisadora no que diz respeito à sua aquisição de linguagem, aquisição fonológica, leitura, escrita e interpretação de texto. Também

foi entregue um questionário aos pais, no qual responderam sobre as queixas auditivas e de acuidade visual das crianças. Ao todo, participaram da triagem 109 crianças e foram excluídas 19 por apresentarem queixas auditivas, uso de tubo de ventilação, alterações fonológicas na fala não esperadas para a idade e alterações na escrita não esperadas para a idade/ano escolar. Ainda, uma criança foi excluída por ser falante nativa do português europeu. Após a triagem fonoaudiológica, nos casos das crianças nas quais foram encontrados indícios de alterações de fala, escrita e/ou linguagem, foi entregue aos responsáveis um relatório com a descrição dos aspectos observados e uma orientação para que procurassem um profissional especialista.

3.5 PROCEDIMENTO EXPERIMENTAL

Para a aplicação do experimento, a pesquisadora conduzia cada participante individualmente desde sua sala de aula até o espaço destinado à coleta. Durante esse trajeto, eram fornecidas explicações à criança sobre a atividade proposta, incluindo a forma de apresentação dos estímulos e o modo como as respostas deveriam ser realizadas. Ao chegarem à sala da coleta, os sujeitos eram dispostos confortavelmente em frente a um computador com fones de ouvido acoplados, bem próximos à tela, e o procedimento experimental era iniciado.

Por meio do *software OpenSesame* (Mathôt, Schreij; Theeuwes, 2012), foram apresentados, de modo aleatório para cada participante, os estímulos representantes dos quatro atos de fala (cf. Figura 2) nas modalidades AO, VO, AV+ e AV-. Para cada apresentação dos estímulos, o teste solicitava, por meio de tarefa de escolha forçada¹², uma resposta dentre quatro opções: (1) *Declaração*, (2) *Pergunta*, (3) *Ordem* e (4) *Pedido*, que foram representadas pelas imagens da Figura 3. O registro das respostas foi realizado por meio das teclas numéricas do teclado do computador, correspondentes às alternativas de resposta (1–4), previamente definidas no script experimental.

¹² Tarefa experimental na qual o participante deve escolher uma resposta dentre um conjunto limitado e previamente definido de alternativas, permitindo o controle das variáveis perceptivas e a mensuração objetiva do desempenho (Macmillan, 2005; Srinivasan e Massaro, 2003).

Figura 2 – Modalidades de apresentação

Legenda: a ilustração mostra uma mulher jovem de cabelos curtos loiros, de pele branca, usando uma blusa branca, em fundo verde, representando as modalidades de apresentação dos estímulos.

Fonte: Elaboração própria (2025)

Figura 3 – Alternativas de respostas do experimento de percepção dos atos de fala

Legenda: a ilustração mostra as quatro opções de resposta declaração, pergunta, ordem e pedido, ilustradas por mulher branca, cabelo castanho, usando blusa vermelha e calça preta, realizando gestos representativos de cada enunciado em fundo branco

Fonte: Dados da Pesquisa. Imagens foram obtidas em busca no Google (2025).

O experimento foi aplicado em duas etapas: na primeira etapa, os estímulos foram apresentados na condição sem ruído; enquanto na segunda etapa, os estímulos foram apresentados na condição com ruído. Em ambas as etapas, os estímulos foram os mesmos quanto aos atos de fala e às modalidades de apresentação com a diferença de que, na segunda etapa, os estímulos nas modalidades AO, VO, AV+ e AV- foram apresentados com ruído.

Uma etapa de treino foi realizada antes da aplicação do experimento, de modo a familiarizar as crianças com a tarefa. Para tanto, foi feita a apresentação de 4 estímulos, um representante de cada modalidade, todos na condição sem ruído. As respostas na etapa de treino não foram registradas. Após a etapa treino, as crianças foram informadas de que suas respostas

seriam registradas, com o objetivo de favorecer maior atenção e engajamento, e, em seguida, o experimento foi iniciado oficialmente.

Ressalta-se que, uma vez que os experimentos foram realizados com crianças, o uso de ilustrações e de linguagem acessível para caracterizar as opções de resposta foi empregado. Além disso, foi solicitado às crianças menores que apontassem na tela do computador a figura para que a pesquisadora apertasse os botões correspondentes à alternativa escolhida. Já as crianças a partir de 7 anos não necessitaram de auxílio e realizaram a seleção da opção de resposta no computador com autonomia.

De modo geral, os participantes mantiveram-se atentos e engajados na tarefa, uma vez que a atividade foi apresentada como um “jogo no computador”, o que despertou o interesse das crianças. Entre elas, poucas crianças, especificamente as de 4 anos, apresentaram momentos pontuais de dispersão, sendo prontamente redirecionadas à tela do computador pela pesquisadora.

A modalidade incongruente (AV-) não foi previamente informada nem explicada pela pesquisadora, com o objetivo de não induzir as crianças a se apoiarem em uma pista específica, auditiva ou visual. Observou-se, nesse caso, um certo estranhamento por parte das crianças, refletido em um tempo ligeiramente maior para a tomada de decisão; ainda assim, elas respondiam apontando ou apertando as teclas, algumas ainda diziam “*acho que é essa alternativa*”. As crianças tiveram um breve intervalo entre as etapas, com duração aproximada de três minutos. Nesse momento, a pesquisadora explicou que, na etapa seguinte, correspondente à condição com ruído, haveria um som semelhante a “chuva” nos fones de ouvido e solicitou que mantivessem atenção redobrada.

A seguir descrevem-se detalhadamente os estímulos do experimento.

3.6 ESTÍMULOS

Durante a etapa de preparação do design experimental, foram elaborados os estímulos a serem utilizados no teste de percepção. Para tanto, uma atriz, falante nativa do interior de São Paulo, de idade adulta, foi convidada a interpretar a oração “*Vamos*” como declaração, pergunta, ordem e pedido. A oração “*Vamos*” foi selecionada após discussão com Prof. Dr. Marc Swerts, da *Tilburg University*, por se tratar de uma forma linguística do Português Brasileiro de mesma constituição fonológica segmental e composição sintática e que possibilitou a diferenciação enunciativa dos atos de fala alvos de investigação exclusivamente por suas características prosódicas. Ou seja, a diferenciação dos atos de fala na produção de

“vamos” se manifestou especificamente pelas características prosódicas e não pelo uso distinto de palavras e/ou de construções sintáticas.

Além disso, a escolha do enunciado prototípico “vamos” justificou-se pela necessidade de otimizar as condições experimentais da investigação perceptiva, uma vez que se trata de um item lexical curto, frequente e estruturalmente estável, permitindo o controle de variações léxico-sintáticas potencialmente distratoras. Assim, as diferenças interpretativas entre os atos de fala decorrem principalmente da modulação prosódica, gestual e semântico-pragmática do enunciado, e não de alterações formais, o que garante maior controle experimental e favorece a análise perceptiva das pistas auditivas e visuais (Swerts e Krahmer, 2008). Outra vantagem da utilização de enunciados idênticos reside na facilitação das manipulações audiovisuais, que consistiram na criação de estímulos audiovisuais incongruentes (Swerts e Krahmer, 2008).

Para a gravação dos estímulos, foram fornecidos à atriz contextos comunicativos para situar a produção dos distintos atos de fala. Na dinâmica adotada na gravação, de modo a fornecer esses contextos comunicativos e elicitar a produção dos atos de fala pela atriz, a pesquisadora dirigia um enunciado de fala à atriz, que, posicionada em frente à câmera, produzia “Vamos” como resposta ao enunciado a ela dirigido. A partir dessa dinâmica, a atriz produziu a forma linguística “Vamos” ora como declaração, ora como pergunta, ora como ordem ou como pedido em resposta ao enunciado produzido pela pesquisadora. O Quadro 2 apresenta os contextos oferecidos à atriz para a produção dos atos de fala usados como estímulo.

Quadro 2 – Descrição contextos usados na gravação dos estímulos para elicitar os atos de fala

Contexto discursivo	Enunciado dirigido à atriz	Resposta da Atriz	Ato de fala
Duas amigas conversam pelo telefone	Nós vamos ao cinema hoje à noite?	Vamos.	Declaração
Duas amigas conversam na saída da aula	Você não vai acreditar! Consegui os ingressos e nós vamos ao show da Madonna!	Vamos?	Pergunta com surpresa
Mãe e filha conversam antes de sair de casa	Mãe, eu não quero ir para aula hoje.	Vamos. (na posição de mãe)	Ordem
Mãe e filha conversam após chegada do trabalho	Filha, estou muito cansada, não vou mais ao cinema.	Vamos. (na posição de filha)	Pedido

Fonte: Dados da Pesquisa (2025).

Do Quadro 2, destaca-se que o ato de fala “pergunta” apresentado nos estímulos se configurou como *pergunta com surpresa* em decorrência do contexto oferecido à atriz para a produção do ato de fala. De acordo com Castilho (2010), a manifestação da surpresa em estruturas interrogativas atua como um modalizador intersubjetivo discursivo, cuja função

consiste em evidenciar o estado epistêmico e a reação emocional do locutor ao realizar a pergunta diante do interlocutor. Assim, a pergunta com surpresa situa-se numa posição limítrofe entre a gramática e a pragmática discursiva, configurando-se como um ato de fala de elevada densidade pragmática.

Para os fins de serem utilizados como estímulos no experimento de percepção, os atos de fala produzidos pela atriz foram descritos no que diz respeito a aspectos semântico-pragmáticos, à caracterização entoacional e à caracterização gestual. Essa descrição é apresentada na próxima seção.

A partir da dinâmica descrita acima, os estímulos foram gravados em ambiente de estúdio, com microfone SONY e câmera de alta resolução. A gravação foi realizada em formato audiovisual. A partir dessa gravação audiovisual, os arquivos de cada ato de fala foram editados por meio do software *Filmora* para obtenção de quatro modalidades de apresentação dos estímulos: (i) somente auditiva (AO), na qual a edição extraiu o áudio dos vídeos gravados para a obtenção de estímulos que apresentassem somente o áudio; (ii) somente visual (VO), na qual a edição silenciou o áudio dos vídeos gravados para a obtenção de estímulos que apresentassem somente a imagem dos vídeos; (iii) audiovisual congruente (AV+), na qual foram mantidas as gravações originais de cada ato de fala; e (iv) audiovisual incongruente (AV-), na qual foi realizada a combinação da informação visual e da informação auditiva advindas de gravações originalmente distintas, formando estímulos audiovisuais nos quais o áudio e a imagem não eram correspondentes entre si. Para tanto, declaração e pergunta foram pareadas, de um lado, e ordem e pedido foram pareados de outro lado, conforme detalha o quadro 4.

Quadro 3 - Detalhamento dos estímulos

Ato de fala	Modalidades dos Estímulos		Especificação
Declaração	Unimodal	AO	Somente áudio da declaração
		VO	Somente vídeo da declaração
	Bimodal	AV+	Áudio e vídeo da declaração
		AV-	Áudio da declaração + vídeo da pergunta
Pergunta com surpresa	Unimodal	AO	Somente áudio da pergunta
		VO	Somente vídeo da pergunta
	Bimodal	AV+	Áudio e vídeo da pergunta
		AV-	Áudio da pergunta + vídeo da declaração
Ordem	Unimodal	AO	Somente áudio da ordem
		VO	Somente vídeo da ordem
	Bimodal	AV+	Áudio e vídeo da ordem
		AV-	Áudio da ordem + vídeo de pedido
Pedido	Unimodal	AO	Somente áudio do pedido
		VO	Somente vídeo do pedido
	Bimodal	AV+	Áudio e vídeo do pedido
		AV-	Áudio do pedido + vídeo da ordem

Fonte: Dados da Pesquisa (2025).

Como mostra o Quadro 3, foram obtidos 16 estímulos (quatro atos de fala x quatro modalidades de apresentação dos estímulos). Por fim, em última etapa da edição dos estímulos, foram obtidos mais 16 estímulos por meio de inserção de ruído branco nos estímulos descritos no Quadro 3. Desse modo, o experimento contou com 16 estímulos sem ruído e 16 estímulos com ruído, totalizando 32 estímulos. Como já afirmado, em termos de procedimento, o experimento foi dividido em duas etapas: na primeira etapa, foram apresentados os 16 estímulos sem ruído e, na segunda etapa, foram apresentados os 16 estímulos com ruído.

3.7 DESCRIÇÃO DOS ESTÍMULOS

3.7.1 Descrição semântico-pragmática

Conforme já explicitado, os estímulos do experimento de percepção eram constituídos, sintaticamente, pela sentença “Vamos”. Essa forma sintática foi produzida como quatro diferentes atos de fala – declaração, pergunta com surpresa, ordem e pedido – e a diferenciação desses atos de fala foi marcada na materialidade linguística por características prosódicas.

A explicação para o fato de que uma mesma estrutura sintática pode ser identificada como diferentes atos de fala, implicando, assim, que essa estrutura pode ser interpretada semântica e pragmaticamente de maneiras distintas, é dada por Castilho (2010). De acordo com o autor, a distinção entre atos de fala não se reduz à estrutura sintática, mas resulta da articulação entre *modus* e *dictum*, sendo o *modus* o responsável por indicar o ato de fala que se manifesta num dado *dictum*.

No caso, a sentença “*vamos*” é um enunciado verbal correspondente à forma do verbo *ir* na primeira pessoa do plural do presente do modo indicativo, com sujeito não expreso. Embora no plano do *dictum*, esse enunciado apresente um conteúdo proposicional cuja referência se identifica na situação enunciativa, no plano do *modus* esse enunciado não codifica um valor ilocucionário fixo, podendo, assim, ser identificado como diferentes atos de fala, tais como uma declaração afirmativa, uma pergunta, uma ordem ou um pedido. Para tanto, os recursos prosódicos sinalizam no plano auditivo, especialmente pela entoação, e no plano visual, por meio de gestos faciais e movimentos de cabeça, o ato de fala implicado na produção do dizer e cujo *modus* permite identificar.

A sentença “Vamos” nos estímulos do teste experimental representa o que Castilho (2010) descreve quanto ao ato ilocucionário, uma vez que a realização de um ato de fala não

depende exclusivamente da forma verbal, mas emerge da articulação entre *modus*, prosódia, contexto discursivo e da relação entre os interlocutores. Considerando que “vamos” é uma forma verbal do modo indicativo, modo que, como afirma Castilho (2010), é tradicionalmente ligado à asserção, tem-se que, na produção da atriz quando da gravação dos estímulos, o indicativo assumiu valor de ato da fala diretivo tanto de ordem quanto de pedido. A diferenciação entre esses atos pode ser melhor especificada quando lançamos mão de conceitos de natureza semântica para fins de uma análise semântico-pragmática dos estímulos alvos de avaliação perceptual.

Do ponto de vista semântico, Pietroforte e Lopes (2014) mostram como a Semântica Lexical fornece uma abordagem descritiva sobre o modo como o significado é organizado e processado no léxico da língua. Partindo da ideia de que o significado das palavras não é único nem se limita a uma referência direta, os autores esclarecem que, nessa perspectiva, cada unidade lexical carrega, internamente, componentes semânticos mínimos – os semas – que constituem seus traços semânticos básicos. Na Semântica Lexical, semas são traços discretos de significado que, em combinação, distinguem as possibilidades de uso e de relação semântica entre unidades lexicais, como em redes de sinonímia, antonímia, hiponímia e hierarquia conceitual.

Pietroforte e Lopes (2014) destacam que a Semântica Lexical não se limita à atribuição de definições de dicionário, mas busca compreender como o significado se estrutura internamente e como esse arranjo semântico influencia a interpretação pragmática do enunciado. Diante disso, semas como obrigatoriedade, polidez, intensidade e hierarquia não constituem propriedades isoladas, mas são componentes dinâmicos que interagem com fatores contextuais e discursivos para caracterizar, semanticamente, um enunciado. Os autores enfatizam que a ativação desses traços semânticos ocorre tanto em nível lexical, ao acessar o significado de palavras específicas, quanto em nível enunciativo, ao interpretar intenções comunicativas mais complexas no plano pragmático, como em ordens, pedidos, perguntas e declarações.

Portanto, considerando a abordagem da Semântica Lexical de que semas são unidades semânticas mínimas que permitem, pela combinação, caracterizar e distinguir nuances semânticas de enunciados linguísticos, apresenta-se como, em termos semânticos, considerando o contexto comunicativo oferecido para a produção da atriz, a forma verbal “vamos” se comportou como diferentes atos de fala nos estímulos testados no experimento de percepção. Como já afirmamos, essa diferenciação semântica e pragmática foi possível, uma vez que, como esclareceu Castilho (2010), a atriz, enquanto falante da língua, atribuiu diferentes *modus* ao

dictum.

Para apresentar os semas que foram identificados na produção dos atos de fala pela atriz, foi elaborado um quadro-resumo da análise semântico-pragmática do enunciado “*vamos*”. Nesse quadro (Quadro 3), explicitam-se os contextos de produção, a forma verbal e os traços semânticos (semas) responsáveis pela distinção entre os diferentes atos de fala que funcionaram como estímulos no experimento.

Quadro 4 - Descrição sêmica dos atos de fala utilizados como estímulos

Ato de fala	Contexto Pragmático de Produção	Formal Verbal	Semas
Declaração	<i>Duas amigas conversam pelo telefone:</i> Amiga 1: — Nós vamos ao cinema hoje à noite? Amiga 2: — Vamos.	Indicativo	[+ ação] [+ futuro] [+ responsabilidade compartilhada] [+ confirmação] [- surpresa]
Pergunta c/ surpresa	<i>Duas amigas conversam na saída da aula:</i> Amiga 1: — Consegui os ingressos e nós vamos ao show da Madonna! Amiga 2: — Vamos?	Indicativo	[+ ação] [+ futuro] [+ responsabilidade compartilhada] [+ solicitação de confirmação] [+ surpresa]
Ordem	<i>Mãe e filha conversam antes de sair de casa:</i> Filho: — Eu não quero ir para a aula hoje. Mãe: — Vamos!	Indicativo	[+ ação] [+ presente] [+ autoridade] [- polidez] [- negociação]
Pedido	<i>Mãe e filha conversam em casa após chegada do trabalho:</i> Mãe: — Estou muito cansada, não vou mais ao cinema. Filha: — Vamos.	Indicativo	[+ ação] [+ presente] [- autoridade] [+ polidez] [+ negociação]

Fonte: Elaboração própria (2025)

É possível observar no quadro 3 como o enunciado “*vamos*”, a despeito da forma verbal no modo indicativo, realiza atos de fala distintos em função da ativação diferenciada de semas lexicais e pragmáticos, os quais se organizam conforme o contexto de produção e as relações hierárquicas entre os interlocutores.

Na declaração, o enunciado ocorre em um contexto não hierárquico, no qual não se estabelece relação de autoridade entre as interlocutoras. Nessa configuração, “*vamos*” ativa os semas [+ação] e [+futuro], associados à projeção de uma atividade compartilhada, bem como o sema [+responsabilidade compartilhada], que distribui igualmente o comprometimento entre as

falantes, identificadas como amigas que conversam sobre uma ida ao cinema. O sema [+confirmação] indica que o enunciado funciona como resposta afirmativa a uma proposição prévia, enquanto a ausência de surpresa, ou seja, o sema [-surpresa] reforça seu valor informativo, sem imposição ou direcionamento hierárquico.

Na pergunta, embora o enunciado mantenha os semas [+ação], [+futuro] e [+responsabilidade compartilhada], observa-se uma reconfiguração semântico-pragmática relevante. O traço [+confirmação] dá lugar ao sema [+solicitação de confirmação], o que desloca a força ilocucionária do enunciado para a interrogação. Além disso, a ativação do sema [+surpresa] indica uma superação das expectativas da falante diante da informação apresentada no turno anterior pela interlocutora, o que diferencia esse uso tanto da declaração quanto da pergunta neutra, indicando que o ato de fala produzido se configura como uma pergunta com surpresa, na qual, além da ilocução relativa à interrogação, tem-se implicado no ato de fala a expressão de uma atitude de surpresa.

Na ordem, o enunciado “*vamos*” ocorre em um contexto marcado por assimetria hierárquica, no qual a mãe ocupa uma posição de autoridade em relação à filha. Nessa configuração, os semas [+ação] e [+presente], associados ao ato de fala diretivo de ordem produzido por uma mãe com o objetivo de fazer com que a filha realize a ação de ir à escola, indicam a exigência de realização imediata da ação. De forma associada ao tempo imediato da ação, a ativação do sema [+autoridade], combinada aos traços [-polidez] e [-negociação], confere ao enunciado uma força ilocucionária diretiva elevada, caracterizando-o como imposição. O significado de “*vamos*” como ato de fala de ordem é, assim, reorientado para a obrigatoriedade, sem abertura para contestação ou ajuste de negociação na interação da filha com a mãe.

Por fim, no pedido, observa-se uma reorganização dos traços semânticos em uma configuração de ato de fala diretivo menos impositivo. Embora permaneçam os semas [+ação] e [+presente], o traço [-autoridade] indica uma posição inferior na hierarquia social de quem fala em relação ao interlocutor, como no caso da filha em relação à mãe. Assim, por não ter autoridade sobre a mãe e, ao contrário, estar em relação a ela numa posição hierarquicamente inferior, o ato de fala diretivo não assume condição para a imposição de uma ação a ser realizada pelo outro. Por consequência, os semas [+polidez] e [+negociação] introduzem um valor atenuado de obrigatoriedade à solicitação de realização apresentada no ato de fala diretivo, permitindo que o enunciado funcione como um pedido, e não como uma ordem. Nesse caso, “*vamos*” expressa uma tentativa de engajamento do outro por quem fala – no caso do contexto comunicativo utilizado, de engajamento da mãe por parte da filha com o pedido para irem ao

cinema –, deixando espaço para recusa ou ajuste do interlocutor, o que suaviza a força ilocucionária do ato diretivo.

De modo geral, a descrição apresentada no Quadro 3 demonstra que a distinção entre os atos de fala não decorre da forma verbal em si, mas da combinação dinâmica dessa forma com semas lexicais como autoridade, polidez, surpresa, negociação e responsabilidade compartilhada, que, em interação com o contexto pragmático e com os marcadores prosódicos, direcionam a interpretação do dictum “*vamos*” ao lhe conferir um *modus*. No português brasileiro, conforme afirmação de Castilho (2010), o *modus* pode ser codificado por diferentes meios linguísticos, incluindo recursos morfossintáticos, lexicais e, de modo central, por recursos prosódicos, especialmente a entoação. De acordo com a perspectiva multimodal assumida neste trabalho, os recursos prosódicos que sinalizam e diferenciam atos de fala se manifestam tanto por pistas auditivas quanto por pistas visuais. Para demonstrar o papel dessas pistas como marcadores prosódicos de diferentes atos de fala, a seguir apresenta-se tanto a descrição acústica quanto a descrição gestual dos estímulos.

3.7.2 Descrição acústica

Os estímulos auditivos representativos de cada ato de fala foram analisados acusticamente por meio do software *Praat* (Boersman; Weenink, 2009). Foi realizada análise entoacional, assim como foi realizada a mensuração de parâmetros de F0, duração e intensidade.

Para a análise entoacional, tomou-se como referência o modelo de Fonologia Entoacional Autossegmental e Métrica (Pierrehumbert, 1980; Ladd, 1996, 2008) de forma integrada à hierarquia prosódica (Nespor; Vogel, 1986). Para além da análise entoacional, foram mensurados os valores de F0, intensidade e duração nos estímulos. Na mensuração dos parâmetros acústicos de F0 e intensidade, foram calculados os valores médios de suas variações de acordo com o movimento entoacional dentro da sílaba tônica. Já para a análise da duração, foi considerado o valor total da produção do enunciado fonológico em segundos. Ressalta-se que a análise entoacional foi realizada pela pesquisadora e posteriormente revisada pela orientadora deste estudo.

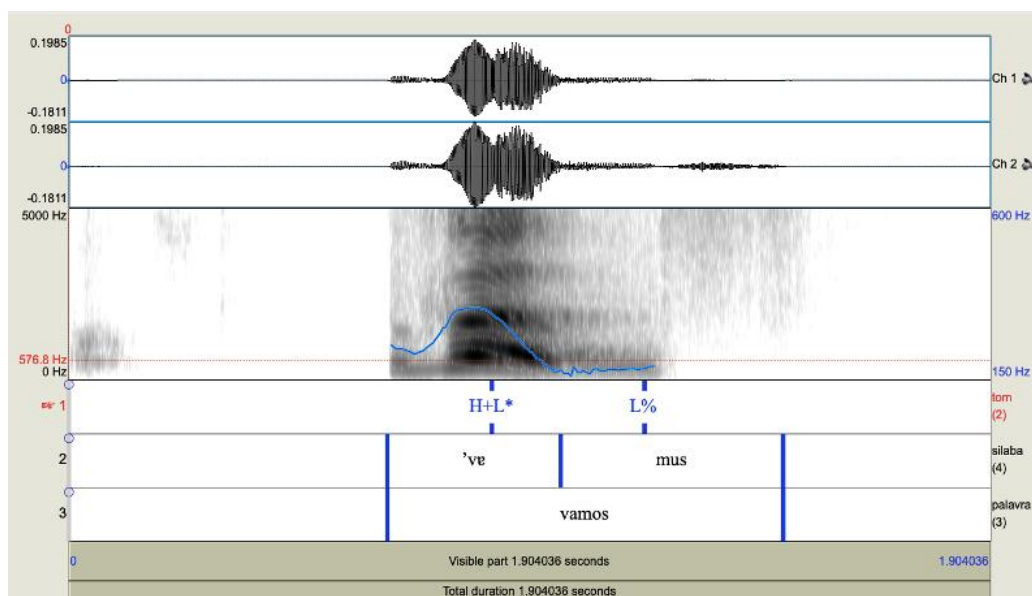
A seguir, a Tabela 1 dispõe as configurações entoacionais identificadas e os valores dos parâmetros acústicos mensurados em cada ato de fala. Por sua vez, as Figuras 4, 5, 6 e 7 apresentam as imagens acústicas da modulação de F0 dos atos de fala obtidas no software *Praat* (Boersman; Weenink, 2009), a partir das quais foi feita a análise da configuração entoacional.

Tabela 1 – Análise entoacional dos atos de fala do português brasileiro

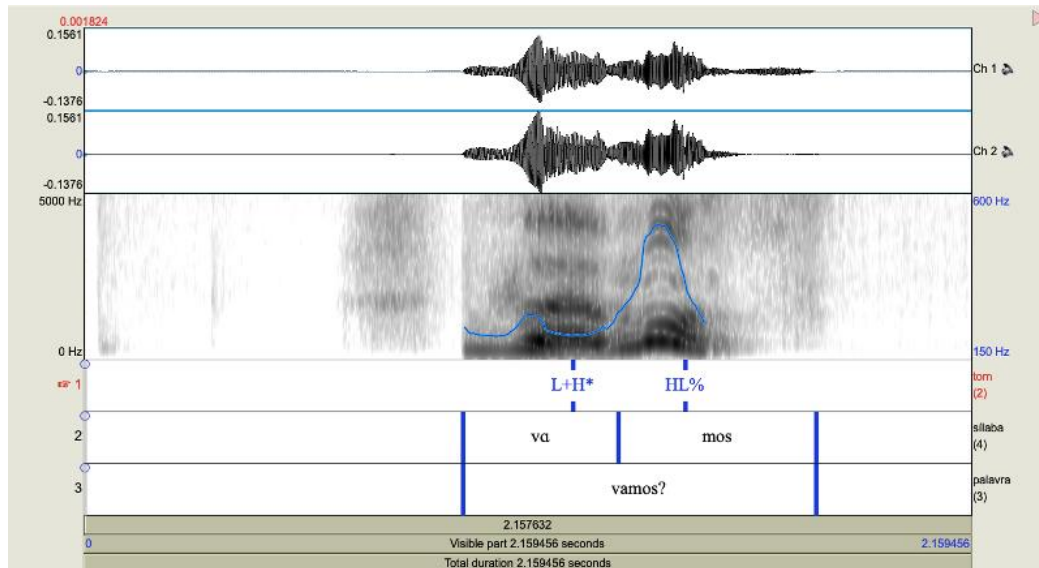
Ato de fala	Configuração entoacional	F0 (Hz)	Duração (s)	Intensidade (db)
Declaração	H+L* L%	335.3 Hz – 203.5 Hz	0.817494	70.94
Pergunta com surpresa	L+H* HL%	217.1 Hz – 263.95	0.859864	66.28
Ordem	H+L* L%	376.2 Hz – 253.6 Hz	0.761125	79.3
Pedido	H* L%	371.1	1.067253	72.26

Fonte: elaboração própria (2025)

A Figura 4 ilustra a configuração entoacional da declaração, enquanto a Figura 5 ilustra a configuração entoacional da pergunta.

Figura 4 – Análise entoacional ato de fala declarativo

Fonte: Elaboração própria (2025)

Figura 5 – Análise entoacional ato de fala pergunta

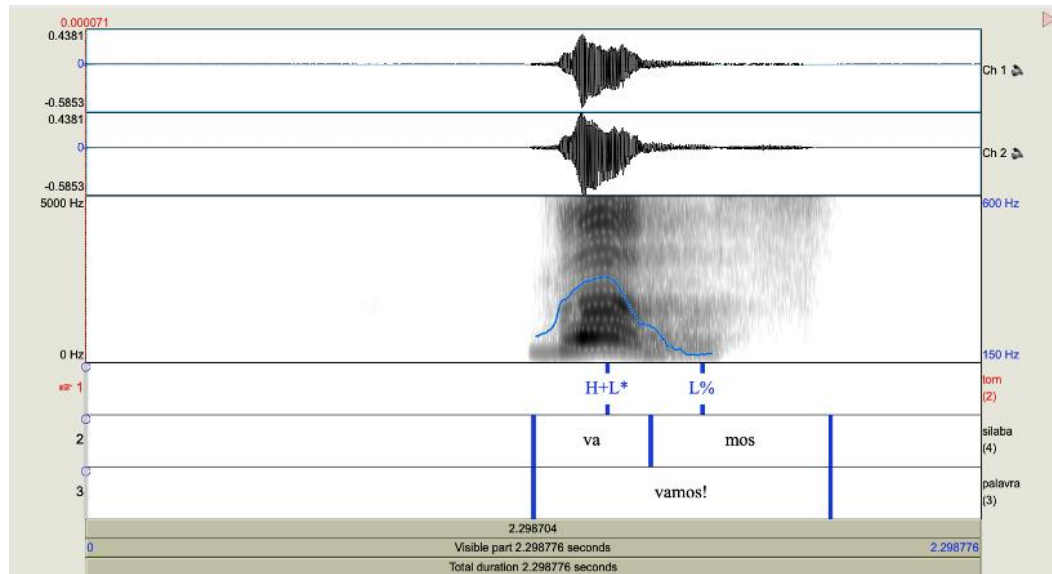
Fonte: Elaboração própria (2025)

Considerando os dados descritivos na tabela 1 e as figuras acima, podemos observar que, entoacionalmente, o ato de fala declarativo é marcado por uma queda de F0 na sílaba tônica, movimento que caracteriza o seu acento nuclear H+L*, o qual é seguido por tom de fronteira L%. Já o ato de fala pergunta apresentou, de forma oposta, um movimento de F0 ascendente, com subida acentuada de F0 na sílaba tônica, onde se identifica o acento nuclear L+H*, seguido por tom de fronteira descendente bitonal HL%. Deve-se destacar que tanto a elevada ascendência quanto o tom de fronteira HL% são características entoacionais que identificam esse ato de fala como uma pergunta com surpresa ou, em outras palavras, uma pergunta com mudança de expectativa. Ainda, destaca-se que o ato de fala declarativo apresentou duração menor em comparação ao ato de fala pergunta com surpresa.

É possível interpretar a caracterização entoacional e acústica da declaração e da pergunta como marcadores das nuances semânticas desses atos de fala (cf. Quadro 3) na medida em que o ato de fala declarativo, com acento nuclear descendente (H+L*) e tom de fronteira baixo (L%), associa-se aos semas de [+confirmação] e [-surpresa], que, ainda, são reforçados por menor duração, haja vista a assertividade ligada a esse ato. Já a pergunta, com acento ascendente (L+H*) e tom de fronteira bitonal (HL%), ativa semas de [+confirmação] e [+surpresa], ambos intensificados pela grande magnitude da ascensão de F0 e pela maior duração do enunciado.

A Figura 6 ilustra a configuração entoacional do ato de fala *ordem* e, na sequência, a Figura 7 apresenta a configuração entoacional do ato de fala *pedido*.

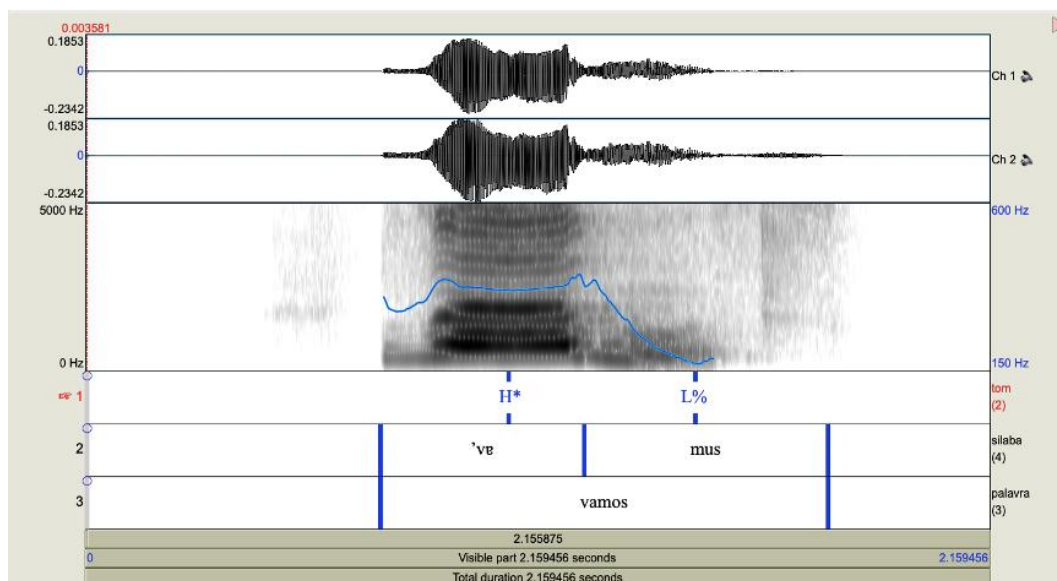
Figura 6 – Análise entoacional ato de fala ordem



Fonte: Elaboração própria (2025)

De forma similar à declaração, a ordem apresentou um movimento de queda de F0 na sílaba tônica com acento nuclear H+L* e tom de fronteira baixo L%. Apesar da mesma configuração entoacional que o ato de fala declarativo, a ordem se diferenciou em relação aos demais atos de fala por apresentar o menor valor de duração, indicando que esse ato foi produzido de forma mais rápida no que diz respeito ao tempo, e o maior valor de intensidade, indicando uma produção mais forte em termos de energia acústica; além disso, a ordem apresentou maior faixa de F0 em relação ao ato de fala declarativo, com picos mais altos de F0, indicando que, ainda que a configuração entoacional desses atos de fala seja a mesma, o tom H do acento tonal nuclear é mais agudo na ordem em relação à declaração.

Sob a perspectiva da semântica lexical, conforme apresentada em Pietroforte e Lopes (2014), a entoação da ordem pode ser interpretada como marcador do sema [+ autoridade] por meio da intensificação prosódica. Embora compartilhe a configuração H+L* L% com a declaração, a ordem se distingue por uma menor duração e por maiores valores de intensidade e de picos de F0, o que amplia o valor diretivo da força ilocucionária e sinaliza uma nuance semântica mais impositiva. A distinção entre os atos, portanto, não se apoia apenas no contorno entoacional, mas no grau de saliência prosódica que ativa semas específicos.

Figura 7 – Análise entoacional ato de fala pedido

Fonte: Elaboração própria (2025)

Em relação ao ato de fala pedido, este se caracterizou por apresentar um tom alto que se mantém na sílaba tônica, caracterizando o acento nuclear H*, com queda brusca para o tom L% na fronteira prosódica. Além disso, trata-se do ato de fala com maior tempo de duração, que se identifica especialmente devido ao alongamento da sílaba tônica. Esse ato de fala também apresentou pico de F0 mais alto e maior intensidade em relação à declaração e à pergunta, mas menores em relação à ordem.

À luz da semântica lexical, esse padrão prosódico identificado no pedido pode ser interpretado como marcador de semas que atenuam a imposição, ao sinalizar [+ polidez] [- autoridade] e [+ negociação]. Particularmente, a elevação melódica se relaciona à expressão da polidez enquanto o alongamento da sílaba tônica acentua o caráter de solicitação, quase com nuances de súplica, representando expectativa de negociação com o interlocutor.

Além da descrição acústica, foi realizada a descrição gestual dos respectivos atos de fala apresentada a seguir.

3.7.3 Descrição gestual

Participaram dessa análise 3 juízes, integrantes do LAAC da UNESP/Marília e com formação em análise prosódica por realizarem pesquisa sobre prosódia multimodal no interior do Núcleo de Pesquisa em Prosódia e Aquisição, decorrente do projeto FAPESP 2020/10144-3) coordenado pela orientadora do presente trabalho. Os juízes julgaram os estímulos na

modalidade congruente de cada ato de fala. Para a análise visual, o presente trabalho se baseou na *Theory of Entrained Manual and Speech Systems* (TEMSS) proposta por Rusiewicz (2011).

À luz dessa abordagem, os estímulos audiovisuais congruentes foram analisados considerando a coordenação entre fala e gesto para cada ato de fala. Para tanto, tomou-se como critério norteador identificar os gestos que se sincronizam à produção do acento nuclear associado à sílaba tônica de “Vamos” na produção de cada ato de fala. A Figura 8 apresenta imagens congeladas de gestos realizados durante a produção desses atos.

Figura 8 – Análise gestual dos tipos de atos de fala



Legenda: a ilustração mostra uma mulher jovem de cabelos curtos loiros, pele branca, usando uma blusa branca, em fundo verde, representando os atos de fala envolvidos no experimento.
Fonte: Elaboração própria (2025)

De modo geral, o resultado da análise mostrou que os atos de fala se distinguiram em relação aos gestos a eles associados.

No ato de fala declarativo, observaram-se gestos faciais e corporais associados à afirmação, como movimento afirmativo da cabeça, protrusão dos ombros e sorriso discreto. À luz da semântica lexical, nesse padrão gestual, o movimento afirmativo de cabeça no sentido vertical funciona como marcador predominantemente da assertividade identificada na declaração que ativa os semas [+confirmação] e [– surpresa] (cf. Quadro 3). Ainda, o sorriso discreto identificado pode estar associado à atitude de aprovação da ação cuja realização foi confirmada pela falante: a ida ao cinema conforme o contexto comunicativo oferecido.

O ato pergunta caracterizou-se por gestos que incluíram sobrancelhas arqueadas, olhos mais arregalados e lábios mais ocluídos e arredondados. Esses recursos visuais, principalmente o simultâneo arqueamento das sobrancelhas com o arregalar dos olhos, se relacionam aos semas de [+ solicitação de confirmação] e [+ surpresa] desse ato de fala, os quais convocam o interlocutor para apresentar uma resposta de confirmação à nova expectativa criada pelo falante.

Na ordem, foram identificados diferentes marcadores visuais para o valor diretivo mais impositivo desse ato de fala, são eles: sobranceiras franzidas, articulação labial exagerada com ampla abertura, fechamento das aletas nasais e um movimento brusco vertical da cabeça acompanhado de protrusão dos ombros. Esses elementos corporais configuram a realização multimodal dos semas de [+ autoridade], [-polidez] e [-negociação], central na organização semântica desse ato de fala, uma vez que a ordem projeta a ação requerida como necessária e não negociável, instaurando uma relação assimétrica entre os interlocutores (Pietroforte; Lopes, 2014).

Por fim, pedido apresentou gestos sutis que, semanticamente, se associam à atenuação do valor diretivo desse ato de fala ao marcarem os semas [- autoridade], [+polidez] e [+negociação]. Esses gestos foram: leve inclinação lateral da cabeça, abaixamento e protrusão moderada dos ombros e lábios semi-ocluídos. De acordo com a semântica lexical, a polidez opera como um traço semântico que atenua a força ilocucionária e reduz o grau de imposição do enunciado, reconfigurando a ação proposta como dependente do consentimento do interlocutor. Assim, os gestos observados funcionam como marcas visuais de atenuação, uma vez que, também esses gestos por se apresentarem de forma sutil em comparação aos gestos identificados na ordem, representam a suavização da obrigatoriedade da ação e, assim, distinguem o pedido da ordem.

3.8 VALIDAÇÃO DOS ESTÍMULOS

Como parte dos procedimentos metodológicos e do rigor científico, os estímulos passaram, antes da elaboração do experimento, por validação perceptual quanto às suas diferenças. Para tanto, os estímulos audiovisuais congruentes de cada ato de fala foram organizados e enviados em formulário criado pela pesquisadora no *Google Forms* a 13 falantes adultos, de ambos os gêneros, residentes do interior de São Paulo e não especialistas em linguagem, para que julgassem, de forma cega, se os estímulos representavam declaração, pergunta, ordem ou pedido. Cada vídeo representativo de cada ato de fala foi apresentado duas vezes, em ordem aleatória. Os juízes foram orientados a responder, em ambiente silencioso, por meio de seus próprios celulares, utilizando fones de ouvido. Como resultado dessa etapa de validação dos estímulos gravados pela atriz, obteve-se um nível de concordância de 100% entre os juízes na identificação de cada ato de fala, demonstrando, portanto, que os estímulos obtidos na gravação foram considerados pelos juízes como representativos dos atos de fala alvos de

investigação e, estando, assim, validados quanto à sua distinção perceptual para o uso nos experimentos.

3.9 FORMA DE ANÁLISE DOS RESULTADOS

Ao final da coleta, o estudo contou com um *corpus* de 3520 dados de percepção para análise (16 estímulos apresentados em cada etapa do experimento, totalizando 32 estímulos avaliados por 110 participantes).

Na análise desses dados, foram contabilizadas por participante as seguintes variáveis dependentes: (i) a *acurácia perceptual*, mensurada pela porcentagem de acertos nas respostas atribuídas aos estímulos nas modalidades AO, VO e AV+; (ii) a *preferência perceptual* (auditiva ou visual) nas respostas atribuídas aos estímulos na modalidade AV-; e (iii) o *índice de efeito visual*, obtido pela diferença entre o percentual de acerto na modalidade AV+ e o percentual de acerto na modalidade AO, representado pela fórmula $IEF = (\%AV+) - (\%AO)$. Essa fórmula foi adaptada dos estudos de Massaro e Cohen (1995) e Massaro (1998), que a propuseram para a mensuração da influência visual em dados de percepção de contrastes fonológicos no plano segmental. Por não assumirmos a ideia de influência visual dos autores, adaptamos a fórmula para a mensuração do que aqui chamamos de efeito da pista visual na percepção de contrastes no plano prosódico relativos a distintos atos de fala.

Essas variáveis foram mensuradas e organizadas, considerando os seguintes fatores de análise: (a) tipo de ato de fala (declarativo, pergunta, ordem e pedido); (b) modalidade dos estímulos (AO, VO, AV+ e AV-); (c) condição de apresentação (sem e com ruído); e (d) o grupo etário (4-5 anos, 6-7 anos, 8-9 anos, 10-12 anos, 18-25 anos).

3.10 ANÁLISE ESTATÍSTICA

Os dados relativos à percepção foram analisados por meio de procedimentos estatísticos descritivos e inferenciais, de natureza quantitativa.

Primeiramente, com o objetivo de comparar a acurácia perceptual entre os diferentes grupos etários na identificação dos atos de fala diferenciados prosodicamente (Objetivo 1), foi realizada uma ANOVA de Medidas Repetidas, tendo como variável dependente a acurácia perceptual geral, expressa pelo percentual de acertos, como variável independente intragrupo o tipo de ato de fala e, como variável independente intergrupo, o grupo etário.

Em seguida, visando avaliar o efeito das pistas prosódicas auditivas e visuais na percepção dos atos de fala pelos diferentes grupos etários (Objetivo 2), foram conduzidas análises separadas para cada ato de fala. Para a acurácia perceptual, realizou-se, para cada ato de fala, uma ANOVA de Medidas Repetidas, considerando a acurácia perceptual como variável dependente, a modalidade de apresentação dos estímulos (audiovisual congruente – AV+, apenas auditiva – AO e apenas visual – VO) como variável independente intragrupo e o grupo etário como variável independente intergrupo.

Ainda, visando maior detalhamento sobre o efeito das pistas prosódicas auditivas e visuais percepção dos grupos para responder ao Objetivo 2, com o propósito de examinar a preferência perceptual dos participantes diante de estímulos audiovisuais incongruentes, foi realizada, para cada ato de fala, uma ANOVA de Medidas Repetidas, considerando a preferência perceptual (auditiva ou visual) como variável dependente intragrupo e o grupo etário como variável intergrupo.

Por fim, com o objetivo de avaliar se a pista visual exerceria um efeito compensatório na identificação dos diferentes atos de fala em condições de estímulo com ruído (Objetivo 3), foram realizadas, para cada ato de fala, ANOVAs de Medidas Repetidas, tendo como variável dependente o índice de efeito visual, como variável independente intragrupo a condição de apresentação dos estímulos (sem ruído e com ruído) e, como variável independente intergrupo, o grupo etário.

Para todos os testes inferenciais, considerou-se o nível de significância $\alpha < 0,05$ e o intervalo de confiança de 95%. A seguir, serão apresentados os resultados encontrados na pesquisa.

4 RESULTADOS

Nesta seção, são apresentados os resultados obtidos a partir da análise descritiva e inferencial dos dados. Para melhor organização da seção, os resultados são apresentados em subseções. Inicialmente, são apresentados os resultados relacionados ao primeiro objetivo, que buscou comparar a acurácia perceptual de diferentes grupos etários na identificação de atos de fala diferenciados prosodicamente por pistas auditivas e visuais. Sequencialmente, especificam-se os resultados referentes ao segundo objetivo deste trabalho, que buscou avaliar o efeito de pistas prosódicas auditivas e visuais na percepção dos diferentes atos de fala pelos grupos etários. Por último, apresenta-se o resultado referente ao terceiro objetivo da pesquisa, que buscou verificar se, em condição de estímulos com ruído, haveria efeito compensatório das pistas visuais na identificação dos diferentes atos de fala.

4.1 ACURÁCIA PERCEPTUAL GERAL DOS GRUPOS NA IDENTIFICAÇÃO DOS ATOS DE FALA DO PB

Primeiramente, em relação à análise descritiva, a Tabela 2 apresenta os resultados referentes à acurácia perceptual geral dos grupos por ato de fala.

Tabela 2 - Média e desvio padrão da acurácia perceptual dos grupos por ato de fala

Grupo	Ato de Fala							
	Declaração		Pergunta		Ordem		Pedido	
	M (%)	DP	M (%)	DP	M (%)	DP	M (%)	DP
G1: 4 – 5 anos	0,3	0,16	0,9	0,22	9,2	0,20	8,5	0,19
G2: 6 – 7 anos	3,4	0,24	9,8	0,24	51,0	0,25	9,3	0,19
G3: 8 – 9 anos	1,6	0,14	3,2	0,19	7,6	0,12	2,2	0,19
G4: 10 – 12 anos	7,8	0,23	1,4	0,19	9,0	0,11	6,7	0,20
G5: 18 – 25 anos	6,2	0,18	3,1	0,18	4,3	0,15	3,7	0,20
Geral	9,5	0,24	7,6	0,25	8,1	0,21	0,0	0,21

Fonte: Dados da Pesquisa (2025).

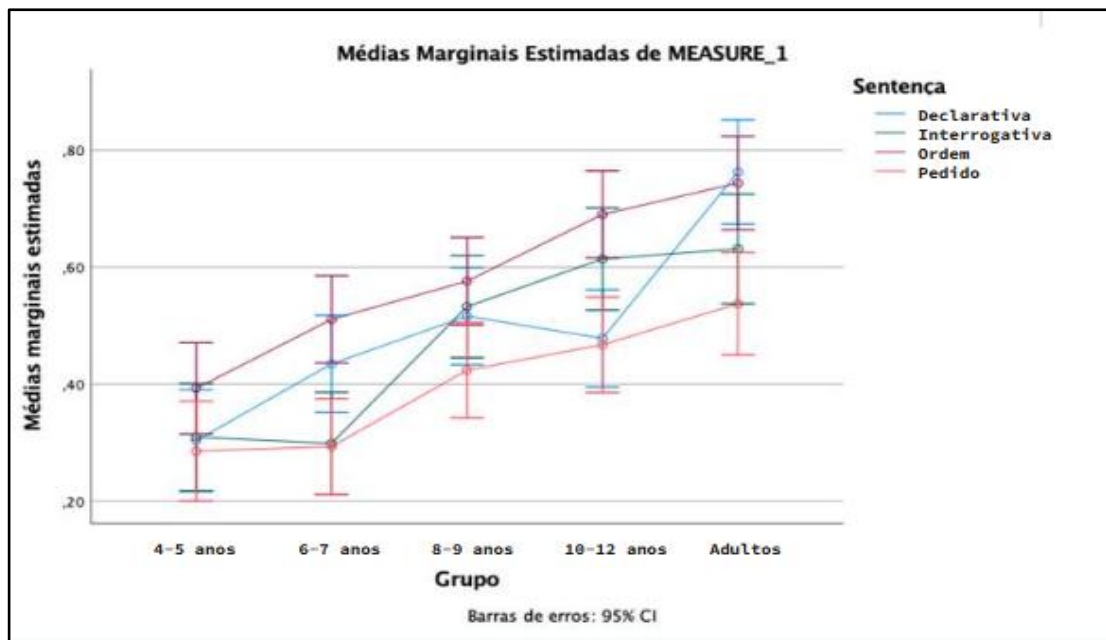
Observa-se, na coluna destinada à porcentagem da média de identificação dos atos de fala, um aumento progressivo da acurácia conforme o aumento da idade em todos os atos de fala analisados, ainda que esse crescimento não ocorra de forma linear.

Quanto à análise inferencial comparativa da acurácia perceptual, a ANOVA de

Medidas Repetidas indicou efeito estatisticamente significativo para o tipo ato de fala ($F(3,315)= 16,009$; $p = 0,000$), para grupo, ($F(4,105)= 38,177$; $p = 0,000$) e, ainda, para a interação entre ato de fala e grupo ($F(12,315)= 1,836$; $p = 0,042$).

O resultado estatístico acima referente ao grupo mostra que a idade é um fator que impacta na identificação mais acurada dos diferentes atos de fala diferenciados prosodicamente. Por sua vez, o resultado estatístico referente ao tipo de ato de fala mostra que identificar declarações, perguntas, ordens e pedidos por pistas prosódicas não é similar, já que a identificação dos participantes foi mais acurada para alguns desses atos de fala em relação a outros, conforme ilustra o gráfico a seguir.

Gráfico 1- Desempenho perceptual na identificação dos tipos de atos de fala por grupo etário



Fonte: Elaboração própria (2025)

É possível verificar no gráfico 1 acima que a acurácia perceptual de identificação do ato de fala ordem foi maior para todos os grupos. Em contrapartida, a acurácia de identificação do ato de fala pedido foi a menor para todos os grupos. Já a acurácia de identificação do ato de fala pergunta foi significativamente maior a partir de 8 anos e, por fim, a acurácia do ato de fala declarativo variou entre os grupos de crianças e mostrou-se significativamente maior para o grupo dos adultos.

Quanto à interação entre tipo de ato de fala e grupo, destaca-se o resultado da ANOVA, citado anteriormente, o qual indica que o desempenho perceptual dos grupos etários

é dependente do ato de fala a ser identificado.

Portanto, a análise da acurácia perceptual auditiva e visual dos grupos, apresentada a seguir, foi realizada considerando, separadamente, cada tipo de ato de fala: declarativo, pergunta, ordem e pedido.

4.2 ACURÁCIA PERCEPTUAL AUDITIVA E VISUAL DOS GRUPOS POR ATO DE FALA

Tendo em vista que a acurácia dos grupos foi diferente a depender do tipo de ato de fala, foi aplicado uma ANOVA de Medidas Repetidas, considerando como variáveis independentes intra e intergrupo a modalidade (AO, VO e AV) e o grupo etário, respectivamente, e, como variável dependente, o percentual de acerto para cada ato de fala especificamente.

Primeiramente, apresentam-se os resultados obtidos para o ato de fala declarativo.

4.2.1 Declaração

A Tabela 3 apresenta os resultados descritivos da acurácia perceptual da declaração.

Tabela 3- Média e desvio padrão da acurácia perceptual dos grupos por modalidade na identificação da declaração

Grupo	Modalidade					
	AO		VO		AV+	
	(%)	P	(%)	P	(%)	P
G1: 4 – 5 anos	8,5	0,37	5,7	0,39	6,1	0,33
G2: 6 – 7 anos	9,1	0,36	4,7	0,35	8,7	0,38
G3: 8 – 9 anos	6,9	0,30	9,1	0,33	7,3	0,26
G4: 10 – 12 anos	1,3	0,41	1,3	0,38	0,8	0,36
G5: 18 – 25 anos	2,5	0,25	7,5	0,30	2,5	0,33

Fonte: Dados da Pesquisa (2025).

Com relação a esses dados, a ANOVA indicou efeito significativo para a modalidade ($F(2,210) = 6,818$; $p = 0,001$) e para o grupo ($F(4, 105) = 12,737$; $p < 0,001$), mas não indicou efeito significativo para a interação entre modalidade e grupo ($F(8,201) = 1,317$; $p = 0,236$), indicando que o efeito da modalidade foi equivalente para todos os grupos.

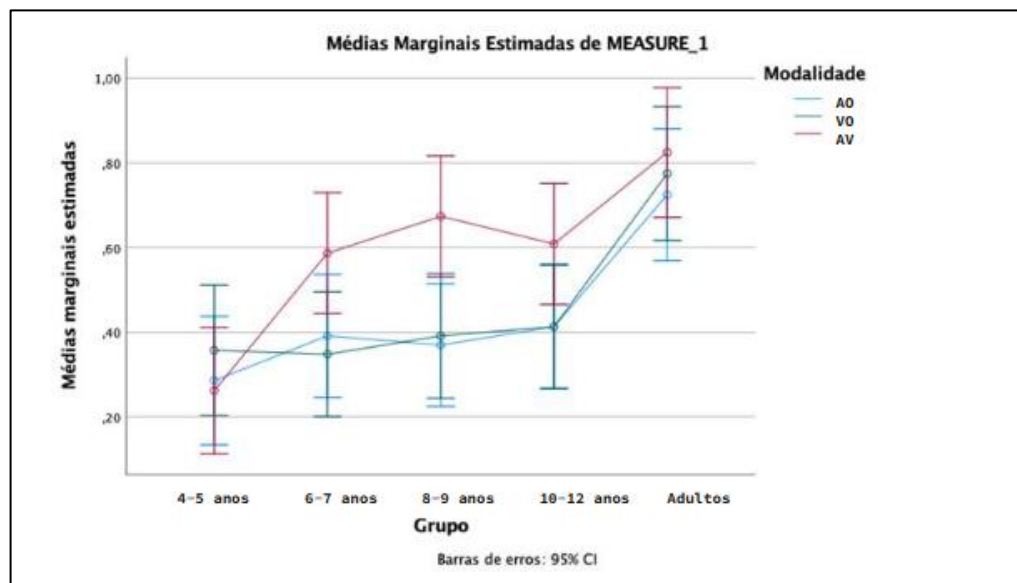
No que diz respeito à variável grupo, após resultado positivo da ANOVA, realizou-se análise complementar por meio do teste Post Hoc Bonferroni. Trata-se de um teste de comparação múltipla dos pares um a um. Os resultados indicaram que o grupo de adultos se

diferenciou estatisticamente de forma significativa de todos os demais grupos ($p < 0,001$) como o grupo que apresentou a maior acurácia perceptual (Média de identificação geral = 77,5%). Em contrapartida, os quatro grupos de crianças não se diferenciam entre si (para todas as comparações, $p > 0,05$) e apresentaram acurácia perceptual similar na identificação da declaração (Médias gerais iguais a: 4-5 anos = 30,2%; 6-7 anos = 44,2%; 8-9 e 10-12 = 47,8%), conforme a tabela 3.

No que diz respeito à variável modalidade, o teste Post Hoc Bonferroni indicou que a modalidade AV+ se diferenciou significativamente em relação às modalidades AO e VO (para ambas as comparações, $p < 0,05$), uma vez que favoreceu maior acurácia perceptual na identificação do ato de fala declarativo (Média geral dos grupos na modalidade AV+ = 59%). Por sua vez, as modalidades AO e VO não se diferenciam entre si ($p = 1,00$), uma vez que, em ambas, a identificação da declaração para todos os grupos foi similar (Média geral em AO = 43% (AO) e Média geral em VO = 45%).

O Gráfico 2 ilustra esses resultados:

Gráfico 2– Acurácia perceptual dos grupos etários por modalidade na identificação da declaração



Fonte: Elaboração própria (2025)

4.2.2 Pergunta

Para a pergunta, a Tabela 4 apresenta os resultados descritivos. Pode-se observar um aumento gradual na média do percentual de identificação desse ato de fala a partir dos 8 anos, com maior acurácia na idade adulta.

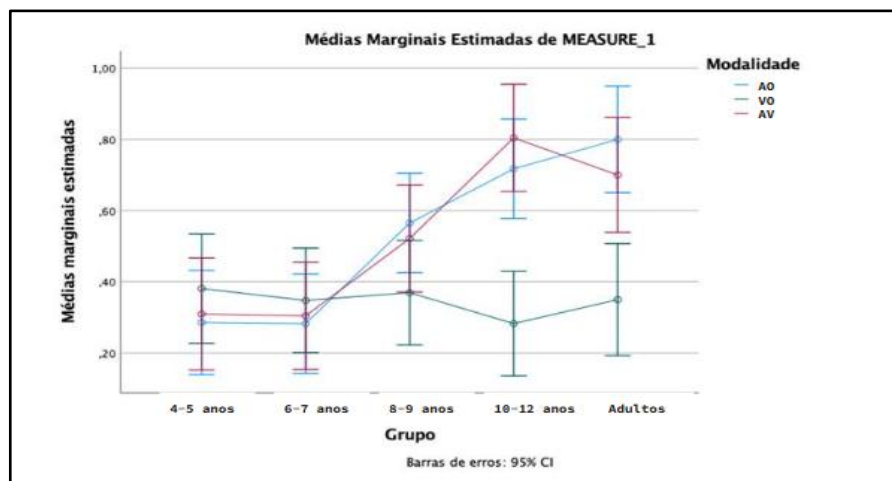
Tabela 4- Média e desvio padrão da acurácia perceptual dos grupos por modalidade na identificação do ato de fala pergunta

Grupo	Modalidade					
	AO		VO		AV+	
	(%)	P	(%)	P	(%)	P
G1: 4 – 5 anos	8,5	0,33	8,1	0,35	0,9	0,37
G2: 6 – 7 anos	8,2	0,39	4,7	0,43	0,4	0,39
G3: 8 – 9 anos	6,5	0,31	6,9	0,34	2,1	0,38
G4: 10 – 12 anos	1,7	0,33	8,2	0,25	0,4	0,29
G5: 18 – 25 anos	0,0	0,29	5,0	0,36	0,0	0,37

Fonte: Dados da pesquisa (2025).

Com os dados desse ato de fala, a ANOVA indicou efeito significativo para a modalidade ($F(2,105) = 11,393$; $p < 0,001$), para o grupo ($F(4,105) = 8,774$; $p < 0,001$), e também para a interação modalidade e grupo ($F(8,210) = 11,393$; $p < 0,001$), indicando, assim, que o efeito da modalidade de apresentação dos estímulos foi diferente a depender do grupo. O Gráfico 3 ilustra esse resultado.

Gráfico 3– Acurácia perceptual dos grupos etários por modalidade na identificação da pergunta



Fonte: Elaboração própria (2025)

Conforme ilustra o Gráfico 3, a modalidade de apresentação dos estímulos não impactou a acurácia perceptual de identificação do ato de fala pergunta nos grupos de crianças de 4-5 e 6-7 anos, já que o percentual de identificação desse ato de fala, independentemente da modalidade, foi próximo, estando num intervalo entre 28 e 38%. Ou seja, a acurácia perceptual desses grupos na identificação da pergunta foi estatisticamente similar nas diferentes modalidades de apresentação dos estímulos.

Diferentemente dos grupos de crianças de 4-5 e 6-7 anos, observa-se que a partir dos 8 anos, a acurácia perceptual foi otimizada quando os estímulos foram apresentados nas modalidades AO e AV; observa-se ainda que as crianças de 10-12 anos apresentaram acurácia perceptual auditiva e/ou audiovisual similar e não se diferenciaram em relação à acurácia dos adultos (Médias AO iguais a 71% e 80% para 10-12 anos e adultos, respectivamente; Médias AV iguais a 80% e 70% para 10-12 anos e adultos, respectivamente).

Considerando-se a baixa acurácia na modalidade VO (média=34,6% de acertos) e a não diferenciação estatística entre AO e AV ($p=1,000$), tem-se que a modalidade AV não favoreceu a identificação da pergunta, indicando que, na percepção desse ato de fala, para sujeitos a partir de 8 anos, a informação auditiva se mostrou mais robusta.

4.2.3 Ordem

Para a ordem, a Tabela 5 apresenta os resultados descritivos. Pode-se observar um aumento gradual na média do percentual de identificação da ordem durante a aquisição em todas as modalidades apresentadas.

Tabela 5- Média e desvio padrão da acurácia perceptual dos grupos por modalidade na identificação do ato de fala ordem

Grupo	Modalidade					
	AO		VO		AV+	
	(%)	P	(%)	P	(%)	P
G1: 4 – 5 anos	8,5	0,37	2,8	0,36	1,9	0,35
G2: 6 – 7 anos	3,4	0,40	7,3	0,41	6	0,36
G3: 8 – 9 anos	4,3	0,33	5,2	0,31	2,6	0,28
G4: 10 – 12 anos	5,2	0,35	4,7	0,27	1,3	0,19
G5: 18 – 25 anos	5,0	0,34	5,0	0,15	5,0	0,15

Fonte: Dados da Pesquisa (2025).

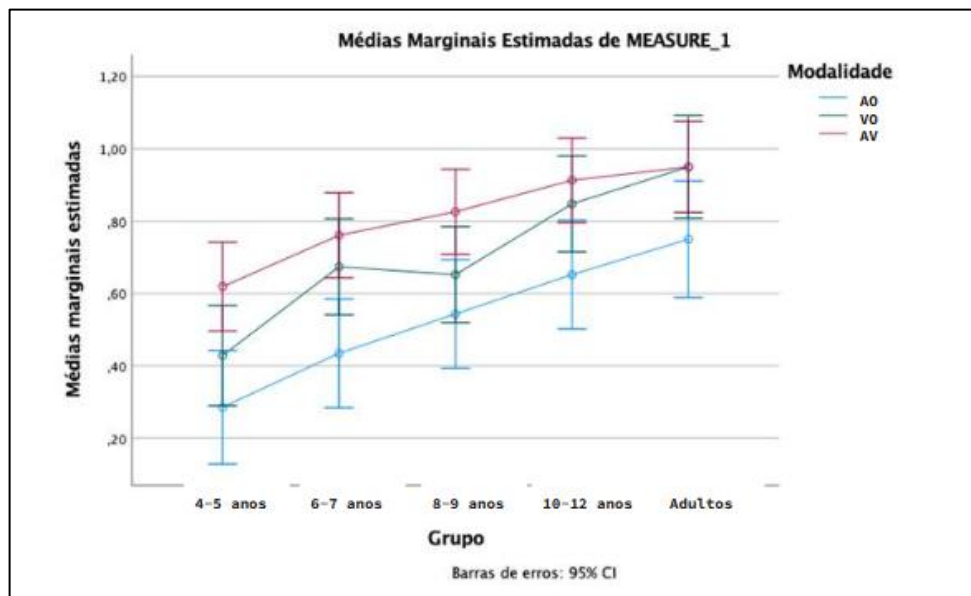
No que diz respeito à análise estatística inferencial, os resultados da ANOVA indicaram diferença estatística significativa para a modalidade ($F(2,210)=24,359$; $p=0,000$) e

também para os grupos ($F(4,210)=13,488$; $p= 0,000$); contudo, não mostrou diferença significativa na interação entre modalidade e grupo ($F(8,210)=0,465$ $p= 0,880$), ou seja, os resultados estatísticos encontrados para a modalidade equivalem para todos os grupos.

No que diz respeito à variável grupo, o teste Post Hoc Bonferroni mostrou que os grupos de crianças de 4-5 e 6-7 anos apresentaram desempenho similar, contudo, as crianças de 4-5 anos se diferenciam de todos os demais grupos ($p<0,05$) por apresentar menor acurácia (média= 44% de acertos). O teste mostrou ainda que o grupo de adultos apresentou maior acurácia em relação aos grupos de crianças de 4-5, 6-7 e 8-9 anos, porém não se diferenciou do grupo de crianças mais velhas de 10-12 anos ($p=1,000$) (média=88% de acertos). Por fim, os grupos de crianças de 8-9 e 10-12 anos não se diferenciaram estatisticamente ($p=0,392$).

Quanto à variável modalidade, o teste Post Hoc Bonferroni indicou que a modalidade AO (Média geral em AO = 53%) se diferenciou significativamente em relação às modalidades VO e AV (para ambas as comparações, $p < 0,05$), e as modalidades VO (Média geral em VO = 71%) e AV (Média geral em AV = 81%) não se diferenciam estatisticamente entre si ($p>0,005$). Destaca-se, desse resultado, o papel preponderante da informação visual na identificação do ato de fala ordem, conforme gráfico 4:

Gráfico 4– Acurácia perceptual dos grupos etários por modalidade na identificação de ordem



Fonte: Elaboração própria (2025)

4.2.4 Pedido

Sobre o ato de fala pedido, a tabela 6 apresenta os resultados descritivos da acurácia perceptual. Pode-se observar os valores das médias do percentual de identificação do pedido bem próximos entre os grupos de 8-9 anos, 10-12 anos e adultos, diferenciando-se dos grupos de 4-5 anos e 6-7 anos, que apresentaram médias de percentual de identificação menores.

Tabela 6- Média e desvio padrão da acurácia perceptual dos grupos por modalidade na identificação do ato de fala pedido

Grupo	Modalidade					
	AO		VO		AV+	
	(%)	P	(%)	P	(%)	P
G1: 4 – 5 anos	5,7	0,32	8,5	0,33	0,9	0,37
G2: 6 – 7 anos	1,3	0,38	3,0	0,27	1,3	0,35
G3: 8 – 9 anos	8,7	0,38	6,0	0,29	2,1	0,38
G4: 10 – 12 anos	8,7	0,35	0,0	0,39	0,0	0,36
G5: 18 – 25 anos	0,0	0,34	5,0	0,35	2,5	0,34

Fonte: Dados da Pesquisa (2025).

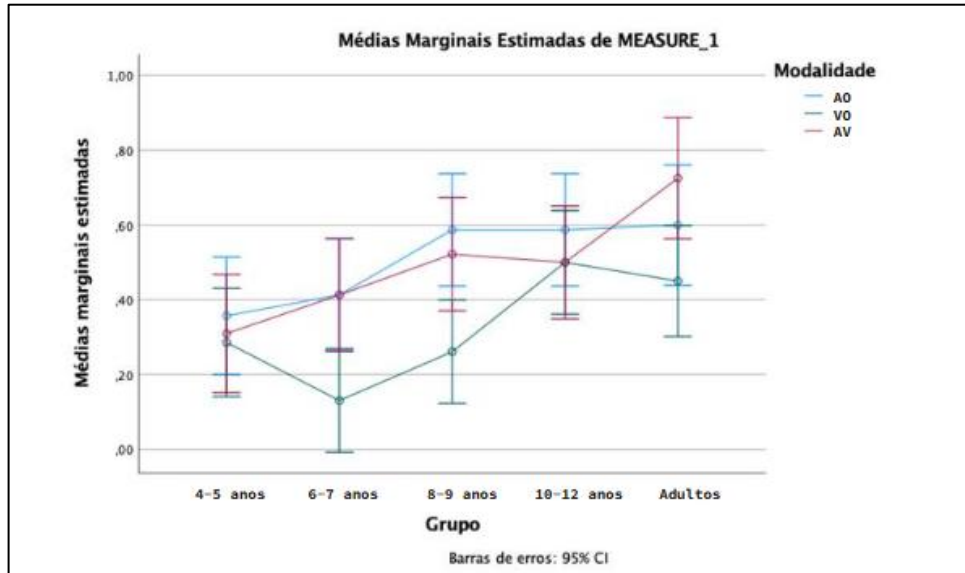
Na análise inferencial, a ANOVA indicou diferença estatística significativa para a modalidade ($F(2,210)=9,562$; $p=0,000$) e para o grupo ($F(4,210)=6,974$; $p=0,000$), contudo não mostrou diferença estatística na interação entre modalidade e grupo $F(8,210)=1,263$, $p=0,265$), ou seja, os resultados estatísticos encontrados para a modalidade equivalem para todos os grupos.

No que diz respeito à variável grupo, o teste Post Hoc Bonferroni indicou que o grupo de adultos se diferenciou estatisticamente apenas dos grupos de crianças de 4-5 e 6-7 anos ($p=0,001$) com média de 59% de acertos. Os demais grupos não se diferenciam entre si e atingiram os percentuais: 4-5 anos média de 31% de acertos, 6-7 anos média de 31% de acertos, 8-9 anos média de 45% de acertos e 10-12 anos média de 52% de acertos na identificação do ato de fala pedido.

No que diz respeito à variável modalidade, o teste Post Hoc Bonferroni indicou que a modalidade VO (Média geral em VO = 32%) se diferenciou estatisticamente de forma significativa em relação às modalidades AO e AV (para ambas as comparações, $p < 0,05$), com média geral em AO = 50% e média geral em AV = 49%. Portanto, pode-se observar que a modalidade VO esteve associada a um desempenho perceptual menos acurado, em contrapartida, as modalidades AO e AV promoveram desempenho mais acurado para perceber

o pedido. O Gráfico 5 ilustra esses resultados.

Gráfico 5 - Acurácia perceptual dos grupos etários por modalidade na identificação do pedido



Fonte: Elaboração própria (2025)

Tendo em vista que o efeito da modalidade de apresentação dos estímulos na acurácia perceptual dos grupos etários foi diferente a depender de cada tipo de ato de fala, buscou-se investigar, então, qual seria a preferência perceptual, auditiva ou visual, por parte dos grupos para cada ato de fala quando as informações auditiva e visual estão em conflito, como é característico na modalidade AV-. Ainda, buscou-se responder se essa preferência perceptual mudaria durante o processo de aquisição prosódica.

4.3 PREFERÊNCIA PERCEPTUAL NA IDENTIFICAÇÃO DOS ATOS DE FALA

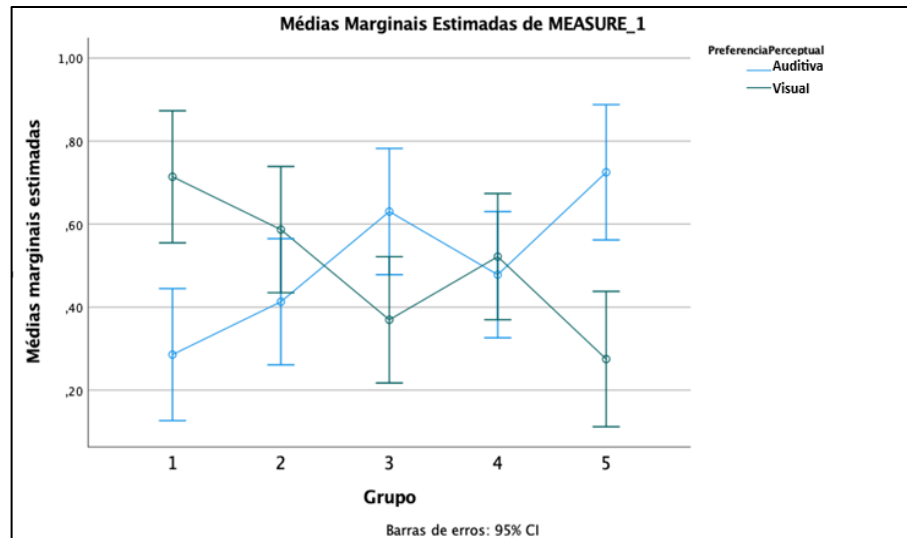
A análise da preferência perceptual auditiva (PA) ou preferência visual (PV) dos grupos foi realizada para cada ato de fala. Inicia-se com a apresentação dos resultados referentes ao ato de fala declarativo.

4.3.1 Declaração

Para o ato de fala declarativo, como resultado, a ANOVA não indicou diferença estatística para a preferência perceptual $F(1,105)=0,034$; $p= 0,854$), porém indicou diferença estatística para grupo ($F(4,105)=0,000$; $p = 0,000$) e para a interação entre a preferência perceptual e grupo ($F(4,105)=4,716$; $p= 0,002$), ou seja, a preferência perceptual foi diferente a depender do grupo. O gráfico 6 ilustra esse resultado.

O gráfico 6 ilustra esse resultado.

Gráfico 6– Preferência perceptual na identificação do ato de fala declarativo por parte dos grupos



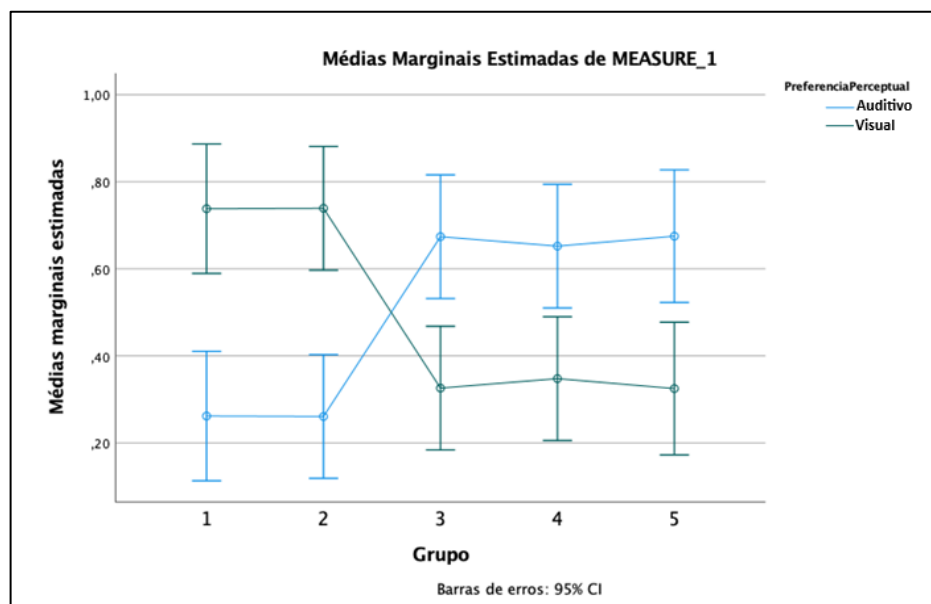
Fonte: Elaboração própria (2025)

Pode-se observar a partir do gráfico acima que a preferência perceptual na identificação do ato de fala declarativo muda ao decorrer do processo aquisitivo. Nota-se, a princípio, uma preferência visual; no entanto, a partir dos 8 anos, identificou-se uma crescente preferência auditiva, que se intensifica no grupo de idade adulta (que apresentou médias PA= 72,5% e PV= 27,5%). Portanto, pode-se entender que há uma inversão da preferência das pistas na percepção do ato de fala declarativo durante a aquisição.

4.3.2 Pergunta

Do mesmo modo, para o ato de fala pergunta, como resultado, a ANOVA não indicou diferença estatística para a preferência perceptual $F(1,105)=0,021$; $p=0,885$), porém indicou diferença estatística para grupo $F(4,105)=0,000$; $p=0,000$) e para a interação entre preferência perceptual e grupo $F(4,105)=9,194$; $p=0,000$), ou seja, a preferência perceptual foi diferente a depender do grupo etário.

Gráfico 7– Preferência perceptual na identificação do ato de fala pergunta por parte dos grupos



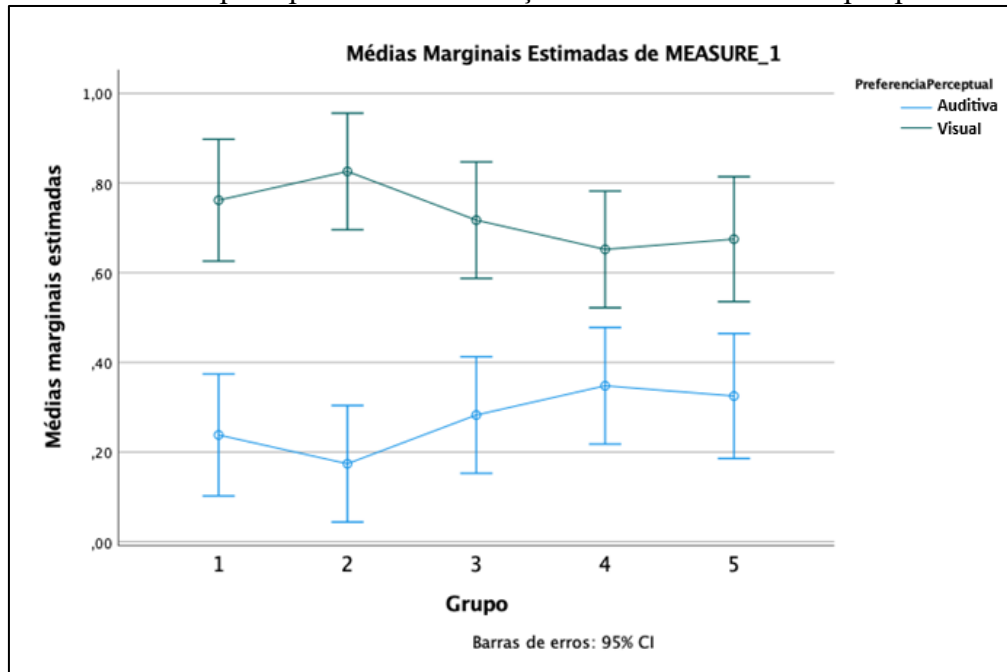
Fonte: Elaboração própria (2025)

Observa-se a partir do gráfico que a preferência perceptual na identificação do ato de fala pergunta, de forma similar ao que se observou no ato declarativo, muda ao decorrer do processo aquisitivo prosódico. Também se percebe, a princípio, uma preferência visual, no entanto, a partir dos 8 anos, identificou-se uma preferência auditiva (com médias PA= 67,5% e PV=32,5%) que se mantém até a idade adulta. Em termos comparativos, essa mudança parece ser mais categórica perceptualmente para o ato de fala pergunta em relação ao ato de fala declarativo, pois o alto índice de preferência auditiva se identifica já aos 8 anos no ato de fala pergunta, enquanto no ato de fala declarativo, ela se estabelece na idade adulta.

4.3.3 Ordem

Na análise do ato de fala ordem, como resultado, a ANOVA indicou diferença estatística significativa para preferência perceptual $F(1,105=56,885; p= 0,000)$ e para grupo $F(4,105)=0,000; p = 0,000)$, porém não mostrou efeito estatístico significativo para a interação entre preferência perceptual e grupo $F(4,105=1,103; p= 0,359)$, ou seja, a preferência perceptual foi a mesma na identificação do ato de fala ordem para todos os grupos ao decorrer da aquisição, como ilustra o gráfico 8.

Gráfico 8- Preferência perceptual na identificação do ato de fala ordem por parte dos grupos



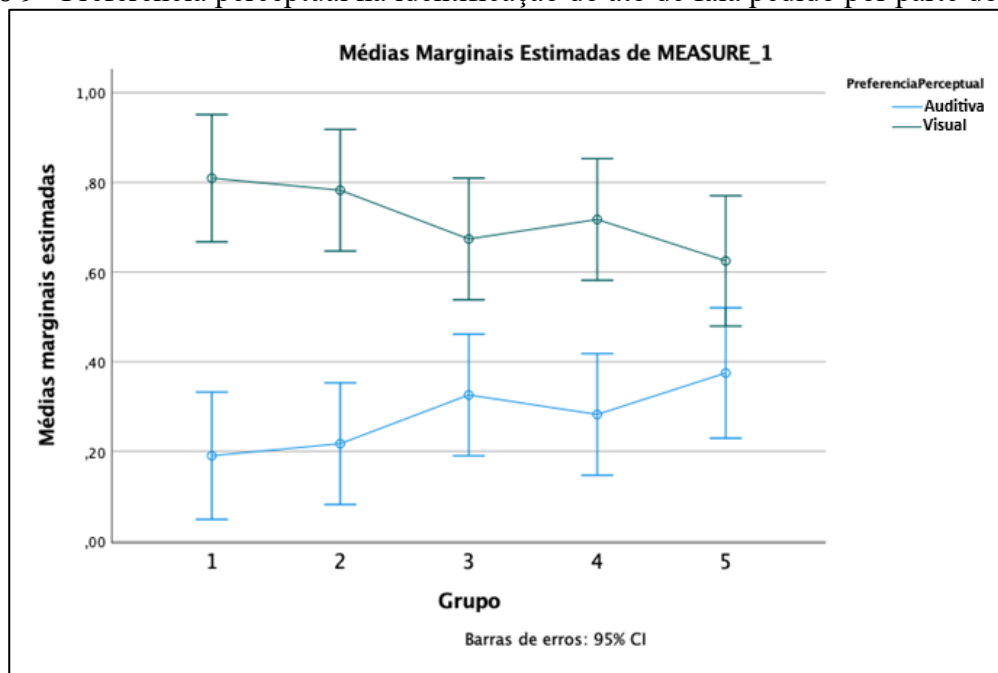
Fonte: Elaboração própria (2025)

O gráfico revela que, durante o processo de aquisição, as pistas visuais foram preferidas para a identificação do ato de fala de ordem. Os dados mostram que, para todos os grupos etários, há uma preferência visual na percepção do ato de fala ordem, ainda que haja variação de médias entre os grupos. Ainda, o teste indicou diminuição da diferença entre preferência visual e auditiva nos dois grupos de maior faixa etária (por exemplo, para G1, PA= 23,8% e PV= 76,2%, enquanto para G4 e G5, aproximadamente, PA = 32,5% e PV = 67,5%).

4.3.4 Pedido

Do mesmo modo, para o ato de fala pedido, como resultado, a ANOVA indicou diferença estatística para a preferência perceptual $F(1,105=50,034; p= 0,000)$ e para grupo $F(4,105)=0,000; p = 0,000)$, porém não mostrou efeito estatístico significativo na interação entre preferência perceptual e grupo $F(4,105=1,131; p= 0,346)$, ou seja, a preferência perceptual foi a mesma para todos os grupos, conforme gráfico 9.

Gráfico 9– Preferência perceptual na identificação do ato de fala pedido por parte dos grupos



Fonte: Elaboração própria (2025)

Nota-se, a partir do gráfico, que a preferência perceptual na identificação do ato de fala pedido se manteve como visual no decorrer do processo aquisitivo. Os resultados indicam, assim como no ato de fala ordem, que desde os 4 anos, há uma preferência visual, a qual se mantém até a idade adulta. Em termos da magnitude da preferência, os três grupos de maior faixa etária se diferenciaram do G1 e do G2 devido à diminuição da diferença entre as preferências visual e auditiva (para G1 e G2, aproximadamente, PA = 20% e PV= 80%, enquanto para G3, G4 e G5, aproximadamente, PA = 32% e PV= 66%).

4.4 ÍNDICE DE EFEITO VISUAL NA IDENTIFICAÇÃO DOS ATOS DE FALA

No que diz respeito ao terceiro objetivo, relativo a verificar se, em condição de estímulos com ruído, as pistas visuais exerceriam efeito compensatório na identificação dos atos de fala, os resultados dos testes ANOVA não indicaram diferença estatisticamente significativa para a condição de apresentação dos estímulos, nem para o grupo e para a interação entre condição e grupo, conforme se identifica na coluna referente ao “valor de p” na tabela 7. Em outras palavras, esses resultados mostram que, do ponto de vista estatístico, o índice de efeito visual não foi maior na condição com ruído assim como não foi maior a depender grupo, indicando, assim, que as pistas visuais não exerceram efeito compensatório na identificação dos atos de fala para todos os grupos.

Tabela 7 – Índice de efeito visual por condição de apresentação dos estímulos

Ato de fala	Média Geral dos Grupos por Condição (%)		Valor de F	Valor de p
	Sem ruído	Com ruído		
Declaração	0,1818	0,1364	Condição (1,105) = 0,400; Grupo (4,105) = 1,564	Condição = 0,528 Grupo = 0,189
Pergunta	0,0455	-0,0455	Condição (1,105) = 1,740; Grupo (4,105) = 0,698	Condição = 0,190 Grupo = 0,595
Ordem	0,2545	0,3091	Condição (1,105) = 0,589; Grupo (4,105) = 0,378	Condição = 0,444 Grupo = 0,824
Pedido	0	-0,0364	Condição (1,105) = 0,362; Grupo (4,105) = 0,699	Condição = 0,549 Grupo = 0,594

Fonte: Elaboração própria (2025)

Por fim, com o intuito de sistematizar os principais resultados apresentados, o quadro 5 sintetiza os achados centrais do estudo por objetivo, os quais são discutidos na seção seguinte.

Quadro 5 - Resumo dos resultados

Objetivos	Principais achados
1) Acurácia perceptual	A acurácia aumentou com a idade e variou conforme o ato de fala. Ordem foi identificada com maior acurácia em todos os grupos, enquanto pedido foi identificado com menor acurácia. A identificação da pergunta melhorou significativamente a partir dos 8 anos e a declaração foi identificada com maior acurácia de forma significativa apenas no grupo adulto.
2) Preferência perceptual	Observou-se mudança da preferência perceptual na identificação dos atos de fala ao longo da aquisição, especialmente para declaração e pergunta com transição de uma preferência visual para a auditiva a partir dos 8 anos. Para ordem e pedido, a preferência visual manteve-se estável desde os quatro anos.
3) Efeito visual	Não houve efeito compensatório da pista visual em condição de apresentação dos estímulos com ruído.

Fonte: elaboração própria (2025)

5 DISCUSSÃO

Tendo em vista os resultados encontrados no presente estudo, a discussão é feita em função dos objetivos propostos.

5.1 COMPARAÇÃO DA ACURÁCIA PERCEPTUAL DOS GRUPOS ETÁRIOS NA IDENTIFICAÇÃO DOS DIFERENTES ATOS DE FALA

Inicialmente, os resultados encontrados no presente estudo corroboraram as hipóteses levantadas e indicaram que a acurácia perceptual na identificação de atos de fala diferenciados prosodicamente varia conforme a idade e o tipo de ato de fala, havendo interação entre esses fatores. Isso sugere que a aquisição da prosódia ocorre de forma gradual e não uniforme, dependendo da complexidade dos enunciados e do desenvolvimento dos ouvintes.

Esse achado está alinhado com a literatura, que aponta para um aumento progressivo das habilidades prosódicas ao longo da infância e adolescência (Kalathottukaren; Purdy, 2017; Filipe et al., 2017), com avanços importantes entre 6 e 9 anos e consolidação mais tardia (Wells; Peppé; Goulondris, 2004; Filipe; Vicente; Frota, 2012; Sousa et al., 2020). Em línguas entoacionais como o português, a distinção entre atos de fala depende fortemente da entoação, exigindo maior maturação perceptiva (Frota; Butler, 2018).

Sob a perspectiva amodal proposta por Rosenblum (2008), a percepção da fala não

resulta da simples soma de pistas auditivas e visuais, mas da captação de informações articulatórias comuns às diferentes modalidades. Assim, o ouvinte acessa propriedades amodais do sinal de fala, integrando diferentes fontes de informação de maneira unificada. O desenvolvimento da acurácia perceptual pode, portanto, ser entendido como o refinamento progressivo da capacidade de detectar essas invariantes amodais ao longo da experiência linguística.

Essa evolução integrada e multimodal encontra respaldo nas proposições de McNeill (2005) e de Esteve-Gibert e Prieto (2018). Na perspectiva de McNeill (2005), palavras e gestos (incluindo as expressões faciais) não são elaborados por canais cognitivos separados; ambos emergem conjuntamente a partir de uma mesma origem mental, denominada ponto de crescimento (*growth point*). Nesse sentido, o desenvolvimento ontogenético da linguagem reflete a coevolução inseparável entre fala e gesto. No contexto da aquisição, Esteve-Gibert e Prieto (2018) sustentam que as competências prosódico-comunicativas se constituem como um sistema integrado desde os primeiros anos de vida, no qual as variações melódicas da fala e os marcadores visuais corporais atuam de maneira sincronizada para expressar intenções pragmáticas.

A importância de articular a estrutura fonológica ao contexto pragmático para a consolidação dessa aquisição é abordada por Cole (2015). A autora argumenta que os contornos e as pistas prosódicas orientam o processamento cognitivo do ouvinte ao codificarem diretamente propriedades gramaticais, o estatuto informacional e os atos de fala. Na arquitetura teórica proposta por Cole (2015), a compreensão da mensagem depende de um modelo suficientemente robusto para relacionar, em tempo real, as variações entoacionais aos seus respectivos contextos situacionais.

Desse modo, as diferentes trajetórias de aquisição observadas entre as categorias pragmáticas investigadas neste estudo podem ser compreendidas a partir das distintas exigências que cada ato de fala impõe a esse sistema integrado e contextualizado. Enquanto a ordem apresenta pistas gestuais-acústicas mais salientes, favorecendo um reconhecimento precoce e consistente, o pedido demanda maior esforço cognitivo. Por depender de nuances pragmáticas mais sutis de polidez e atenuação (Bodolay, 2009), o pedido requer que a criança alcance um nível mais avançado de maturação para interpretar e integrar adequadamente essas sutilezas amodais. De maneira semelhante, a diferenciação entre declarações e perguntas revela maior instabilidade nas fases iniciais do desenvolvimento, em razão da necessidade de um mapeamento refinado dos contornos melódicos transitórios e de suas correspondências faciais, estabilizando-se apenas à medida que o ouvinte amplia seu domínio da prosódia funcional

(Zanchi; Peppé; Zampini, 2022).

Assim, a aquisição da prosódia dos atos de fala do Português Brasileiro configura-se como um processo gradual e heterogêneo, no qual a criança adquire progressivamente a capacidade de identificar, de forma simultânea e preditiva, as coordenadas acústicas e gestuais que compõem a arquitetura unificada da fala (Cole, 2015; Esteve-Gibert; Prieto, 2018; McNeill, 2005; Rosenblum, 2008).

5.3 EFEITO DAS PISTAS PROSÓDICAS AUDITIVAS E VISUAIS NA PERCEPÇÃO DOS ATOS DE FALA

No que se refere ao efeito das pistas prosódicas auditivas e visuais na percepção dos atos de fala ao longo da aquisição, os resultados encontrados corroboraram as hipóteses levantadas. A expectativa de maior acurácia na modalidade audiovisual congruente confirmou-se para os atos de fala de declaração e ordem, nos quais a integração entre pistas auditivas e visuais favoreceu significativamente a identificação perceptual. Em contrapartida, para os atos de pergunta e pedido, a modalidade audiovisual não apresentou vantagem em relação à modalidade exclusivamente auditiva, evidenciando a predominância das pistas acústicas nesses atos de fala.

Quanto às diferentes predominâncias de pistas na percepção dos atos de fala por parte dos grupos em contextos de incongruência audiovisual, os resultados mostraram que tais predominâncias das pistas não variaram de forma linear com a idade, mas em função do tipo de ato de fala. Observou-se predominância auditiva crescente para os atos de declaração e perguntas a partir das faixas etárias mais avançadas, enquanto, para o ato de fala ordem, a predominância manteve-se consistentemente visual desde a infância até a idade adulta. Na percepção do pedido, observou-se uma predominância da pista auditiva, a qual mostrou-se estável ao longo da aquisição.

Em conjunto, esses achados indicam que os efeitos da integração multimodal estão diretamente relacionados à configuração prosódico-pragmática de cada ato de fala. A predominância da pista visual na ordem pode ser interpretada à luz de McNeill (2005), que concebe fala e gesto como partes de um mesmo sistema integrado. Nessa perspectiva, gesto e prosódia constituem uma “unidade de pensamento” (*Growth Point*), sendo coexpressivos e sincronizados. Assim, a expressividade facial associada à ordem não atua como elemento complementar, mas como componente constitutivo de sua força ilocucionária, permitindo ao ouvinte apreender a intenção diretiva como um evento semântico unificado. Ou seja, a acurácia

perceptual pode ser entendida como o refinamento da capacidade de detectar essas invariantes amodais e captar a integridade dessa unidade de pensamento ao longo da experiência linguística (McNeill, 2005).

Essa interpretação converge com a abordagem amodal de Rosenblum (2008), ao sugerir que a percepção da fala não se baseia no processamento isolado de canais, mas na captação integrada do evento articulatório e expressivo. No caso da ordem, isso implica que as propriedades de obrigatoriedade são percebidas de forma particularmente eficiente porque se manifestam de maneira convergente tanto na voz quanto na face do locutor.

Por outro lado, as mudanças na predominância das pistas auditivas e visuais para a distinção entre declaração e pergunta evidenciam a natureza maleável do sistema perceptivo ao longo do desenvolvimento ontogenético, conforme proposto por Zanchi; Peppé; Zampini (2022). O aumento da preferência pela pista auditiva na percepção em idades mais avançadas sugere que, com a experiência, os ouvintes passam a atribuir maior peso às pistas entoacionais, especialmente às variações de frequência fundamental, que são mais robustas para esses contrastes no português (Frota; Butler, 2018).

Nesse sentido, os resultados indicam que a integração multimodal não ocorre de forma automática, mas envolve um processo de aprendizagem no qual o ouvinte passa a explorar seletivamente as pistas mais informativas em cada contexto (Krahmer; Swerts, 2005; Bodolay, 2009; Cole, 2015; Filipe et al., 2017). A variabilidade observada em atos como o pedido sugere que, ao longo da aquisição, a criança ainda está refinando esse sistema e buscando identificar regularidades em um domínio de alta complexidade prosódico-pragmática (Wells; Peppé; Goulondris, 2004; Bodolay, 2009; Cole, 2015; Filipe et al., 2017).

Tendo em vista os dados descritos acima, é possível propor uma cronologia de estabilização da acurácia perceptual dos atos de fala do PB: ordem > declaração > pergunta > pedido. Essa hierarquia sugere que a facilidade na identificação da ordem decorre da transparência de sua unidade de pensamento (McNeill, 2005) e da redundância amodal de suas pistas (Rosenblum, 2008), que a tornam saliente desde os estágios iniciais da aquisição.

A consolidação subsequente da percepção dos atos de fala declaração e pergunta reflete o período de ajuste adaptativo em que o ouvinte, munido de maior experiência linguística, passa a priorizar as variações de frequência fundamental como pistas soberanas (Frota; Butler, 2018). Por fim, o pedido posiciona-se como o ato de fala de aquisição mais tardia, o que pode ser explicado por sua natureza atenuada e pela alta variabilidade pragmática (Bodolay, 2009). Nesses casos, a menor saliência de marcas formais e a dependência de sutilezas contextuais impõem ao ouvinte um desafio maior na extração de regularidades

prosódicas, exigindo um refinamento cognitivo e social que só se estabiliza em idades mais avançadas.

Portanto, a competência prosódica no português brasileiro deve ser compreendida não apenas como resultado da maturação biológica, mas como um processo contínuo de ajuste à variabilidade do input linguístico e às demandas sócio-pragmáticas de interpretação das diferentes forças ilocucionárias (Bodolay, 2009; Filipe; Vicente; Frota, 2012; Cole, 2015; Miranda; Moraes; Rilliard, 2020).

5.5 EFEITO COMPENSATÓRIO DA PISTA VISUAL NA IDENTIFICAÇÃO DOS ATOS DE FALA EM CONDIÇÃO DE RUÍDO

Por fim, os resultados encontrados não corroboraram as hipóteses levantadas para o terceiro objetivo deste trabalho, que buscou verificar se pistas visuais exerceriam efeito compensatório para a identificação dos atos de fala em condição de ruído. Os achados da análise de variância (ANOVA) mostraram um cenário diferente daquele normalmente encontrado na literatura tradicional sobre percepção multimodal. Diferentemente de estudos clássicos e contemporâneos que apontam melhora significativa na precisão perceptual quando a imagem é adicionada ao sinal acústico degradado (Barkhuysen; Krahmer; Swerts, 2010; Miranda et al., 2020; Sumby; Pollack, 1954), os dados desta pesquisa não revelaram diferenças estatisticamente significativas nem para a condição de apresentação dos estímulos, nem para o grupo etário, tampouco para a interação entre essas variáveis (Tabela 7).

No entanto, à luz das abordagens teóricas mais recentes, esse resultado pode ser interpretado sob um novo prisma conceitual. Conforme discutido por Van Engen et al. (2022), a investigação da percepção audiovisual da fala precisa superar visões reducionistas que enxergam a modalidade visual como um mero facilitador aditivo ou um recurso de emergência acionado unicamente sob mascaramento acústico. As autoras argumentam que o processamento linguístico em condições desafiadoras de escuta é intrinsecamente multimodal e preditivo, operando a partir de restrições mútuas entre os sinais auditivo e visual. Sob essa perspectiva, a pista visual não funciona apenas como um recurso complementar ou auxiliar, conforme mostram os resultados obtidos, mas como parte constitutiva do próprio ato de fala.

Esse raciocínio converge diretamente para a proposta de unidade gesto-fala de McNeill (2005) e para os modelos integrados de aquisição da linguagem de Esteve-Gibert e Prieto (2018), que sustentam que fala e gesto não são processados separadamente em canais paralelos. A força ilocucionária do enunciado é construída como uma unidade audiovisual

integrada (Miranda et al., 2024). Assim, tanto adultos quanto crianças não recorrem à informação visual apenas em situações casos de situação ruidosa; os elementos corporais e faciais já fazem parte da própria representação linguística do ato de fala na cognição do ouvinte.

Essa hipótese se fortalece quando observamos o comportamento contrastante entre as categorias pragmáticas. Na ordem, a média aumentou de 0,2545 (sem ruído) para 0,3091 (com ruído), enquanto na declaração houve uma estabilização em 0,1364 na condição ruidosa. Esses resultados indicam dinâmicas perceptivas distintas: enquanto ordens são caracterizadas por marcas faciais e corporais altamente salientes que dão corpo gestual à entoação impositiva (Swerts; Krahmer, 2010), a declaração configura-se como a categoria não marcada ou de base (ou seja, neutra) do sistema linguístico.

Dessa forma, diante do mascaramento acústico, o ouvinte não precisou mudar abruptamente de estratégia perceptiva para a declaração; na escassez de pistas acústicas e gestuais específicas, o sistema cognitivo tendeu a convergir para a interpretação do enunciado neutro estrutural (Cole, 2015). Na ordem, por outro lado, a robustez e a saliência visual dos traços faciais associados à dinamicidade do ato diretivo foram suficientes para ancorar e expandir o reconhecimento, mesmo sob restrição do canal auditivo (Carnaval, 2021).

Por outro lado, os valores negativos observados na pergunta (-0,0455) e no pedido (-0,0364) em condição de ruído reforçam a ideia de que certas categorias pragmáticas dependem de uma combinação consideravelmente mais sutil entre pistas acústicas primárias, como as flutuações de F0, e sinais faciais de suavização ou polidez (Cole, 2015; Miranda et al., 2021). Segundo Altieri e Wenger (2013) e Altieri e Hudock (2014), em situações de alta ambiguidade estrutural, o cérebro pode apresentar um custo de processamento cognitivo elevado quando tenta integrar sinais auditivos e visuais parcialmente obscurecidos pelo ruído.

Nesse caso, o ruído ambiental não rompe a natureza bimodal da fala, mas sabota a capacidade do ouvinte de realizar previsões temporais e discriminações fonéticas finas a partir do rosto do falante, um mecanismo que Van Engen et al. (2022) apontam como crucial para o aproveitamento do ganho visual. Quando essas pistas dinâmicas delicadas sofrem interferência, o ganho perceptivo esperado é severamente reduzido.

O mecanismo subjacente a essa quebra perceptiva envolve a incapacidade do ouvinte de utilizar os movimentos articulatórios pré-fonatórios da face para restringir o tempo de ocorrência e a forma do sinal acústico subsequente (Van Engen et al., 2022). Em contextos de escuta limpa, o córtex visual capta as pistas cinésicas macro e microfaciais milissegundos antes da emissão da onda sonora, permitindo ao cérebro "preparar" e acelerar o processamento no córtex auditivo por meio de um mecanismo de *feedforward* visual (Van Wassenhove; Grant;

Poeppel, 2005). Na pergunta e no pedido, cujas distinções pragmáticas dependem, respectivamente, de pistas estritas de contorno melódico e de expressões gestuais de suavização altamente efêmeras, o ruído degrada a fidelidade acústica necessária para que ocorra o alinhamento com a pista visual correspondente (Cole, 2015; Miranda et al., 2021). Sem essa calibração precisa, os limites da janela de integração multissensorial do ouvinte são desafiados, impedindo que os sinais cinésicos sutis funcionem como âncoras preditivas robustas (Van Engen et al., 2022; Wallace et al., 2004). Como consequência, o sistema cognitivo enfrenta um severo aumento no custo de processamento decorrente da ambiguidade estrutural dos estímulos; em vez de um ganho perceptual facilitador, a tentativa de integrar sinais parcialmente obscurecidos gera uma sobrecarga na capacidade de carga de trabalho do ouvinte, mitigando o benefício esperado e resultando nos índices de acerto negativos observados para esses atos de fala sob mascaramento auditivo (Altieri; Wenger, 2013).

Desse modo, a ausência de um efeito estatístico global na condição multimodal não deve ser entendida como uma separação funcional entre habilidades auditivas e visuais, seja em crianças ou em adultos. Pelo contrário, os resultados reforçam a ideia de que a prosódia multimodal no Português Brasileiro possui natureza constitutiva e integrada. Portanto, o funcionamento do sistema perceptivo audiovisual é governado pelas demandas linguísticas e pragmáticas específicas de cada ato de fala.

Esse comportamento valida o entendimento de que a entonação exerce uma função de ancoragem primária no Português Brasileiro, atuando em consonância com o modelo de arquitetura formal proposta por Cole (2015), o qual prevê que as flutuações melódicas e temporais da fala codificam categorias linguísticas discretas e guiam o processamento cognitivo do ouvinte através do contexto, enquanto a modalidade visual opera de forma indissociável para refinar (e, portanto, não compensar) a interpretação da força ilocucionária (Miranda et al., 2024).

6 CONSIDERAÇÕES FINAIS

Os achados deste estudo demonstraram que a acurácia perceptual na identificação dos atos de fala do Português Brasileiro varia conforme a idade e o tipo de ato de fala, evidenciando que a aquisição da prosódia ocorre de forma gradual e heterogênea (Kalathottukaren; Purdy, 2017; Filipe et al., 2017; Sousa et al., 2020). Houve melhora progressiva das habilidades perceptivas ao longo do desenvolvimento, especialmente entre a infância e a adolescência, em consonância com estudos prévios sobre aquisição prosódica (Wells; Peppé; Goulondris, 2004; Filipe; Vicente; Frota, 2012). A interpretação dos dados apoia modelos teóricos multimodais, segundo os quais fala, gestos e expressões faciais formam um sistema integrado de comunicação, no qual o ouvinte aprende progressivamente a identificar pistas acústicas e visuais associadas às intenções pragmáticas (Rosenblum, 2008; McNeill, 2005; Esteve-Gibert; Prieto, 2018).

Os resultados também mostraram que a integração entre pistas auditivas e visuais varia de acordo com o ato de fala. Declarações e ordens apresentaram benefício significativo da modalidade audiovisual, enquanto perguntas e pedidos dependeram predominantemente das pistas acústicas (Frota; Butler, 2018; Cole, 2015). A ordem destacou-se como o ato de fala mais facilmente reconhecido, devido à alta saliência de suas marcas gestuais e prosódicas, ao passo que o pedido revelou aquisição mais tardia, por exigir interpretação de nuances pragmáticas mais sutis, como polidez e atenuação (McNeill, 2005; Bodolay, 2009). A cronologia de estabilização perceptual observada foi: ordem > declaração > pergunta > pedido (Zanchi; Peppé; Zampini, 2022).

Em relação à avaliação de um efeito compensatório da pista visual em condição de ruído, os resultados não confirmaram a hipótese de maior ancoragem no visual como forma de compensar a qualidade ruidosa da informação auditiva, diferentemente do que apontam estudos clássicos sobre percepção multimodal da fala (Sumby; Pollack, 1954; Barkhuysen; Krahmer; Swerts, 2010; Miranda et al., 2020). Ainda assim, os dados sugerem que a ausência desse efeito não indica separação entre os canais auditivo e visual, mas reforça a natureza constitutiva e integrada da prosódia multimodal (Van Engen et al., 2022; McNeill, 2005). Em atos de fala como ordens, a saliência visual contribuiu para sustentar o reconhecimento mesmo sob mascaramento acústico (Swerts; Krahmer, 2010; Carnaval, 2021), enquanto perguntas e pedidos apresentaram maior vulnerabilidade ao ruído devido à dependência de pistas entoacionais e pragmáticas mais delicadas (Cole, 2015; Miranda et al., 2021).

Assim, conclui-se que a competência prosódica no Português Brasileiro resulta de

um processo contínuo de refinamento perceptivo, guiado tanto pela maturação cognitiva quanto pela experiência linguística e pelas demandas pragmáticas específicas de cada ato de fala (Cole, 2015; Miranda; Moraes; Rilliard, 2020).

Nesse sentido, a pesquisa contribui para a compreensão do desenvolvimento prosódico no português brasileiro, oferecendo dados de referência que podem subsidiar tanto a teorização linguística quanto práticas aplicadas, como a avaliação e a intervenção fonoaudiológica em casos de aquisição de linguagem desviante.

Em suma, o conjunto de achados deste estudo permite reposicionar a prosódia, superando a visão tradicional que a subordina à sintaxe para concebê-la como um sistema central, heterogêneo e constitutivo da linguagem (Cole, 2015). A aquisição dessa competência revela-se intrinsecamente dependente da experiência social e pragmática, visto que demanda a habilidade de decodificar nuances comunicativas complexas. Nesse sentido, os dados demonstram que tal domínio não se estabiliza de forma precoce, mas percorre uma trajetória prolongada de aquisição, alcançando maior consolidação apenas ao longo da adolescência.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALLAN, K.; LAMARQUE, P.; ASHER, R. Speech act theory: overview. **Concise Encyclopedia of Philosophy of Language**, p. 454–466, 1997.
- ALMEIDA, R. A. S. **A prosódia e o processamento on-line de sentenças ambíguas do português brasileiro**. 2017. 182 f. Tese (Doutorado em Letras e Linguística) – Faculdade de Letras, Universidade Federal de Alagoas, Maceió, 2017. Disponível em: Repositório Institucional da UFAL.
- ALMEIDA, R. A. S.; SANTOS, R. S. A prosódia em sentenças sintaticamente ambíguas do Português Brasileiro. **Alfa: Revista de Linguística**, São José do Rio Preto, v. 62, n. 1, p. 11-37, 2018. Disponível em: SciELO.
- ALTIERI, N.; WENGER, M. J. Neural dynamics of audiovisual speech integration under variable listening conditions: an individual participant analysis. **Frontiers in Psychology**, v. 4, art. 615, p. 1-13, 2013. Disponível em: doi.org.
- ALTIERI, N.; HUDOCK, D. Assessing variability in audiovisual speech integration skills using capacity and accuracy measures. **International Journal of Audiology**, v. 53, n. 10, p. 710-718, 2014. Disponível em: doi.org.
- ARMENGAUD, F. **A pragmática**. Tradução de Marcos Marcionilo. São Paulo: Parábola, 2006.
- ARVANITI, A. The phonetics of prosody. In: **Oxford Research Encyclopedia of Linguistics**. Oxford: Oxford University Press, 2020.
- ARVANITI, A. Measuring speech rhythm. In: KNIGHT, R.-A.; SETTER, J. (orgs.). **The Cambridge Handbook of Phonetics**. Cambridge: Cambridge University Press, 2022. p. 312–335.
- AUSTIN, J. L. **Quando dizer é fazer: palavras e ação**. Tradução de Danilo Marcondes de Souza Filho. Porto Alegre: Artes Médicas, 1990.
- BARBOSA, P. A. **Prosódia**. São Paulo: Parábola, 2019.
- BARKHUYSEN, P.; KRAHMER, E.; SWERTS, M. Audiovisual perception of communication problems. **Speech Communication**, v. 52, n. 6, p. 510-520, 2010.
- BARKHUYSEN, P.; KRAHMER, E.; SWERTS, M. The Role of the Auditory and Visual Modalities in the Perception of Statement–Question Contours. **Language and Speech**, v. 63, n. 1, p. 115-139, 2020.
- BATES, E. et al. From gesture to the first word: On cognitive and social prerequisites. In: LEWIS, M.; ROSENBLUM, L. (ed.). **Origins of behavior: communication, pride and fear**. New York: Wiley, 1975. p. 247-303.
- BATES, E. et al. **The emergence of symbols: cognition and communication in infancy**. New York: Academic Press, 1979.

BIAU, E.; SOTO-FARACO, S. Beat gestures modulate auditory integration in rhythmic speech. **Frontiers in Psychology**, v. 4, n. 101, p. 1-10, 2013.

BISOL, L. **Introdução a estudos de fonologia do português brasileiro**. 4. ed. Porto Alegre: Edipucrs, 2005.

BOERSMA, P.; WEENINK, D. **Praat**: doing phonetics by computer (Version 5.0.30). 2009. Disponível em: <http://www.praat.org/>.

BODOLAY, A. N. **Pragmática da entonação**: a relação prosódia/contexto em atos diretivos no português. 2009. 210 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: Repositório UFMG.

BODOLAY, A. O papel da prosódia em enunciados de ordens e pedidos. **Revista Vozes dos Vales**, n. 5, p. 1-28, 2014.

BORGES, G. C. **Percepção da marcação de proeminência em crianças falantes do Português Brasileiro**. 2025. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2025.

CALDAS, V. G. Prosody and sentence processing in Brazilian Portuguese: A visual world paradigm study. **Revista da ABRALIN**, v. 23, n. 1, p. 161-188, 2024.

CAMPBELL, R. The audiology of-looking: as we joint the dots, what do we see? **International Journal of Audiology**, [S. l.], v. 47, n. 2, p. S3-S16, 2008.

CARNAVAL, T. S. **A percepção e a produção prosódica de ordens, pedidos, súplicas e ameaças em diálogos reais e encenados**. 2021. 235 f. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

CARNAVAL, M.; MORAES, J. A.; RILLIARD, A. Focus types in Brazilian Portuguese: multimodal production and perception. **DELTA: Documentação de Estudos em Linguística Teórica e Aplicada**, v. 38, n. 3, p. 1-25, 2022.

CARNAVAL, T.; MIRANDA, L. S.; SOUZA, S. R. Influência visual na identificação de atos de fala no português brasileiro. **Cadernos de Educação e Pesquisa**, v. 1, n. 1, p. 1-15, 2022.

CARNAVAL, M. et al. Funções dos gestos faciais na prosódia audiovisual. In: AVELAR, M.; PACHECO, V.; OLIVEIRA, M. (orgs.). **Linguística e Estudos de Gestos**: interfaces. Campinas: Pontes Editores, 2023. p. 11-39.

CARPENTER, M.; NAGELL, K.; TOMASELLO, M. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. **Monographs of the Society for Research in Child Development**, v. 63, n. 4, 1998.

CASTILHO, A. T. de. **Nova gramática do português brasileiro**. São Paulo: Contexto, 2010.

CAVALCANTE, M. C. **Gesto e prosódia**: uma análise da matriz multimodal da linguagem. João Pessoa: Editora da UFPB, 2012.

CAVALCANTE, M. C. Da unidade gesto-fala à multimodalidade da linguagem: um estudo de caso. **Bakhtiniana**, São Paulo, v. 10, n. 2, p. 54-73, 2015. Disponível em: SciELO.

CAVALCANTE, M. C.; SANTOS, S. L.; SCARPA, E. M. Multimodalidade e aquisição de linguagem: o papel dos gestos e da prosódia. **Cadernos de Letras da UFF**, v. 27, n. 54, p. 111-131, 2017.

CHAFE, W. L. Polyphonic topic development. In: GIVÓN, T. (ed.). **Conversation: Cognitive, communicative and social perspectives**. Amsterdam: John Benjamins, 1997. p. 41-53. (Obra original publicada em 1994).

CHANDRASEKARAN, C.; TRUBANOVA, A.; STILLITTANO, S.; CAPLIER, A.; GHAZANFAR, A. A. The Natural Statistics of Audiovisual Speech. **PLoS Computational Biology**, [S. l.], v. 5, n. 7, p. e1000436, 2009.

CHEN, A. et al. **Development of phrase-level prosody from infancy to late childhood**. 2020.

CHRISTOPHE, A. et al. Bootstrapping lexical and syntactic acquisition. **Language and Speech**, v. 51, n. 1-2, p. 61-89, 2008.

CLARK, E. V. **First language acquisition**. 2. ed. Cambridge: Cambridge University Press, 2009.

COLE, J. Prosody in context: a review. **Laboratory Phonology**, v. 6, n. 1, p. 1-37, 2015.

COLLETTA, J.-M. et al. Multimodal explanation in 6- and 10-year-old French children. **Journal of Child Language**, v. 42, n. 4, p. 850-881, 2015.

CRUZ, M.; SWERTS, M.; FROTA, S. The role of intonation and visual cues in the perception of sentence types: evidence from European Portuguese varieties. **Laboratory Phonology**, v. 8, n. 1, p. 1-24, 2017.

CUTLER, A.; CARTER, D. The predominance of strong initial syllables in the English vocabulary. **Computer Speech and Language**, v. 2, n. 3, p. 133-142, 1987.

DO NASCIMENTO, M. G. R. et al. Desempenho auditivo e visual de crianças na percepção de funções prosódicas emotivas e atitudinais. **Caderno Pedagógico**, v. 21, n. 10, p. e9617, 2024.

EKMAN, P.; FRIESEN, W.; HAGER, J. **The Facial Action Coding System**. 2. ed. Salt Lake City, Utah: Research Nexus, 2002.

ESCARPA, E. M. **Pragmática e aquisição da linguagem**. São Paulo: EDUC/Pontes, 1991.

ESTEVE-GIBERT, N.; GUELLAÏ, B. Prosody in the auditory and visual domains: a developmental perspective. **Frontiers in Psychology**, v. 9, p. 338, 2018.

ESTEVE-GIBERT, N.; PRIETO, P. Prosodic structure shapes the temporal realization of intonation and gesture coordination. **Journal of Speech, Language, and Hearing Research**, v. 56, n. 3, p. 850-864, 2013.

ESTEVE-GIBERT, N.; PRIETO, P. Infants temporally coordinate gesture-speech combinations before they produce their first words. **Speech Communication**, v. 57, p. 301-

316, 2014.

ESTEVE-GIBERT, N. et al. Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. **Infancy**, v. 22, n. 4, p. 435-459, 2017.

ESTEVE-GIBERT, N.; PRIETO, P. Early development of the prosody-meaning interface. In: PRIETO, P.; ESTEVE-GIBERT, N. (ed.). **The Development of Prosody in First Language Acquisition**. Amsterdam: John Benjamins, 2018. p. 227–246.

FERNALD, A. Intonation and communicative intent in mothers' speech to infants: Is the melody the message? **Child Development**, v. 60, n. 6, p. 1497–1510, 1989.

FERNANDES, F. R. **Ordem, focalização e preenchimento em português: sintaxe e prosódia**. 2007. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2007. Disponível em: Repositório Unicamp.

FERNANDES, F. D. M. et al. Desempenho auditivo e visual de crianças na percepção de funções prosódicas emotivas e atitudinais. **Cadernos de Educação e Pesquisa**, [S.l.], v. 20, p. 1-15, 2015.

FILIPPE, M.; VICENTE, S.; FROTA, S. Aquisição prosódica entre os 4 e os 18 anos de idade. In: INTERNATIONAL CONFERENCE OF PSYCHOLOGY AND EDUCATION, 12., 2012. **Proceedings...** [S.l.: s.n.], 2012. p. 444–458.

FILIPPE, M. G. et al. Prosodic development in European Portuguese from childhood to adulthood. **Applied Psycholinguistics**, v. 38, n. 5, p. 1045–1070, 2017.

FÓNAGY, I. **La modulation de la parole**. Paris: Éditions de l'Albatros, 1991.

FÓNAGY, I. Physei/Thesei: l'aspect évolutif d'un debate millénaire. **Faits de Langues**, v. 1, n. 1, p. 29–45, 1993.

FÓNAGY, I. Des fonctions de l'intonation: essay de synthèse. In: **Flambeau**, Tokyo, n. 29, 2003. p. 1-20.

FÓNAGY, I. **Languages within language: an evolutive approach**. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2001.

FONTES, C. S. C. **Desempenho prosódico verbal e gestual em crianças com e sem transtorno fonológico**. 2020. 138 f. Tese (Doutorado em Fonoaudiologia) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Marília, 2020. Disponível em: Repositório Unesp.

FROTA, S. et al. Early Prosodic Development. In: CONFERENCE OF THE STUDENT ORGANIZATION OF LINGUISTICS IN EUROPE (ConSOLE), 22., 2014, Veneza. **Proceedings...** Veneza: University of Venice, 2014. p. 1-20. Disponível em: repositorio.ulisboa.pt.

FROTA, S. et al. **P-ToBI: tools for the transcription of Portuguese prosody**. Lisboa: Laboratório de Fonética, CLUL/FLUL, 2015.

FROTA, S. et al. Infotonal: an interactive database of Romance intonation and its acquisition.

In: INTERNATIONAL CONFERENCE ON SPEECH PROSODY, 8., 2016, Boston. **Proceedings...** Boston: Boston University, 2016. p. 1-5.

FROTA, S.; BUTLER, J. Early development of intonation: perception and production. In: PRIETO, P.; ESTEVE-GIBERT, N. (ed.). **The Development of Prosody in First Language Acquisition**. Amsterdam: John Benjamins, 2018. p. 145–164.

FROTA, S.; NAME, C. Questões de percepção em língua materna. In: FREITAS, M. J.; SANTOS, A. L. (ed.). **Aquisição de língua materna e não materna: questões gerais e dados do português**. Berlin: Language Science Press, 2017. p. 35–50.

FROTA, S.; PRIETO, P.; CRUZ, M.; ESTEVE-GIBERT, N. Intonation in Romance: an overview and new perspectives. In: FÉRY, C.; ISHIKAWA, S.; SHATTUCK-HUFNAGEL, S. (ed.). **The Oxford Handbook of Language Prosody**. Oxford: Oxford University Press, 2022. p. 693–715.

FROTA, S.; SANTOS, R. S. Aquisição da prosódia no português. In: BISOL, L.; COLLISCHONN, G. (org.). **Fonologia do Português e Interfaces**. Londres: Routledge, 2023. p. 80–109.

FROTA, S.; VIGÁRIO, M. Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. **Actas do XV Encontro da Associação Portuguesa de Linguística**, v. 1, p. 533–555, 2000.

GILI-FIVELA, B. Multimodal analyses of audio-visual information: some methods and issues in prosody research. In: FELDHAUSEN, I.; FLIESSBACH, J.; VANRELL, M. M. (ed.). **Methods in Prosody: a Romance Language Perspective**. Berlin: Language Science Press, 2018. p. 83–122.

GLEITMAN, L. R.; WANNER, E. **Language acquisition: the state of the art**. Cambridge: Cambridge University Press, 1982.

GOLDIN-MEADOW, S. Pointing sets the stage for learning language — and creating language. **Child Development**, v. 78, n. 3, p. 741–745, 2007.

GOWDA, D. K. et al. Robust audiovisual speech recognition using multimodal fusion. **IEEE Access**, v. 9, p. 108343–108354, 2021.

GUSSENHOVEN, C. **The phonology of tone and intonation**. Cambridge: Cambridge University Press, 2004.

GUSSENHOVEN, C.; CHEN, A. (ed.). **The Oxford Handbook of Language Prosody**. Oxford: Oxford University Press, 2020. 752 p.

HALLIDAY, M. A. K. **Intonation and Grammar in British English**. The Hague: Mouton, 1967.

HÖHLE, B. Bootstrapping mechanisms in first language acquisition. **Linguistics**, v. 47, n. 2, p. 359–382, 2009.

HOUSE, D. Intonational and visual cues in the perception of interrogative mode in Swedish. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (INTERSPEECH), 2002. **Proceedings...** [S.l.: s.n.], 2002. p. 1957–1960.

HÜBSCHER, I.; ESTEVE-GIBERT, N.; IGUALADA, A.; PRIETO, P. Prosody in the Auditory and Visual Domains. **Frontiers in Psychology**, v. 9, n. 338, p. 1-13, 2018.

IGUALADA, A. et al. Language Development in Early Childhood: The Role of Prosody and Gesture. In: PRIETO, P.; ESTEVE-GIBERT, N. (ed.). **The development of prosody and gesture in early language**. Amsterdam: John Benjamins, 2018. p. 1-17.

IVERSON, J. M.; THELEN, E. Hand, mouth and mind: the genesis of language and gesture. **Frontiers in Developmental Biology**, v. 6, n. 1, p. 19-40, 1999.

IVERSON, J. M.; FAGAN, M. K. Infant vocal-motor coordination: Precursor to the gesture-speech system? **Child Development**, v. 75, n. 4, p. 1053-1066, 2004.

IVERSON, J. M.; GOLDIN-MEADOW, S. Gesture paves the way for language development. **Psychological Science**, v. 16, n. 5, p. 367-371, 2005.

JUSCZYK, P. W. **The discovery of spoken language**. Cambridge, MA: MIT Press, 1997.

KALATHOTTUKAREN, R.; PURDY, S. Prosody perception in typically developing school-aged children. **Journal of Psycholinguistic Research**, v. 46, p. 1211-1229, 2017.

KERBRAT-ORECCHIONI, C. **Os atos de linguagem no discurso: teoria e funcionamento**. Niterói: EdUFF, 2005.

KEMLER NELSON, D. G.; HIRSH-PASEK, K.; JUSCZYK, P. W.; WRIGHT CASSIDY, K. How the prosodic cues in motherese might assist language learning. **Journal of Child Language**, v. 16, n. 1, p. 55-68, 1989.

KENDON, A. **Gesture: visible action as utterance**. Cambridge: Cambridge University Press, 2004.

KIM, J. et al. Seeing prosody: how visual cues contribute to the perception of prominence and intonation. In: INTERNATIONAL CONFERENCE ON SPEECH PROSODY, 7., 2014, Dublin. **Proceedings...** Dublin: Trinity College, 2014. p. 110-114.

KRAHMER, E.; SWERTS, M. The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. **Journal of Memory and Language**, v. 57, n. 3, p. 396-414, 2007.

KRAHMER, E.; SWERTS, M. Audiovisual prosody — introduction to the special issue. **Language and Speech**, v. 52, n. 2-3, p. 129-133, 2009.

Kuhl PK, Andruski JE, Chistovich IA, Chistovich LA, Kozhevnikova EV, Ryskina VL, et al. Cross-language analysis of phonetic units in language addressed to infants. **Science**. 1997;277(5326):684-686.

LADD, D. R. **Intonational Phonology**. Cambridge: Cambridge University Press, 1996. (ou 2. ed., 2008).

LADD, D. R. Defining prosody. In: LADD, D. R. **Simultaneous Structure in Phonology**. Oxford: Oxford University Press, 2014. p. 57-84.

LEITE, A. M. **Processamento de orações relativas ambíguas no Português Brasileiro: a influência de pistas prosódicas**. 2011. 110 f. Dissertação (Mestrado em Linguística) – Universidade Federal da Paraíba, João Pessoa, 2011. Disponível em: Repositório Institucional da UFPB.

LEVINSON, S. C. **Pragmatics**. Cambridge: Cambridge University Press, 1983. (ou ed. 2007).

LIBERMAN, M. Y. **The Intonational System of English**. 1975. Tese (Doutorado em Linguística) – Massachusetts Institute of Technology, Cambridge, 1975.

MACMILLAN, N. A.; CREELMAN, C. D. **Detection theory: a user's guide**. 2. ed. Mahwah: Lawrence Erlbaum Associates, 2005.

MAIA, M.; LOURENÇO-DUARTE, L.; OLIVEIRA, C. V.; ALMEIDA, L. S. Early and late closure in PB: Prosody and parsing. In: INTERNATIONAL CONFERENCE ON SPEECH PROSODY, 2., 2004, Nara. **Proceedings...** Nara: ISCA, 2004. p. 433-436.

MASSARO, D. W.; COHEN, M. M. Perceiving talking faces. **The Journal of the Acoustical Society of America**, v. 97, n. 5, p. 3308–3308, 1995.

MASSARO, D. W. **Perceiving Talking Faces: from speech perception to a behavioral principle**. Cambridge, MA: MIT Press, 1998.

MATHEW, P. M. et al. Visual speech cues in the perception of prosody. In: INTERNATIONAL CONFERENCE ON SPEECH PROSODY, 7., 2014, Dublin. **Proceedings...** Dublin: Trinity College, 2014. p. 1-5.

MATHEW, P. M. et al. The role of visual speech in the perception of prosodic contrast. **The Journal of the Acoustical Society of America**, v. 141, n. 5, p. 3550-3561, 2017.

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

McCANN, D.; PEPPÉ, S. Prosody in autism spectrum disorders: a critical review. **International Journal of Language & Communication Disorders**, v. 38, n. 4, p. 325-338, 2003.

McGURK, H.; MacDONALD, J. Hearing lips and seeing voices. **Nature**, v. 264, n. 5588, p. 746–748, 1976.

McNEILL, D. So you think gestures are nonverbal? **Psychological Review**, v. 92, n. 3, p. 350–371, 1985.

MCNEILL, D. **Hand and Mind: what gestures reveal about thought**. Chicago: University of Chicago Press, 1992.

McNEILL, D. **Gesture and Thought**. Chicago: University of Chicago Press, 2005.

MEHLER, J. et al. A precursor of language acquisition. **Cognition**, v. 29, n. 2, p. 143-178, 1988.

MIRANDA, L. S. **Estudo fonético-perceptivo da entoação de enunciados assertivos, interrogativos e exclamativos no português do Brasil: uma análise multimodal**. 2019. 240 f. Tese (Doutorado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.

MIRANDA, L. S.; MORAES, J. A.; RILLIARD, A. Percepção audiovisual da entoação modal do português do Brasil. **Gradus — Revista Brasileira de Fonologia de Laboratório**, v. 5, n. 1, p. 47–70, 2020.

MIRANDA, L. S. et al. The role of the auditory and visual modalities in the perceptual identification of Brazilian Portuguese statements and echo questions. **Language and Speech**, v. 64, n. 1, p. 3–23, 2021.

MIRANDA, L. S.; MORAES, J. A.; RILLIARD, A. Percepção audiovisual de perguntas-QU e exclamações-QU no português brasileiro. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 32, n. 1, p. 145–178, 2024.

MORAES, J. A. The pitch accents in Brazilian Portuguese declarative and interrogative utterances. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4., 2008, Campinas. **Proceedings...** Campinas: UNICAMP, 2008. p. 1-4.

MORAES, J. A.; COLAMARCO, C. Percepção e produção da entoação de sentenças interrogativas no português do Brasil. In: CONGRESSO INTERNACIONAL DA ASSOCIAÇÃO DE LINGÜÍSTICA E FILOGIA DA AMÉRICA LATINA (ALFAL), 15., 2008, Montevideu. **Anais...** Montevideu: ALFAL, 2008. p. 1-12.

MORAES, J. A.; RILLIARD, A. Prosody and emotion in Brazilian Portuguese. In: ARMSTRONG, M.; HENRIKSEN, N.; VANRELL, M. M. (ed.). **Intonational Grammar in Ibero-Romance: approaches across linguistic subfields**. Amsterdam: John Benjamins, 2016. p. 135–152.

MORGAN, J. L.; DEMUTH, K. (ed.). **Signal to syntax: bootstrapping from speech to grammar in early acquisition**. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.

MUNHALL, K. G. et al. Visual prosody: Facial movement and speech acoustics. **Journal of the Acoustical Society of America**, v. 100, n. 4, p. 2424-2425, 1996.

NAZZI, T.; BERTONCINI, J.; MEHLER, J. Language discrimination by newborns: toward an understanding of the role of rhythm. **Journal of Experimental Psychology: Human Perception and Performance**, v. 24, n. 3, p. 756–766, 1998.

NAZZI, T.; JUSCZYK, P. W.; JOHNSON, E. K. Language discrimination by English-learning 5-month-olds: effects of rhythm and familiarity. **Journal of Memory and Language**, v. 43, n. 1, p. 1–19, 2000.

NESPOR, M.; VOGEL, I. **Prosodic Phonology**. Dordrecht: Foris Publications, 1986.

NESPOR, M.; VOGEL, I. **Prosodic Phonology: with a new foreword**. Berlin: Walter de Gruyter, 2007.

OLIVEIRA, J. S. N.; PACHECO, V.; OLIVEIRA, M. Análise perceptual das frases exclamativas e interrogativas realizadas por falantes de Vitória da Conquista/BA. **Signum:**

Estudos da Linguagem, v. 17, n. 2, p. 354–388, 2014.

OLIVEIRA, T. S. **A percepção e a produção prosódica de ordens e pedidos no português brasileiro**. 2019. 145 f. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.

ÖZÇALIŞKAN, Ş.; GOLDIN-MEADOW, S. Gesture is an auxiliary to mouth and mind. **Child Development**, v. 76, n. 6, p. 1141-1156, 2005.

ÖZÇALIŞKAN, Ş.; GOLDIN-MEADOW, S. Is there a delay between producing a gesture and adding it to speech? Exploring the gesture-speech mismatch. **Journal of Child Language**, v. 38, n. 4, p. 825-844, 2011.

ÖZÇALIŞKAN, Ş. et al. Do parents' gestures facilitate children's learning of gesture-speech combinations? **Developmental Science**, v. 16, n. 5, p. 734-745, 2013.

PARLADÉ, M. V.; IVERSON, J. M. The development of coordinated gesture–speech combinations across the transition to language. **Journal of Child Language**, v. 38, n. 4, p. 790-824, 2011.

PEPPÉ, S.; MCCANN, J. Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. **Clinical Linguistics & Phonetics**, v. 17, n. 4–5, p. 345–354, 2003.

PERES, D. O.; NETTO, W. F.; MEDEIROS, B. R. O papel do estímulo visual na percepção da prosódia: um estudo experimental. **Revista Virtual de Estudos da Linguagem**, v. 8, n. 15, p. 256–269, 2010.

PERES, D. O. et al. The role of the visual stimuli in the perception of prosody in Brazilian Portuguese. In: LABORATORY APPROACHES TO ROMANCE PHONOLOGY (LARP), 5., 2011. **Proceedings...** [S.l.: s.n.], 2011. p. 136–141.

PIERREHUMBERT, J. B. **The Phonology and Phonetics of English Intonation**. 1980. Tese (Doutorado em Linguística) – Massachusetts Institute of Technology, Cambridge, 1980.

PIETROFORTE, A. V. S.; LOPES, I. C. Semântica lexical. In: FIORIN, J. L. (org.). **Introdução à linguística II: princípios de análise**. 2. ed. São Paulo: Contexto, 2014. p. 111-133.

PILLING, M.; THOMAS, R. The role of visual speech in the perception of prosody. In: INTERNATIONAL CONFERENCE ON AUDITORY-VISUAL SPEECH PROCESSING (AVSP), 2011, Volterra. **Proceedings...** Volterra: [s. n.], 2011. p. 1-6.

PINKER, S. **Language Learnability and Language Development**. Cambridge, MA: Harvard University Press, 1984.

PRIETO, P. et al. The Role of Visual Pointers in Intonational Perception. **Language and Speech**, v. 56, n. 4, p. 555-573, 2013.

PRIETO, P.; ESTEVE-GIBERT, N. The development of prosody and gesture. In: PRIETO, P.; ESTEVE-GIBERT, N. (ed.). **The development of prosody and gesture in early language**. Amsterdam: John Benjamins, 2018. p. 1-17.

PRINCE, E. F. Toward a taxonomy of given-new information. In: COLE, P. (ed.). **Radical Pragmatics**. New York: Academic Press, 1981. p. 223-255.

RILLIARD, A. et al. Multimodal perception of French social attitudes. **Language and Speech**, v. 57, n. 2, p. 211-235, 2014.

RUSIEWICZ, H. L. Synchronization of prosodic stress and gesture: a dynamic systems perspective. In: GESTURE AND SPEECH IN INTERACTION (GESPIN), 2., 2011. **Proceedings...** [S.l.: s.n.], 2011. p. 109-114.

RUSIEWICZ, H. L.; ESTEVE-GIBERT, N. Set in time: temporal coordination of prosody and gesture in the development of spoken language production. In: PRIETO, P.; ESTEVE-GIBERT, N. (ed.). **The Development of Prosody in First Language Acquisition**. Amsterdam: John Benjamins, 2018. p. 103-124.

SANTOS, P. S. **Ordem ou pedido?** Como os brasileiros interpretam atos de fala produzidos por aprendizes de PLE. 2009. 152 f. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

SCHWARTZ, J.-L.; BERTHOMMIER, F.; SAVARIAUX, C. Seeing to hear better: evidence for early audio-visual interactions in speech intonational perception. **Cognition**, v. 93, n. 2, p. B63-B78, 2004.

SEARLE, J. R. **Speech Acts**: an Essay in the Philosophy of Language. Cambridge: Cambridge University Press, 1969.

SEARLE, J. R. **Intentionality**: an Essay in the Philosophy of Mind. Cambridge: Cambridge University Press, 1981.

SELKIRK, E. O. On the major class features and syllable theory. In: ARONOFF, M.; OEHRLE, R. T. (ed.). **Language Sound Structure**: studies in Phonology Presented to Morris Halle by His Teacher and Students. Cambridge, MA: MIT Press, 1984. p. 107-136.

SERRA, C. R. **Realização e percepção de fronteiras prosódicas no português do Brasil**: fala espontânea e leitura. 2009. 235 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

SERRA, C. R. A interface prosódia-sintaxe e o fraseamento prosódico no português do Brasil. **JOSS - Journal of Speech Science**, v. 5, n. 1, p. 47-86, 2016.

SILVA, J. C. B. et al. Percepção de interrogativas globais entre variedades do português do Brasil. **Filologia e Linguística Portuguesa**, v. 20, p. 11-25, 2017.

SILVA, C. G. C. **A influência das fronteiras prosódicas no processamento de sentenças ambíguas**: uma investigação experimental. 2015. 248 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2015.

SILVA, C. G.; CARNAVAL, M.; MORAES, J. A. Atos de fala diretivos em português e em espanhol: uma análise acústica comparativa. In: CONGRESSO INTERNACIONAL DA ALFAL, 19., 2020. **Anais...** [S.l.: s.n.], 2020.

SILVA, M. C. C. P.; NAME, M. C. L. A interface entre prosódia e aquisição do léxico. **Revista**

de Estudos da Linguagem, Belo Horizonte, v. 22, n. 1, p. 11-40, 2014.

SILVA, L. S.; MORAES, J. A.; RILLIARD, A. Percepção audiovisual de atos de fala modais no português brasileiro. **Alfa: Revista de Linguística**, São José do Rio Preto, v. 64, e11534, p. 1-28, 2020.

SILVA, Í. O. et al. A sensibilidade de bebês brasileiros a pistas prosódicas de fronteiras de sintagma entoacional na fala dirigida à criança. **Letrônica**, v. 7, n. 1, p. 4-25, 2014.

SILVERMAN, K. E. A. et al. ToBI: a standard for labeling English prosody. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP), 1992, Banff. **Proceedings...** Alberta: University of Alberta, 1992. p. 867-870.

SNEDEKER, J.; YUAN, S. Effects of prosodic and lexical constraints on parsing in young children (and adults). **Journal of Memory and Language**, v. 58, n. 2, p. 574-608, 2008.

SNOW, C. E. Issues in the study of input: fine-tuning, universality, individual and developmental differences, and necessary causes. In: FLETCHER, P.; MACWHINNEY, B. (org.). **The handbook of child language**. Oxford: Blackwell, 1995. p. 180-193.

SONCIN, G. Prosodic boundary in Brazilian Portuguese: the relation between auditory perception and phonetic cues. In: VII CONGRESO INTERNACIONAL DE FONÉTICA EXPERIMENTAL (CIFE), 2017, Madrid. **Tendencias actuales en fonética experimental...** Madrid: UNED, 2017. p. 83-86.

SONCIN, G.; TENANI, L. E. Variações de F0 e configurações de frase entoacional: análise de estruturas contrastivas. **Domínios de Linguagem**, v. 10, n. 2, p. 534-558, 2016.

SONCIN, G. C.; TENANI, L. E.; BERTI, L. C. Percepção da distinção entre os atos de fala de afirmação e de interrogação por crianças com desenvolvimento fonológico típico e atípico. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 25, n. 4, p. 1969-1996, 2017.

SOUSA, L. S. et al. Percepção audiovisual de atos de fala modais no português brasileiro. **Alfa: Revista de Linguística**, São José do Rio Preto, v. 64, e11534, p. 1-28, 2020.

SOUZA, L. S. M. **Percepção audiovisual de modalidades entoacionais no português brasileiro**. 2018. 142 f. Dissertação (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

SPLENDORÉ, K. M. et al. Investigação da prosódia e da linguagem na interação mãe-bebê. **Working Papers em Linguística**, v. 20, n. 1, p. 172-188, 2019.

SRINIVASAN, R. J.; MASSARO, D. W. Perceiving prosody from the face and voice: distinguishing statements from echoic questions in English. **Language and Speech**, v. 46, n. 1, p. 1-22, 2003.

SUMBY, W. H.; POLLACK, I. Visual contribution to speech intelligibility in noise. **Journal of the Acoustical Society of America**, v. 26, n. 4, p. 583-587, 1954.

SWERTS, M.; KRAHMER, E. Audiovisual prosody and feeling of knowing. **Journal of Memory and Language**, v. 53, n. 1, p. 81-94, 2005.

SWERTS, M.; KRAHMER, E. Facial expression and prosodic prominence: effects of modality and facial area. **Journal of Phonetics**, v. 36, n. 2, p. 219-238, 2008.

SWERTS, M.; KRAHMER, E. Visual prosody as a cue to prominence: evidence from a reaction-time experiment. **Journal of Phonetics**, v. 38, n. 2, p. 198-205, 2010.

TENANI, L. E. **Domínios prosódicos no português**. 2002. 210 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2002.

TENANI, L. E.; SONCIN, G.; BERTI, L. C. Representação fonológica e percepção de fala: o papel da pausa. **Diacrítica**, v. 33, n. 2, p. 4-22, 2019.

TOMASELLO, M. **Constructing a language**: a usage-based theory of language acquisition. Cambridge: Harvard University Press, 2003.

VAN WASSENHOVE, V.; GRANT, K. W.; POEPPPEL, D. Visual speech speeds up the neural processing of auditory speech. **Proceedings of the National Academy of Sciences**, v. 102, n. 4, p. 1181-1186, 2005.

VIGÁRIO, M.; FROTA, S. Aquisição da prosódia I: uma categorização das produções sonoras de e para a criança. **Análise Psicológica**, v. 10, n. 4, p. 457-478, 1992.

Vihman, M. M. (2019). Advances in early speech production: Interactions with maturation, perception and learning. In J. S. Horst & J. von Koss Torkildsen (Eds.), **International handbook of language acquisition** (pp. 173-194). Routledge/Taylor & Francis Group.

VOLTERRA, V. et al. From gesture to language: the emergence of symbolic abilities. In: KITA, S. (ed.). **Pointing**: where language, culture, and cognition meet. Mahwah, NJ: Lawrence Erlbaum Associates, 2005. p. 31-47.

WAGNER, P.; MALISZ, Z.; KOPP, S. Gesture and Speech in Interaction: an overview. **Speech Communication**, v. 57, p. 209-232, 2014.

WALLACE, M. T. et al. Unifying multisensory signals across time and space. **Experimental Brain Research**, v. 158, n. 2, p. 252-258, 2004.

WELLS, B.; PEPPÉ, S.; GOULANDRIS, N. Intonation development from five to thirteen. **Journal of Child Language**, v. 31, n. 4, p. 749-778, 2004.

ZANCHI, P.; PEPPÉ, S.; ZAMPINI, L. Prosodic development from 4 to 10 years: data from the Italian adaptation of the PEPS-C. **Speech Communication**, v. 144, p. 10-19, 2022.

ANEXOS

ANEXO 1 - Aprovação do Comitê de Ética em Pesquisa (Processo nº 6.018.084).

	UNESP - FACULDADE DE FILOSOFIA E CIÊNCIAS - CAMPUS DE MARÍLIA									
PARECER CONSUBSTANCIADO DO CEP										
DADOS DO PROJETO DE PESQUISA										
Título da Pesquisa: Percepção auditiva e visual de sentenças declarativas, interrogativas e imperativas na aquisição do Português Brasileiro										
Pesquisador: LIVIA MAYRA ROQUE DE OLIVEIRA										
Área Temática:										
Versão: 1										
CAAE: 68240823.4.0000.5406										
Instituição Proponente: Centro de Estudos da Educação e Saúde										
Patrocinador Principal: UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO										
DADOS DO PARECER										
Número do Parecer: 6.018.084										
Apresentação do Projeto:										
O projeto intitulado "Percepção auditiva e visual de sentenças declarativas, interrogativas e imperativas do Português Brasileiro" foi submetido para análise ética deste CEP.										
O projeto apresenta todos os elementos textuais para análise ética, com boa contextualização do tema e embasado cientificamente com bibliografia atualizada, deixa claro hipótese e justificativa do estudo. Também apresenta os procedimentos metodológicos, os quais estão bem descritos, delineados e de acordo com os objetivos que pretende alcançar.										
Trata-se de um estudo experimental, transversal, quantitativo e prospectivo.										
A amostra será composta de 50 participantes de ambos os gêneros, organizados em grupos de diferentes faixas etárias. Participarão 40 crianças com idades entre quatro anos e 12 anos e 11 meses e 10 adultos com idades entre 18 anos e 30 anos.										
A coleta será realizada no Centro de Estudos da Educação e da Saúde/ Unesp/Marília, devidamente autorizado.										
Os experimentos consistirão em testes de identificação de tipos de sentenças do Português Brasileiro. Os participantes estarão dispostos em frente a um computador com fones de ouvido e serão apresentadas diferentes sentenças em quatro condições (auditivo, visual, audiovisual congruente, audiovisual incongruente) para que ele faça sua opção de resposta de acordo com que ouviu.										
<table border="0" style="width: 100%;"> <tr> <td style="width: 50%;">Endereço: Avenida Hygino Muzzi Filho, 737.</td> <td style="width: 50%;">CEP: 17.525-900</td> </tr> <tr> <td>Bairro: Campus Universitário</td> <td></td> </tr> <tr> <td>UF: SP</td> <td>Município: MARÍLIA</td> </tr> <tr> <td>Telefone: (14)3402-1346</td> <td>E-mail: cep.marilia@unesp.br</td> </tr> </table>			Endereço: Avenida Hygino Muzzi Filho, 737.	CEP: 17.525-900	Bairro: Campus Universitário		UF: SP	Município: MARÍLIA	Telefone: (14)3402-1346	E-mail: cep.marilia@unesp.br
Endereço: Avenida Hygino Muzzi Filho, 737.	CEP: 17.525-900									
Bairro: Campus Universitário										
UF: SP	Município: MARÍLIA									
Telefone: (14)3402-1346	E-mail: cep.marilia@unesp.br									



UNESP - FACULDADE DE
FILOSOFIA E CIÊNCIAS -
CAMPUS DE MARÍLIA



Continuação do Parecer: 6.018.084

Objetivo da Pesquisa:

1) Objetivo principal

Objetivo geral será comparar o desempenho perceptual auditivo e visual de crianças de diferentes faixas etárias na identificação de sentenças declarativas, interrogativas e imperativas do Português Brasileiro (PB)

2) Objetivos específicos

Objetivos específicos são: a) Avaliar o papel de pistas auditivas e visuais (auditiva ou visual) na identificação de sentenças declarativas, interrogativas e imperativas do PB de modo comparativo entre grupos de crianças de diferentes faixas etárias; b) Avaliar a robustez da pista visual na identificação de sentenças declarativas, interrogativas e imperativas do PB de modo comparativo entre os grupos etários; c) Avaliar se o desempenho perceptual auditivo e visual dos grupos etários seria diferente a depender do tipo de sentença a ser identificado.

Avaliação dos Riscos e Benefícios

Página 2 / 5

1) Riscos

Segundo a pesquisadora não haverá riscos potenciais, porém em caso de incomodo, constrangimento e/ou desconforto a pesquisadora refere que o participante poderá desistir sem nenhum ônus ou prejuízo.

2) Benefícios

As pesquisadoras afirmam que os resultados obtidos pela pesquisa trarão benefícios para diferentes frentes de conhecimento linguístico e, acredita-se, proporcionarão desdobramentos significativos para a área de avaliação fonoaudiológica. A equipe escolar será convidada a conhecer os resultados da pesquisa.

Comentários e Considerações sobre a Pesquisa:

A pesquisa analisada trará contribuições científicas para as áreas de Linguística e Fonoaudiologia, pois seus resultados poderão fomentar discussões mais aprofundadas por profissionais dessas áreas a fim de que estes compreendam o impacto da prosódia no desenvolvimento da linguagem falada do Português Brasileiro.

Endereço: Avenida 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



UNESP - FACULDADE DE
FILOSOFIA E CIÊNCIAS -
CAMPUS DE MARÍLIA



Continuação do Parecer: 6.018.084

Considerações sobre os Termos de apresentação obrigatória:

1)O Projeto de pesquisa "Percepção auditiva e visual de sentenças declarativas, interrogativas e imperativas do Português Brasileiro" está bem delineado e contém todos os elementos textuais necessários para análise ética deste CEP.

Todos os documentos necessários encontram-se redigidos e apresentados em conformidade as resoluções vigentes para a aprovação no CEP.

2)O documento "autorização da instituição" onde será realizada a pesquisa, contém o nome da pesquisa e o nome do pesquisador e, em seu conteúdo a responsável autoriza a pesquisa, dando fé com sua assinatura e carimbo institucional.

3)Os documentos do TCLE e TALE estão redigidos adequadamente, com linguagem clara e acessível de forma clara, o nome e objetivo da pesquisa, também informa ao participante as intervenções/procedimentos que terá que participar.

4)O documento "Cronograma" a pesquisadora esclarece que os participantes serão recrutados somente após a aprovação do CEP. As datas apresentadas estão adequadas quanto a tramitação da documentação. Assim sendo, considero que está adequado e exequível.

Página 3 / 5

Recomendações:

Nada consta

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 20/04/2023, após acatar o parecer do membro relator previamente aprovado para o presente estudo e atendendo a todos os dispositivos das resoluções 466/2012, 510/2016 e complementares, bem como ter aprovado o Termo de Consentimento Livre e Esclarecido como também todos os anexos incluídos na pesquisa, resolve APROVAR a pesquisa "Percepção auditiva e visual de sentenças declarativas, interrogativas e imperativas na aquisição do Português Brasileiro".

Endereço: Avenida Hygino Muzzi Filho, 737.

Bairro: Campus Universitário

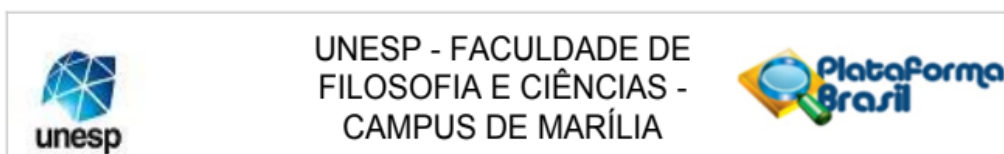
CEP: 17.525-900

UF: SP

Município: MARILIA

Telefone: (14)3402-1346

E-mail: cep.marilia@unesp.br



Continuação do Parecer: 6.018.084

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2064975.pdf	20/03/2023 23:47:22		Aceito
Cronograma	cronograma_.pdf	20/03/2023 23:46:10	LIVIA MAYRA ROQUE DE OLIVEIRA	Aceito
Projeto Detalhado / Brochura Investigador	Projeto.pdf	13/03/2023 21:08:54	LIVIA MAYRA ROQUE DE OLIVEIRA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE.pdf	13/03/2023 21:07:25	LIVIA MAYRA ROQUE DE OLIVEIRA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_ilustrado.pdf	13/03/2023 21:07:09	LIVIA MAYRA ROQUE DE OLIVEIRA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Projeto.pdf	13/03/2023 21:06:56	LIVIA MAYRA ROQUE DE OLIVEIRA	Aceito
Folha de Rosto	Folha_de_Rosto.pdf	13/03/2023 21:05:45	LIVIA MAYRA ROQUE DE OLIVEIRA	Aceito
Outros	autorizacaoescola.pdf	22/02/2023 00:16:14	LIVIA MAYRA ROQUE DE OLIVEIRA	Aceito
Solicitação registrada pelo CEP	autorizacao_pesquisa.pdf	22/02/2023 00:08:28	LIVIA MAYRA ROQUE DE OLIVEIRA	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

MARILIA, 24 de Abril de 2023

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

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

ANEXO 2 - Termo de Consentimento Livre e Esclarecido – TCLE

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Estamos realizando uma pesquisa na EMEF Prof Eduardo Romualdo de Souza Ribeirão Preto/SP e na Faculdade de Filosofia e Ciências (FFC) – UNESP/Marília, intitulada **“Percepção auditiva e visual de sentenças declarativas, interrogativas e imperativas na aquisição do Português Brasileiro”** e gostaríamos que participasse da mesma. O objetivo desta é comparar o desempenho perceptual auditivo e visual de crianças entre 4 anos e 12 anos e 11 meses na identificação de sentenças declarativas, interrogativas e imperativas do Português Brasileiro, bem como avaliar a) o papel de pistas auditivas e visuais (auditiva ou visual) na identificação das sentenças de modo comparativo entre grupos de crianças de diferentes faixas etárias; b) avaliar a robustez da pista visual na identificação das sentenças de modo comparativo entre os grupos etários; e c) avaliar se o desempenho perceptual auditivo e visual dos grupos etários seria diferente a depender do tipo de sentença a ser identificado.

A fim de investigar a percepção auditiva e visual de diferentes tipos de sentenças, trabalharemos também com um grupo controle, formado por falantes adultos do Português Brasileiro com faixa etária entre 18 anos e 30 anos, que realizarão os mesmos testes que os grupos de crianças.

Convidamos você a participar da pesquisa a fim de que possamos desenvolvê-la e atender os objetivos delineados. Participar desta pesquisa é uma opção e, no caso de não aceitar participar ou de desistir em qualquer fase da pesquisa, fica assegurado o atendimento de sua escolha imediatamente.

Caso aceite participar da pesquisa, esclarecemos e enfatizamos que: a) as crianças e os adultos participarão da pesquisa em três etapas: para realização de triagem fonoaudiológica, para os experimentos de percepção, que serão realizados na EMEF Prof Eduardo Romualdo de Souza Ribeirão Preto/SP e na Unesp- Marília; b) a divulgação dos resultados desta pesquisa a identidade dos sujeitos participantes; c) a identidade dos participantes será totalmente preservada; d) será garantida devolutiva para as crianças triadas ou avaliadas e que apresentarem alterações.

- No caso de responsável por criança participante:

Solicitamos que preencha o termo abaixo, caso autorize a participação.

Eu, _____, portador do RG _____, responsável pelo(a) participante _____ (comunidade) _____, o autorizo a participar da pesquisa intitulada **“Percepção auditiva e visual de sentenças declarativas, interrogativas e imperativas na aquisição do Português Brasileiro”** a ser realizada na EMEF Prof Eduardo Romualdo de Souza Ribeirão Preto/SP. Declaro ter recebido as devidas explicações sobre a referida pesquisa e concordo que minha desistência poderá ocorrer em qualquer momento sem que ocorram 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 do responsável

Nome da criança

ANEXO 3 - Termo de Assentimento Livre e Esclarecido – TALE

TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO

Convido você a participar da pesquisa chamada “Percepção auditiva e visual de sentenças declarativas, interrogativas e imperativas na aquisição do Português Brasileiro”. Seus pais permitiram que você participe. Esta é uma pesquisa que busca entender melhor a forma como as crianças percebem diferentes tipos de frases em diferentes situações de comunicação.

As crianças que irão participar dessa pesquisa têm entre 4 anos e 12 anos e 11 meses de idade; irão participar crianças que não tem nenhuma dificuldade de ouvir, ver e/ou falar. É importante que crianças como você participem para que nós, pesquisadores, possamos estudar e entender melhor a percepção das crianças. Se não houver crianças que participem, não é possível fazer o estudo.

A pesquisa será feita na EMEF Prof Eduardo Romualdo de Souza Ribeirão Preto/SP. Ao aceitar participar da pesquisa, você se sentará na frente de um computador e usará um fone de ouvido para responder, entre duas alternativas, qual a tipo de frase apresentada. É bastante simples. Ninguém saberá que você está participando da pesquisa. Suas respostas ficarão gravadas no computador e serão analisadas para tirarmos algumas conclusões de como é a percepção dos diferentes tipos de frases e publicaremos esses resultados em revistas científicas, mas ninguém saberá quais foram as crianças que participaram da pesquisa.

Lembro que participar ou não é um direito seu. Se não aceitar, basta me informar. Mas, caso você aceite, será excelente, pois você ajudará a produzir conhecimento científico. Além disso, caso você tenha alguma dúvida, pode me perguntar em qualquer momento. Esse é o meu telefone: (16) 991943538; sou a pesquisadora Lívia Mayra Roque de Oliveira.

Caso aceite participar, coloque o seu nome na declaração abaixo:

Declaração de participação do participante:

Eu _____ aceito participar da pesquisa “Percepção auditiva e visual de sentenças declarativas, interrogativas e imperativas na aquisição do Português Brasileiro”, que tem o objetivo de comparar o desempenho perceptual audiovisual em diferentes grupos na identificação de diferentes tipos de frase do Português Brasileiro. Entendi como o processo de participar da pesquisa acontecerá. Entendi também que posso dizer “sim” e participar, mas que, a qualquer momento, se eu mudar de ideia, posso desistir e meus dados serão retirados da pesquisa. Os pesquisadores tiraram minhas dúvidas e conversaram com os meus responsáveis.

Recebi uma cópia deste termo de assentimento, li e concordo em participar da pesquisa. Ribeirão Preto, ____ de _____ de _____.

Assinatura do menor

Assinatura do(a) pesquisador(a)

ANEXO 4 - Termo de Assentimento Livre e Esclarecido ilustrado – TALE

Termo de Assentimento Livre e Esclarecido^(Anexo)

4 a 7 anos ou crianças não alfabetizadas

"PERCEÇÃO AUDITIVA E VISUAL DE SENTENÇAS DECLARATIVAS, INTERROGATIVAS E IMPERATIVAS NA AQUISIÇÃO DO PORTUGUÊS BRASILEIRO"

Existem atividades como as crianças percebem diferentes tipos de frases em diferentes situações de comunicação.

Se conversarmos com sua mãe e sua mãe e uma professora que, se você quiser, você pode nos ajudar.

Caso não queira nos ajudar, não tem problema algum. Caso tenha a idade de idade, pode desistir a qualquer momento.

Se decidir nos ajudar:



Você estará feliz, logo após a gente e que você na tela do computador. Sua resposta será gravada.

Você quer participar?

✓ ☐ Nome da criança
✗ ☐ Nome da criança

Responsáveis pelo estudo

Luiz Mauro Roque de Oliveira (0209044-020)

Coordenadora Carolina Rocha Soares (0100000-000)

Coordenador de Eliza em Português (000100000-000)

Termo de Assentimento Livre e Esclarecido^(Anexo)

8 a 12 anos ou crianças alfabetizadas

"PERCEÇÃO AUDITIVA E VISUAL DE SENTENÇAS DECLARATIVAS, INTERROGATIVAS E IMPERATIVAS NA AQUISIÇÃO DO PORTUGUÊS BRASILEIRO"

Existem atividades como as crianças percebem diferentes tipos de frases em diferentes situações de comunicação.

Se conversarmos com sua mãe e sua mãe e uma professora que, se você quiser, você pode nos ajudar.

Caso não queira nos ajudar, não tem problema algum. Caso tenha a idade de idade, pode desistir a qualquer momento.

Se decidir nos ajudar:



Você estará feliz, logo após a gente e que você na tela do computador. Sua resposta será gravada.

Você quer participar?

✓ ☐ Nome da criança

Responsáveis pelo estudo

Luiz Mauro Roque de Oliveira (0209044-020)

Coordenadora Carolina Rocha Soares (0100000-000)

Coordenador de Eliza em Português (000100000-000)