

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA**  
**FACULDADE DE LETRAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

**Aline Cristina Cardoso**

**A Prosódia Facial e Vocal na Interação Bebê-Cuidador:** Explorando a Multimodalidade na  
Aquisição da Linguagem

Juiz de Fora

2025

**Aline Cristina Cardoso**

**A Prosódia Facial e Vocal na Interação Bebê-Cuidador:** Explorando a Multimodalidade na  
Aquisição da Linguagem

Trabalho de Dissertação de Mestrado apresentado  
ao Programa de Pós-graduação em Linguística da  
Universidade de Juiz de Fora, como requisito  
parcial à obtenção do título de Mestre em  
Psicolinguística.

Orientadora: Profa. Dr<sup>a</sup>. Cristina Name

Coorientadores: Profa. Dr<sup>a</sup>. Sandra Madureira (PUCSP)

Prof. Dr. Mario Augusto de Souza Fontes (PUCSP)

Juiz de Fora

2025

**Aline Cristina Cardoso**

**A Prosódia Facial e Vocal na Interação Bebê-Cuidador:** Explorando a Multimodalidade na Aquisição da Linguagem

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de mestra em Linguística. Área de concentração: Linguística.

Aprovada em 26 de setembro de 2025.

**BANCA EXAMINADORA**

**Profª Drª Maria Cristina Lobo Name - Orientadora**

Universidade Federal de Juiz de Fora

**Profª Drª Sandra Madureira - Coorientadora**

Pontifícia Universidade Católica de São Paulo

**Prof. Dr. Mario Augusto de Souza Fontes - Coorientador**

Pontifícia Universidade Católica de São Paulo

**Profª Drª Marianne Carvalho Bezerra Cavalcante**

Universidade Federal da Paraíba

**Prof. Dr. Paulo Vinícius Ávila Nóbrega**

Universidade Estadual da Paraíba

Juiz de Fora, 04/09/2025.



Documento assinado eletronicamente por **Maria Cristina Lobo Name, Professor(a)**, em 29/09/2025, às 17:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Paulo Vinícius Ávila Nóbrega, Usuário Externo**, em 06/10/2025, às 13:17, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sandra Madureira, Usuário Externo**, em 06/10/2025, às 18:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marianne Carvalho Bezerra Cavalcante, Usuário Externo**, em 07/10/2025, às 09:34, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mario Augusto de Souza Fontes, Usuário Externo**, em 09/10/2025, às 11:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Uffj ([www2.uffj.br/SEI](http://www2.uffj.br/SEI)) através do ícone Conferência de Documentos, informando o código verificador **2601013** e o código CRC **2006606B**.

---

## AGRADECIMENTOS

Neste momento de realização, meu coração se enche de gratidão por todos que, de alguma forma, tornaram esta jornada possível e significativa. A conclusão deste mestrado é o resultado de um esforço coletivo, e a cada um de vocês, dedico meu mais profundo reconhecimento.

Aos meus orientadores, minha eterna gratidão. À professora Cristina, agradeço pela paciência e por suas inúmeras lições, que transcenderam o conteúdo acadêmico. Sua serenidade foi um porto seguro em meio às minhas inquietações, me acalmando e me fazendo enxergar a profissional incrível que ela é. À professora Sandra, por cada reunião iluminada por seu sorriso contagiante, que enchia meu coração de esperança e renovava minha paixão por aprender. Cada palavra sua, cada momento, eram aulas que guardarei para sempre na memória. Ao professor Mario, por todo o carinho e pela paixão com que me apresentou às análises do *FaceReader*, desvendando um universo novo e inspirador que habitava meu imaginário desde a série "Lie to Me".

Às minhas maiores inspirações, as razões pelas quais busco ser uma pessoa melhor a cada dia, meus filhos, Guilherme e Esther. Vocês são a força que me move e o futuro que desejo construir.

Ao Caetano, que com seu nascimento deu origem a todas as perguntas que movem este trabalho.

De forma mais íntima e especial, meu agradecimento máximo ao meu marido, Randolf. Meu amor, meu companheiro, que nunca largou minha mão. Obrigada por vibrar com minhas conquistas e por me abraçar quando a dúvida surgia, sussurrando em meu ouvido: "Você consegue, meu amor". Sua presença foi fundamental. Obrigada por ser o meu par, aquele que não apenas entendeu, mas encontrou as melhores soluções, desde a gravação das interações até os ajustes que tornaram o universo do mestrado mais leve e suave para mim. A oportunidade de chegar até aqui só se tornou real porque o universo sabia que eu teria você ao meu lado. Obrigada por tanto.

Agradeço imensamente aos membros da banca examinadora, por aceitarem o convite para contribuir nesta etapa crucial de minha trajetória acadêmica.

À Isabella, da secretaria do PPG, minha gratidão por sua constante presteza e eficiência.

Aos professores do programa, que me acolheram e compartilharam seus valiosos conhecimentos, moldando minha formação como pesquisadora.

Aos meus queridos colegas, Fredy, Fernanda, Lorena, Tristão e Graça. Vocês foram a primeira turma que verdadeiramente me acolheu, permitindo que eu fosse eu mesma, como pessoa autista. Em seu convívio, encontrei um espaço de aceitação e pertencimento. Obrigada, Graça por acreditar tanto em mim.

Meu eterno agradecimento e carinho à minha díade. Minha mamãe e bebê que fizeram esse trabalho acontecer. Num país onde a pesquisa às vezes é tão difícil, vocês duas abriram as portas de casa e compartilharam comigo momentos tão lindos e íntimos entre uma mãe e sua bebê.

Por fim, um agradecimento especial ao professor Tiago Torrent. Seu acolhimento teve um impacto que talvez nem ele imagine, mostrando-me que a verdadeira inclusão é a alma que transborda e compreende, sem a necessidade de palavras.

## RESUMO

O projeto investiga a multimodalidade no processo de aquisição da linguagem em português brasileiro, com foco na integração das prosódias facial e vocal durante as interações mãe-bebê. Trata-se de um estudo de caso longitudinal, acompanhando uma única díade em dois momentos cruciais do desenvolvimento: aos 6 e aos 12 meses de idade. A literatura aponta para uma relação estreita entre a prosódia vocal e a prosódia facial, indicando que ambas desempenham papéis fundamentais na expressão de intenções comunicativas (Esteve-Gibert; Guellaï, 2018). No contexto brasileiro, ainda há poucos estudos sobre aquisição da linguagem sob essa perspectiva multimodal, especialmente no primeiro ano de vida.

O estudo busca analisar a evolução da sincronia diádica e as estratégias de correção mútua, interpretando como a performance da mãe e a resposta do bebê se reconfiguram ao longo do tempo. Para isso, foram coletados dados em contextos naturalísticos, e as análises multimodais foram conduzidas utilizando softwares especializados como PRAAT e *FaceReader*.

Os resultados sugerem uma transição de uma interação baseada na sincronia afetiva face a face aos 6 meses, para uma dinâmica mais complexa de negociação da atenção compartilhada em um contexto triádico aos 12 meses. Espera-se que estes achados contribuam para uma compreensão mais abrangente do desenvolvimento comunicativo dos bebês, explorando como os gestos faciais e vocais co-constroem a base para a linguagem. Adicionalmente, este estudo contribui para a pesquisa sobre a percepção da prosódia facial em bebês adquirindo o português brasileiro, oferecendo subsídios para futuras investigações sobre a aquisição da linguagem em uma abordagem multimodal.

**Palavras-chave:** Aquisição da linguagem; prosódia facial; prosódia vocal; multimodalidade; interação mãe-bebê.

## ABSTRACT

This research investigates multimodality in the language acquisition process of Brazilian Portuguese, focusing on the integration of facial and vocal prosody during mother-infant interactions. It is a longitudinal case study, following a single dyad at two crucial developmental moments: 6 and 12 months of age. The literature points to a close relationship between vocal and facial prosody, indicating that both play fundamental roles in the expression of communicative intentions (Esteve-Gibert; Guellai, 2018). In the Brazilian context, there are still few studies on language acquisition from this multimodal perspective, especially in the first year of life.

The study aims to analyze the evolution of dyadic synchrony and mutual co-regulation strategies, interpreting how the mother's performance and the infant's response are reconfigured over time. To this end, data were collected in naturalistic contexts, and multimodal analyses were conducted using specialized software such as PRAAT and FaceReader.

Results suggest a transition from an interaction based on face-to-face affective synchrony at 6 months to a more complex dynamic of negotiating joint attention in a triadic context at 12 months. These findings are expected to contribute to a more comprehensive understanding of infants' communicative development, exploring how facial and vocal gestures co-construct the foundation for language. Furthermore, this study contributes to the research on the perception of facial prosody in infants acquiring Brazilian Portuguese, offering insights for future investigations into language acquisition from a multimodal approach.

**Keywords:** Language acquisition; facial prosody; vocal prosody; multimodality; mother-infant interaction.



## **LISTA DE ILUSTRAÇÕES**

Figura 1 - Forma da onda, contorno de frequência fundamental e transcrição de enunciado	28
Figura 2 - Análise de Expressões Faciais e Emoções a Partir das Unidades de Ação (AUs)	29
Figura 3 - Momento de afirmação de uma interlocução bem-sucedida	36
Figura 4 - Ativação das AUs (Bebê)	40
Figura 5 - Perfil do PRAAT	43
Figura 6 - Mãe em Análise do FaceReader	46
Figura 7 - Bebê em Análise do FaceReader	47
Figura 8 - Anotação no PRAAT da mãe aos 12 meses	56

## LISTA DE GRÁFICOS

Gráfico 1 - Gráfico de Valência (Mãe)	34
Gráfico 2 - Gráfico de Arousal (Mãe)	34
Gráfico 3 - Gráfico de expressões emocionais (Mãe)	35
Gráfico 4 - Gráfico de orientação do olhar (Mãe)	36
Gráfico 5 - Gráfico de batimentos cardíacos (Mãe)	37
Gráfico 6 - Gráfico direcionamento do Olhar (Bebê)	37
Gráfico 7 - Gráfico direcionamento da Cabeça (Bebê)	38
Gráfico 8 - Gráfico de Ativação do Organismo (Bebê)	39
Gráfico 9 - Gráfico de Valência (Bebê)	39
Gráfico 10 - Batimentos cardíacos (Bebê)	41
Gráfico 11 - Gráfico de Valência (Mãe) aos 12 meses	51
Gráfico 12 - Gráfico de Arousal ou Ativação do Organismo (Mãe) aos 12 meses	51
Gráfico 13 - Gráfico de expressões emocionais da mãe	52
Gráfico 14 - Gráfico de orientação de cabeça da mãe	52
Gráfico 15 - Gráfico de orientação do olhar da mãe	53
Gráfico 16 - Gráfico de Frequência Cardíaca da mãe	53
Gráfico 17 - Gráfico de orientação da cabeça do bebê	57
Gráfico 18 - Gráfico de ângulo do olhar	58
Gráfico 19 - Gráfico de Valência da bebê	58
Gráfico 20 - Gráfico de Arousal (Ativação do Organismo) da bebê	59
Gráfico 21 - Gráfico de Expressões da bebê	59
Gráfico 22 - Gráfico de frequência cardíaca	61
Gráfico 23 - Gráfico de frequência cardíaca da bebê	61
Gráfico 24 - Gráfico de Batimentos Cardíacos da bebê e o momento em que ela volta sua atenção para mãe	62
Gráfico 25 - Gráfico de BPM da bebê	63

## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO</b>	<b>12</b>
<b>2.</b>	<b>AQUISIÇÃO DA LINGUAGEM</b>	<b>15</b>
2.1.	ETAPAS INICIAIS DO DESENVOLVIMENTO LINGUÍSTICO	15
2.2.	INTERAÇÃO ADULTO-BEBÊ	17
2.3.	A AQUISIÇÃO MULTIMODAL DA LINGUAGEM	19
2.4.	PROSÓDIA	20
2.4.1.	Prosódia vocal	21
2.4.2.	Prosódia facial	23
2.5.	Conclusão	23
<b>3.</b>	<b>27</b>	
3.1.	27	
3.1.1.	Participantes	25
3.1.2.	Procedimento	25
3.2.	30	
<b>4.</b>	<b>ANÁLISE DOS DADOS</b>	<b>31</b>
4.1.	34	
4.1.1.	Descrição Detalhada do Vídeo	32
4.1.2.	Análise Acústica e Prosódica a partir do PRAAT	42
4.1.3.	Análise da Sincronia Diádica	45
4.1.4.	Conclusão	48
4.2.	52	
4.2.1.	Descrição Detalhada do Vídeo da Mãe	49
4.2.2.	Dados da Mãe: A Performance Comunicativa	50
4.2.3.	Dados da Bebê: A Resposta em um Contexto de Atenção Dividida	57
4.2.4.	Conclusão	63
4.3.	68	

<b>5.</b>	<b>71</b>
<b>REFERÊNCIAS.</b>	<b>69</b>
<b>APÊNDICE A</b>	<b>75</b>
<b>APÊNDICE B</b>	<b>78</b>
<b>ANEXO A</b>	<b>82</b>
<b>ANEXO B</b>	<b>83</b>

## 1. INTRODUÇÃO

Esta pesquisa investiga a multimodalidade no processo de aquisição da linguagem em português brasileiro, com foco na integração das prosódias facial e vocal durante as interações mãe-bebê. Embora a presente pesquisa, por sua natureza de estudo de caso, analise a interação de uma díade mãe-bebê, adota-se neste trabalho a terminologia 'bebê-cuidador' no título como um reconhecimento das dinâmicas familiares contemporâneas, que transcendem a configuração tradicional. Desse modo, a escolha do termo 'cuidador' busca abranger a diversidade de arranjos parentais, incluindo casais homoafetivos, pais solo, avós ou outros guardiões primários que assumem papel central no desenvolvimento infantil. A interação analisada, portanto, serve como um modelo para compreender a complexa dança comunicativa que pode ocorrer entre um bebê e sua figura de apego principal, independentemente de gênero ou laço biológico.

A aquisição da linguagem é um processo multimodal, no qual os bebês integram informações de diferentes modalidades sensoriais, incluindo a audição, a visão e o tato. Com base em uma revisão de estudos, Guellaï et al. (2014) apontam que a prosódia vocal e a prosódia visual, da qual faz parte a prosódia facial, interagem de forma complexa para transmitir sentidos e intenções. Cavalcante (no prelo), em seus trabalhos, ressalta a importância dessa integração multimodal, especialmente em bebês, que dependem fortemente de pistas visuais e auditivas para interpretar o significado dos enunciados, corroborando os achados sintetizados na revisão de literatura de Esteve-Gibert e Guellaï (2018).

A prosódia vocal, componente essencial da comunicação humana, desempenha um papel importante na aquisição da linguagem, transcendendo a mera decodificação lexical e configurando-se como um sistema de sinalização multifacetado que orchestra a melodia da interação verbal (Crystal, 2003). Através da modulação, a prosódia confere nuances de sentido, emoção e intenção à mensagem, revelando-se essencial para a compreensão da intencionalidade e da subjetividade na comunicação (Paulmann; Titone; Pell, 2012).

A prosódia visual refere-se ao conjunto de movimentos expressivos e rítmicos do corpo, particularmente da face, cabeça e mãos, que acompanham e complementam a fala, conferindo-lhe nuances de significado e intenção. Tal como a prosódia vocal, que modula a altura, a intensidade e o ritmo da voz, a prosódia visual utiliza a dinâmica do movimento, a amplitude e a orientação espacial para estruturar a informação, enfatizar elementos específicos e expressar emoções. Essa "melodia do movimento" é relevante para a comunicação face a face, contribuindo para a clareza do discurso e a interpretação da atitude do falante. McNeill (1992) destaca a importância dos gestos manuais como parte integrante do processo de fala,

demonstrando que eles não são meros adornos, mas uma parte fundamental da expressão do pensamento. Kendon (2004) aprofunda essa ideia, argumentando que a gesticulação e outras formas de movimento corporal formam um sistema semiótico paralelo e interconectado com a fala, contribuindo para a construção do sentido na interação. É importante notar que a própria visão de Kendon sobre essa relação evoluiu. Se inicialmente a gestualidade era vista como um sistema 'paralelo', seus trabalhos posteriores (e.g., Kendon, 2011) avançam para uma concepção de embricamento, onde fala e gesto não são apenas interconectados, mas co-constitutivos e inseparáveis na produção do enunciado. A unidade de análise, portanto, não é a palavra ou o gesto isolado, mas a unidade gesto-fala como um todo indissociável. Esteve-Gibert e Guellaï (2018) apresentam uma revisão abrangente da literatura sobre o papel da prosódia visual na percepção da fala e na aquisição da linguagem, evidenciando sua importância na segmentação do fluxo da fala, na identificação de unidades prosódicas e na compreensão da intenção comunicativa, especialmente em estágios iniciais da aquisição da linguagem.

A prosódia facial desempenha um papel essencial na percepção da fala pelos bebês, auxiliando na compreensão de estados emocionais do interlocutor (Peltola et al., 2008). As expressões faciais sincronizadas com a fala fornecem pistas visuais que reforçam a intencionalidade comunicativa (Bristow et al., 2009; Lewkowicz, 1996). A prosódia facial está envolvida na regulação da interação social, sendo um componente crucial para o desenvolvimento das habilidades comunicativas na infância (Grossmann; Johnson, 2007).

A jornada de um bebê na aquisição da linguagem começa muito antes do que imaginamos, ainda no útero materno. Durante a fase final da gestação, o sistema auditivo do feto já está desenvolvido o suficiente para captar sons e informações da fala da mãe (Lecanuet et al., 1993 apud Gervain; Kisilevsky et al., 2000 apud Gervain, 2018). Esses sons, que atravessam o líquido amniótico e os tecidos uterinos (Querleu et al., 1988 apud Gervain, 2018), permitem que o feto se familiarize com a voz materna e comece a construir representações da língua. Trazendo essa bagagem pré-natal para o mundo pós-natal, o bebê encontra na interação com sua mãe o cenário ideal para continuar essa jornada, onde as pistas prosódicas se tornam visíveis e audíveis de forma integrada.

O principal objetivo deste estudo é investigar como o bebê processa as prosódias facial e vocal durante as interações com sua mãe. Os objetivos específicos são (i) examinar a sincronia entre expressões faciais e vocais na fala dirigida à criança (FDC); (ii) avaliar a resposta do bebê a diferentes padrões prosódicos em interações naturais (iii) identificar padrões de expressão utilizados pela mãe e seus efeitos na percepção do bebê; e (iv) contribuir para a compreensão

da aquisição da linguagem no contexto do português brasileiro. Para isso, a investigação empírica se pauta na observação de interações espontâneas que permitem capturar a forma como o bebê utiliza essas informações para compreender e se engajar na comunicação.

Foi realizado um estudo de caso longitudinal, acompanhando, por meio de visitas periódicas, a interação entre um bebê do sexo feminino dos 5 aos 12 meses de idade e sua mãe. Os dados naturalísticos coletados foram analisados utilizando-se o FaceReader para a análise dos dados faciais e o Praat para a análise dos dados vocais.

A pesquisa busca contribuir para uma compreensão mais aprofundada do desenvolvimento comunicativo humano, lançando luz sobre a trajetória desenvolvimental que vai da sintonização afetiva à negociação da atenção compartilhada, desenvolvidos pelos bebês de modo a se comunicarem de forma eficaz em um mundo multimodal.

Este estudo está estruturado da seguinte forma: no próximo capítulo, será apresentada uma revisão da literatura sobre aquisição da linguagem, com ênfase em aspectos do desenvolvimento perceptual inicial, da interação adulto-bebê e na aquisição multimodal. Tem, ainda, uma seção específica sobre prosódia – vocal e facial – a fim de entendermos melhor seu papel nesse processo. Em seguida, o capítulo 3 descreve a metodologia empregada na investigação empírica, detalhando os procedimentos de coleta e de análise dos dados. O capítulo 4 apresenta a análise dos dados coletados aos 6 e aos 12 meses de idade da bebê, e discute como a comparação entre esses dois momentos revela a trajetória de desenvolvimento da comunicação diádica, partindo de um modelo de sintonização afetiva para um de negociação da atenção compartilhada, processo fundamental para a aquisição da linguagem. Na conclusão, são retomados os objetivos do estudo, apresentadas suas principais contribuições e as limitações da pesquisa.

## 2. AQUISIÇÃO DA LINGUAGEM

Assumimos que a aquisição da linguagem não se limita apenas à aquisição da fala, mas também engloba a compreensão e a expressão de emoções, intenções e sentidos por meio de múltiplas modalidades, incluindo expressões faciais, gestos e prosódia. Este processo é fundamentalmente diádico e interacional, no qual a díade mãe-bebê funciona como um sistema dinâmico e auto organizado, onde o significado é co-construído momento a momento. Apresentamos, neste capítulo, uma breve revisão de estudos que sustentam tal perspectiva. Na seção 2.1, trataremos das etapas iniciais do desenvolvimento linguístico, sobretudo, das capacidades perceptuais do bebê. Em seguida (seção 2.2), discutiremos sobre o papel crucial da interação entre adulto e bebê para o desenvolvimento linguístico, cognitivo e socioemocional da criança. Por fim, na seção 2.3, esboçamos como entendemos a aquisição multimodal da linguagem.

### 2.1. ETAPAS INICIAIS DO DESENVOLVIMENTO LINGUÍSTICO

Pesquisas mostram que fetos no terceiro trimestre de gestação (31-40 semanas) conseguem não só reconhecer a voz da mãe (De Casper; Fifer, 1980; Kisilevsky et al., 2009), mas também diferenciar a língua materna de uma língua estrangeira, sendo sensíveis aos padrões de entonação (Kisilevsky et al., 2009).

Essa sensibilidade à prosódia da fala se mantém após o nascimento. Recém-nascidos demonstram uma clara preferência por estímulos de fala em comparação com outros sons (Vouloumanos; Werker, 2007). Curiosamente, a melodia do choro dos bebês recém-nascidos também é influenciada pela entonação característica de sua língua materna. Mampe et al. (2009) observaram que bebês franceses recém-nascidos tendem a chorar com uma entonação ascendente, enquanto bebês alemães de 2 a 5 dias choram com uma entonação descendente, seguindo os padrões de suas respectivas línguas. Esses resultados sugerem que a experiência auditiva pré-natal tem um impacto importante no desenvolvimento da percepção da linguagem, e que os bebês já nascem com uma capacidade de processar informações linguísticas.

Poulin-Dubois e Graham (2007) destacam a importância da percepção da fala na aquisição da linguagem, uma vez que as características acústicas e distribucionais percebidas pelos bebês permitem que eles descubram as propriedades da linguagem, como a estrutura gramatical e o léxico. Nos primeiros meses de vida, os bebês demonstram uma capacidade notável de discriminar contrastes fonéticos de todas as línguas, diferenciando sons de categorias distintas, uma habilidade conhecida como percepção categórica (Eimas Et Al., 1971; Werker;



Lalonde, 1988). A partir dos seis meses de idade, ocorre uma mudança perceptual significativa, com a diminuição da capacidade de discriminação dos bebês para vogais que não pertencem ao repertório fonêmico de sua língua materna, e, por volta dos dez meses, essa diminuição se estende às consoantes (Werker; Tees, 1984 apud Name, 2022; Kuhl, 1991 apud Name, 2022; Rivera-Gaxiola et al., 2005 apud Name, 2022).

A sensibilidade a pistas prosódicas permite que a criança segmente a fala em unidades significativas, como frases e palavras, diferenciando elementos prosódicos como entonação, ritmo e pausas. A partir dessa segmentação, a criança identifica elementos linguísticos como sílabas, palavras e frases, construindo representações fonológicas e lexicais. Além disso, a percepção de pistas prosódicas auxilia na compreensão da intenção comunicativa do falante, permitindo que a criança identifique diferentes atos de fala, como perguntas, afirmações e ordens (Name, 2022).

Ainda, Cavalcante (1999), Scarpa (1999) e Barros (Barros; Scarpa; Cavalcante, 2024) destacam a importância das pistas prosódicas na fala adulta dirigida a crianças (FDC) (ver seção 2.2) para a aquisição da prosódia. As autoras observaram o desenvolvimento prosódico em quatro estágios principais da produção infantil: balbucio, jargão, primeiras palavras e blocos de enunciados.

Analizando o desenvolvimento prosódico de uma criança, as autoras identificam uma progressão complexa e não linear ao longo de distintos estágios, que incluem o balbucio, o jargão e a emergência das primeiras palavras, culminando na construção de blocos de enunciados. Inicialmente, no balbucio, observa-se a predominância de tons simples ascendentes (L+H) ou descendentes (H+L), carentes de acento primário e sem correspondência a padrões de uma comunidade linguística específica. A transição para o jargão é marcada por uma maior diversificação tonal e por contornos e ritmos que mimetizam padrões prosódicos adultos, a ponto de serem interpretados como sentenças pelo interlocutor, mesmo que o conteúdo segmental permaneça ininteligível, revelando a primazia do segmento prosódico-rítmico na interpretação. Com o surgimento das primeiras palavras, por volta de 1 ano e 4 meses, nota-se um aumento na variação de altura e o surgimento de sequências silábicas mais curtas, reconhecíveis como pertencentes a um léxico primitivo. Essas primeiras palavras já carregam intenções comunicativas distintas, como perguntas, afirmações e pedidos, indicando um desenvolvimento significativo na capacidade prosódica de expressar atos de fala, sendo recorrentes e produtivas, não mais efêmeras. Posteriormente, observa-se a construção de blocos de enunciados, demonstrando a capacidade da criança de produzir um *continuum* entonacional,

em vez de palavras isoladas. Aos 1 ano e 6 meses, as autoras registram uma alteração na altura do tom H e uma ampliação da gama tonal, com variações ascendentes, descendentes, ascendentes-descendentes e descendentes-ascendentes, evidenciando um refinamento crescente do sistema prosódico. As autoras ressaltam que esses estágios não são estanques, mas concomitantes, havendo uma sobreposição temporal entre eles. Além disso, elas destacam o papel fundamental do "manhês" ou FDC nesse processo, argumentando que as características prosódicas exageradas da fala dirigida à criança, como F0 mais elevada e prolongamento silábico, são cruciais para captar a atenção infantil e guiar a criança na aquisição da linguagem.

## 2.2. INTERAÇÃO ADULTO-BEBÊ

A interação adulto-bebê desempenha um papel crucial no desenvolvimento linguístico, cognitivo e socioemocional da criança. É neste contexto que as capacidades perceptuais inatas do bebê, descritas anteriormente, são postas em prática e aprimoradas. A díade mãe-bebê é aqui entendida como um sistema único, cujos membros influenciam-se mutuamente em um processo contínuo de correção (Fogel, 1993). As interações sociais, especialmente aquelas que envolvem linguagem e comunicação, proporcionam aos bebês oportunidades valiosas para aprender sobre o mundo ao seu redor, desenvolver habilidades de comunicação e estabelecer laços afetivos com seus cuidadores. A qualidade das interações adulto-bebê tem um impacto significativo no desenvolvimento da criança. Interações responsivas, nas quais os cuidadores se sintonizam com os sinais e necessidades do bebê, promovem um desenvolvimento saudável e positivo. Essa sintonização, ou "sincronia diádica", refere-se à coordenação temporal ou não de comportamentos entre os parceiros (mãe-bebê), criando um ritmo compartilhado que organiza a interação (Feldman, 2007). É importante distinguir, contudo, a sincronia da mera simultaneidade. Enquanto a simultaneidade é a ocorrência de ações ao mesmo tempo, a sincronia implica uma conexão funcional e uma regulação mútua, onde a ação de um parceiro está dinamicamente ligada à do outro, formando um padrão coeso.

Nessa perspectiva, a interação precoce é o berço da intersubjetividade. Colwyn Trevarthen (1979) foi pioneiro ao postular a existência de uma intersubjetividade primária, uma sintonia inata entre bebê e cuidador que se manifesta por volta dos dois meses, onde o foco é puramente a conexão socioemocional face a face. Para ele, o bebê já nasce com uma motivação para o engajamento interpessoal. Autores como Beebe e Lachmann (2002), por meio de microanálises de vídeos, aprofundaram essa visão, demonstrando como essa "dança" interativa é um processo contínuo de regulação mútua, onde mãe e bebê ajustam seus comportamentos

em frações de segundo. Essa regulação, no entanto, não é perfeita. Tronick (1989), com seu "Modelo de Regulação Mútua", destaca que a interação é, na verdade, um fluxo constante de 'desencontros' e 'reparos', e é justamente na capacidade da díade de reparar as falhas comunicativas que reside a chave para o desenvolvimento da resiliência e da confiança. Por fim, é nesse cenário que o conceito de *scaffolding* (andaime), proposto por Bruner (1975), se torna central. O cuidador oferece um suporte estruturado, o andaime, para que a criança consiga realizar uma tarefa que ainda não dominaria sozinha, retirando esse suporte gradualmente conforme a competência da criança aumenta.

Ávila-Nóbrega (2023) a partir dos trabalhos de Tomasello (Carpenter et al., 1998; Tomasello, 1999 [2023]) destaca a importância da atencionalidade nesse processo, definindo-a como a capacidade da criança de compartilhar a atenção com o adulto em relação a um objeto ou evento. Essa habilidade de se engajar em atenção conjunta e de compreender o outro como um agente intencional é crucial para o desenvolvimento da comunicação e da interação social.

A atencionalidade se manifesta em diferentes formas na interação adulto-bebê. O compartilhamento de atenção ocorre quando a criança e o adulto olham para o mesmo objeto ou evento, demonstrando um foco de atenção compartilhada (Barros; Scarpa; Cavalcante, 2024). Por exemplo, a mãe pode mostrar um brinquedo ao bebê e dizer o seu nome, e o bebê responde olhando para o brinquedo e para a mãe. A intencionalidade se manifesta quando a criança demonstra, por meio de gestos, vocalizações e/ou olhares, que deseja compartilhar a atenção com o adulto em relação a algo, como quando o bebê aponta para um objeto e olha para o adulto em seguida. A reciprocidade ocorre quando o adulto responde à tentativa de compartilhamento de atenção da criança, criando uma interação recíproca. Por exemplo, o adulto pode se aproximar, olhar para o objeto e interagir com a criança que está apontando para algo. O engajamento conjunto se manifesta quando a criança e o adulto se engajam em uma atividade conjunta, mantendo a atenção compartilhada e interagindo de forma recíproca, como em um jogo de encaixe, onde ambos se concentram na atividade, trocam o brinquedo e se comunicam.

Outro ponto de destaque na interação adulto-bebê é a Fala Dirigida à Criança (FDC), com suas características prosódicas específicas, como *pitch* mais alto, pausas mais longas e vogais alongadas. Estudos empíricos têm apontado uma correlação entre seu uso e o desenvolvimento linguístico infantil (Fernald, 1987; Name; Sosa, 2020, 2022a, 2022b). A prosódia da FDC, em especial, com seu contorno entonacional marcado e sua expressividade

facial acentuada, pode auxiliar os bebês a segmentarem a fala, percebendo as fronteiras entre as palavras, e a compreender as intenções comunicativas dos adultos.

O estudo de Schwarz et al. (2023), sobre a interação de mães e pais de bebês suecos, oferece insights valiosos sobre como a prosódia se manifesta na interação entre pais e seus filhos durante os primeiros meses de vida. Os resultados da pesquisa indicam que tanto mães quanto pais utilizam características prosódicas distintas em suas interações com os bebês, refletindo um padrão de expressividade emocional que é fundamental para o desenvolvimento da linguagem. Ao longo do primeiro ano de vida, a valência afetiva da fala das mães apresenta uma trajetória que varia, com os níveis mais altos de afeto positivo observados em interações com bebês de 3 meses, diminuindo aos 9 meses e aumentando novamente aos 12 meses. Essa dinâmica sugere que a prosódia não apenas serve para transmitir informações, mas também para estabelecer e manter vínculos emocionais entre pais e filhos (Schwarz et al., 2023). Assim, a prosódia não apenas enriquece a interação entre pais e bebês, mas também parece desempenhar um papel importante na formação das bases para a aquisição da linguagem e o desenvolvimento social na infância.

É importante ressaltar, contudo, que o engajamento conjunto nem sempre é um estado estável ou contínuo. Especialmente à medida que a autonomia motora e perceptual da criança se desenvolve, emergem cenários de atenção dividida, nos quais a criança direciona seu foco para um estímulo concorrente no ambiente (um som, um objeto fora do alcance) enquanto o cuidador tenta manter o engajamento na atividade principal. Esses momentos, longe de serem falhas comunicativas, representam oportunidades importantes para o cuidador empregar estratégias de reparo e andaimeio (*scaffolding*) mais complexas, negociando ativamente o foco atencional. A capacidade da díade de navegar e resolver esses episódios de atenção dividida, como será explorado na análise aos 12 meses, pode ser um forte indicador do amadurecimento da competência comunicativa.

### 2.3. A AQUISIÇÃO MULTIMODAL DA LINGUAGEM

O processo de aquisição da linguagem é um fenômeno multifacetado que vai além da aquisição fonológica, do léxico e da sintaxe, abrangendo também aspectos multimodais que integram as prosódias facial e vocal (ver seção 1.4). Desde o período pré-natal, os fetos são sensíveis à prosódia vocal da língua materna (Mampe et al., 2009), evidenciando que a experiência auditiva e a exposição precoce à fala desempenham papéis relevantes na construção das bases para a percepção e produção linguística. A partir dessas bases, os bebês desenvolvem

habilidades como a percepção categórica e a segmentação da fala, que são fundamentais para a formação de representações fonológicas e lexicais. Mas não podemos esquecer que as interações linguísticas se constroem em situações sociocomunicativas e, nesse sentido, a perspectiva multimodal permite ampliar a compreensão da aquisição da linguagem, ao destacar como a integração de sinais vocais e visuais enriquece a experiência comunicativa, permitindo que os bebês atribuam sentidos a partir de diferentes modalidades. Como aponta Cavalcante (2009), é nessas rotinas interativas que o significado é co-construído, não apenas transmitido do adulto para a criança.

Estudos pioneiros, como o de Lewkowicz (1996), investigaram a resposta dos bebês aos aspectos auditivos e visuais da fala materna. Esse estudo revelou que os bebês são sensíveis tanto ao conteúdo linguístico quanto à prosódia da fala, e que a combinação de informações visuais e auditivas facilita a compreensão e o processamento da linguagem pelos bebês.

Outras pesquisas, como o estudo de Grossmann e Johnson (2010), examinaram as respostas do córtex pré-frontal de bebês em relação à atenção conjunta. Como já visto, a atenção conjunta ocorre quando um bebê e um adulto compartilham o foco de atenção em um objeto ou evento específico. Os resultados mostraram que o córtex pré-frontal dos bebês apresentava respostas seletivas à atenção conjunta, evidenciando a importância da prosódia facial na interação social precoce.

Um estudo conduzido por Bristow, Dehaene-Lambertz e Mattout (2008) investigou como o cérebro do bebê combina a face que vê com a fala que ouve. Os resultados indicaram que o cérebro infantil exibía respostas específicas que relacionavam a face vista com a fala ouvida, destacando a importância da prosódia facial na integração de informações sociais e linguísticas.

Além disso, a prosódia facial também pode influenciar as respostas emocionais dos bebês. Peltola, Leppänen e Hietanen (2008) examinaram as respostas cardíacas e atencionais de bebês a faces temerosas. O estudo constatou que os bebês apresentavam respostas cardíacas e atencionais aumentadas a faces temerosas, indicando a sensibilidade dos bebês às pistas emocionais transmitidas por meio da prosódia facial.

Esses estudos e outros trabalhos subsequentes têm demonstrado a importância da prosódia facial na comunicação infantil. A combinação de expressões faciais, gestos e entonação vocal fornece aos bebês uma rica fonte de informações sociais, emocionais e linguísticas. A pesquisa tem avançado na compreensão dos mecanismos neurais subjacentes à

integração dessas informações na comunicação infantil, fornecendo insights valiosos sobre o desenvolvimento socioemocional e linguístico dos bebês.

Para entendermos melhor a importância da prosódia multimodal na comunicação humana e, particularmente, no desenvolvimento linguístico, sociocomunicativo e emocional da criança, discorreremos sobre a prosódia, vocal e facial na próxima seção.

## 2.4. PROSÓDIA

A prosódia, elemento intrínseco à comunicação humana, desempenha um papel que transcende a mera decodificação lexical, configurando-se como um sistema de sinalização multifacetado que orchestra a melodia da interação verbal (Crystal, 2003). Através da modulação rítmica, melódica e da intensidade da fala, a prosódia confere nuances de significado, emoção e intenção à mensagem, revelando-se essencial para a compreensão da intencionalidade e da subjetividade na comunicação (Gili Fiol; Fernández Planas, 2011). Para além de sua função estritamente linguística, a prosódia atua como um componente pragmático essencial, refletindo atitudes proposicionais, papéis sociais e o estado emocional do falante. Como destaca Barbosa (2012), a prosódia não se limita a aspectos como "acentuação, fronteira de constituinte, ênfase, entoação e ritmo", mas engloba também as dimensões sociocognitivas da interação. A prosódia contribui para a construção de sentido na comunicação, atuando como uma interface entre a estrutura linguística e os aspectos sociocognitivos da interação (Hirst; Di Cristo, 1998).

A complexidade da prosódia reside em sua capacidade de modular e complementar o significado das palavras, expressar intenções e enquadrar a interação em uma perspectiva pragmática (Couper-Kuhlen, 2009). A entonação, por exemplo, pode sinalizar diferentes modalidades de frase, como interrogativas, exclamativas ou declarativas, além de denotar surpresa, ironia, dúvida ou ênfase. O ritmo e a intensidade, por sua vez, podem transmitir entusiasmo, hesitação, raiva ou surpresa, contribuindo para a expressão de diferentes estados emocionais (Barbosa, 2012).

A prosódia desempenha múltiplas funções na comunicação humana (Cohen; Macaulay; Young, 2004). A função emotiva expressa as emoções e sentimentos do falante, como alegria, tristeza, raiva, medo etc. (Scherer, 2003). A função informativa auxilia na transmissão de informações, destacando elementos importantes da mensagem, organizando o discurso e sinalizando relações entre as diferentes partes do texto. A função social revela informações sobre a identidade social do falante, como classe social, grupo étnico, gênero etc., e contribui

para a construção e manutenção das relações interpessoais. Por fim, a função estética confere musicalidade e beleza à fala, sendo explorada em manifestações artísticas como poesia, música e teatro.

A prosódia está intimamente relacionada aos processos cognitivos (Cutler; Dahan; Van Dongen, 1997), influenciando e sendo influenciada pelo processamento da fala, memória, atenção e emoção. A prosódia auxilia na segmentação da fala em unidades linguísticas, na identificação de palavras e na compreensão da sintaxe (Mattys; Davis, 2012).

É importante destacar que a prosódia não se manifesta apenas na produção oral da fala, mas também através da modulação das expressões faciais, da velocidade e intensidade dos movimentos e configuração das mãos, braços, cabeça e tronco (Cavalcante, 2023). Essa perspectiva multimodal ressalta a importância da prosódia como um sistema de sinalização universal, presente em diferentes línguas e modalidades de comunicação.

A investigação da prosódia em diferentes contextos comunicativos, como a aquisição da linguagem, a interação adulto-criança, o discurso político e a comunicação intercultural, permite aprofundar a compreensão dos mecanismos subjacentes à comunicação humana, revelando as nuances e complexidades que permeiam a interação verbal. No caso específico deste estudo, a análise prosódica vocal e facial contribui para o entendimento mais completo e abrangente do processo inicial de aquisição da linguagem, levando-se em conta não apenas o código linguístico, mas também os aspectos sociocognitivos e afetivos da interação adulto-bebê.

#### **2.4.1. Prosódia vocal**

A prosódia vocal, componente essencial da comunicação oral, refere-se às variações na entonação, ritmo e aspectos de continuidade da fala como pausas, qualidade de voz, taxa de elocução e taxa de articulação em produções de fala. O ritmo diz respeito à organização temporal e sequencial da fala. A taxa de elocução, um componente do ritmo, engloba a duração das sílabas e a extensão das pausas entre elas, influenciando diretamente a percepção da fluidez e a dinâmica da articulação. Em outras palavras, é a variação na duração desses elementos que contribui para a sensação de velocidade ou lentidão na fala e para a criação de padrões rítmicos distintos. O acento é a proeminência dada a uma sílaba em relação às demais na palavra. A entoação refere-se à variação da altura (*pitch*) ao longo da frase, que pode ser classificada como ascendente, descendente ou circunflexa. As pausas são interrupções na fala, que podem ser silenciosas ou preenchidas por sons como “ah” e “hum”. Crystal (2003) destaca a importância

da prosódia vocal na transmissão de nuances de significado que vão além das palavras em si. Por exemplo, uma elevação na entonação no final de uma frase pode transformar uma afirmação em uma pergunta, enquanto um aumento no volume pode indicar ênfase ou intensidade emocional.

Além desses elementos, a qualidade de voz, que se refere à configuração específica do aparelho fonador durante a fala, também é um componente prosódico crucial para a transmissão de afetos e atitudes. Ajustes como a voz sussurrada (*whispery voice*), que combina escape de ar com a vibração das pregas vocais, podem ser utilizados para expressar ternura e confidencialidade, enquanto ajustes do trato vocal, como a laringe levantada, podem agudizar a voz, sinalizando delicadeza, como observado nas análises deste trabalho (ver capítulo 4, seção 4.1.21). Para além de sua função estritamente linguística, a prosódia atua como um componente pragmático essencial, refletindo atitudes proposicionais, papéis sociais e o estado emocional do falante.

Existem diferentes modelos teóricos para analisar a prosódia. O modelo autossegmental-métrico (Pierrehumbert, 1980) propõe que a prosódia é organizada em diferentes níveis hierárquicos, desde o nível segmental (sons) até o nível da frase. O modelo de contornos entoacionais (Jun, 2005) analisa a forma global da curva de entonação da frase, identificando padrões característicos de diferentes tipos de frases e emoções. O modelo de produção e percepção da fala investiga os mecanismos fisiológicos e cognitivos envolvidos na produção e percepção da prosódia como o modelo fonte-filtro da produção da fala (Fant, 1960) e a teoria motora da percepção da fala (Liberman ; Mattingly, 1985).

Um dos pilares para a análise da produção e percepção da fala é o modelo fonte-filtro, proposto por Gunnar Fant (1960). Este modelo, que revolucionou a fonética acústica, postula que a produção dos sons da fala pode ser entendida como um processo de duas etapas independentes. A fonte sonora é gerada pela vibração das pregas vocais na laringe, criando uma sequência de harmônicos múltiplos, sendo o primeiro correspondente à frequência fundamental (F0) que percebemos como o pitch ou a melodia da voz. As frequências geradas pela fonte são modificadas pelo trato vocal (faringe, cavidade oral e nasal), que atua como um filtro acústico. A forma como articulamos a língua, os lábios e o véu palatino altera o formato desse filtro, fazendo com que certas frequências sejam reforçadas (os formantes) e outras atenuadas. É a ação de ressonância desse filtro que molda o som gerado pela fonte e o transforma nos diferentes sons de vogais e consoantes que reconhecemos. A grande contribuição de Fant foi demonstrar



que podemos analisar esses dois componentes — a fonte, ligada ao vozeamento, e o filtro, como ressoador.

Em uma vertente que busca explicar a percepção da fala, a Teoria Motora da Percepção da Fala, desenvolvida por Alvin Liberman e seus colaboradores (1985), propõe uma outra ideia: nós percebemos a fala ao inferir ou simular internamente os gestos articulatórios que o falante realizou para produzir os sons. A teoria parte do 'problema da ausência de invariância', que observa que o sinal acústico de um mesmo fonema varia enormemente dependendo do contexto. Por exemplo, o som do /d/ é acusticamente muito diferente em 'di' e 'du'. Como, então, nosso cérebro os percebe como o mesmo som? A resposta da Teoria Motora é que o nosso cérebro não percebe o som em si, mas reconhece que, para produzir tanto 'di' quanto 'du', o falante precisou executar o mesmo gesto motor intencional: o de levar a ponta da língua até o alvéolo. Portanto, para essa teoria, o objeto da percepção da fala não é o sinal acústico, mas os gestos fonéticos do falante. Essa perspectiva reforça uma visão profundamente multimodal e corporificada da comunicação, sugerindo que o cérebro humano possui um módulo especializado que integra intrinsecamente a informação auditiva com a informação motora. No contexto da interação mãe-bebê, isso implica que as pistas visuais, como os movimentos dos lábios e as expressões faciais que acompanham a produção vocal, não são meros complementos, mas parte integral do sinal que o bebê processa para decodificar a intenção comunicativa

A prosódia apresenta variações em diferentes níveis (Crystal, 2003). A variação regional refere-se às diferenças prosódicas entre diferentes regiões geográficas que falam a mesma língua. A variação social relaciona-se às diferenças prosódicas entre diferentes grupos sociais, como classes sociais, grupos étnicos e faixas etárias (Lopes et al., 2014). A variação estilística diz respeito às diferenças prosódicas entre diferentes estilos de fala, como fala formal e informal.

#### **2.4.2. Prosódia facial**

A prosódia facial, componente essencial na comunicação multimodal, compreende os movimentos musculares e expressões faciais que acompanham a fala, adicionando camadas de significado e nuances à mensagem verbal. Ekman (1992), pioneiro no estudo da expressão facial, aponta que a prosódia facial abrange uma variedade de componentes, desde microexpressões fugazes até gestos faciais mais amplos e prolongados. Pesquisas atuais, como a de Pacheco (2011), indicam que os sinais de cabeça contribuem para a segmentação do contínuo da fala, desempenhando um papel importante no processamento audiovisual da fala.

A prosódia facial/visual é um componente importante na comunicação infantil, que envolve a sincronização das informações auditivas e visuais transmitidas pelos cuidadores. Essa forma de comunicação combina as expressões faciais e os gestos corporais com a prosódia vocal, proporcionando aos bebês uma rica fonte de informações sobre as intenções e emoções dos falantes. É neste processo que a díade mãe-bebê opera como um sistema dinâmico, no qual os membros se influenciam mutuamente em um processo contínuo de correção (Fogel, 1993). A "dança" interativa que se estabelece é marcada pela sincronia diádica, que, para além da mera simultaneidade de ações, refere-se à coordenação temporal e funcional dos comportamentos, criando um ritmo compartilhado que organiza a interação e possibilita a emergência de "momentos de encontro" (Stern, 1985; Feldman, 2007).

## 2.5. Conclusão

Neste capítulo, apresentamos a aquisição da linguagem assumindo uma perspectiva multimodal, como um processo complexo que se inicia muito antes da produção das primeiras palavras. A atencionalidade, a intencionalidade e a reciprocidade nas interações adulto-bebê mostram-se essenciais para a aquisição da linguagem, oferecendo contextos ricos para o aprendizado e a construção de vínculos emocionais. A Fala Dirigida à Criança (FDC), com suas características prosódicas específicas, surge como um elemento facilitador nesse processo, auxiliando na segmentação da fala e na interpretação de intenções comunicativas. Dessa forma, exploramos o papel central da prosódia como elemento essencial na comunicação humana e suas implicações no processo de interação social e linguística. Assim, a prosódia transcende a função estritamente linguística, configurando-se como um sistema multimodal que integra variações vocais e faciais para enriquecer a troca de informações.

Ao desdobrar as funções da prosódia vocal, foi possível observar como elementos como entonação, ritmo e pausas orquestram o significado além das palavras, alinhando-se a aspectos pragmáticos e sociocognitivos. A prosódia facial, por sua vez, reforça a comunicação multimodal ao adicionar camadas visuais que complementam a expressão vocal. Os estudos revisados apontam que tanto a prosódia vocal quanto a prosódia facial desempenham um papel determinante na segmentação do discurso, na atenção conjunta e na integração de informações sociais e linguísticas, particularmente em contextos de interação adulto-bebê.

Essas evidências sublinham a relevância da prosódia no desenvolvimento infantil, indicando que bebês não apenas respondem às pistas prosódicas vocais e faciais, mas também as utilizam como base para construir suas competências socioemocionais e linguísticas. A

integração dessas informações promove o desenvolvimento de habilidades importantes para a comunicação, reforçando a ideia de que a prosódia não opera de forma isolada, mas como um sistema integrado que conecta dimensões cognitivas, sociais e emocionais.

Assim, este capítulo estabelece os fundamentos para compreender a aquisição da linguagem como um fenômeno multifacetado e multimodal. Esses insights servirão como base para aprofundar a investigação da interação adulto-bebê e os mecanismos que sustentam o desenvolvimento linguístico e social no primeiro ano de vida, conforme explorado nos capítulos subsequentes desta dissertação.

No próximo capítulo, apresentamos a investigação empírica, um estudo longitudinal que acompanhou uma díade mãe-bebê em situações de interação comunicativa, durante x meses. Buscamos observar se e como o bebê processa as prosódias facial e vocal durante as interações com sua mãe.

### 3. METODOLOGIA

Este estudo, de caráter longitudinal, investiga interações entre mãe e bebê no período de 6 a 12 meses de idade. Estudos longitudinais apontam para mudanças na interação adulto-bebê em função do crescimento da criança, tanto na fala e gestual do adulto, quanto nas expressões vocais e gestuais do bebê (Cavalcante; Barros, 2012; Name; Sosa, 2020, 2022). O período em torno de 6 meses de idade tem sido considerado momento-chave para mudanças nessa relação (Saint-Georges et al., 2013), assim como ao fim do primeiro ano de vida, daí a escolha desse período de vida do bebê. Nesta dissertação, apresentaremos a análise dos dados coletados durante uma diáde aos 6 meses e outra aos 12 meses de idade do bebê, embora a coleta tenha acompanhado seu desenvolvimento ao longo de todo o período.

A investigação se alinha a pesquisas anteriores que destacam a sincronia entre expressões faciais e vocais na comunicação entre bebês e adultos (Cavalcante; Barros, 2012). Durante esse período, bebês demonstram maior sensibilidade à entonação, ritmo da fala e gestos faciais (Guellaï et al., 2016). Desse modo, nossos objetivos específicos são (i) examinar a sincronia entre expressões faciais e vocais na fala dirigida à criança (FDC); (ii) avaliar a resposta do bebê a diferentes padrões prosódicos em interações naturais; e (iii) identificar padrões de expressão utilizados pela mãe e seus efeitos na percepção do bebê.

#### 3.1. COLETA DE DADOS

A coleta de dados para este estudo foi conduzida a partir de um delineamento de caso longitudinal, acompanhando sistematicamente uma única diáde mãe-bebê em seu ambiente residencial, a fim de garantir a validade ecológica dos registros. As interações, focadas em atividades espontâneas como a leitura de livros, foram capturadas por meio de gravações audiovisuais de alta qualidade, utilizando-se um aparato com duas câmeras e microfones sensíveis. Este procedimento foi planejado para permitir a captação detalhada dos múltiplos canais comunicativos — faciais, vocais e gestuais —, assegurando um corpus robusto para a análise multimodal da evolução da comunicação no primeiro ano de vida.

### 3.1.1. Participantes

Foi recrutada uma díade bebê-mãe, localizada por meio de contato via networking. A bebê, que iniciou sua participação no estudo com 5 meses e 3 dias de vida, foi selecionada de acordo com critérios estabelecidos no projeto. Poderiam participar da pesquisa responsáveis – pais e/ou cuidadores maiores de 18 anos – junto com bebês de 5 a 10 meses de idade adquirindo o PB como língua materna. Foi dada preferência à participação da mãe como cuidadora. A cuidadora assinou os Termos de Consentimento Livre e Esclarecido (TCLE), conforme diretrizes éticas (Ver Apêndice A e B).

### 3.1.2. Procedimento

A coleta de dados ocorreu por meio de gravações em vídeo, permitindo um registro detalhado dos sinais faciais e vocais produzidos durante as interações espontâneas. As interações foram registradas de maneira sistemática. Foram coletadas 45 interações entre bebê e mãe desde outubro de 2024, com coletas realizadas quinzenalmente ou mensalmente.

Para sistematizar a investigação desses múltiplos canais, a análise dos dados foi conduzida com base em um roteiro (Anexo B), desenvolvido por Madureira para esta pesquisa (2025), que integra a avaliação de parâmetros vocais, faciais e acústicos.

Esse roteiro orienta a análise de diversas variáveis, incluindo: (i) na expressão vocal, as modalidades de contornos melódicos, as características prosódicas de *pitch* e *loudness*, a taxa de elocução e os ajustes de qualidade de voz; (ii) na expressão facial, as Unidades de Ação (AUs), a Valência (positiva/negativa) e o Arousal (alta ativação/baixa ativação do organismo); e (iii) nos parâmetros acústicos, as medidas de frequência fundamental ( $F_0$ ), e de intensidade

Para garantir a melhor qualidade na captura das expressões faciais dos participantes, foram utilizadas duas câmeras de alta definição, posicionadas estrategicamente de forma que o enquadramento adequado dos rostos das participantes fosse alcançado. As duas câmeras empregadas foram: CANON VIXIA HS R800 e SONY HANDYCAM DCR-SR 40. As câmeras foram posicionadas a uma distância aproximada de 1,0 metro da díade, com o enquadramento cuidadosamente ajustado para capturar os rostos de forma clara e o mais frontal possível. Esse cuidado no posicionamento visou proporcionar condições ideais para a análise facial, permitindo que o software de reconhecimento facial, ou *FaceReader*, realizasse uma avaliação precisa das expressões faciais durante as interações observadas.

Para garantir a estabilidade e a consistência na captura das imagens ao longo das interações, utilizaram-se dois tripés. Além disso, microfones de captação sensível acoplados às

câmeras foram empregados para possibilitar uma análise detalhada da prosódia vocal. A extração e avaliação dos parâmetros acústicos foram realizadas utilizando o software PRAAT, enquanto o FACS (*Facial Action Coding System*) foi utilizado para a codificação das expressões faciais. Esses recursos permitem a sincronização e análise multimodal dos dados.

É fundamental ressaltar que, para garantir a espontaneidade e a naturalidade das interações, a cuidadora não foi informada previamente sobre os detalhes específicos da análise. Embora ciente do objetivo geral da pesquisa — observar a comunicação mãe-bebê — ela não possuía conhecimento técnico sobre os parâmetros que seriam extraídos com os softwares *FaceReader* e PRAAT. Essa medida foi essencial para evitar qualquer tipo de performance consciente ou enviesada, assegurando que os dados coletados refletissem genuinamente as rotinas interativas da díade.

Antes de cada sessão de gravação, a mãe era orientada a brincar livremente com a bebê, utilizando um livro infantil como objeto de mediação para a interação ou qualquer outra brincadeira.

Em algumas sessões de gravação, optou-se pela utilização adicional de um gravador de celular Xiaomi Redmi Note 14 Pro. Essa medida teve como objetivo assegurar a qualidade do áudio, funcionando como um recurso de apoio em caso de falhas nos microfones das câmeras. O uso do gravador contribuiu para preservar a integridade dos dados sonoros, fator essencial para as análises acústicas conduzidas posteriormente.

Após as gravações, os vídeos foram fragmentados conforme as necessidades analíticas. A segmentação foi guiada por critérios como a ocorrência de episódios de atencionalidade e intencionalidade, aspectos centrais para a compreensão das dinâmicas interativas. Para a fragmentação dos vídeos, utilizou-se o editor *CapCut*, que proporcionou agilidade e precisão na extração dos trechos relevantes, sem comprometer a qualidade visual das gravações.

Com o intuito de reduzir ruídos de fundo e melhorar a inteligibilidade das vozes, especialmente nos casos em que a acústica do ambiente poderia interferir na análise, foi empregado o recurso de inteligência artificial da plataforma *Eleven Labs* para a limpeza e o aprimoramento do áudio. Esse processo permitiu a obtenção de registros sonoros mais claros e estáveis, facilitando a análise posterior no software PRAAT.

A abordagem multimodal adotada no estudo busca compreender como os bebês reagem às prosódias facial e vocal da mãe durante suas interações. Um resumo de estudos longitudinais (ver Saint-Georges et al., 2013, para revisão), indicam que, nos primeiros meses de vida, os

bebês demonstram uma crescente atenção às expressões faciais e vocais, o que é fundamental para o seu desenvolvimento comunicativo.

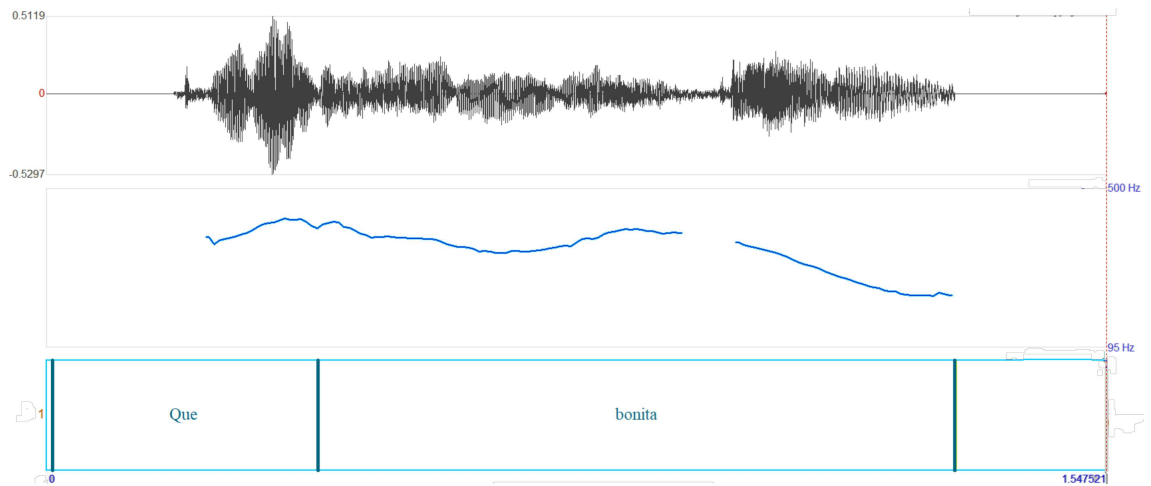
### 3.2. FERRAMENTAS DE ANÁLISE

Para sistematizar a investigação desses múltiplos canais, a análise dos dados foi conduzida com base em um roteiro detalhado, desenvolvido por Madureira para esta pesquisa (2025), que integra a avaliação de parâmetros vocais, faciais e acústicos. Este protocolo orienta a análise de diversas variáveis, incluindo: (i) na expressão vocal, as modalidades de contornos melódicos, as características prosódicas de *pitch* e *loudness*, a taxa de elocução e os ajustes de qualidade de voz; (ii) na expressão facial, as Unidades de Ação (AUs), a valência e o *arousal*; e (iii) nos parâmetros acústicos, as medidas de Frequência Fundamental ( $F_0$ ), e intensidade.

Os dados coletados foram analisados com ferramentas especializadas para a extração e quantificação desses parâmetros. O *software* PRAAT foi utilizado para examinar a prosódia vocal da fala da mãe; o *FaceReader* foi empregado para a codificação das expressões faciais da mãe e da bebê, permitindo a identificação de padrões emocionais e comunicativos.

O PRAAT é um software amplamente utilizado para a análise da prosódia vocal, permitindo a extração de parâmetros acústicos como frequência fundamental ( $F_0$ ), duração, intensidade e formantes. Desenvolvido por Paul Boersma e David Weenink na Universidade de Amsterdã, o PRAAT é uma ferramenta gratuita e de código aberto, amplamente adotada em estudos fonéticos e psicolinguísticos (Boersma; Weenink, 2021). No contexto deste estudo, o PRAAT será utilizado para analisar as características acústicas da frequência fundamental da fala da bebê e da mãe em situações de interações naturais. Estudos anteriores já demonstraram a eficácia do PRAAT na análise da fala dirigida à criança (Fernald et al., 1989; D'odorico; Jacob, 2006).

Figura 1 - Forma da onda, contorno de frequência fundamental e transcrição de enunciado



Fonte: Elaborado pela autora (2025).

O *software FaceReader* aplica o algoritmo desenvolvido por Viola e Jones (2001) para detectar rostos e realiza uma modelagem precisa em 3D utilizando o método *Active Appearance* (Cootes; Walker; Taylor, 2000 apud Fontes; Madureira, 2022). Esse modelo é treinado com um banco de dados composto por imagens anotadas. Para construir a malha facial, mais de 500 pontos-chave são utilizados, conectando áreas como os contornos da face, olhos, cantos dos olhos, sobrancelhas, cantos da boca e a ponta do nariz. Após isso, o modelo facial é ajustado em posição, tamanho e escala para alinhar-se à face real do entrevistado (Loijens; Krips, 2018).

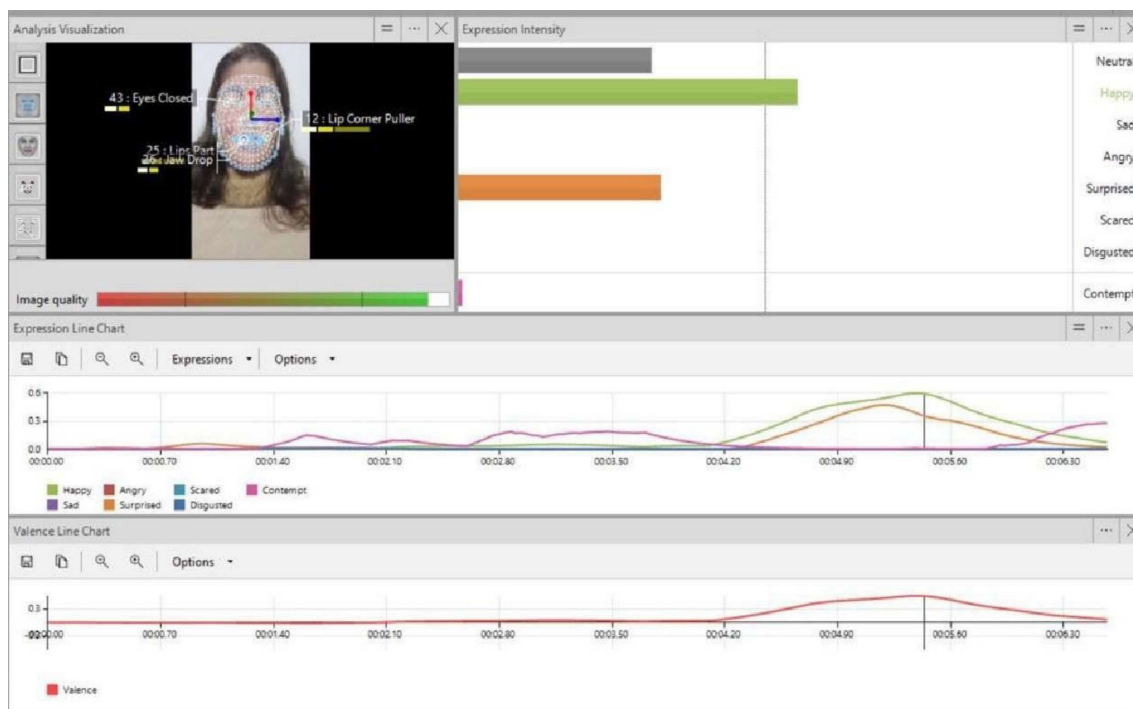
A classificação das expressões faciais é realizada por uma rede neural artificial (Bishop, 1995 apud Fontes; Madureira, 2022). O treinamento da rede utilizou mais de 10.000 imagens anotadas manualmente por especialistas capacitados. Com o método *Deep Face*, o *FaceReader* classifica a face diretamente com base nos pixels da imagem, utilizando a rede neural para reconhecer padrões. Esse processo de classificação, baseado em *Deep Learning*, combinado com o *Active Appearance Model*, aprimora a precisão da análise das expressões faciais. A Figura 2, extraída do *FaceReader*, versão 8.1, ilustra a malha sobre a face de uma participante do sexo feminino, mostrando as Unidades de Ação (AUs) e as emoções associadas às expressões faciais.

A emoção de "alegria" é a mais proeminente, destacando-se na análise. Os gráficos apresentados indicam a evolução das variações dos estados afetivos e da valência. Nos gráficos de evolução emocional, observa-se inicialmente um baixo nível de desprezo (*Contempt*), enquanto, no momento destacado pela linha cinza horizontal, há um aumento significativo da "alegria" (*Happiness*). No gráfico de valência, o traçado inicialmente



acompanha a linha de base, mas no momento focalizado, ele se eleva acima dessa linha, indicando uma valência positiva.

Figura 2 - Análise de Expressões Faciais e Emoções a Partir das Unidades de Ação (AUs)



Fonte: Elaborado pela própria autora (2025)

Tanto o FaceReader como o BabyFacs utilizam o Facial Action Coding System (FACS), desenvolvido por Paul Ekman e Wallace Friesen (1978), que é um sistema amplamente utilizado para a codificação e análise de expressões faciais. Ele permite identificar e classificar movimentos musculares específicos da face, conhecidos como Unidades de Ação Facial (Action Units - AUs). Essas unidades de ação são fundamentais para a análise da comunicação não verbal, pois refletem emoções e intenções comunicativas. Cada AU corresponde à contração ou relaxamento de um ou mais músculos faciais específicos. Por exemplo, a combinação das AUs 6 (elevação das bochechas) e 12 (puxar o canto dos lábios) está associada à expressão de felicidade.

De forma mais específica, a categoria 'alegria' ou felicidade no FaceReader é acionada pela detecção do que Ekman e Friesen (1982) definiram como o 'Sorriso de Duchenne', o marcador de uma emoção genuína. Tecnicamente, ele é identificado pela ativação combinada e síncrona de duas Unidades de Ação principais: a AU 6 (Levantador de Bochecha), que eleva as bochechas e cria rugas ao redor dos olhos, e a AU 12 (Puxador do Canto do Lábio), que eleva

os cantos da boca. É a presença conjunta desses marcadores musculares que permite ao sistema classificar a expressão como uma experiência de alegria autêntica.

O FACS é amplamente utilizado em diversas áreas, incluindo pesquisa científica, animação e reconhecimento facial por computador. Na pesquisa científica, ele permite uma análise objetiva e confiável das expressões faciais, sendo útil em estudos de psicologia, comunicação e comportamento humano.

No contexto deste estudo, o FACS será utilizado para analisar as expressões faciais da bebê e de sua mãe durante as interações, permitindo a consideração sobre a relação entre as expressões faciais e os estados emotivos manifestados. A análise das AUs, em conjunto com a prosódia vocal e gestual, fornecerá uma visão abrangente da comunicação multimodal entre cuidadores e bebês, contribuindo para a compreensão dos mecanismos subjacentes ao desenvolvimento comunicativo nos primeiros meses de vida.

Toda a metodologia empregada, incluindo os procedimentos de coleta e análise de dados, foi submetida e aprovada pelo Comitê de Ética em Pesquisa (CEP) da Universidade Federal de Juiz de Fora, sob o parecer de número 7.025.159, garantindo a conformidade ética e o respeito aos participantes do estudo.

No próximo capítulo, serão apresentadas e discutidas as análises dos dados de prosódia visual da mãe e da bebê, e da prosódia vocal da mãe, em duas situações de interação mãe e bebê aos 6 meses e aos 12 meses de vida.

4. ANÁLISE DOS DADOS

Este capítulo apresenta a investigação empírica que constitui o cerne desta dissertação. A partir da metodologia multimodal detalhada anteriormente, realizaremos uma microanálise das interações da díade em dois momentos cruciais do desenvolvimento: aos 6 e aos 12 meses de idade. O objetivo é demonstrar, por meio da análise integrada dos dados comportamentais, fisiológicos e acústicos, a trajetória evolutiva da comunicação diádica, processo fundamental para a aquisição da linguagem. Segue quadro resumitivo das interações:

Quadro 1 – Detalhamento dos Segmentos Analisados

Idade da Bebê	Data da Gravação	Duração Total da Gravação	Fragmento Analisado	Breve Descrição da Atividade
6 meses	22/10/2024	3 min 22 seg	33.74 segundos	A mãe está sentada no chão, de frente para a bebê, narrando de forma expressiva a história do livro "Bambi", apontando para as figuras. O foco é a sintonização afetiva.
12 meses	14/04/2025	2 min e 24 seg	29.81 segundos	A mãe está sentada em uma cadeira em frente à cadeira de alimentação onde a bebê se encontra. A mãe tenta engajar a bebê em um jogo de identificação de figuras em um livro, mas a atenção da criança é disputada por um som externo, exigindo estratégias de reparo atencional.

Fonte: Elaborado pela autora (2025).

4.1. ANÁLISE DA INTERAÇÃO DIÁDICA AOS 6 MESES: A EMERGÊNCIA DA ATENÇÃO COMPARTILHADA

A comunicação entre mãe e bebê nos primeiros meses de vida é um campo fértil para o estudo da expressividade vocal e facial. As modificações que os adultos, especialmente os cuidadores, promovem em sua fala ao se dirigirem a bebês são um fenômeno robusto e transcultural. Essas alterações, conhecidas como "manhês" ou *Infant-Directed Speech* (Fala Dirigida à Criança), não são meramente casuais, mas funcionam como ferramentas eficazes para o engajamento, a regulação emocional e o próprio desenvolvimento linguístico da criança. (Fernald et al., 1989; Snow, 1977).

Longe de ser um monólogo do cuidador, essa interação é uma "dança" finamente orquestrada, um diálogo multimodal onde olhares, sorrisos, gestos e vocalizações se entrelaçam em um processo de correção mútua. Este capítulo se dedica a uma microanálise de um desses momentos, buscando detalhar os múltiplos canais comunicativos em foco para revelar a complexidade da sincronia diádica.

O objeto de nossa seção é um segmento de 33.74 segundos de uma interação naturalística entre uma mãe e sua filha de seis meses, durante a leitura do livro infantil "Bambi". A escolha dessa atividade não é acidental: a narração de uma história é um cenário ideal para observar como um cuidador utiliza intencionalmente a expressividade para modular o afeto, capturar a atenção e convidar o bebê a participar de uma experiência compartilhada.

Para detalhar as camadas dessa complexa interação, o percurso analítico deste capítulo será organizado em três grandes eixos, que se complementam para construir uma visão integrada do fenômeno:

1. **A Análise da Situação Interativa:** Iniciaremos com uma descrição qualitativa e detalhada da cena. O objetivo é apresentar o ambiente, o posicionamento da díade e o desenrolar narrativo da interação, estabelecendo o cenário sobre o qual as análises mais aprofundadas irão operar.
2. **A Análise da Interação a partir do *FaceReader*:** Em um segundo momento, a análise se aprofundará no nível microcomportamental e fisiológico, utilizando dados extraídos do software *FaceReader* (NOLDUS). Nesta seção, examinaremos a expressividade facial (Valência, *Arousal*, Unidades de Ação) e os dados de frequência cardíaca de mãe e bebê, buscando averiguar se há dados relativos a sincronia, contingência temporal e processos de regulação mútua que definem a "dança" interativa.
3. **A Análise Acústica e Prosódica a partir do PRAAT:** Finalmente, o terceiro eixo analítico se concentrará no canal vocal da mãe, com o auxílio do software PRAAT (Boersma & Weenink, 2022). Investigaremos a prosódia de sua fala por meio de análise de parâmetros acústicos (frequência fundamental e intensidade) e de análise perceptiva da qualidade e a dinâmica vocais, comparando a Fala Dirigida à Criança com seu perfil vocal de base. O objetivo é verificar em que medida a mãe esculpe ativamente a sonoridade de sua voz, transformando-a em uma poderosa ferramenta para o engajamento afetivo.

#### 4.1.1. Descrição Detalhada do Vídeo

A cena se passa em um ambiente interno, a sala de uma casa, o que confere um tom pessoal. A mãe está sentada no chão, com a bebê deitada confortavelmente num tapete de EVA, de frente para ela. Essa posição permite um contato visual direto e constante entre as duas.

A Protagonista e o Ambiente:

O foco principal é a mãe. Seu rosto é extremamente expressivo; ela se inclina para a frente, em direção à bebê, com os olhos bem abertos e um semblante de grande entusiasmo.

A Narração da História:

A mãe segura um livro infantil grande e colorido, aberto em seu colo, e o apresenta para a bebê. Ela está narrando a história com uma voz calorosa, animada e cheia de inflexões dramáticas. A história é a de Bambi.

O vídeo acompanha a sua narração enquanto ela vira as páginas:

1. **A Chuva e o Resgate:**(aproximadamente entre 1:24 e 1:26 da interação original). Ela começa apontando para uma ilustração e diz: "Ih, um dia choveu, olha a chuva, olha!". Sua voz transmite urgência. A qualificação perceptiva de 'urgência' ancora-se em marcadores acústicos e prosódicos específicos identificados na análise. O enunciado 'Ih, um dia choveu, olha a chuva, olha!' é caracterizado por um aumento abrupto na frequência fundamental média (*pitch*), uma expansão da faixa dinâmica com picos de intensidade (*loudness*) e uma aceleração na taxa de elocução, em nítido contraste com o ritmo mais cadenciado do restante da narrativa. Essa combinação de agudeza, volume e velocidade funciona como um potente atrator de atenção, sinalizando para o ouvinte a importância e a dramaticidade do evento narrado. Ela conta que o personagem (Bambi) se perdeu na chuva e foi encontrado por outro veado, que o devolveu para a mãe.
2. **O Inverno e o Caçador:**( aproximadamente entre 1:36 da interação original). Ela vira a página e sua expressão muda para uma de espanto. "Olha a neve!", exclama. Ela mostra uma paisagem de inverno e anuncia com uma voz tensa: "Ficou inverno... Ih, o caçador!". Ela aponta para a figura de um caçador no livro. "O caçador chegou e vai atirar no Bambi!".
3. **A Tragédia:** (aproximadamente 1:48 da interação original). Sua expressão se torna triste e sua voz, pesarosa. "Ah, tadinho... Matou a mamãe do Bambi", ela diz, mostrando a página correspondente. Ela explica que Bambi ficou sozinho com os outros animais da floresta.
4. **O Crescimento de Bambi:** (aproximadamente 1:56 da interação original) A energia da cena muda novamente. Com um sorriso e voz animada, ela vira a página e mostra

uma nova ilustração. "Ih, ele cresceu! Olha como é que ele ficou grande!". Ela aponta com entusiasmo para a imagem de Bambi já adulto. "Nasceu o chifrinho no Bambi!".

5. **A Descoberta:**(aproximadamente 2:06 da interação original). Ela aproxima o livro da bebê, apontando para os chifres recém-crescidos do personagem. "Nasceu! Olha aqui, ó! Olha só!". Neste momento, a bebê, fora do campo da câmera, vocaliza. A mãe diz: "É, nasceu!" e sorri em concordância.

6. **O Retorno da Ameaça:**(aproximadamente 2:12 da interação original). No final do clipe, ela vira mais uma página e, com um sobressalto e olhos arregalados, exclama dramaticamente: "Ih! O caçador de novo!", encerrando o vídeo em um momento de suspense.

#### 4.1.2. Análise da Interação a partir do *FaceReader* (Noldus)

Nesta seção, aprofunda-se o estudo da interação, movendo o foco do comportamento individual para a díade como uma unidade sistêmica e corregulada. A partir dos dados simultâneos da mãe e da bebê, e em consonância com o arcabouço teórico apresentado, examinamos a micro sincronia e a contingência momento a momento que caracterizam a "dança" interativa (Stern, 1985). A utilização de protocolos do *software* de reconhecimento facial e de dados fisiológicos (frequência cardíaca) para examinar interações precoces alinha-se a uma tradição de microanálise que busca investigar a complexidade da comunicação pré-verbal (Beebe & Lachmann, 2002). Este método fornece uma "anatomia" rica e quantificável do episódio interativo. A análise microcomportamental e fisiológica corrobora, com dados de alta resolução temporal, a existência de um circuito de sintonização afetiva (Stern, 1985) entre mãe e bebê.

##### 4.1.1.1. Dados da Mãe: A "Oferta" Comunicativa

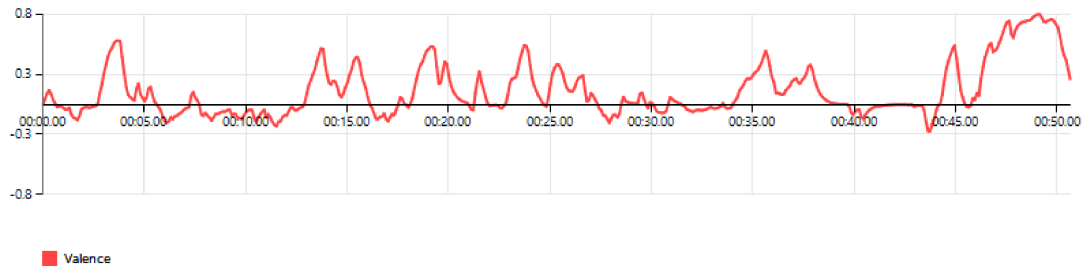
A performance vocal e facial da mãe não é meramente expressiva, mas uma oferta comunicativa deliberadamente orquestrada para capturar a atenção e modular o estado afetivo da bebê. De acordo com Bruner (1975), esta função de andaime (*scaffolding*) é fundamental nas interações iniciais. A análise dos dados da mãe no momento da narração (~12.70 segundos) revela os dados abaixo explicados.

#### Intensidade Afetiva (Valência e Ativação do Organismo)

Os gráficos de Valência e *Arousal* exibem picos agudos. A valência da mãe atinge seu ponto mais positivo, enquanto o *arousal* demonstra um estado de alta ativação. Isso indica que

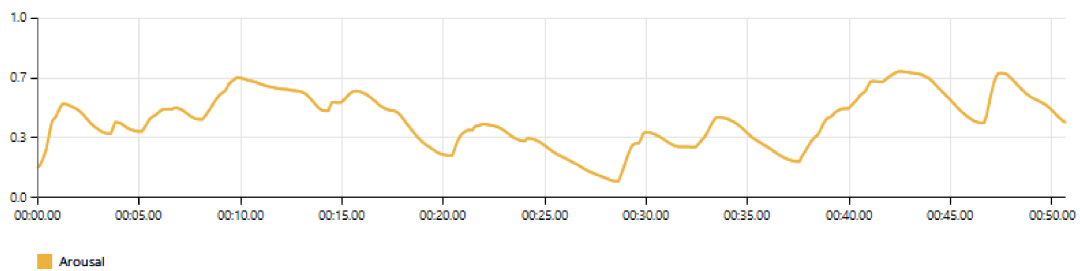
a mãe investe um esforço significativo para tornar o estímulo saliente e atraente, um processo de marcação afetiva (Stern, 1985).

Gráfico 1 - Gráfico de Valência (Mãe)



Fonte: Elaborado pela autora (2025)

Gráfico 2 - Gráfico de Arousal (Mãe)

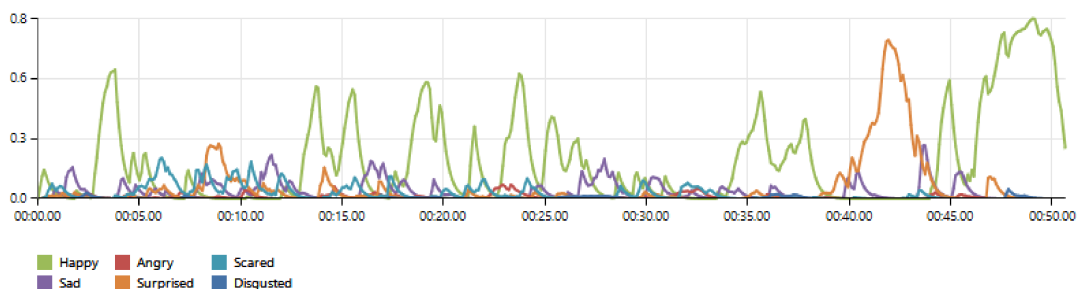


Fonte: Elaborado pela autora (2025)

### Modulação Emocional ao Longo da Narração

O gráfico de expressões emocionais demonstra a rica e dinâmica modulação afetiva da mãe ao longo da narrativa. As diferentes linhas representam a intensidade das seis emoções básicas detectadas pelo *FaceReader*: "Feliz" (verde), "Triste" (roxo), "Surpresa" (laranja), "Assustado" (azul claro), "zangado" (vermelho) e "Com nojo" (azul escuro).

Gráfico 3 - Gráfico de expressões emocionais (Mãe)



Fonte: Elaborado pela autora (2025).

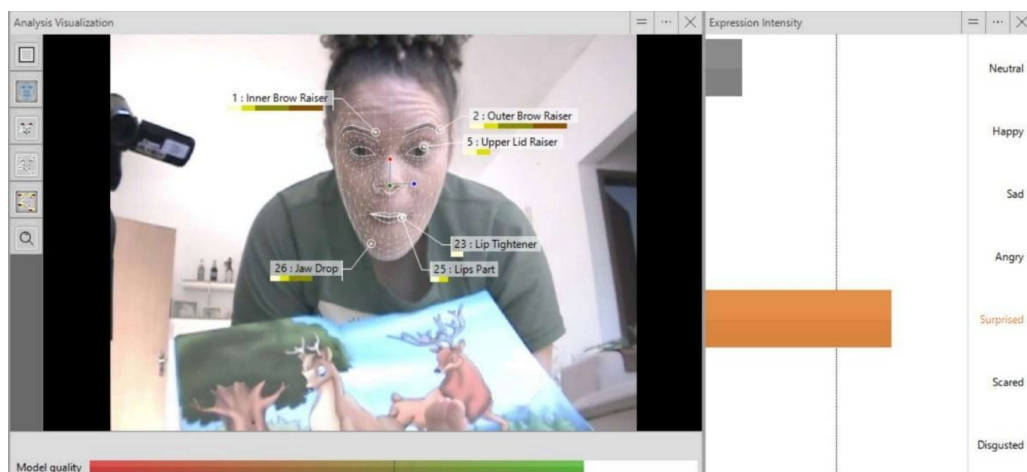


Observa-se que a expressão predominante é a de felicidade (*Happy*, linha verde), que apresenta múltiplos picos de alta intensidade, correspondendo aos momentos de entusiasmo ao mostrar o livro para a bebê. No entanto, a performance não é monotemática. Destaca-se um pico notável da expressão de surpresa (*Surprised*, linha laranja) por volta dos 42 segundos, que se alinha perfeitamente com o momento em que a mãe narra a parte mais trágica da história ("Matou a mamãe do Bambi"). As flutuações menores de tristeza (*Sad*, linha roxa) também coincidem com as viradas de página e as exclamações ("Ih, o caçador!"). Este padrão demonstra que a mãe não apenas sente, mas utiliza ativamente as expressões emocionais como uma ferramenta narrativa, modulando seu rosto para contar a história e guiar a experiência afetiva da filha.

### A Anatomia da Expressão Feliz (Unidades de Ação)

A análise específica da expressão "Feliz" revela que ela é constituída pela coativação de alta intensidade das Unidades de Ação AU6 (Elevador das Bochechas; *Cheek Raiser*) e AU12 (Extrator do Canto Labial; *Lip Corner Puller*). A mãe produz um sorriso de Duchenne claro e intenso (ver seção 4.1.2.2), um sinal social que pode ser interpretado como indicador de emoção positiva e convite à interação (Ekman, 1992).

Figura 3 - Momento de afirmação de uma interlocução bem-sucedida



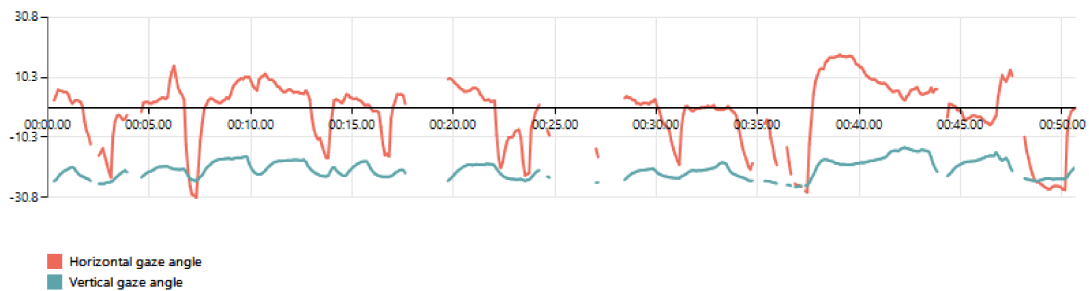
Fonte: Elaborado pela autora (2025)

### Engajamento Visual (Direcionamento do Olhar)

A análise do direcionamento do olhar da mãe corrobora sua performance comunicativa intencional. O gráfico abaixo exhibe a variação dos ângulos de olhar horizontal (em vermelho) e vertical (em azul) ao longo do tempo.



Gráfico 4 - Gráfico de orientação do olhar (Mãe)



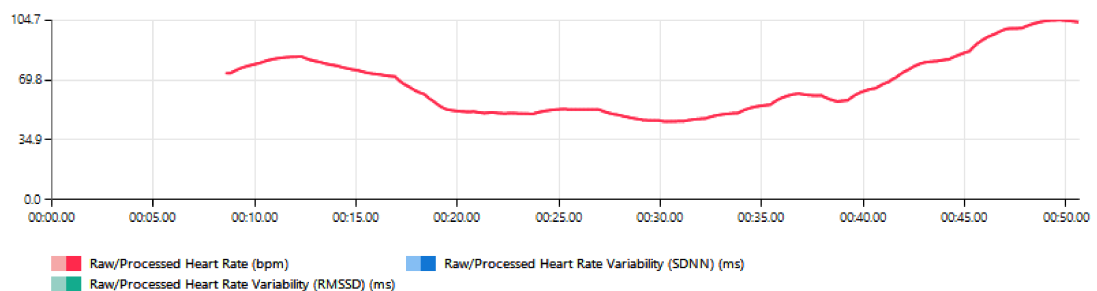
Fonte: Elaborado pela própria autora (2025)

As flutuações nas duas linhas indicam o movimento dinâmico da cabeça e dos olhos da mãe. Observa-se que a mãe não mantém um olhar fixo; pelo contrário, ela alterna seu foco visual. Os picos e vales no gráfico correspondem aos momentos em que ela olha para o livro, para a bebê e possivelmente ao redor, em uma "dança" visual que acompanha sua narração. Essa movimentação é crucial para checar a atenção da bebê, direcionar o foco dela para as ilustrações e manter o circuito de atenção compartilhada ativo. A variação no olhar é, portanto, mais um componente do seu andaime (*scaffolding*), demonstrando que seu engajamento é ativo e constantemente ajustado para manter a conexão com a filha.

#### Engajamento Fisiológico (Frequência Cardíaca)

A análise da frequência cardíaca da mãe revela engajamento também no nível fisiológico. Sua frequência cardíaca se eleva durante esses picos expressivos, confirmando que a performance envolve uma mobilização de todo o seu organismo para a tarefa de se conectar com a bebê.

Gráfico 5 - Gráfico de batimentos cardíacos (Mãe)

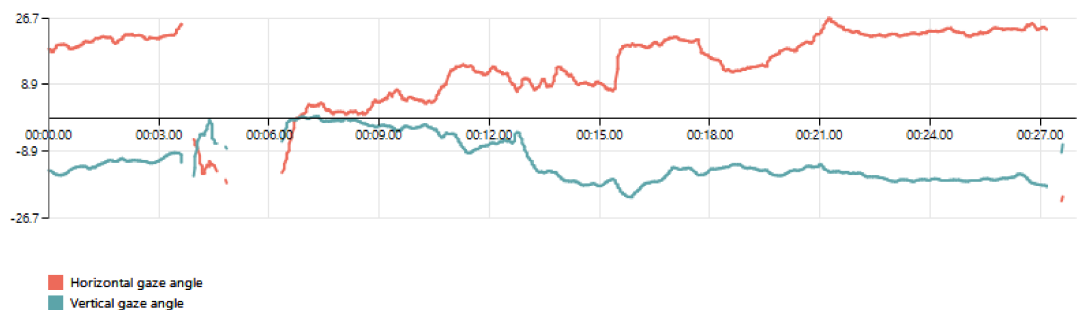


Fonte: Elaborado pela autora (2025)

#### 4.1.2.2. Dados da Bebê: Recepção e Resposta Afetiva

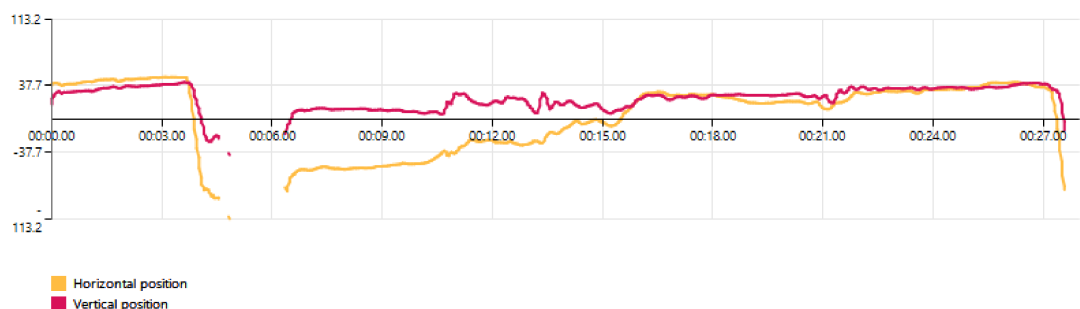
A análise qualitativa apontou que a bebê se orientava para a voz da mãe. Os dados dos gráficos *Head Position* e *Gaze Angles* (posição da cabeça e orientação do olhar) quantificam este comportamento. Observa-se que, nos momentos que antecedem a vocalização mais enfática da mãe, que ocorre por volta dos 19 segundos, há mudanças sutis, porém mensuráveis, nos eixos de rotação da cabeça e nos ângulos do olhar. Isso sugere fortemente que a bebê está em um estado de escuta ativa, ajustando sua orientação para otimizar a recepção do estímulo social materno, um comportamento que reflete a motivação inata do infante para o engajamento interpessoal (Trevarthen, 1979).

Gráfico 6 - Gráfico direcionamento do Olhar (Bebê)



Fonte: Elaborado pela autora (2025)

Gráfico 7 - Gráfico direcionamento da Cabeça (Bebê)



Fonte: Elaborado pela própria autora (2025)

O clímax da sintonização afetiva ocorre em resposta à mudança na prosódia materna. A análise dos dados permite detalhar este momento:

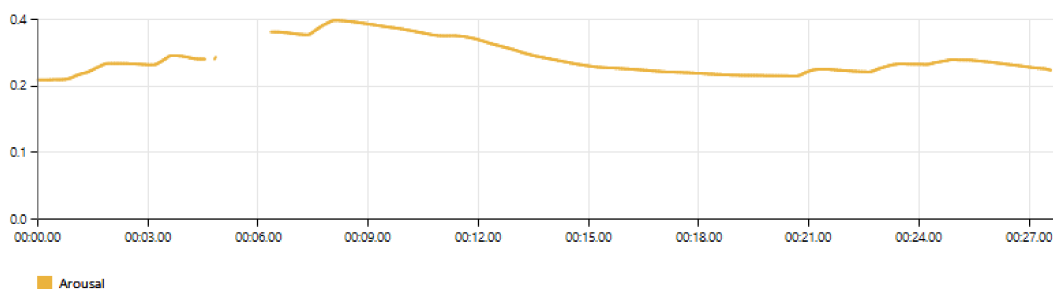
### Ativação do Organismo e Valência

Os gráficos mostram um pico acentuado e simultâneo nos eixos de Valência (agradabilidade) e Arousal (ativação do organismo). O padrão geral revelado no gráfico de

ativação do organismo é o de uma linha de base predominantemente calma, indicando que a bebê se encontra em um estado de receptividade confortável. Sobre essa base, contudo, emergem picos de ativação agudos e temporalmente localizados, que demonstram uma forte contingência com a performance narrativa da mãe. O pico de maior proeminência, por exemplo, ocorre por volta dos 41 segundos e coincide precisamente com o momento em que a mãe narra o evento mais dramático da história: "Matou a mamãe do Bambi". Isso sugere que, embora o conteúdo afetivo seja negativo, a intensidade dramática da enunciação materna provoca uma forte reação de ativação no organismo da bebê. Outros picos de menor magnitude também se alinham a momentos de exclamação e surpresa na fala da mãe, como "Olha a neve!" e "Ih, o caçador!" (aproximadamente 1:36 da interação original), funcionando como disparadores que modulam o estado de alerta da criança.

A análise da ativação do organismo evidencia que a bebê não atua como uma ouvinte passiva. Seu estado fisiológico é diretamente influenciado pela modulação dramática da mãe, tornando visível a "dança" da correção mútua. Este dado complementa a análise das expressões emocionais discretas, ao isolar a dimensão da intensidade da experiência e quantificar o impacto da performance narrativa materna no organismo da bebê, demonstrando a eficácia das estratégias vocais e gestuais para capturar e manter o engajamento da bebê.

Gráfico 8 - Gráfico de Ativação do Organismo (Bebê)



Fonte: Elaborado pela autora

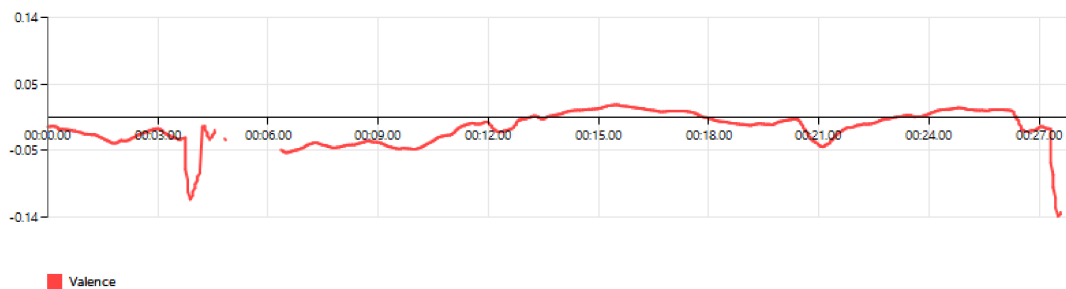
O gráfico da Valência da bebê demonstra um padrão notavelmente claro e significativo. A linha de base de sua experiência situa-se predominantemente em um nível neutro, próximo a zero, o que sugere um estado de observação calma e atenta, sem grande expressão de prazer ou desprazer.

Este estado de neutralidade é, contudo, dramaticamente rompido por um único e proeminente pico de valência positiva, que se manifesta na fala por *pitch* agudizado e *loudness* forte. Este evento não é aleatório; ele está temporalmente alinhado com um momento específico

da performance materna: o clímax positivo da narrativa, quando a mãe exclama de forma entusiástica "Nasceu! Olha aqui, ó! Olha só!". A emergência deste pico de prazer no exato momento da "oferta" afetiva da mãe é a evidência mais contundente da sintonização afetiva (*affective attunement*, Stern, 1985) entre a díade.

Portanto, a análise da valência nos certifica a interação bem-sucedida e os esforços e estratégias da mãe. A mãe não apenas consegue modular o nível de ativação do organismo da bebê, mas também consegue evocar uma resposta afetiva congruente e de qualidade positiva. É neste terreno de valência positiva compartilhada que os alicerces da confiança e do prazer na interação social são estabelecidos.

Gráfico 9 - Gráfico de Valência (Bebê)

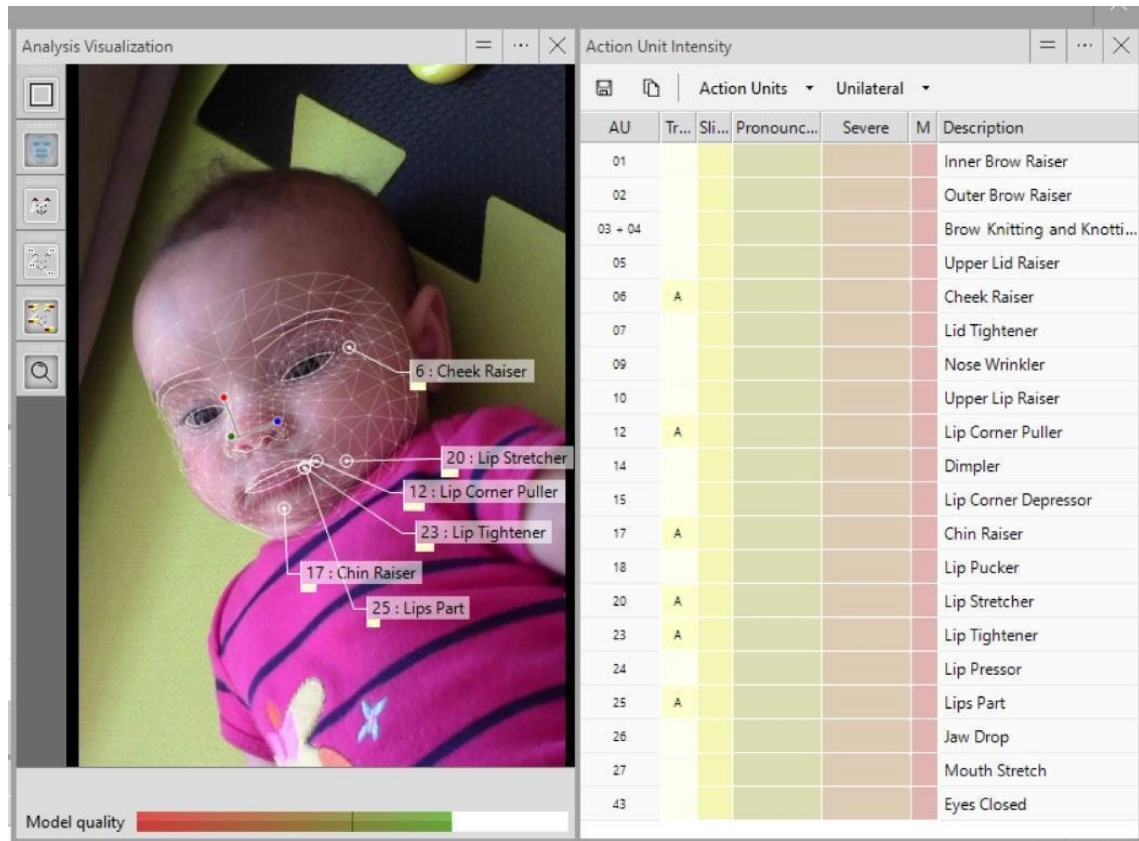


Fonte: Elaborado pela autora

### A "Anatomia" do Sorriso (*Action Units*)

Uma análise aprofundada das Unidades de Ação Facial (AUs) da bebê, baseada no Facial Action Coding System (FACS) e extraída pelo software FaceReader, revela uma expressão facial de grande complexidade, que vai além de um simples sorriso. A imagem abaixo detalha a intensidade de cada ativação muscular no momento da resposta da bebê.

Figura 4 - Ativação das AUs (Bebê)



Fonte: Elaborado pela autora

O núcleo da expressão é o Sorriso de Duchenne, indicado pela ativação intensa e simultânea da AU12 (Extrator do Canto Labial; *Lip Corner Puller*), que eleva os cantos da boca, e da AU6 (Elevador das Bochechas; *Cheek Raiser*), que levanta as bochechas. Como estabelecido por Ekman e Friesen (1982), essa combinação é o marcador de uma experiência de alegria genuína.

No entanto, outras AUs ativas adicionam camadas cruciais de significado a essa expressão de prazer. Observa-se uma ativação igualmente forte da AU 25 (Lábios Entreabertos; *Lips Part*) e da AU 26 (Queda do Queixo; *Jaw Drop*), que indicam que a boca da bebê está aberta. Essa configuração é característica não apenas de um sorriso, mas de uma exclamação de alegria ou de uma risada incipiente, demonstrando um alto grau de engajamento participativo.

Adicionalmente, ativações de menor intensidade são detectadas na parte superior do rosto. A AU 2 (Elevador Externo da Sobrancelha; *Outer Brow Raiser*) e a AU 5 (Elevador da Pálpebra Superior; *Upper Lid Raiser*), responsáveis por levantar as sobrancelhas externas e as pálpebras superiores, respectivamente, e conferem à expressão um componente de interesse

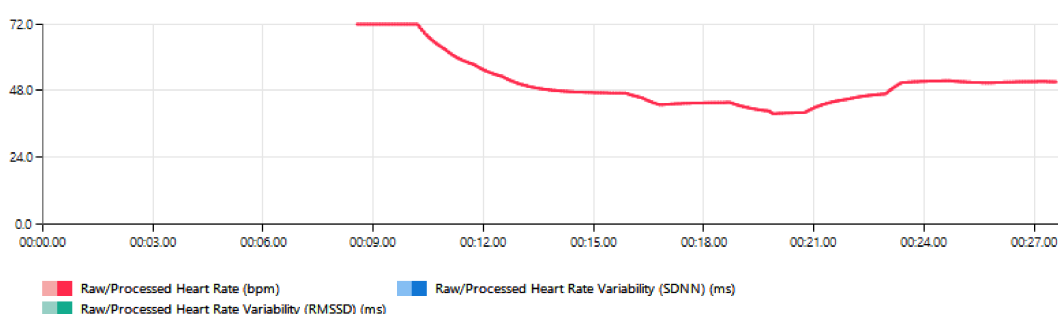
aguçado e talvez até de feliz surpresa. A leve ativação da AU10 (Elevador do lábio superior; *Upper Lip Raiser*) complementa a amplitude do sorriso.

A expressão facial da bebê não é apenas "feliz". Trata-se de uma rica constelação de movimentos que, em conjunto, comunicam uma alegria engajada, participativa e expressiva. A análise completa das AUs permite validar objetivamente que a resposta da bebê à mãe é uma reação complexa e multimodal, combinando o prazer genuíno (AU6+12) com a vocalização aberta (AU25+26) e o interesse intenso (AU2+5).

### Engajamento Fisiológico da bebê

A análise dos dados fisiológicos da bebê oferece uma janela para as respostas de seu organismo que não são visíveis a olho nu, validando o engajamento em um nível biológico fundamental. O gráfico abaixo apresenta sua Frequência Cardíaca (*Heart Rate*) e a Variabilidade da Frequência Cardíaca (*Heart Rate Variability*). A frequência cardíaca mede o número de batimentos por minuto, funcionando como um indicador direto do estado de ativação do sistema nervoso autônomo.

Gráfico 10 - Batimentos cardíacos (Bebê)



Fonte: Elaborado pela autora

O gráfico revela um evento fisiológico notável e temporalmente (aos 9 segundos no gráfico) localizado. Enquanto a linha de base da frequência cardíaca da bebê demonstra um ritmo relativamente estável, observa-se um pico agudo e significativo de aceleração cardíaca em um momento específico da interação. Este aumento súbito nos batimentos por minuto é um correlato fisiológico clássico de um estado de atenção focada e de processamento de um estímulo saliente.

De forma crucial, este pico de aceleração cardíaca está temporalmente (aos 9 segundos do gráfico) alinhado com os momentos de maior intensidade na interação com a mãe e com as

respostas afetivas da própria bebê (os picos de ativação do organismo e valência positiva, e a expressão do Sorriso de Duchenne). Isso indica que a "oferta" comunicativa da mãe – sua voz entusiástica, seu sorriso e seu olhar – não apenas evoca uma resposta comportamental (o sorriso), mas também provoca uma reação fisiológica mensurável. O organismo da bebê responde em uníssono com sua expressão facial e aceleração cardíaca que, nesse contexto, não sinaliza estresse, mas uma mobilização de recursos fisiológicos para o processamento de uma interação social positiva e prazerosa.

A análise da frequência cardíaca transcende a observação do comportamento e fornece um indicador objetivo do engajamento profundo da bebê. Ela mostra que a sincronia diádica não é apenas uma "dança" comportamental, mas também uma ressonância fisiológica, onde a modulação afetiva do cuidador tem o poder de regular, momento a momento, o estado fisiológico do bebê.

#### **4.1.2. Análise Acústica e Prosódica a partir do PRAAT**

Esta subseção apresenta uma análise detalhada da expressão vocal da mãe ao interagir com seu bebê. A análise foca nos parâmetros prosódicos e acústicos da fala, contrastando a fala dirigida ao bebê (*Infant-Directed Speech*) com a voz usual da falante, com base em um protocolo detalhado de análise (ver Anexo B).

Para compreender a magnitude das adaptações vocais da mãe na interação com a bebê, é essencial primeiro caracterizar sua voz em um contexto de fala usual, não dirigida à criança. Para tanto, será utilizado um protocolo de análise perceptivo-auditiva desenvolvido por Madureira e Fontes (2025) para traçar este perfil de base.

A partir deste protocolo, serão avaliados os seguintes aspectos da voz da mãe: as características prosódicas de *pitch* e *loudness* (incluindo suas médias, extensões e variabilidades); a taxa de elocução; e os ajustes de qualidade de voz, que englobam configurações do trato vocal, de tensão muscular e de fonação. A caracterização deste perfil servirá como um ponto de referência crucial para contrastar com as modulações vocais observadas na interação com a bebê, que serão detalhadas adiante.

##### *4.1.2.1. Perfil Vocal de Base*

A Frequência Fundamental ( $F_0$ ) apresenta uma média de 135 Hz, valor considerado baixo para a média de referência feminina de 220 Hz. Seus valores mínimo e máximo em fala usual são de 82 Hz e 250 Hz, respectivamente. Perceptivamente, o *pitch* é classificado como

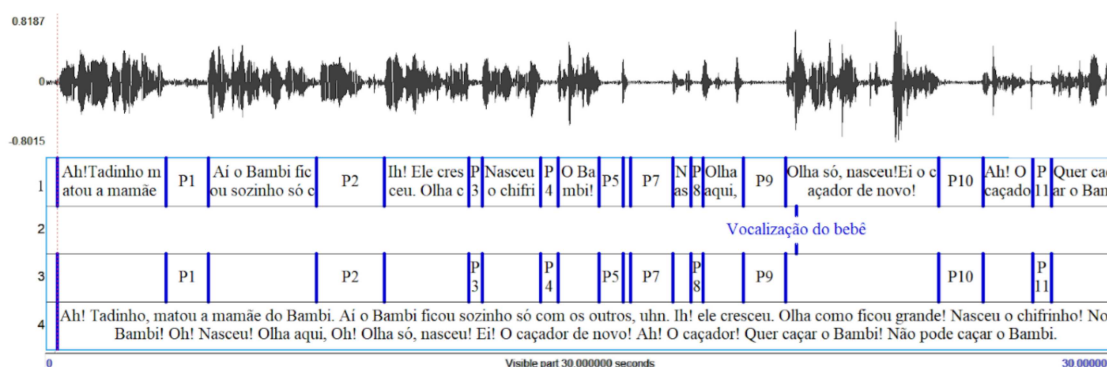


baixo, com extensão e variabilidade minimizadas. Sua voz usual exibe ajustes de qualidade como *Creaky Voice* (grau 2), *Harsh* (grau 2), *Whispery voice* intermitente e Laringe Abaixada (grau 1), de acordo com o roteiro elaborado por Madureira (2025).

#### 4.1.2.2. Análise da Expressão Vocal na Interação

A fala da mãe no trecho analisado é marcada por uma expressividade acentuada, operando como um atrator de atenção para o bebê. A seguir, detalhamos os componentes dessa performance vocal.

Figura 5 - Perfil do PRAAT



Fonte: Elaborado pela autora

### Contornos Melódicos e Modalidades

A análise do segmento revela um predomínio de contornos exclamativos, sendo três ocorrências de contornos assertivos e sete de contornos exclamativos (ver Anexo B). A entoação exclamativa confere vivacidade à fala por expressar entusiasmo e estados de ânimo exaltados, funcionando como um potente atrator de atenção para o ouvinte.

#### Pitch

O *pitch*, correlato perceptivo da frequência fundamental, é o parâmetro que sofre a modificação mais drástica. A análise perceptiva classifica sua média como alta, com extensão maximizada e variabilidade alta (ver Anexo B). Essa variabilidade elevada confere vivacidade à fala, opondo-se a uma fala monótona, que se caracteriza por pouca variação de *pitch*. Acusticamente, a F0 da mãe varia de um mínimo de 79 Hz a um máximo de 538 Hz. Essa elevação pode ser interpretada à luz do Código de Frequência de Ohala (1984), que argumenta haver correlações entre vocalizações agudas, tamanho pequeno e delicadeza. Ao agudizar a voz,



a mãe projeta uma persona vocal percebida como menos impositiva, ideal para a interação com bebês.

### **Loudness**

A *Loudness*, correlato perceptivo de intensidade, é classificada como média e de extensão regulares, mas com variabilidade alta (ver Anexo A). Seus valores variam de uma mínima de 45.54 dB a uma máxima de 85.71 dB. A alta variabilidade de *loudness*, que implica uma alternância entre volumes fortes e fracos, cria expectativas e destaca pontos de culminância na fala, que podem se constituir em pontos de atratividade para a atenção da bebê.

### **Elocução e Qualidade Vocal**

A taxa de fala é classificada como lenta. Em um segmento com duração total de 33.74 segundos, as pausas silenciosas somam 8.9 segundos, correspondendo a 26.04% do tempo total (ver Anexo A). Essa presença marcante de pausas reflete o caráter interativo da comunicação, no qual a mãe cria espaços para a participação da bebê, transformando a contação de história em uma interlocução. Foram identificados ajustes de qualidade de voz (VPA - Laver, 1980) como a Laringe Levantada (Grau 1), que diminui o tamanho do trato vocal e reforça frequências altas, gerando uma percepção de voz mais aguda e delicada. Adicionalmente, o ajuste de fonação, *Whispery Voice* (sussurrada), que combina escape de ar com vibração das pregas vocais, é utilizado para suavizar a emissão, expressando gentileza e ternura.

### **Estratégias Interativas e Gatilhos de Atenção:**

A análise dos eventos interativos revela estratégias deliberadas para atrair e manter o foco da bebê (ver Anexo A). A mãe utiliza recorrentemente 5 interjeições ("Ah!", "Ih!", "Oh!", "Ei!") e 3 repetições da palavra "Olha" como atratores de atenção.

Um evento relevante ocorre quando a bebê produz uma vocalização alinhada com a fala da mãe ("Olha só, nasceu"). A reação da mãe é imediata: ela aumenta ainda mais o *pitch* e a extensão de sua  $F_0$ . No trecho que segue a reação da bebê, a  $F_0$  da mãe atinge seu pico máximo de 539.8 Hz e um mínimo de 179 Hz. Isso demonstra um ciclo de retroalimentação positiva, onde a reação da bebê motiva a mãe a intensificar sua expressividade, refletindo um aumento da ativação do organismo (*arousal*) de ambas.

#### **4.1.3. Análise da Sincronia Diádica**

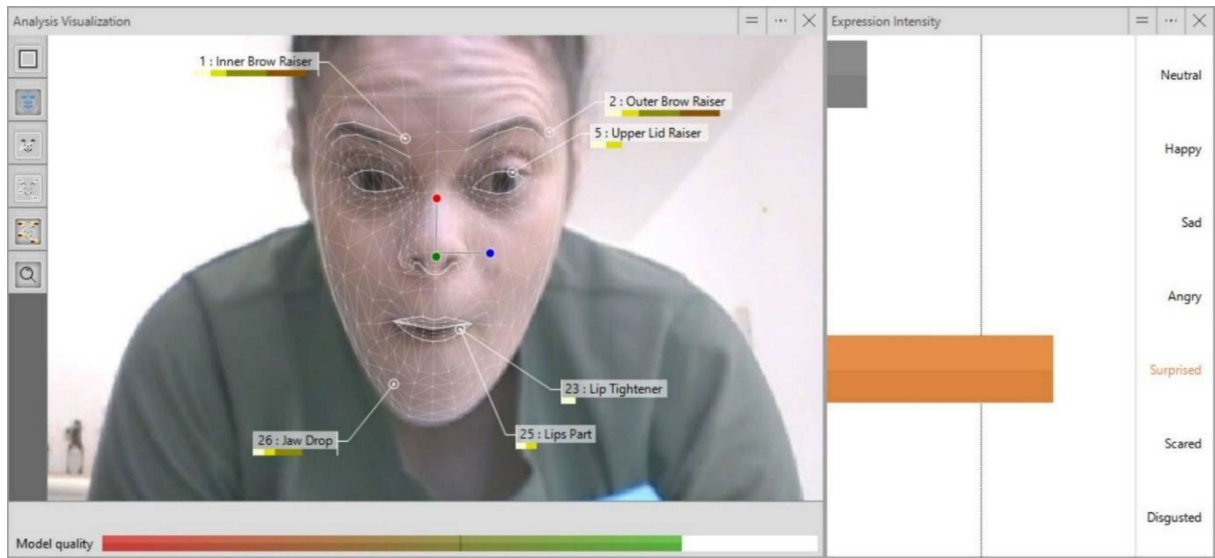
A emissão do enunciado responsivo da bebê atesta três competências cruciais: o alinhamento atencional, que confirma que seu foco perceptual e cognitivo estava alinhado ao da mãe; a compreensão pragmática, que evidencia a compreensão da intenção comunicativa da mãe de celebrar um evento narrativo; e, finalmente, a agência comunicativa, que posiciona a bebê não como receptor passivo, mas como interlocutor ativo, que assume seu turno e co-constrói o momento interativo.

A análise da sincronia diádica foca em como a interação é estruturada e nos mecanismos que a mãe utiliza para construí-la. A arquitetura da interação observada ancora-se em uma estrutura triádica (mãe-bebê-livro) que, embora inicial, segue em direção às formas mais consolidadas de atenção compartilhada. O livro, neste contexto, opera como um pivô referencial, um pretexto que legitima e organiza a performance comunicativa da mãe. Ao posicionar o objeto no campo visual do bebê, a mãe executa um ato comunicativo intencional, cuja finalidade pragmática é a de recrutar a bebê para uma experiência perceptual e narrativa comum. Segundo Tomasello (2003), a habilidade de participar de tais tríades é um divisor de águas no desenvolvimento infantil, marcando a emergência da compreensão do outro como um agente intencional.

Para estabelecer essa interação triádica, as estratégias de andamento (*scaffolding*) utilizadas pela mãe constituem um repertório sofisticado de ações comunicativas. A sua performance multimodal, em que os canais vocal, gestual e facial se integram, reflete o postulado de McNeill (2005) de que fala e gesto são facetas de um único processo cognitivo e expressivo. O gesto de apontar, por exemplo, é empregado com precisão dêitica para direcionar a atenção visual do bebê. Ele não é um mero acompanhamento, mas um componente integral da unidade enunciativa, coexpressando o significado em um continuum de fala e gesto, sendo um precursor fundamental das formas mais explícitas de atenção compartilhada.

Ao sobrepor temporalmente os dados de mãe e bebê, a sincronia torna-se visível e mensurável, um fenômeno cuja importância para o desenvolvimento tem sido extensivamente documentada (Feldman, 2007). A interação segue um padrão de contingência com um atraso mínimo (em frações de segundo), característico de interações afetivas bem sucedidas e ajustadas.

**Iniciação Materna (T<sub>0</sub>):** A mãe produz o pico de expressividade multimodal (prosódia, sorriso de Duchenne, arousal, olhar direcionado). Este é o sinal de partida.

Figura 6 - Mãe em Análise do *FaceReader*

Fonte: Elaborada pela autora

**Recepção e Resposta do Bebê (T<sub>1</sub>):** Imediatamente após a oferta da mãe, a análise do bebê mostra sua resposta: a orientação da cabeça, seguida pelo pico de valência positiva, *arousal*, e a ativação das AUs 6 e 12.

Figura 7 - Bebê em Análise do *FaceReader*

Fonte: Elaborado pela própria autora

Essa sequência temporalmente conectada é a essência da regulação mútua (Tronick, 1989). A mãe, ao modular sua expressão, regula ativamente o estado interno da bebê. A bebê, ao responder de forma congruente, não só demonstra que foi regulada, mas também confirma o sucesso da comunicação para a mãe, regulando o comportamento materno e mantendo o circuito interativo em funcionamento.

A validação do sucesso interacional materializa-se na resposta do bebê, que confirma o fechamento do circuito comunicativo. A vocalização do bebê e a resposta materna "É, nasceu!", representa um turno em uma interação comunicativa finamente orquestrada. Essa contingência temporal, com tempo de resposta extremamente curto e quase imediato, exemplifica a "sincronia diádica" (Feldman, 2007), um processo rítmico de regulação mútua que organiza a interação e é crucial para o desenvolvimento de habilidades socioemocionais.

A emissão do enunciado responsivo da bebê atesta três competências cruciais: o alinhamento atencional, que confirma que seu foco perceptual e cognitivo estava alinhado ao da mãe; a compreensão pragmática, que evidencia a compreensão da intenção comunicativa da mãe de celebrar um evento narrativo; e, finalmente, a agência comunicativa, que posiciona o bebê não como receptor passivo, mas como um interlocutor ativo, que assume seu turno e co-constrói o momento interativo.

#### **4.1.4. Conclusão**

A análise convergente dos dados qualitativos, instrumentais (*FaceReader*) e acústicos (PRAAT) oferece um quadro relevante de uma interação bem-sucedida. Fica demonstrado que a mãe age com uma clara intencionalidade comunicativa, utilizando sua expressividade como uma ferramenta de andamento (*scaffolding*) para guiar a bebê a um estado afetivo compartilhado.

A contingência temporal entre os picos afetivos da mãe e da bebê é a evidência da sincronia diádica. Não são dois indivíduos agindo em paralelo, mas um sistema único "dançando" no mesmo ritmo emocional. A resposta da bebê é comportamentalmente visível (orientação da cabeça), anatomicamente precisa (ativação das AUs 6 e 12, formando um sorriso de Duchenne) e emocionalmente quantificável (picos de valência positiva e ativação do organismo). A mãe "convida" a bebê para um estado emocional, e a bebê "aceita" o convite, respondendo com todo o seu ser.

A análise vocal, por sua vez, revela as ferramentas que a mãe utiliza para orquestrar essa "dança", abandonando seu perfil de base para adotar um estilo comunicativo expressivo, com

pitch elevado, intensidade variável, elocução lenta e qualidades de voz que conferem delicadeza e ternura. A reação contingente à vocalização do bebê evidencia a natureza síncrona da interação, onde a expressividade de um participante amplifica a do outro.

Conforme apresentado anteriormente, a capacidade de se engajar nessas trocas sincrônicas e mutuamente reguladas aos seis meses é o alicerce fundamental sobre o qual as competências de atenção compartilhada são edificadas. A díade primeiro aprende a "sentir junto" para depois poder "agir junto". A análise integrada e quantitativa da díade demonstra que, muito antes da emergência de atos enunciativos (compreensão e/ou produção linguística) pela bebê, a comunicação humana floresce no terreno fértil da sincronia afetiva, em um diálogo corporal e emocional ricamente orquestrado entre mãe e bebê.

#### 4.2. ANÁLISE DA INTERAÇÃO DIÁDICA AOS 12 MESES: CONSOLIDAÇÃO DA ATENÇÃO COMPARTILHADA

A transição dos seis para os doze meses de vida marca um ponto de inflexão no desenvolvimento sociocognitivo da criança. Nesta fase, a interação transcende a díade face a face e passa a incorporar sistematicamente um terceiro elemento, um objeto ou evento no mundo. Conforme aponta Tomasello (2003), a habilidade de participar de interações triádicas (mãe-bebê-objeto) é um divisor de águas, indicando que o bebê começa a compreender os outros como agentes com intenções e focos de atenção que podem ser compartilhados e influenciados.

Aos doze meses, a criança se torna cada vez mais uma agente ativa, capaz de direcionar sua própria atenção e explorar o mundo com maior autonomia. Esta seção analisa a interação diádica neste período de desenvolvimento, focando em um desafio comum e ecologicamente válido: a competição por atenção. A análise se debruça sobre uma cena em que a mãe tenta engajar a filha em uma atividade com um livro, enquanto a atenção da bebê é capturada por um estímulo sonoro concorrente no ambiente.

Este cenário oferece uma oportunidade para investigar as estratégias de andaimento (*scaffolding*) (Bruner, 1975) que a mãe emprega para recrutar e manter a atenção da filha. A análise se apoiará em uma abordagem multimodal, examinando como os canais visual, facial, fisiológico e vocal da bebê são orquestrados pela mãe em sua tentativa de estabelecer um quadro de atenção compartilhada (Tomasello, 2003). Em contrapartida, os dados da bebê serão analisados para compreender seu estado de atenção dividida e sua resposta à performance materna. A estrutura desta seção seguirá, portanto, uma lógica que vai da descrição do cenário

à análise detalhada da participante (bebê), culminando em uma conclusão que sintetiza a complexa dinâmica observada.

Esta seção se dedica à microanálise de uma interação neste período crucial. O objetivo é investigar se e como as estratégias comunicativas da mãe se adaptam a essa nova competência da criança e se e como a "dança" interativa (Stern, 1985) se reconfigura para acomodar a atenção compartilhada. A análise buscará detalhar os mecanismos de andaimeiro (*scaffolding*) (Bruner, 1975) e a sincronia diádica e triádica (Feldman, 2007) em um contexto em que o diálogo não é mais apenas a procura por uma conexão afetiva, mas também sobre um tópico externo, estruturado pela diáde.

#### 4.2.1. Descrição Detalhada do Vídeo da Mãe

A cena se passa em um ambiente interno, similar ao da análise anterior. A mãe está sentada na cadeira, e a bebê, agora com doze meses e maior autonomia motora, está sentada à sua frente, na “cadeirinha de papar”. Essa configuração permite um contato visual direto e constante entre as duas.

O foco da interação é um livro infantil. Diferentemente da narração contínua observada aos seis meses, a atividade agora se assemelha a um jogo de identificação de figuras, com turnos mais definidos. A mãe aponta para as imagens, nomeia-as e convida a bebê a participar ativamente, pedindo que ela localize os objetos.

O vídeo acompanha os seguintes momentos da interação:

1. **Nomeação e Som:** A mãe aponta para uma figura e nomeia: "Trator. Vamos ver o trator, ó. O trator...". Em seguida, ela produz um som onomatopéico associado ao objeto.
2. **Descrição da Cena:** Ela vira a página, aponta para uma nova imagem e descreve a cena para a bebê: "Olha a vaquinha... está comendo a graminha. O fazendeiro tá tirando o leite da vaca.".
3. **Convite à Ação:** A mãe busca ativamente a participação da filha, solicitando uma ação dêitica (apontar). Ela pergunta diretamente: “Mostra pra mamãe aqui, (nome da bebê). Mostra pra mamãe, cadê o trator aqui? ”
4. **Insistência e Chamado:** Diante da aparente ausência de resposta, a mãe insiste no convite, chamando o nome da bebê para recapturar sua atenção e direcioná-la para a tarefa solicitada: “ (nome da bebê), cadê o trator? Cadê o trator? ”

#### 4.2.2. Dados da Mãe: A Performance Comunicativa



A análise da performance materna aos 12 meses revela uma notável adaptação às novas competências sociocognitivas da criança. Se aos 6 meses sua performance se assemelhava à de uma "narradora afetiva", cujo principal objetivo era modular e sintonizar o estado da bebê, agora seu papel evolui para o de uma "interlocutora" ativa.

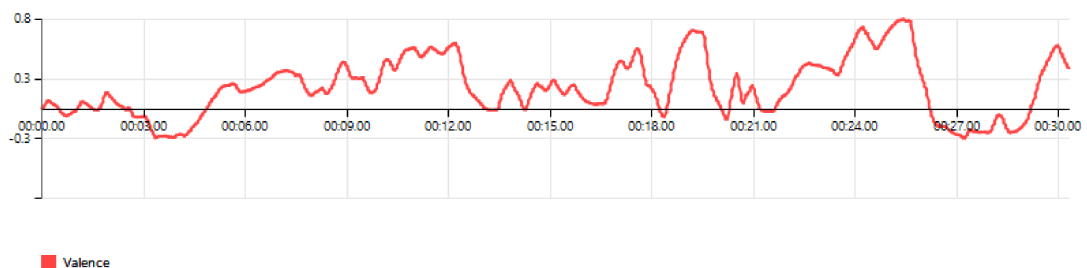
#### 4.2.2.1. *Análise da Interação a partir do FaceReader*

A performance da mãe aos 12 meses transcende a mera expressão afetiva e se caracteriza por uma clara intencionalidade comunicativa. Suas ações, como apontar para figuras específicas e fazer perguntas diretas ("Cadê o trator?"), são estratégias deliberadas que revelam seu objetivo de guiar a atenção da filha para um foco externo. Conforme aponta Tomasello (2003), é precisamente essa capacidade de ler e direcionar as intenções dos outros que fundamenta a atenção compartilhada. Portanto, ao analisar a performance da mãe, estamos observando as manifestações de sua intencionalidade em criar uma experiência triádica (mãe-bebê-livro)."

#### **Intensidade Afetiva (Valência e Ativação do Organismo)**

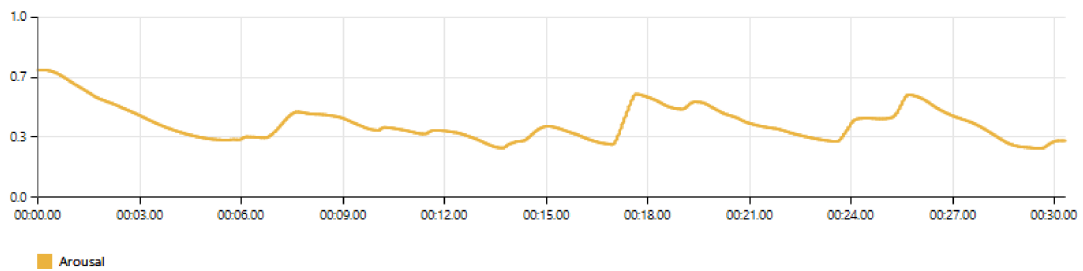
A análise da intensidade afetiva da mãe na interação de 12 meses, mensurada através dos eixos de Valência e Ativação do Organismo, oferece um panorama quantitativo de seu estado emocional durante a brincadeira. O gráfico de Valência revela um padrão de picos positivos recorrentes e de alta intensidade. Isso sugere uma estratégia de oferecer um ambiente afetivo consistentemente alegre. Corroborando essa observação, o gráfico de Ativação do Organismo exibe picos de ativação que parecem ocorrer em sincronia com os picos de valência positiva.

Gráfico 11 - Gráfico de Valência (Mãe) aos 12 meses



Fonte: Elaborada pela autora

A ocorrência de alta valência (prazer) e alto *arousal* (ativação do organismo) é a assinatura de estados emocionais como entusiasmo e alegria expressiva.

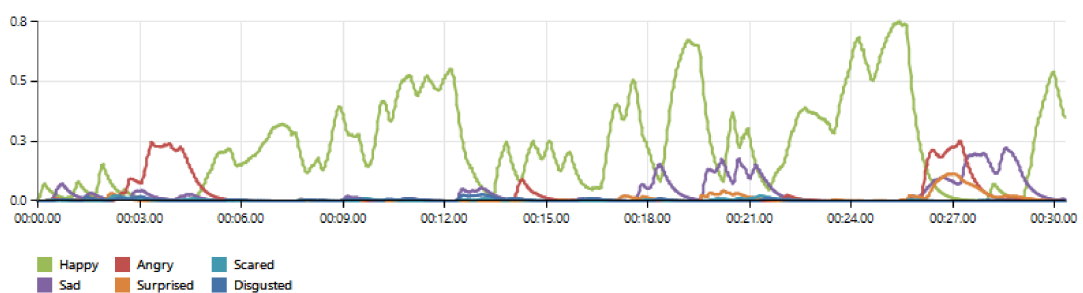
Gráfico 12 - Gráfico de *Arousal* ou Ativação do Organismo (Mãe) aos 12 meses

Fonte: Elaborado pela autora

### Modulação Emocional

A análise do gráfico de expressões emocionais da mãe permite decompor sua performance afetiva. O dado mais proeminente é o domínio da expressão de felicidade (*Happy*), que funciona como o tom emocional de base. Sobre essa base, observam-se picos de surpresa (*Surprised*), utilizados como ferramenta para gerar picos de interesse. Notavelmente, as emoções de valência negativa estão praticamente ausentes, indicando que a mãe constrói um espaço interativo de segurança e positividade.

Gráfico 13 - Gráfico de expressões emocionais da mãe



Fonte: Elaborado pela autora

### Engajamento Visual (Orientação da Cabeça e Olhar)

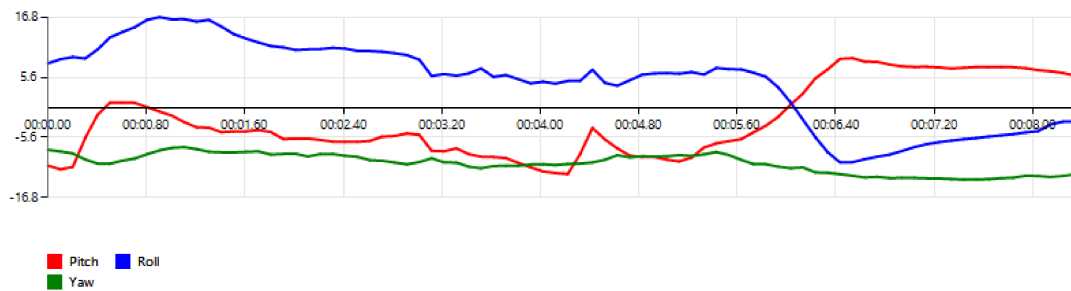
Os gráficos de orientação da cabeça e do ângulo do olhar demonstram o engajamento visual ativo da mãe. As flutuações constantes nas linhas indicam que ela não mantém uma postura estática, mas move dinamicamente sua cabeça e seu olhar. Este comportamento sugere uma "dança" visual, na qual ela alterna o foco entre a brincadeira e o rosto da bebê.

Essa performance visual é um componente central de sua estratégia multimodal, que integra diferentes canais (neste caso, o visual e o gestual) para construir a interação. Conforme



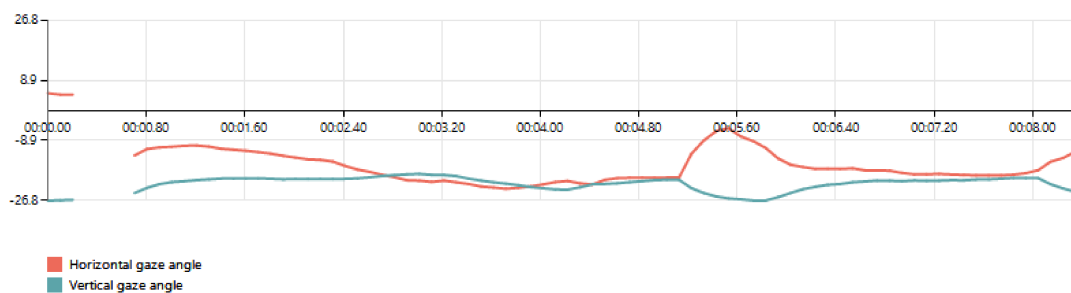
apontado por Madureira e Fontes (2023), as impressões e os significados em uma interação são formados a partir da integração desses múltiplos canais. Portanto, essa "dança" visual é uma estratégia fundamental e intencional para monitorar, direcionar e manter a atenção compartilhada com a criança.

Gráfico 14 - Gráfico de orientação de cabeça da mãe



Fonte: Elaborado pela própria autora

Gráfico 15 - Gráfico de orientação do olhar da mãe

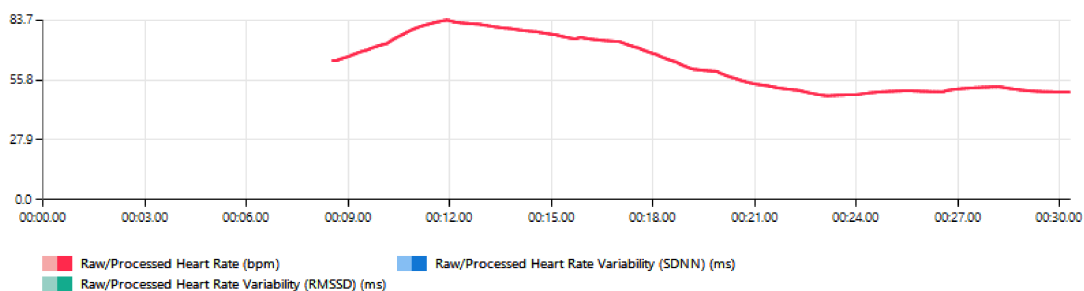


Fonte: Elaborado pela autora

### Engajamento Fisiológico (Frequência Cardíaca)

A análise da frequência cardíaca da mãe corrobora seu engajamento ativo. O gráfico mostra um padrão de estabilidade, mas com acelerações pontuais que coincidem com os momentos de maior intensidade da interação. Isso indica que sua performance comunicativa envolve uma mobilização de todo o seu organismo para se conectar com a bebê.

Gráfico 16 - Gráfico de Frequência Cardíaca da mãe



Fonte: Elaborado pela autora

#### 4.2.2.2. Análise Acústica e Prosódica a partir do PRAAT

Após a análise da performance visual e fisiológica da mãe na diáde, o foco se volta para o canal auditivo, especificamente para a expressão vocal da mãe. A voz materna é uma das mais poderosas ferramentas de andaime (*scaffolding*) e regulação afetiva nas interações precoces. As modulações na prosódia e na qualidade da voz não são meros acompanhamentos da fala, mas constituem uma estratégia complexa para capturar a atenção, transmitir emoção e estruturar o diálogo com a criança.

Esta seção se dedica, portanto, a uma análise detalhada dos parâmetros acústicos e prosódicos da fala da mãe, extraídos com o auxílio do software PRAAT. O objetivo é investigar objetivamente como ela esculpe a sonoridade de sua voz para se comunicar com uma bebê de 12 meses, que já demonstra maior compreensão e capacidade participativa.

Para compreender a magnitude das adaptações vocais da mãe, o perfil vocal de base previamente analisado serve como um ponto de referência relevante. Conforme análise anteriormente feita (ver seção 4.1.2), sua voz usual se caracteriza por uma frequência fundamental ( $F_0$ ) média de 135 Hz, com *pitch* baixo e variabilidade minimizada.

Este perfil de base se opõe diretamente aos achados da análise da interação aos 6 meses, na qual a performance vocal da mãe foi marcadamente expressiva. Naquela interação, seu *pitch* foi classificado como alto, com extensão e variabilidade maximizadas, e a Frequência Fundamental atingiu picos de até 538 Hz. A comparação entre os dois cenários evidencia, portanto, a adaptação intencional e significativa que a mãe realiza ao se dirigir à bebê.

A fala da mãe no segmento (clipe) de 12 meses demonstra uma adaptação estratégica às novas competências da bebê. A performance vocal é dinâmica e complexa, refletindo a transição de uma narração para um diálogo mais interativo.

Diferentemente da interação aos 6 meses, que era dominada por contornos exclamativos, a análise deste segmento revela uma maior variedade de modalidades. Foram

identificadas 2 ocorrências assertivas, 5 exclamativas e, crucialmente, 4 interrogativas (3 de questão total e 1 de questão parcial) (ver Anexo B). A introdução de perguntas diretas ("Cadê o trator?") é um marcador claro de que a mãe agora posiciona a bebê como uma parceira capaz de responder e participar ativamente da construção de sentido.

O *pitch*, correlato perceptivo da frequência fundamental ( $F_0$ ), é o parâmetro que sofre a modificação mais drástica na fala dirigida à criança. A análise perceptiva classifica sua média como alta, com extensão e variabilidade maximizadas. A sensação perceptiva de *pitch* alto corresponde, em nível acústico, a valores aumentados de frequência fundamental e, em nível de produção, ao índice de vibração das pregas vocais e à tensão nos músculos da laringe. Acusticamente, os dados confirmam essa percepção, mostrando que a  $F_0$  da mãe na interação de 12 meses atinge um valor máximo de 597 Hz e um mínimo de 147 Hz.

A ocorrência de um *pitch* tão agudizado pode ser interpretada sob a perspectiva do simbolismo sonoro (Jakobson & Waugh, 1979; Fónagy 1983; Tsur 1992; Hinton, Nichols, Ohala, 1994; Ohala, 1994, 1995; Nobile, 2019, entre outros). Considerando o papel do som na atribuição de sentidos, que se dá pelos mapeamentos não arbitrários entre as propriedades fonéticas e as perceptivas de seus referentes (Svantesson, 2017; Knoeferle, 2017), podemos analisar essa ocorrência no âmbito do Código de Frequência (Ohala, 1984).

O Código de Frequência, que se apoia na observação de vocalizações de animais, argumenta haver correlações entre vocalizações graves, tamanho grande e comportamento ameaçador, e entre vocalizações agudas, tamanho pequeno, fragilidade e delicadeza. Para interagir com bebês, os adultos produzem alterações no modo de produção da fala que ocasionam, no plano acústico, a elevação da frequência fundamental e, no plano perceptivo, o *pitch* agudizado (Ishi, Ishiguro e Hagita, 2009). Ao modular sua voz dessa forma, a mãe projeta uma persona vocal percebida como menos impositiva e mais delicada, ideal para a interação afetiva.

Essa potencialidade expressiva das características dos sons da fala tem bases intersubjetivas (Tsur, 1992). A função expressiva da fala decorre das características indexicais (que permitem atribuir aos falantes estados afetivos e características físicas) e simbólicas (que permitem expressar diferentes efeitos de sentido com a mesma sequência fônica) das propriedades da materialidade sonora (Madureira, 2000; Abelin, 1999). Finalmente, a alta variabilidade de *pitch* observada nos dados confere grande vivacidade à fala da mãe, opondo-se a uma fala monótona e funcionando como um potente atrator de atenção.

A intensidade da fala materna mantém o padrão de média e extensão regulares, mas com alta variabilidade, variando entre 54.6 dB e 85.3 dB (ver Anexo B). Essa variação dinâmica de volume continua a ser uma ferramenta para destacar informações e manter a atenção da bebê, que agora explora ativamente o ambiente.

Uma mudança notável ocorre na taxa de elocução, que passa de "lenta" (aos 6 meses) para regular (ver Anexo B). Isso sugere que a mãe ajusta sua fala à capacidade de processamento mais rápida da bebê de 12 meses. Apesar da taxa de fala mais rápida, a proporção de pausas aumenta para 31.46% do tempo total do segmento (9.38s de pausa em 29.81s). Esse aumento sugere que, embora fale mais rápido, a mãe cria pausas ainda mais longas e significativas, esperando ativamente a contribuição da filha na interação. O ajuste de qualidade de voz de laringe levantada diminui o tamanho do trato vocal, reforçando frequências altas que, perceptivamente são avaliadas como agudas.

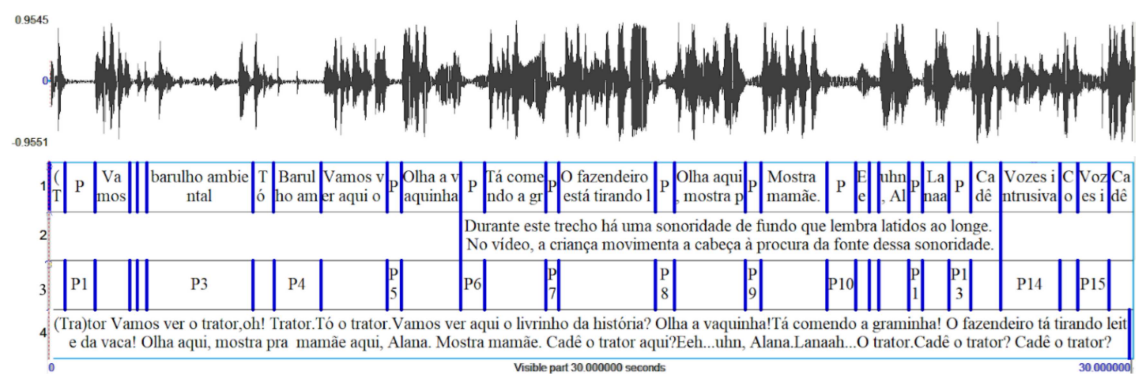
Vozes agudas são consideradas delicadas enquanto vozes graves impressionam por força e falta de suavidade.

O efeito de diminuição do trato vocal também ocorre em ajustes de qualidade de voz de lábios estirados. Em Madureira e Fontes (2023) foram encontradas associações entre ajustes de laringe levantada, lábios estirados, elevação de *pitch* com expressões de emoções faciais de alegria. O ajuste de Lábios Estirados pode corresponder ao sorriso.

A classificação dos ajustes de qualidade de voz é feita de acordo com os rótulos do sistema VPA (*Voice Profile Analysis*) desenvolvido por Laver e Mackenzie Beck, 2007 e fundamentado teoricamente em Laver (1980). Os graus no VPA variam de 1 (mais fraco) a 6 (mais forte).

O ajuste de qualidade de voz “*Whispery*” combina escape de ar com vibração das pregas vocais. O escape de ar entre as pregas vocais suaviza a emissão de fala e é utilizado para expressar, entre outros efeitos de sentido, gentileza e ternura (Ishi, Ishiguro e Hagita, 2009).

Figura 8 - Anotação no PRAAT da mãe aos 12 meses



Fonte: Elaborado pela autora

Nota: De cima para baixo na Figura, forma da onda, camada de segmentação dos enunciados com transcrição ortográfica, camada com comentário sobre presença de sonoridade à distância, camada destacando as pausas e camada com transcrição ortográfica do trecho de fala.

Estratégias Interativas e Gatilhos de Atenção

As estratégias da mãe aos 12 meses são marcadamente mais interativas e multimodais. A performance materna, que integra os canais vocal, gestual e facial, reflete uma complexa orquestração de sinais para construir sentido com a criança, como apontado nos estudos sobre impressões multimodais de Madureira e Fontes (2023). Além das repetições e do uso de onomatopeias (som do trator) para engajar a criança, a principal estratégia observada é o uso de perguntas diretas para solicitar a participação da bebê ("Mostra pra mamãe... Cadê o trator?").

Este tipo de questionamento funciona como uma forma mais sofisticada de andaimeito (*scaffolding*), conforme o conceito de Bruner (1975). A mãe não está mais apenas sustentando a atenção da filha, mas agora a orienta para que ela realize uma ação específica no mundo. Isso transforma a atividade de leitura em um jogo colaborativo e em um verdadeiro exercício de atenção compartilhada. Ao perguntar "Cadê o trator?", a mãe não apenas apresenta o mundo, mas convida a filha a agir sobre ele, verificando e guiando seu foco atencional dentro de uma cena triádica (mãe-bebê-livro), o que Tomasello (2003) descreve como um comportamento fundamental para a compreensão de intenções.

4.2.3. Dados da Bebê: A Resposta em um Contexto de Atenção Dividida

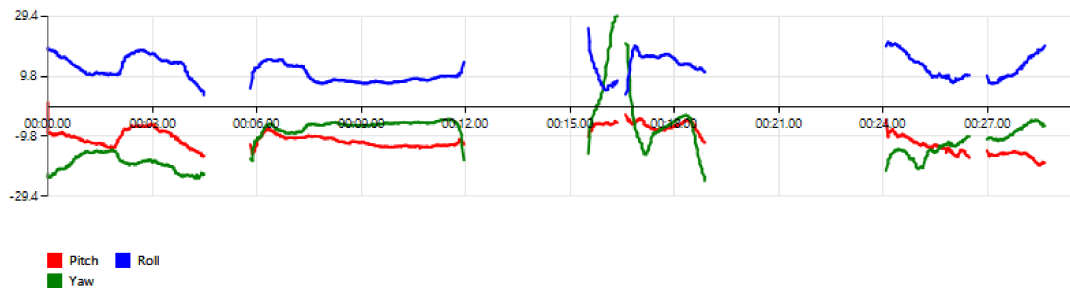
#### 4.2.3.1. *Análise da Interação a partir do FaceReader*

A análise dos dados da bebê de 12 meses revela um comportamento complexo, marcado não pela sincronia com a mãe, mas pela sua capacidade de direcionar e sustentar a atenção em um estímulo de seu próprio interesse, evidenciando sua crescente autonomia como agente no mundo.

#### **Engajamento Visual (Orientação da Cabeça e Olhar)**

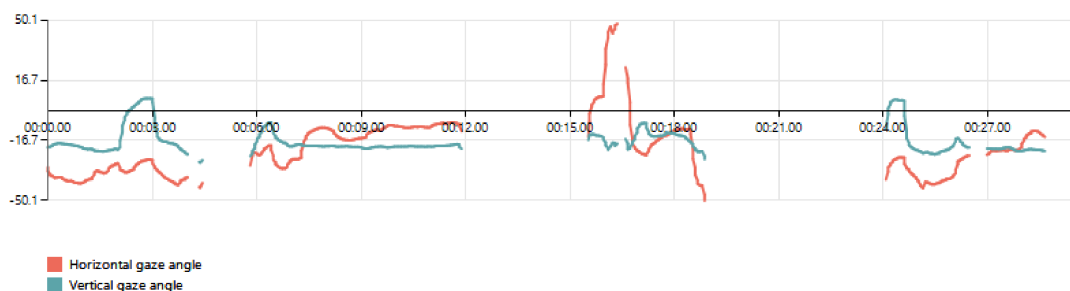
Os gráficos de orientação da cabeça e do ângulo do olhar são as evidências mais claras do foco de atenção da bebê. Eles quantificam a observação qualitativa de que a criança vira a cabeça para o lado. Os dados mostram movimentos de rotação lateral da cabeça que são sustentados por vários segundos, com o ângulo do olhar acompanhando essa nova direção. Este comportamento sugere uma busca ativa pela fonte de uma sonoridade externa que competia com a fala da mãe. A análise do engajamento visual, neste momento, não mostra o sucesso da díade em estabelecer um foco comum no livro, mas sim a capacidade da bebê de exercer sua autonomia e selecionar um foco de atenção independente, um aspecto fundamental na construção das interações (Cavalcante, 1994).

Gráfico 17 - Gráfico de orientação da cabeça do bebê



Fontes: Elaborado pela autora

Gráfico 18 - Gráfico de ângulo do olhar

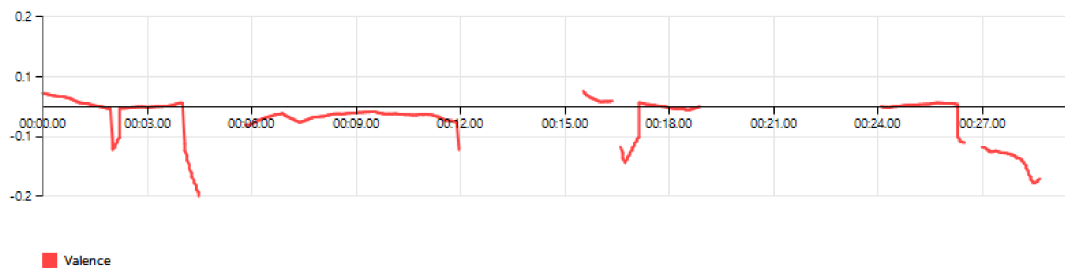


Fonte: Elaborado pela autora

### Intensidade e Expressão afetiva (Valência, Ativação do Organismo e Emoções)

A análise afetiva da bebê corrobora seu estado de atenção dividida. O gráfico de Valência permanece em um nível predominantemente neutro ou ligeiramente negativo, sem os picos positivos de prazer que marcariam uma sintonização com a brincadeira proposta pela mãe.

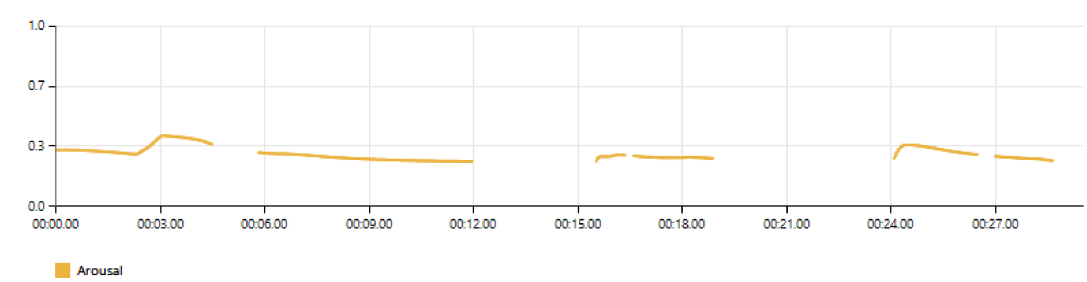
Gráfico 19 - Gráfico de Valência da bebê



Fonte: Elaborado pela autora

Da mesma forma, o *Arousal* (Ativação do Organismo) se mantém em um nível baixo e estável, indicando um estado de calma e observação, mas sem a excitação que caracterizaria um engajamento lúdico.

Gráfico 20 - Gráfico de *Arousal* (Ativação do Organismo) da bebê



Fonte: Elaborado pela autora

O gráfico de linha das Expressões confirma isso, com a ausência de picos significativos da expressão de felicidade (*Happy*). Em conjunto, esses dados sugerem que, embora a bebê não esteja em um estado de desconforto, seu foco afetivo não está conectado com a "oferta" entusiástica da mãe durante todo o tempo.

Gráfico 21 - Gráfico de Expressões da bebê



Fonte: Elaborado pela autora (2025)

A análise da linha do tempo das Unidades de Ação (AUs) da bebê oferece uma visão objetiva de sua expressão facial, momento a momento. A análise das expressões faciais da bebê



corroborar a observação de que a felicidade não é a emoção predominante nesta interação. Conforme aponta a literatura do desenvolvimento, em situações de engajamento ativo e direcionado por tarefas, como as perguntas feitas pela mãe, é comum que expressões de atenção e seriedade se sobreponham às de alegria. Este padrão pode ser interpretado não como uma ausência de prazer, mas como um indicador do esforço cognitivo da criança, que está concentrada em processar a solicitação da mãe e em participar da troca comunicativa.

A análise revela a ausência do Sorriso: A AU12 (*Lip Corner Puller*), que é o músculo fundamental para o ato de sorrir (puxar os cantos da boca), apresenta atividade nula ou insignificante durante todo o clipe. Sem a AU12, não é possível configurar uma expressão de alegria ou felicidade.

As AUs que de fato mostram ativação são as da parte superior do rosto, principalmente as AU5 (*Upper Lid Raiser*) ou o levantamento das pálpebras superiores e as AU1 e AU2 (*Inner e Outer Brow Raiser*) que seria o levantamento das sobrancelhas.

Essa combinação de movimentos é a assinatura clássica de um estado de surpresa, interesse ou atenção focada. A bebê está com os olhos bem abertos, uma expressão de quem está tentando captar e processar um estímulo saliente.

Observa-se também a ativação da AU25 (*Lips Part*), indicando que a boca da bebê permanece entreaberta, um sinal que frequentemente acompanha estados de concentração.

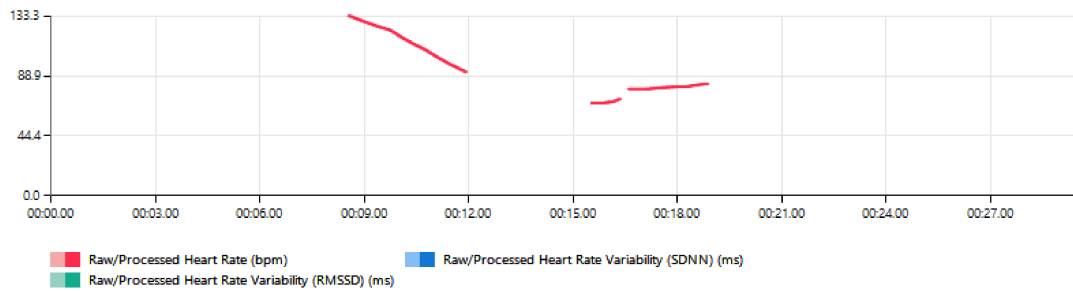
As AUs ativas (olhos e sobrancelhas levantadas, boca entreaberta) e a ausência da AU12 (o sorriso) pintam um quadro claro. A expressão dominante da bebê é intensa concentração e talvez curiosidade ou surpresa.

Esta evidência corrobora e fortalece imensamente a análise de que a bebê estava com a atenção dividida, ora o foco é no estímulo sonoro, ora na mãe/na cena com a mãe. Seu rosto não expressa a alegria da brincadeira com a mãe, mas sim o esforço cognitivo de tentar localizar e entender o som que ouvia à distância.

### **A Dinâmica Fisiológica da Atenção Dividida**

A análise dos dados fisiológicos da bebê oferece uma janela para a dinâmica de sua atenção dividida, revelando a narrativa de seus estados internos momento a momento.

Gráfico 22 - Gráfico de frequência cardíaca

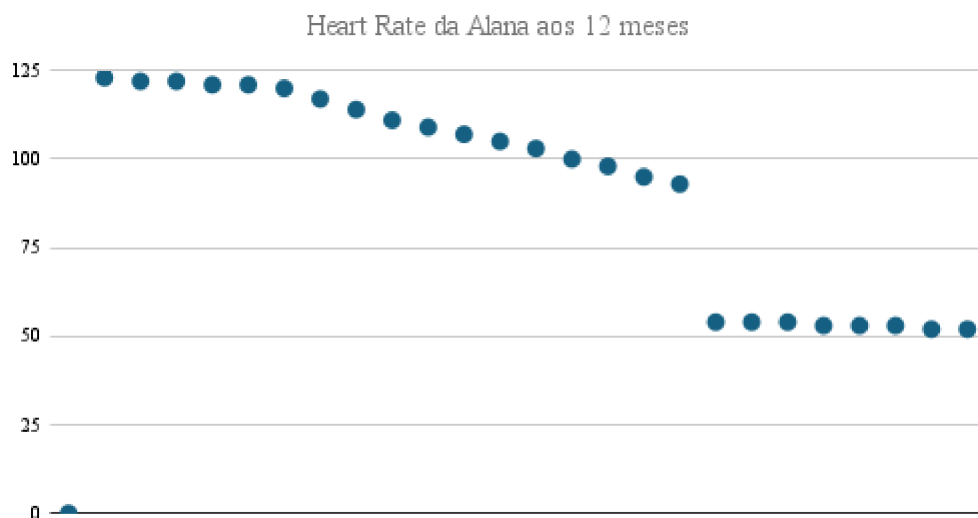


Fonte: Elaborado pela autora

Inicialmente, enquanto seu comportamento visual, observado nos gráficos de orientação da cabeça, indica um foco claro no som externo, seu ritmo cardíaco se encontra em um patamar elevado, porém com um padrão de desaceleração gradual. Este padrão fisiológico é um correlato clássico de atenção sustentada. Isso sugere que a bebê não está simplesmente distraída, mas ativamente engajada no processamento do estímulo sonoro concorrente, dedicando recursos cognitivos e fisiológicos a essa tarefa.

É neste contexto de foco externo que a intervenção da mãe se torna crucial. Ao perceber o desvio de atenção, a mãe intensifica suas "ofertas" comunicativas, chamando o nome da filha com uma prosódia marcada (o tratoorr!!!!) e repetindo as perguntas. O gráfico revela o sucesso dessa estratégia: por volta dos 9 segundos, ocorre uma aceleração cardíaca súbita e acentuada.

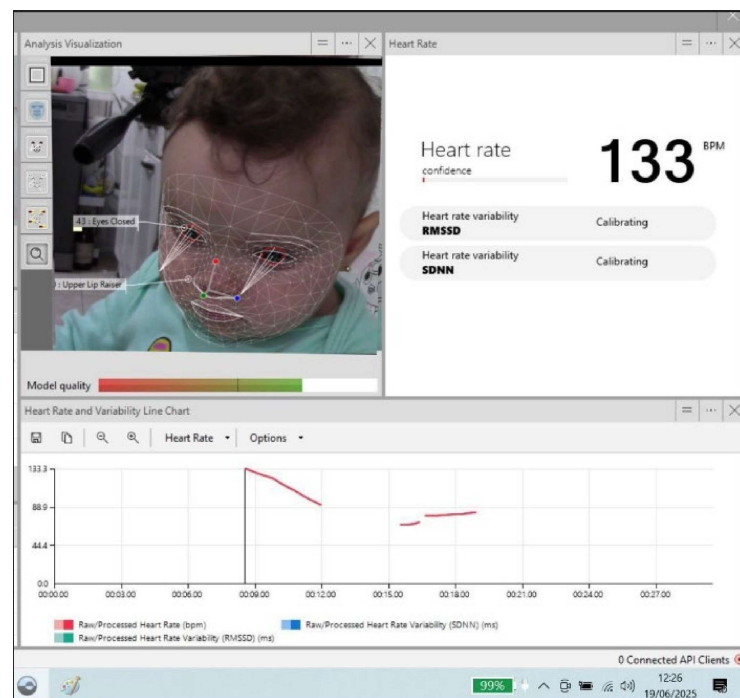
Gráfico 23 - Gráfico de frequência cardíaca da bebê



Fonte: Elaborado pela autora

Este pico de aceleração é o marcador fisiológico de uma "resposta de orientação" (*orienting response*), o momento exato em que a atenção da bebê é finalmente recapturada pela mãe e realocada de volta para a interação diádica. Este evento fisiológico ocorre em sincronia com o movimento de sua cabeça, que, como visto na análise visual, se vira de volta para a mãe neste instante.

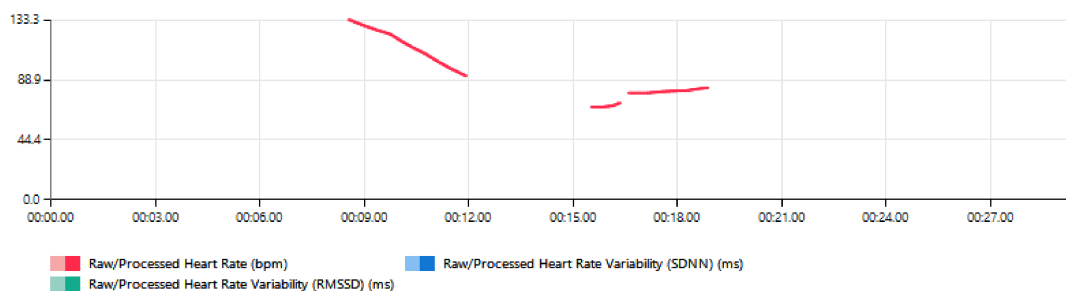
Gráfico 24 - Gráfico de Batimentos Cardíacos da bebê e o momento em que ela volta sua atenção para mãe



Fonte: Elaborado pela autora

Portanto, a análise da frequência cardíaca não mostra um engajamento passivo ou ausente, mas sim uma narrativa fisiológica da competição pela atenção. Ela quantifica o estado de atenção inicial da bebê, o momento preciso da virada atencional provocada pela mãe e, finalmente, a retomada da sincronia diádica, demonstrando a eficácia das estratégias maternas em regular não apenas o comportamento, mas também o estado biológico da filha.

Gráfico 25 - Gráfico de BPM da bebê



Fonte: Elaborado pela autora

#### 4.2.4. Conclusão

Na interação aos 6 meses, o que emerge é uma sintonização afetiva imediata e bem-sucedida. A mãe, em um papel de narradora, orchestra uma performance expressiva, modulando sua prosódia vocal e facial para guiar a experiência emocional da bebê. A resposta da criança, evidenciada por picos de valência positiva, ativação fisiológica e a produção de sorrisos genuínos, confirma o sucesso dessa sintonização. Nesta fase, a interação é um sistema de regulação mútua focado na conexão afetiva, onde o principal objetivo comunicativo é "sentir junto".

A análise integrada da interação aos 12 meses revela uma dinâmica diádica substancialmente mais complexa do que a observada aos 6 meses. O que emerge não é uma sintonia afetiva imediata, mas um retrato realista da comunicação nesta fase: uma negociação contínua pelo foco de atenção, na qual a mãe e a bebê atuam como agentes com focos por vezes divergentes.

De um lado, a performance da mãe pode ser compreendida como um "pacote" comunicativo rico e persistente. Ela mobiliza um repertório multimodal sofisticado alterando suas prosódias para incluir perguntas diretas, intensificando sua expressividade facial positiva e mantendo um engajamento visual e fisiológico ativo com a clara intenção de estabelecer um quadro de atenção compartilhada em torno do livro (Tomasello, 2003). Suas estratégias de andaimento (Bruner, 1975) adaptam-se para solicitar a participação de uma criança que ela já percebe como uma parceira mais competente e obtém sucesso.

Do outro lado, a análise dos dados da bebê sugere que ela não é uma receptora passiva desta oferta, mas uma participante com sua própria autonomia. Seus comportamentos visuais e seu estado fisiológico indicam um engajamento ativo e sustentado, porém direcionado, a princípio, a outro estímulo. A ausência de uma resposta afetiva positiva e contingente à mãe

não deve ser interpretada como uma falha na comunicação, mas como uma evidência de sua capacidade de selecionar e explorar seu ambiente de forma independente.

É neste momento que analisar a interação como um "continuum" se mostra fundamental. A aparente falta de sincronia temporal não invalida a conexão da díade. Pelo contrário, ela expõe as etapas do processo de reconexão. A evidência mais forte disso reside na análise fisiológica: a aceleração súbita na frequência cardíaca da bebê, que ocorre após as insistentes chamadas da mãe, pode ser interpretada como o primeiro passo para essa reconexão. Este alerta fisiológico, ou "resposta de orientação", precede qualquer mudança afetiva.

Portanto, a interação aos 12 meses ilustra que a sincronia se reconfigura. Ela deixa de ser primariamente uma ressonância afetiva imediata para se tornar um processo mais elaborado de negociação, que envolve "desencontros" e tentativas de reparo. O que se observa é o alicerce de uma forma de diálogo mais madura, onde o sucesso não está apenas em "sentir junto", mas em aprender a navegar e a negociar um mundo de interesses compartilhados.

#### 4.3. Discussão Final

A análise comparativa das interações mãe-bebê aos 6 e 12 meses, realizada através de uma abordagem multimodal que integra dados qualitativos, comportamentais, fisiológicos e acústicos, oferece um panorama detalhado da evolução da comunicação diádica no primeiro ano de vida. Os resultados sugerem uma transição fundamental: a passagem de uma interação baseada em uma parceria de interlocução afetiva para uma dinâmica mais complexa, centrada na negociação da atenção compartilhada.

Na interação aos 6 meses, o que emerge não é apenas uma "dança" de sintonização afetiva, mas a fundação de uma parceria de interlocução. A mãe, em um papel de narradora, orquestra uma performance expressiva para guiar a experiência emocional da bebê. O momento crucial desta interação ocorre quando a criança, em resposta a um clímax na fala materna, produz uma vocalização. A mãe, por sua vez, valida essa emissão como um turno comunicativo legítimo, respondendo com "É, nasceu!". Este evento é a evidência de que a interação já opera como um diálogo. A resposta da criança, que vai além do sorriso e se manifesta no canal vocal, é tratada pela mãe como uma contribuição significativa. A análise multimodal corrobora a profundidade deste momento: a vocalização é acompanhada por picos de valência positiva, ativação fisiológica (aceleração cardíaca) e a produção de um sorriso genuíno na bebê. Portanto, nesta fase, a interação já é um sistema de regulação mútua que não apenas busca o "sentir junto",

mas que já estabelece as bases de um diálogo co-construído, que irá evoluir para formas mais complexas de negociação aos 12 meses.

Aos 12 meses, o cenário se reconfigura drasticamente. A emergência da capacidade da criança de focar em um estímulo externo introduz um novo desafio: a gestão da atenção. A mãe adapta-se, transitando do papel de narradora para o de interlocutora. Suas estratégias de andaimento (*scaffolding*) (Bruner, 1975) tornam-se mais sofisticadas, incorporando perguntas diretas ("Cadê o trator?") e uma taxa de elocução mais rápida, reconhecendo a maior competência da filha. Sua performance continua a ser intensamente multimodal, combinando o gesto de apontar com a fala para direcionar o foco, uma característica da matriz gesto-fala (McNeill, 2005; Cavalcante, 2014).

Contudo, a resposta da bebê aos 12 meses evidencia sua crescente autonomia. A análise de seus dados revela um "desencontro" atencional, no qual seu foco visual, afetivo e fisiológico se volta para um estímulo sonoro concorrente. Este momento, no entanto, não representa uma falha comunicativa, mas a própria essência da interação nesta fase. A contribuição mais significativa desta análise reside na documentação do processo de reparo desse desencontro. A persistência da mãe culmina em uma "resposta de orientação" fisiológica na bebê (a aceleração cardíaca), o primeiro passo para a retomada da atenção. Este achado apoia a visão da interação como um "contínuum" (Cavalcante, 2009), onde a conexão não é constante, mas um processo dinâmico de rupturas e reparos.

Pode-se argumentar que a performance materna, em ambas as fases, transcende a função de apenas manter o engajamento afetivo, operando também como um mecanismo pedagógico fundamental para a aquisição da linguagem. A mãe não apresenta a língua à bebê como uma série de palavras abstratas, mas como um "pacote" multimodal coeso. Nesse contexto, a expressividade prosódica da mãe — a ampla variabilidade de *pitch*, os contornos melódicos exagerados e as pausas bem marcadas — funciona como primeiras "aulas" sobre a estrutura sonora da língua, um fenômeno conhecido como *bootstrap prosódico*. A manipulação dos "envelopes prosódicos" que Stern (1985) descreve como cruciais para a sintonização afetiva serve, também, como um guia para a segmentação do fluxo contínuo da fala.

A análise comparativa demonstra que, em apenas seis meses, a natureza da comunicação diádica evolui profundamente. A díade transita de um modelo de diálogo baseado na partilha de estados afetivos para um modelo focado na partilha de focos de atenção sobre o mundo, como postulado por Tomasello (2003). A "dança" da sincronia não desaparece, mas se torna mais complexa, incorporando a negociação, o convite à ação e a co-construção de significado

sobre objetos externos. Este estudo, ao integrar múltiplas modalidades de análise (Madureira & Fontes, 2023), oferece uma versão dessa trajetória, ilustrando como as fundações do diálogo e da intencionalidade compartilhada são forjadas nos desafios e sucessos das interações cotidianas.

**Quadro 2 – Síntese Comparativa da Evolução da Comunicação Diádica**

<b>Parâmetro de Análise</b>	<b>Interação aos 6 Meses</b>	<b>Interação aos 12 Meses</b>
<b>Papel Materno</b>	"Narradora Afetiva": Foco na regulação do estado emocional da bebê.	"Interlocutora Ativa": Foco na solicitação de uma ação e na negociação da atenção.
<b>Natureza da Interação</b>	Diádica (foco na relação). Sincronia afetiva imediata.	Triádica (foco em um objeto externo). Negociação da atenção com rupturas e reparos.
<b>Objetivo Comunicativo</b>	"Sentir Junto": Partilha de estados afetivos.	"Agir Junto": Partilha de foco de atenção e intenções sobre o mundo.
<b>Prosódia Vocal Materna</b>	Predomínio de contornos exclamativos. Ritmo lento e pausado (26% de pausas). F <sub>0</sub> máx: 538 Hz.	Introdução de contornos interrogativos. Ritmo mais rápido com pausas mais longas (31%). F <sub>0</sub> máx: 597 Hz.
<b>Resposta da Bebê</b>	Contingente e síncrona: Picos de valência positiva, sorriso de Duchenne (AU6+AU12), vocalização validada como turno.	Autônoma e assíncrona: Foco inicial em estímulo concorrente. Valência neutra. Expressão de concentração (AU1, AU2, AU5).
<b>Achado Fisiológico Chave</b>	Sincronia de ativação (Arousal) e aceleração cardíaca em resposta à performance materna.	"Resposta de orientação": Aceleração cardíaca súbita como marcador do sucesso do reparo atencional pela mãe.

Fonte: Elaborado pela autora (2025).



## 5. Considerações Finais, Limitações e Perspectivas Futuras

A presente dissertação, ao realizar uma análise multimodal detalhada de interações naturalísticas, contribui para a compreensão da aquisição da linguagem no contexto do português brasileiro. O principal achado deste estudo é a documentação da trajetória desenvolvimental da comunicação diádica, que transita de um foco na conexão afetiva para a complexa negociação de focos de atenção sobre o mundo. Demonstrou-se como a mãe adapta suas estratégias multimodais e como a bebê evolui de uma receptora sintonizada para uma agente com foco de atenção independente.

Como estudo de caso único, os resultados aqui apresentados possuem limitações no que tange à generalização. A dinâmica observada reflete as particularidades desta díade específica. No entanto, a profundidade da análise microanalítica e multimodal oferece um modelo metodológico robusto que pode ser replicado em estudos com amostras maiores.

A observação de que mãe e bebê apresentaram elevação conjunta na frequência cardíaca no momento em que o contato visual foi estabelecido oferece uma contribuição significativa à compreensão da comunicação afetiva precoce. Esse tipo de resposta fisiológica coordenada é interpretado pela literatura como um indicativo de sincronia biocomportamental, fenômeno que reflete o alinhamento entre os estados emocionais e atencionais de dois indivíduos em interação (Feldman, 2007; Wass et al., 2020).

Estudos em neurociência do desenvolvimento indicam que sincronias fisiológicas, como a regulação conjunta do ritmo cardíaco, ocorrem frequentemente em contextos de atenção compartilhada e são mediadas por interações olho-no-olho, vocalizações responsivas e expressões afetivas congruentes (Feldman, 2012; Goldstein et al., 2017). Essas respostas simultâneas refletem a ativação de mecanismos de regulação emocional recíproca, fundamentais para o desenvolvimento comportamental e afetivo da bebê.

A elevação simultânea da frequência cardíaca, identificada por meio do FaceReader, pode estar relacionada à liberação de hormônios como a oxitocina, que tem sido associada à promoção de comportamentos pró-sociais e ao fortalecimento do vínculo mãe-bebê (Bornstein & Esposito, 2020). Além disso, esse tipo de resposta emocional conjunta pode indicar estados compartilhados de excitação positiva, que favorecem a aprendizagem social e linguística.

Adicionalmente, o contato visual é um potente organizador da comunicação social desde os primeiros meses de vida. Farroni et al. (2002) demonstraram que recém-nascidos já são capazes de detectar o olhar direto e que essa habilidade é acompanhada por mudanças



atencionais e autonômicas, possivelmente relacionadas à antecipação de interações significativas.

Assim, o aumento conjunto da frequência cardíaca identificado neste estudo pode ser interpretado como um marcador fisiológico de acoplamento emocional e engajamento comunicativo, contribuindo para a construção de uma base relacional na qual se apoia a aquisição da linguagem. Esse fenômeno integra múltiplas modalidades — visual, vocal e fisiológica — reforçando a hipótese de que a comunicação no primeiro ano de vida é profundamente multimodal e incorporada (Guellaï et al., 2014; Esteve-Gibert & Guellaï, 2018).

As perspectivas futuras que emergem deste trabalho são promissoras. Uma continuação natural seria a análise longitudinal desta mesma díade em idades mais avançadas (18, 24, 36 meses), investigando como os padrões de interação aqui observados se relacionam com o desenvolvimento lexical e sintático da criança. Além disso, a aplicação desta metodologia multimodal em estudos comparativos, envolvendo díades com diferentes perfis de desenvolvimento (típico e atípico), poderia levar à identificação de marcadores interacionais precoces para o diagnóstico de transtornos de linguagem. Por fim, a integração desta abordagem com métodos de neurociência social, como o hiperscaneamento, poderia futuramente investigar os correlatos neurais da sincronia diádica, conectando o comportamento observável à atividade cerebral subjacente.

## REFERÊNCIAS

- ABELIN, Å. **Studies in sound symbolism**. 1999. Tese (Doutorado) – Göteborg University, Gotemburgo, 1999.
- ÁVILA-NÓBREGA, P. V. Reflexões sobre intencionalidade e atencionalidade como uma contribuição para a aquisição da linguagem. In: DIEDRICH, M. S.; OLIVEIRA, G. F.; DEL RÉ, A., orgs. **Língua, discurso e suas relações na aquisição da linguagem**. [S.l.]: Pedro & João Editores, 2023.
- BARBOSA, P. A. **Prosódia**. São Paulo: Parábola Editorial, 2012.
- BARBOSA, P. A.; MADUREIRA, S. Experimentação em fonética acústica: prosódia. In: BARBOSA, P. A.; MADUREIRA, S., orgs. **Manual de fonética acústica experimental: aplicações a dados do português**. São Paulo: Cortez, 2015. p. 196–231.
- BEEBE, Beatrice; LACHMANN, Frank M. Infant research and adult treatment: Co-constructing interactions. Hillsdale: **The Analytic Press**, 2002.
- BOERSMA, P.; WEENINK, D. **Praat: doing phonetics by computer** [programa de computador]. Versão 6.1.39, 2021. Disponível em: <http://www.praat.org/>. Acesso em: 23 fev. 2025.
- BORNSTEIN, M. H.; ESPOSITO, G. Biobehavioral synchrony and the development of parent–infant relationships. **Current Opinion in Psychology**, v. 36, p. 49–54, 2020.
- BRISTOW, D. et al. Hearing faces: how the infant brain matches the face it sees with the speech it hears. **Journal of Cognitive Neuroscience**, v. 21, n. 5, p. 905–921, 2009.
- BRUNER, Jerome S. The ontogenesis of speech acts. **Journal of Child Language**, v. 2, n. 1, p. 1-19, 1975.
- CALIL, E.; DEL RÉ, A. Análise multimodal de uma história inventada: o caso da onomatopeia visual. **Revista da ANPOLL**, Belo Horizonte, n. 27, p. 30–45, 2009.
- CAVALCANTE, M. C. B. Rotinas interativas mãe-bebê: constituindo gêneros do discurso. **Revista Investigações**, v. 21, n. 2, p. 153–169, 2009.
- \_\_\_\_\_. Gestos: da marginalidade à proeminência na aquisição da linguagem. In: ENCONTRO NACIONAL DA ANPOLL (ENANPOLL), 35., 2020, Londrina. **Anais do XXXV ENANPOLL**. [S.l.: s.n.], 2020.
- CAVALCANTE, M. C. B.; BARROS, A. T. M. Manhês: qualidade vocal e deslocamentos na dialogia mãe-bebê. **Veredas – Revista de Estudos Linguísticos**, v. especial, p. 25–39, 2022.
- COOTES, T. F.; WALKER, K.; TAYLOR, C. View-based active appearance models. In: **IEEE International Conference on Automatic Face and Gesture Recognition**. Washington, DC: IEEE Computer Society, 2000. p. 227–232.
- CRYSTAL, D. **The Cambridge encyclopedia of the English language**. 2. ed. Cambridge: Cambridge University Press, 2003.
- CUTLER, A.; DAHAN, D.; VAN DONSELAAR, W. Prosody in the comprehension of spoken language: a literature review. **Language and Speech**, v. 40, n. 2, p. 141–201, 1997.

- D'ODORICO, L. Non-segmental features of prelinguistic communication: an analysis of some types of infant cry and non-cry vocalizations. **Journal of Child Language**, v. 11, p. 17–27, 1984.
- DECASPER, A. J.; FIFER, W. P. Of human bonding: newborns prefer their mothers' voices. **Science**, v. 208, n. 4448, p. 1174–1176, 1980.
- EIMAS, P. D. et al. Discrimination of speech by infants. **Science**, v. 171, n. 3968, p. 303–306, 1971.
- EKMAN, P. Facial expressions of emotion: new findings, new questions. **Psychological Science**, v. 3, n. 1, p. 34–38, 1992.
- EKMAN, P.; FRIESEN, W. V. **Facial Action Coding System**. Palo Alto: Consulting Psychologists Press, 1978.
- ESTEVE-GIBERT, N.; GUELLAÏ, B. Prosody in the auditory and visual domains: a developmental perspective. **Frontiers in Psychology**, v. 9, p. 338, 2018.
- \_\_\_\_\_. The role of prosody and gestures in the development of spoken language: a multimodal perspective. **Frontiers in Psychology**, v. 9, p. 1387, 2018.
- FANT, Gunnar. **Acoustic Theory of Speech Production: With Calculations based on X-Ray Studies of Russian Articulations**. Haia: Mouton, 1960.
- FANT, Gunnar. The source-filter concept in voice production. **SSTL-QPSR**, v. 24, n. 1, p. 21–37, 1983.
- FARRONI, T. et al. Eye contact detection in humans from birth. **Proceedings of the National Academy of Sciences**, v. 99, n. 14, p. 9602–9605, 2002.
- FELDMAN, R. Parent-infant synchrony: biological foundations and developmental outcomes. **Current Directions in Psychological Science**, v. 16, n. 6, p. 340–345, 2007.
- \_\_\_\_\_. Parenting behavior as the environment where children grow. **Biological Psychiatry**, v. 72, n. 10, p. 795–801, 2012.
- FERNALD, A. et al. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. **Journal of Child Language**, v. 16, n. 3, p. 477–501, 1989.
- FÓNAGY, I. **La vive voix: essais de psycho-phonétique**. Paris: Payot, 1983.
- FONTE, R. F. L. da et al. Estudos em Aquisição da Linguagem e Multimodalidade no Nordeste brasileiro. **Estudos da Língua(gem)**, v. 20, n. 1, p. 195–218, 2022.
- FONTES, M. A. S.; MADUREIRA, S. Um experimento sobre a linguagem não verbal na detecção de efeitos de sentidos: o questionamento da autenticidade. In: MOUTINHO, L. C. et al., coords. **Estudos em variação linguística nas línguas românicas – 2**. 1. ed. Aveiro: UA Editora, 2022. p. 138–159.
- GERVAIN, J.; MEHLER, J. The statistical basis of speech segmentation for infants. **Language Learning and Development**, v. 6, n. 4, p. 253–275, 2010.
- GERVAIN, J. et al. Near-infrared spectroscopy: a report from the McDonnell infant methodology consortium. **Developmental Cognitive Neuroscience**, v. 1, n. 1, p. 22–46, 2011.

GOLDSTEIN, M. H.; WASS, S. V.; LEONG, V. Synchrony and the development of social communication. **Current Opinion in Psychology**, v. 17, p. 132–139, 2017.

GROSSMANN, T.; JOHNSON, M. H. The development of the social brain in human infancy. **European Journal of Neuroscience**, v. 25, n. 4, p. 909–919, 2007.

GUELLAÏ, B.; STRERI, A.; CHOPIN, A. Prosody and gestures: a developmental perspective. **Frontiers in Psychology**, v. 5, art. 700, 2014.

GUELLAÏ, B. et al. Newborns' sensitivity to prosody in maternal speech: the role of temporal envelope. **Infancy**, v. 21, n. 3, p. 331–354, 2016.

HINTON, L.; NICHOLS, J.; OHALA, J. J., orgs. **Sound symbolism**. Cambridge: Cambridge University Press, 1994.

HIRST, D.; DI CRISTO, A. **Intonation systems: a survey of twenty languages**. Cambridge: Cambridge University Press, 1998.

JAKOBSON, R.; WAUGH, L. R. **The sound shape of language**. Bloomington: Indiana University Press; Harvester Press, 1979.

JUN, S.-A., org. **Prosodic typology: the phonology of intonation and phrasing**. Oxford: Oxford University Press, 2005.

KENDON, A. **Gesture: visible action as utterance**. Cambridge: Cambridge University Press, 2009.

\_\_\_\_\_. Some modern reflections on the history of the study of gesture. **Gesture**, v. 11, n. 3, p. 349–359, 2011.

KISILEVSKY, B. S. et al. Fetal sensitivity to properties of maternal speech and language. **Infant Behavior and Development**, v. 32, n. 1, p. 59–71, 2009.

KNOEFERLE, P. et al. What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. **Scientific Reports**, v. 7, art. 5562, 2017.

LAVER, J. **The phonetic description of voice quality**. Cambridge: Cambridge University Press, 1980.

LAVER, J.; MACKENZIE-BECK, J. **Vocal Profile Analysis Scheme – VPAS. Handout**. Edinburgh: Queen Margaret University College, Research Centre, 2007.

LEWKOWICZ, D. J. Infants' response to the audible and visible properties of the human face. I: role of lexical-syntactic content, temporal synchrony, gender, and manner of speech. **Developmental Psychology**, v. 32, n. 2, p. 347, 1996.

LEWKOWICZ, D. J. Infants' response to the audible and visible properties of the human face: II. discrimination of differences between singing and adult-directed speech. **Developmental Psychobiology**, v. 32, n. 4, p. 261–274, 1998.

LIBERMAN, Alvin M.; MATTINGLY, Ignatius G. The motor theory of speech perception revised. **Cognition**, v. 21, n. 1, p. 1–36, 1985.

LOIJENS, L.; KRIPS, O. FaceReader methodology note. **Behavioral Research Blog**, Noldus Information Technology, 2018. Disponível em: <https://student.hva.nl/binaries/content/assets/serviceplein-a-z-lemmas/media-creatie-en->

informatie/media–communicatie/observatorium/factsheet-facial-coding-reference-manual.pdf?2900513938585. Acesso em: 8 ago. 2021.

LOPES, J. et al. Prosody growth and reading comprehension: a longitudinal study from 2nd through the end of 3er grade. **Revista de Psicodidáctica**, v. 20, n. 1, p. 5–23, 2014.

MADUREIRA, S.; FONTES, M. A. S. The analysis of facial and speech expressivity: tools and methods. In: LAHOZ-BENGOCHEA, J. M.; RAMÓN, R. P., orgs. **Subsidia: tools and resources for speech sciences**. 1. ed. Málaga: Universidade de Málaga, 2019. v. 1, p. 1–150.

MADUREIRA, S. **Fala e expressividade**. Verbetes LBASS, 2020. Disponível em: <http://www.letras.ufmg.br/lbass/>. Acesso em: 17 jul. 2025.

MADUREIRA, S.; FONTES, M. A. S. Multimodal impressions of voice quality settings: the role of vocal and visual symbolism. **Frontiers in Communication**, v. 8, 2023.

MAMPE, B. et al. Newborns' cry melody is shaped by their native language. **Current Biology**, v. 19, n. 23, p. 1994–1997, 2009.

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

MIRANDA, L. S.; MORAES, J. A. de; RILLIARD, A. O papel dos gestos faciais na percepção da entoação da asserção e da questão-eco em áudio limpo e degradado. **Gradus**, v. 5, n. 1, 2022.

NAME, C. Language perception development. In: DEL RÉ, A.; FALASCA, P.; NÁPOLES, J. N., orgs. **From discriminating to discrimination: the influence of language on identity and subjectivity**. São Paulo: Fundação Editora da Unesp, 2022. p. 13–22.

NAME, C.; SOSA, J. M. Prosódia e funções discursivas das interrogativas na fala dirigida à criança. **Estudos da Língua(gem)**, v. 20, n. 1, p. 71–89, 2022.

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

NOBILE, L. Introduction – Sound symbolism in the age of digital orality. A perspective on language beyond «nature» and «culture». **Signifiences (Signifying)**, 2019.

NOLDUS. **Facial Action Coding System (FACS)**. [s.d.]. Disponível em: <https://www.noldus.com/applications/facial-action-coding-system>. Acesso em: 15 dez. 2024.

OHALA, J. J. An ethological perspective on common cross-language utilization of F0 of voice. **Phonetica**, v. 41, n. 1, p. 1–16, 1984.

\_\_\_\_\_. The frequency code underlies the sound symbolic use of voice pitch. In: HINTON, L.; NICHOLS, J.; OHALA, J. J., orgs. **Sound symbolism**. Cambridge: Cambridge University Press, 1994. p. 325–347.

OSTER, H. et al. Pre-conference workshop: Facial Action Coding System for infants and young children (Baby FACS). In: **ICIS 2018 – Building Bridges**. International Society on Infant Studies, 2018. p. 24–25.

PACHECO, V. Movimentos faciais e corporais e percepção de ênfase e atenuação. In: **COLÓQUIO BRASILEIRO DE PROSÓDIA DA FALA**, 3., 2011.

PELTOLA, M. J. et al. Fearful faces modulate looking duration and attention disengagement in 7-month-old infants. **Developmental Science**, v. 11, n. 1, p. 60–68, 2008.

- PIERREHUMBERT, J. B. **The phonetics and phonology of English intonation**. 1980. Tese (Doutorado em Linguística) — Massachusetts Institute of Technology, Cambridge, 1980.
- POULIN-DUBOIS, D.; GRAHAM, S. A. Cognitive processes in early word learning. **Psychological Bulletin**, v. 133, n. 4, p. 628–647, 2007.
- SAINT-GEORGES, C. et al. Motherese in interaction: at the cross-road of emotion and cognition? (a systematic review). **PLOS ONE**, v. 8, n. 10, art. e78103, 2013.
- SCHERER, K. R. Vocal affect expression: a review and a model for future research. **Psychological Bulletin**, v. 99, n. 2, p. 143–165, 1986.
- \_\_\_\_\_. Vocal communication of emotion: a review of research paradigms. **Speech Communication**, v. 40, n. 1–2, p. 227–256, 2003.
- SCHERER, K. R.; ELLGRING, H. Multimodal expression of emotion: affect programs or componential appraisal patterns? **Emotion**, v. 7, n. 1, p. 158–171, 2007.
- SCHWARZ, I.-C. et al. Affect in infant-directed speech of Swedish-speaking mothers and fathers to 3-, 6-, 9-, and 12-month-old infants. **Language Learning and Development**, v. 20, n. 2, p. 145–157, 2023.
- SNOW, K. Mothers' speech research: from input to interaction. In: SNOW, C. E.; FERGUSON, C. A., eds. **Talking to children: language input and acquisition**. Cambridge: Cambridge University Press, 1977.
- SPINELLI, M.; FASOLO, M.; MESMAN, J. Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. **Developmental Review**, v. 44, p. 1–18, 2017.
- SVANTESSON, J.-O. Sound symbolism: the role of word sound in meaning. **WIREs Cognitive Science**, v. 8, n. 5, art. e1441, 2017.
- THIESSEN, E. D.; HILL, D. S.; SAFFRAN, J. R. Infant-directed speech facilitates word segmentation. **Infancy**, v. 7, n. 1, p. 53–71, 2005.
- TOMASELLO, M. Acquiring linguistic constructions. In: KUHN, D.; SIEGLER, R., eds. **Handbook of child psychology**. New York: Wiley, 2006.
- TREVARTHEN, Colwyn. Communication and cooperation in early infancy: A description of primary intersubjectivity. In: BULLOWA, Margaret (ed.). **Before speech: The beginning of interpersonal communication**. Cambridge: Cambridge University Press, 1979.
- TRONICK, Edward Z. Emotions and emotional communication in infants. **American Psychologist**, v. 44, n. 2, p. 112–119, fev. 1989.
- TSUR, R. **What makes sound patterns expressive? The poetic mode of speech perception**. Durham: Duke University Press, 1992.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: **Computer Vision and Pattern Recognition**, 2001.
- VOLOUMANOS, A.; WERKER, J. F. Listening to language at birth: evidence for a bias for speech in neonates. **Developmental Science**, v. 10, n. 2, p. 159–164, 2007.
- WASS, S. V. et al. Interpersonal neural entrainment during early social interaction. **Trends in Cognitive Sciences**, v. 24, n. 4, p. 329–341, 2020.

WERKER, J. F.; LALONDE, C. E. Cross-language speech perception: initial capabilities and developmental change. **Developmental Psychology**, v. 24, n. 5, p. 672–683, 1988.

WITTENBURG, P. et al. ELAN: a professional framework for multimodality research. In: **Proceedings of LREC**, 2006.

## APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO



### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Gostaríamos de convidar você a participar como voluntária (o) da pesquisa Explorando a Prosódia Facial e Vocal na Interação Bebê-Cuidador: Um Estudo da Multimodalidade Gestuo-Vocal na Aquisição da Linguagem. O motivo que nos leva a realizar esta pesquisa é caracterizar como o/a bebê percebe e responde às expressões faciais e sonoras do adulto quando eles interagem durante seu primeiro ano de vida. Nesta pesquisa, buscamos analisar os movimentos faciais e as expressões sonoras produzidas pelo (a) bebê em situações rotineiras de interação com seu cuidador, mãe ou pai. Também analisaremos essas expressões do(a) cuidador(a), de modo a observar a integração delas nas reações do(a) bebê.

Caso você concorde em participar, vamos fazer visitas à sua casa, quando filmaremos momentos em que você interage com seu/sua filho(a) – situações rotineiras, como conversar, brincar ou contar uma história para ele/ela, no período em que ele/ela tiver entre 4 e 10 meses de idade. As visitas poderão ser mensais ou quinzenais, como você preferir, e os dias, horários, duração e atividades também serão de sua escolha. Levaremos e posicionaremos os equipamentos (duas câmeras com tripés) de modo a captar seu rosto e o de seu/sua filho(a), mas evitando ao máximo interferir na interação de vocês e causar algum desconforto. Além de vocês, apenas a pesquisadora responsável estará presente durante a filmagem das atividades. É importante destacar que não se trata de nenhum tipo de avaliação, apenas queremos descrever as expressões que acontecem durante essas interações.

Esta pesquisa tem alguns riscos, que estão relacionados à possibilidade de os participantes serem identificados posteriormente, através de seu rosto e/ou de sua voz. Para diminuir a chance desses riscos acontecerem, adotaremos as seguintes medidas: (i) os arquivos de vídeo e áudio serão armazenados em dispositivo físico guardado em nosso laboratório de pesquisa, acessível apenas pelos participantes do projeto de pesquisa; na divulgação dos



resultados, (ii) os nomes dos participantes serão trocados; (iii) as imagens de face serão camufladas por malhas digitais, e (iv) os dados sonoros serão apresentados em espectograma e na forma de onda, impedindo, respectivamente, a identificação visual e o reconhecimento da voz dos participantes; ainda, (v) nenhum material que indique sua participação será liberado sem a sua permissão.

Esta pesquisa pode ajudar a entender melhor o desenvolvimento comunicativo, emocional e linguístico do bebê, assim como o papel das expressões faciais e sonoras durante a interação adulto-bebê nesse processo. Isso é muito importante, pois existem crianças que têm algumas dificuldades nesse processo. Para entender melhor o que ocorre com elas e poder ajudá-las, primeiro precisamos conhecer melhor como se dão as interações adulto-bebê quando não há nenhuma dificuldade acontecendo.

Para participar deste estudo você não vai ter nenhum custo, nem receberá qualquer vantagem financeira. Apesar disso, se você tiver algum dano causado pelas atividades que fizermos com você nesta pesquisa, você tem direito a buscar indenização.

Você terá todas as informações que quiser sobre esta pesquisa e estará livre para participar ou recusar-se a participar. Mesmo que você queira participar agora, você pode voltar atrás ou parar de participar a qualquer momento. A sua participação é voluntária e o fato de não querer participar não vai trazer qualquer penalidade ou mudança na forma em que você é atendida (o). A pesquisadora não vai divulgar seu nome. Os resultados da pesquisa estarão à sua disposição quando finalizada. Seu nome ou o material que indique sua participação não será liberado sem a sua permissão. Você não será identificada (o) em nenhuma publicação que possa resultar dessa pesquisa

Este termo de consentimento encontra-se impresso em duas vias originais, sendo que uma será arquivada pela pesquisadora responsável e a outra será fornecida a você. Os dados coletados na pesquisa ficarão arquivados com a pesquisadora responsável por um período de 5 (cinco) anos. Decorrido este tempo, a pesquisadora avaliará os documentos para a sua destinação final, de acordo com a legislação vigente. Os pesquisadores tratarão a sua identidade com padrões profissionais de sigilo, atendendo a legislação brasileira (Resolução Nº 466/12 do Conselho Nacional de Saúde), utilizando as informações somente para os fins acadêmicos e científicos.

Declaro que concordo em participar da pesquisa e que me foi dada à oportunidade de ler e esclarecer as minhas dúvidas.

(        ) Autorizo a gravação, mas não autorizo a divulgação de minha imagem e de minha voz.

(        ) Não autorizo a gravação nem a divulgação de minha imagem e de minha voz.

Juiz de Fora, \_\_\_\_\_ de \_\_\_\_\_ de 2024.

---

**Assinatura da (o) Participante**

---

**Assinatura da Pesquisadora**

**Nome da Pesquisadora Responsável: Aline Cristina Cardoso**

**Campus Universitário da UFJF**

**Faculdade de Letras/ Programa de Pós-Graduação em Linguística/ Núcleo de Estudos em Aquisição da Linguagem e Psicolinguística (NEALP)**

**CEP: 36036-900**

**Fone: (32) 991442737**

**E-mail: [04030053610@estudante.ufjf.br](mailto:04030053610@estudante.ufjf.br)**

**APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E  
ESCLARECIDO/RESPONSÁVEIS**



***TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO/RESPONSÁVEIS***

A criança \_\_\_\_\_, sob sua responsabilidade, está sendo convidada como voluntário (a) a participar da pesquisa Explorando a Prosódia Facial e Vocal na Interação Bebê-Cuidador: Um Estudo da Multimodalidade Gestuo-Vocal na Aquisição da Linguagem. O motivo que nos leva a realizar esta pesquisa é caracterizar como o/a bebê percebe e responde às expressões faciais e sonoras do adulto quando eles interagem durante seu primeiro ano de vida. Nesta pesquisa, buscamos analisar os movimentos faciais e as expressões sonoras produzidas pelo(a) bebê em situações rotineiras de interação com seu/sua cuidador(a), mãe ou pai. Também analisaremos essas expressões do(a) cuidador(a), de modo a observar a integração delas nas reações do(a) bebê.

Caso você concorde com a participação da criança, vamos fazer as seguintes atividades com ela/ele. Faremos visitas à sua casa, quando filmaremos momentos em que ela/ele interage com seu/sua cuidador(a) – situações rotineiras, como conversas, brincadeiras ou contação de história, no período em que ela/ele tiver entre 4 e 10 meses de idade. As visitas poderão ser mensais ou quinzenais, de acordo com a sua preferência, e os dias, horários, duração e atividades também serão de sua escolha. Levaremos e posicionaremos os equipamentos (duas câmeras com tripés) de modo a captar o rosto da criança e o do(a) cuidador(a), mas evitando ao máximo interferir na interação de deles/delas e causar algum desconforto. Além da criança e do(a) cuidador(a), apenas a pesquisadora responsável estará presente durante a filmagem das atividades. É importante destacar que não se trata de nenhum tipo de avaliação, apenas queremos descrever as expressões que acontecem durante essas interações.

Esta pesquisa tem alguns riscos, que estão relacionados à possibilidade de os participantes serem identificados posteriormente, através de seu rosto e/ou de sua voz. Para diminuir a chance desses riscos acontecerem, adotaremos as seguintes medidas: (i) os arquivos de vídeo e áudio serão armazenados em dispositivo físico guardado em nosso laboratório de pesquisa, acessível apenas pelos participantes do projeto de pesquisa; na divulgação dos

resultados, (ii) os nomes dos participantes serão trocados, (iii) as imagens de face serão camufladas por malhas digitais, e (iv) os dados sonoros serão apresentados em espectograma e na forma de onda, impedindo, respectivamente, a identificação visual e o reconhecimento da voz dos participantes; ainda, (v) nenhum material que indique a participação de seu/sua filho(a) será liberado sem a sua permissão.

Esta pesquisa pode ajudar a entender melhor o desenvolvimento comunicativo, emocional e linguístico do bebê, assim como o papel das expressões faciais e sonoras durante a interação adulto-bebê nesse processo. Isso é muito importante, pois existem crianças que têm algumas dificuldades nesse processo. Para entender melhor o que ocorre com elas e poder ajudá-las, primeiro precisamos conhecer melhor como se dão as interações adulto-bebê quando não há nenhuma dificuldade acontecendo.

Para participar desta pesquisa, a criança sob sua responsabilidade e você não irão ter nenhum custo, nem receberão qualquer vantagem financeira. Apesar disso, se a criança tiver algum dano por causa das atividades que fizermos com ela/ele nesta pesquisa, ela/ele tem direito a buscar indenização.

Como responsável pela criança, você terá todas as informações que quiser sobre esta pesquisa e estará livre para permitir ou recusar a participação dele(a). Você poderá retirar seu consentimento ou interromper a participação dele(a) a qualquer momento. Mesmo que você queira deixá-lo(a) participar agora, você pode voltar atrás e parar a participação a qualquer momento. A participação dele(a) é voluntária e o fato em não deixá-lo(a) participar não vai trazer qualquer penalidade ou mudança na forma em que ele/ela é atendido(a). Os resultados da pesquisa estarão à sua disposição quando finalizada. O nome ou o material que indique a participação da criança sob sua responsabilidade não será liberado sem a sua permissão. A criança não será identificada em nenhuma publicação.

Este termo de consentimento encontra-se impresso em duas vias originais, sendo que uma será arquivada pela pesquisadora responsável e a outra será fornecida a você. Os dados coletados na pesquisa ficarão arquivados com a pesquisadora responsável por um período de 5 (cinco) anos. Decorrido este tempo, a pesquisadora avaliará os documentos para a sua destinação final, de acordo com a legislação vigente. Os pesquisadores tratarão a sua identidade com padrões profissionais de sigilo, atendendo a legislação brasileira (Resolução N° 466/12 do

Conselho Nacional de Saúde), utilizando as informações somente para os fins acadêmicos e científicos.

Declaro que concordo em deixá-lo(a) participar da pesquisa e que me foi dada à oportunidade de ler e esclarecer as minhas dúvidas.

( ) Autorizo a gravação, mas não autorizo a divulgação da imagem e da voz da criança \_\_\_\_\_, sob minha responsabilidade.

( ) Não autorizo a gravação nem a divulgação da imagem e da voz da criança \_\_\_\_\_, sob minha responsabilidade.

Juiz de Fora, \_\_\_\_\_ de \_\_\_\_\_ de 2024.

\_\_\_\_\_  
Assinatura do(a) Responsável

\_\_\_\_\_  
Assinatura da Pesquisadora

Nome da Pesquisadora Responsável: Aline Cristina Cardoso

Campus Universitário da UFJF

Faculdade de Letras/ Programa de Pós-Graduação em Linguística/ Núcleo de Estudos em Aquisição da Linguagem e Psicolinguística (NEALP)

CEP: 36036-900

Fone: (32) 991442737

E-mail: [04030053610@estudante.ufjf.br](mailto:04030053610@estudante.ufjf.br)

## **ANEXO A**

Este documento é um parecer consubstanciado do Comitê de Ética em Pesquisa (CEP) da Universidade Federal de Juiz de Fora (UFJF) sobre o projeto "Explorando a Prosódia Facial e

Vocal na Interação Bebê-Cuidador: Um Estudo da Multimodalidade Gestuo-Vocal na Aquisição da Linguagem". O parecer, datado de 23 de agosto de 2024, número 7.025.159, aprova o projeto de pesquisa liderado por Aline Cristina Cardoso.

## **ANEXO B – Procedimentos de análise dos dados**

### **I- Análise da expressão vocal**

#### **1) Número de enunciados em cada segmento de análise:**

#### **2) Modalidades de contornos melódicos:**

- ( ) assertivo;
- ( ) interrogativo {questão total};
- ( ) interrogativo {questão parcial};
- ( ) exclamativo.

#### **3) Características prosódicas**

##### **3.1. Pitch**

- Média: ( ) Alta; ( ) Média; ( ) Baixa.
- Extensão: Maximizada ( ); ( ) Regular; ( ) Minimizada.
- Variabilidade: ( ) Alta; ( ) Média; ( ) Baixa.

##### **3.2. Loudness**

- Média: ( ) Alta; ( ) Média; ( ) Baixa.
- Extensão: Maximizada ( ); ( ) Regular; ( ) Maximizada.
- Variabilidade: ( ) Alta; ( ) Média; ( ) Baixa.

##### **3.3 Elocução**

Taxa: ( ) Rápida; ( ) Regular; ( ) Lenta.



#### **4) Ajustes de qualidade de voz**

Ajustes de trato vocal:

Ajustes de tensão muscular:

Ajustes de fonação:

### **II- Análise da expressão facial**

- 1) Unidades de Ação da Face e suas intensidades
- 2) Orientação da cabeça
- 3) Direção do olhar
- 4) Ativação do organismo
- 5) Valência
- 6) Batimento cardíaco

### **III- Análise de parâmetros acústicos**

#### **1) Frequência fundamental (F0)**

- Mínima em Hz (    )
- Máxima em Hz (    )

#### **2) Duração**

- Duração total em ms do segmento de análise:
- Duração total em ms das pausas no segmento de análise:
- Percentagem de pausas silenciosas em relação à duração do segmento de análise:

### **3) Intensidade**

-Variação da Intensidade em dB ao longo do segmento de análise: Mínima (     ); Máxima (     ).

## **IV-            Análise dos Disparadores (Triggers) de Eventos de Interação e Atratores de atenção**

- 1) Transcrição ortográfica do trecho de fala
- 2) Eventos observados