

**Universidade Federal de Pelotas
Centro de Letras e Comunicação
Programa de Pós-Graduação em Letras
Mestrado em Linguagem, Texto e Imagem**



Dissertação de Mestrado

Aquisição de encontros consonantais com *tap* no português brasileiro:

análises acústica e articulatória

Thais Telles Barbieri

Pelotas, 2019

Thais Telles Barbieri

Aquisição de encontros consonantais com *tap* no português brasileiro:

análises acústica e articulatória

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Letras do Centro de Letras e Comunicação da Universidade Federal de Pelotas, como requisito parcial para obtenção do título de Mestre em Letras, área Linguagem, Texto e Imagem.

Orientadora: Prof.^a Dr. Giovana Ferreira Gonçalves

Pelotas, 2019

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

B236a Barbieri, Thais Telles

Aquisição de encontros consonantais com tap no português brasileiro : análises acústica e articulatória / Thais Telles Barbieri ; Giovana Ferreira Gonçalves, orientadora. — Pelotas, 2019.

236 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação em Letras, Centro de Letras e Comunicação, Universidade Federal de Pelotas, 2019.

1. Aquisição fonológica. 2. Fonologia Articulatória. 3. Encontros consonantais. 4. Análise acústica. 5. Ultrassonografia. I. Gonçalves, Giovana Ferreira, orient. II. Título.

CDD : 418

Thais Telles Barbieri

**Aquisição de encontros consonantais com tap no português brasileiro: análises
acústica e articulatória**

Dissertação aprovada, como requisito parcial, para obtenção do grau de Mestre em Letras, Área de Concentração Estudos da Linguagem, do programa de Pós-Graduação em Letras, da Universidade Federal de Pelotas.

Pelotas, 26 de fevereiro de 2019

Banca examinadora:



Profa. Dra. GIOVANA FERREIRA GONCALVES

Orientadora/Presidente da banca

Universidade Federal de Pelotas



Profa. Dra. MÁRCIA KESKE SOARES

Membro da Banca

UNIVERSIDADE FEDERAL DE SANTA MARIA



Profa. Dra. MIRIAN ROSE BRUM DE PAULA

Membro da Banca

Universidade Federal de Pelotas

A minha mãe, por tudo.

Agradecimentos

Os agradecimentos são muitos e o leitor que quiser ir direto ao ponto pode dispensar esta seção sem ressentimentos. No entanto, esta é uma seção não apenas necessária, como também repleta de afeto e boas energias. Um dos momentos mais prazerosos de se escrever este trabalho de pesquisa foi a oportunidade de agradecer a todas as pessoas e organizações que tornaram este trabalho possível. Então, vamos lá.

À CAPES, pela bolsa a mim concedida durante este mestrado.

Ao CNPq, pela bolsa de iniciação científica a mim concedida na graduação, quando iniciei meu caminho na pesquisa em fonologia.

Aos governantes e entidades que apoiaram e valorizaram a educação e a pesquisa brasileiras.

Aos informantes desta pesquisa e aos responsáveis pela criança participante, pela disposição em contribuir com o trabalho.

Às professoras e professores do Centro de Letras e Comunicação (CLC) da UFPel que me incentivaram nos meus percursos da graduação e da pós-graduação. Vocês são exemplos para mim. Cada um de vocês tem parte na acadêmica que sou hoje e que ainda serei.

À professora Roberta Rego Rodrigues, pelas aulas de Ferramentas para Tratamento de Dados Linguísticos nas quais pude ampliar minha experiência com o *Excel*. Agradeço, também, pela gentileza e pelas conversas amigas quando nos víamos pelos corredores. Essas coisas, aparentemente pequenas, fizeram o CLC ser um lugar de acolhimento para mim.

À professora Mirian Rose Brum de Paula, que, no início da minha graduação, me apresentou a uma ciência intrigante chamada Linguística (!) e, anos depois, me orientou em meu estágio de regência na disciplina de Psicolinguística. Professora, muito obrigada por ter participado da minha formação e, também, por todas as contribuições na qualificação e na defesa de dissertação deste trabalho.

À professora Márcia Keske Soares, por todas as contribuições na qualificação e na defesa. A leitura atenta de uma profissional da fonoaudiologia foi essencial para compreender meu próprio trabalho e aprimorá-lo.

À professora Giovana Ferreira Gonçalves, minha orientadora da iniciação científica ao mestrado, por ter me apresentado aos estudos em aquisição fonológica e por ter visto em mim um potencial para a pesquisa. Professora, obrigada por toda a paciência que tiveste comigo, por toda a confiança no meu trabalho, pela sabedoria compartilhada e por todas as orientações em que, volta e meia, estávamos discutindo algum tópico que, por si só, já renderia um outro estudo. Contigo, aprendi a valorizar a minha própria curiosidade. Obrigada.

Aos psicólogos Ana Candida Wetzel Vieira e Felipe Schulte Ferreira Rodrigues. Com o acompanhamento desses grandes profissionais, pude me conhecer ainda melhor em vários aspectos. Eles me ajudaram a lidar com meus medos e inseguranças para que eu pudesse conduzir esta pesquisa até o fim.

À Patrícia, colega querida do laboratório, por toda a boa vontade e disposição em me ajudar nas coletas de dados.

Ao Roberto, por todas as conversas e por todo o apoio que me motivou a conduzir este trabalho, principalmente nas fases finais.

Ao Lucas, pela grande amizade que começamos na graduação e pelo entusiasmo pela Linguística que compartilhamos desde então. Agradeço também pela revisão atenta do *resumen* deste trabalho. Que nós possamos continuar próximos independentemente de onde estivermos.

À Aline, que foi minha colega na graduação, depois se tornou colega de laboratório, depois de mestrado... e acabou se tornando minha grande amiga. Aline, guardo cada conversa, cada risada e cada tarde de trabalho com muito carinho. Obrigada por ter tornado meus dias mais leves. Tu és uma pesquisadora com grande futuro e uma pessoa de coração enorme.

À minha mãe, Suzana Telles, por tantas coisas. Serei eternamente grata por toda a educação e amparo que recebi, por sempre apoiar minhas escolhas, por sempre acreditar no meu valor enquanto pessoa e enquanto profissional e, principalmente, pelo amor e pelo cuidado sem igual, desde sempre. Mãe, nunca terei agradecido o suficiente por tudo. Te amo muito.

Por fim, agradeço a todos os familiares, amigos, colegas e professores que acreditam em mim e se orgulham de estar do meu lado.

A todos vocês, muito, muito obrigada.

*"The only person standing in your way is you.
It's time to let her go. **Lose yourself.**"*

*– Thomas Leroy (personagem)
Black Swan (filme, 2010)*

Resumo

BARBIERI, T. T. **Aquisição de encontros consonantais com *tap* no português brasileiro**: análises acústica e articulatória. 236f. Dissertação (Mestrado em Letras). Programa de Pós-Graduação em Letras, Centro de Letras e Comunicação, Universidade Federal de Pelotas, Pelotas, 2019.

Este estudo investiga o processo de aquisição de encontros consonantais formados por obstruente e *tap* em dados de fala acústicos e articulatórios de uma criança adquirindo o português brasileiro (PB) como língua materna. O principal objetivo do estudo é descrever e analisar a emergência de sílabas CCV (consoante + consoante + vogal) com base em pressupostos da Fonologia Articulatória (BROWMAN; GOLDSTEIN, 1989). Os objetivos específicos são i) verificar se a duração de vogais difere significativamente em função do tipo de sílaba (CCV ou CV); ii) verificar se a duração de obstruintes difere em função do mesmo fator; iii) descrever padrões de coordenação gestual envolvidos na produção de CCVs; iv) descrever padrões de coordenação gestual envolvidos na produção do *tap* e v) verificar de que formas dados acústicos e articulatórios se complementam. Para tais propósitos, foram realizadas coletas de dados de fala com dois grupos: i) uma criança adquirindo encontros consonantais e ii) três adultas. A criança realizou oito coletas longitudinais, enquanto cada adulta realizou uma coleta transversal. Os informantes produziram pares mínimos com diferentes sílabas iniciais, CCV ou CV (e.g. “prato” e “pato”). As palavras foram gravadas em áudio e ultrassom, dentro de cabine acústica, com o programa AAA. Esses dados passaram por análise acústica e por análise articulatória qualitativa. Para a análise acústica, as palavras foram organizadas e segmentadas no programa *Praat*, para posterior realização de medidas. A partir destas, foram gerados valores médios que passaram por testes estatísticos paramétricos. Já a análise articulatória consistiu no exame qualitativo do contorno de língua no momento de máxima constrição da ponta de língua na produção do *tap*. Os resultados descritivos demonstraram que, na fala adulta, as durações de obstruintes e vogais são mais longas em CVs do que em CCVs, independentemente de contexto fonológico. As durações de vogais e obstruintes em CCVs e CVs foram comparadas por meio de um teste-T, o qual revelou diferenças significativas na maioria dos contextos. Os resultados mostraram, assim, que diferenças em duração não são exclusivas da fala infantil. Testes One-Way ANOVA compararam durações de obstruente e vogal antes e depois de a criança começar a produzir CCVs. As obstruintes apresentaram diferenças em ambos os grupos; também, a criança apresentou vogais mais longas mesmo após começar a produzir CCVs. Portanto, os dados não parecem evidenciar alongamento compensatório, mas sim ajustes em duração, os quais seguem ocorrendo quando CCVs já são produzidas. As medidas *elemento vocálico + vogal nuclear* e *elemento vocálico + tap + vogal nuclear* foram comparadas a vogais em CVs; a segunda medida mostrou

significâncias para todos os contextos. A inspeção articulatória identificou a produção de diferentes tipos de róticos, como *tap*, vibrante, retroflexa e aproximante. A realização do *tap* mostrou somente gesto de ponta de língua, e os contornos de língua sugerem que há uma coordenação gestual em processo de estabilização. Também, um teste-T comparou contornos de língua na produção do elemento vocálico e da vogal nuclear em CCVs. O teste não mostrou significâncias, o que sugere que elemento vocálico e vogal nuclear são o mesmo segmento, o qual não é interrompido, mas sim *eclipsado* pelo *tap*.

Palavras-chave: aquisição fonológica; Fonologia Articulatória; encontros consonantais; análise acústica; ultrassonografia

Abstract

BARBIERI, T. T. **Acquisition of consonantal clusters with taps in Brazilian Portuguese**: acoustic and articulatory analyses. 236p. Dissertation (Masters in Languages). Postgraduate Program in Languages, Center of Languages and Communication, Federal University of Pelotas, Pelotas, 2019.

This study examines the process of acquisition of consonantal clusters formed by an obstruent and a tap, in acoustic and articulatory speech data of a child acquiring Brazilian Portuguese as a first language. The main aim of this study is to describe and analyze the emergence of CCV (consonant + consonant + vowel) syllables according to principles of Articulatory Phonology (BROWMAN; GOLDSTEIN, 1989). Specific aims of this study are to i) verify whether vowel length differs significantly according to syllable type (CCV or CV); ii) verify whether obstruent length differs significantly according to syllable type; iii) describe patterns of gestural coordination involved in the production of CCV syllables; iv) describe patterns of gestural coordination involved in the production of taps and v) verify in which ways acoustic and articulatory data complement one another. For such purposes, sessions of data collection were conducted with two groups: i) a child acquiring consonantal clusters and ii) three adults. The child attended to eight sessions of longitudinal data collection, whereas each adult attended to one session of transversal data collection. The informants produced minimal pairs with different initial syllables, *i.e.* CCV or CV, such as *prato* and *pato*. The words were recorded in audio and ultrasound inside an acoustic booth, using the software AAA. The data were submitted to acoustic analysis and qualitative articulatory analysis. For the acoustic analysis, the words were organized and segmented in the software Praat, so that length measures could be taken. These measures generated average values, which were submitted to parametrical statistical tests. The articulatory analysis consisted in the qualitative examination of the tongue contour in the moment of maximum tongue tip constriction in the production of the tap. Descriptive results showed that in adult speech, vowel and obstruent lengths are longer in CV than in CCV syllables, regardless of phonological context. The T-test compared vowel and obstruent lengths in CCV and CV syllables, which revealed significant differences in the majority of contexts. Hence, these results demonstrate that length differences are not exclusive of child language. One-Way ANOVAs compared obstruent and vowel lengths before and after the child started producing consonantal clusters. Obstruents presented differences in both groups and the child presented longer vowels even when clusters were already produced. Therefore, the data do not seem to evidence compensatory lengthening, but adjustments in length that still take place when CCV syllables are already produced. The measures *vowel-like element + nuclear vowel* and *vowel-like element + tap + nuclear vowel* were compared to CV vowels. The latter showed significances for all contexts. The articulatory inspection identified the production of

different types of rhotics, such as *tap*, vibrant, retroflex and approximant. The taps were produced only with a tongue tip gesture, and tongue contours suggest that gestural coordination is under a process of stabilization. Also, a T-test comparing tongue shapes in the production of the vowel-like element and the nuclear vowel did not provide significant differences. This suggests that the vowel-like element and the nuclear vowel are the same segment, which is not interrupted, but rather *eclipsed* by the tap.

Keywords: phonological acquisition; Articulatory Phonology; consonantal clusters; acoustic analysis; ultrasonography

Resumen

BARBIERI, T. T. **Adquisición de encuentros consonánticos con *tap* en el portugués brasileiro**: análisis acústico y articulatorio. 236p. Disertación (Maestría en Letras). Programa de Posgrado en Letras, Centro de Letras y Comunicación, Universidad Federal de Pelotas, Pelotas, 2019.

Este trabajo investiga como ocurre el proceso de adquisición de encuentros consonánticos formados por obstruyente y *tap*, en datos de habla acústicos y articulatorios de un niño adquiriendo el portugués brasileiro (PB) como lengua materna. El objetivo principal es describir y analizar la emergencia de sílabas CCV (consonante + consonante + vocal) en base a principios de la Fonología Articulatoria (BROWMAN; GOLDSTEIN, 1989). Los objetivos específicos son: i) verificar si la duración de vocales difiere significativamente en función del tipo de sílaba (CCV o CV); ii) verificar si la duración de obstruyentes difiere en función del mismo factor; iii) describir patrones de coordinación gestual involucrados en la producción de CCVs; iv) describir patrones de coordinación gestual involucrados en la producción del *tap* y v) verificar de que formas datos acústicos y articulatorios se completan. Para esos fines, se realizaron colectas de datos de habla con dos grupos: i) un niño adquiriendo encuentros consonánticos y ii) tres adultas. Con el niño se realizaron ocho colectas longitudinales, y con cada adulta se realizó una colecta transversal. Los informantes produjeron pares mínimos con sílabas iniciales diferentes, CCV o CV (e.g. *prato* y *pato*). Las palabras fueron grabadas en audio y ultrasonido, en cabina acústica, con el programa AAA. Esos datos pasaron por análisis acústico y por análisis articulatorio cualitativo. Para el análisis acústico, las palabras fueron organizadas y segmentadas en el programa *Praat*, para realización de medidas posterior. A partir de esas medidas, se calcularon promedios que pasaron por pruebas estadísticas paramétricas. Para el análisis articulatorio, se examinó el contorno de lengua en el momento de máxima constricción de punta de lengua en la producción del *tap*. Los resultados descriptivos demostraron que, en habla adulta, las duraciones de obstruyentes y vocales son más largas en CVs que en CCVs, independientemente de contexto fonológico. Las duraciones de vocales y obstruyentes en CCVs y CVs fueron comparadas en una prueba t, la cual mostró diferencias significativas en la mayoría de los contextos. Los resultados mostraron, por lo tanto, que diferencias de duración no son exclusivas del habla infantil. Análisis One-Way ANOVA compararon duraciones de obstruyente y vocal antes y después de que el niño empezara a producir CCVs. Las obstruyentes presentaron diferencias en ambos grupos; el niño también produjo vocales más largas, aún tras empezar a producir CCVs. Así, los datos no parecen indicar un alargamiento compensatorio, sino ajustes de duración, que siguen ocurriendo cuando CCVs ya son producidas. Las medidas *elemento vocálico + vocal nuclear* y *elemento vocálico + tap + vocal nuclear* fueron comparadas a vocales en CVs; la segunda medida mostró

significaciones para todos los contextos. La inspección articuladora identificó la producción de diferentes tipos de róticos, como *tap*, vibrante, retrofleja y aproximante. La realización del *tap* mostró solamente el gesto de punta de lengua y los contornos de lengua sugieren que hay una coordinación gestual en proceso de estabilización. Una prueba t comparó contornos de lengua en las producciones del elemento vocálico y de la vocal nuclear en CCVs. La prueba no reveló significaciones; lo cual sugiere que el elemento vocálico y la vocal nuclear son el mismo segmento, que no es interrumpido, sino *eclipsado* por el *tap*.

Palabras clave: adquisición fonológica; Fonología Articuladora; encuentros consonánticos; análisis acústico; ultrasonografía.

Lista de Figuras

Figura 1 – Representação do trato vocal com a localização das respectivas variáveis, reproduzida de Browman e Goldstein (1989, p. 73)	44
Figura 2 – Exemplos de pautas gestuais com base na palavra <i>palm</i> , do inglês, reproduzidos de Browman e Goldstein (1989, p. 76)	48
Figura 3 – Enunciado <i>pea splots</i> , produzido por falante de inglês, com oscilograma e deslocamento vertical dos articuladores envolvidos. Imagem adaptada de Browman e Goldstein (1988, p. 87)	52
Figura 4 – Oscilograma e deslocamento vertical dos articuladores lâmina de língua e lábio inferior para as produções de sentenças nas formas [pi#’C(C)(C)ats], à esquerda, e [pa#’C(C)(C)idz], à direita. Imagem adaptada de Browman e Goldstein (1988, p. 90-91)	53
Figura 5 – Oscilograma e deslocamento vertical dos articuladores lâmina de língua e lábio inferior para as produções de sentenças nas formas [pi#’C(C)(C)ats], à esquerda, e [pa#’C(C)(C)idz], à direita. Imagem adaptada de Browman e Goldstein (1988, p. 92-93)	55
Figura 6 – Oscilograma e deslocamento vertical dos articuladores lâmina de língua e lábio inferior para as produções de sentenças nas formas [pip ‘adz], [pis ‘adz] e [piC#’C(C)ats], à esquerda, e [paC#’C(C)idz], à direita. Imagem adaptada de Browman e Goldstein (1988, p. 94-95)	57
Figura 7 – Oscilograma e espectrograma de uma produção da palavra “prato”. Imagem reproduzida de Silva, Clemente e Nishida (2006, p. 5)	61
Figura 8 – Oscilograma e espectrograma de uma produção da palavra “expor”. Imagem reproduzida de Silva, Clemente e Nishida (2006, p. 11)	62

Figura 9 – Proposta de pauta gestual para a sequência de elemento vocálico, <i>tap</i> e vogal em grupos do tipo obstruinte + <i>tap</i> + vogal. Imagem reproduzida de Silva, Clemente e Nishida (2006, p. 21).....	64
Figura 10 – Proposta de pauta gestual para o <i>tap</i> em codas. Imagem reproduzida de Silva, Clemente e Nishida (2006, p. 22)	65
Figura 11 – Sobreposição de 14 repetições de <i>cop top</i> , produzidas por um falante de inglês. Imagem adaptada de Goldstein <i>et al.</i> (2007, p. 393)	70
Figura 12 – Distribuição dos tipos de erro ocorridos na produção dos sete sujeitos. Imagem adaptada de Goldstein <i>et al.</i> (2007, p. 394)	72
Figura 13 – Repetições de <i>cop top</i> , com transição de 2:1 para 1:1 no acoplamento de dorso de língua e lábio inferior. Imagem adaptada de Goldstein <i>et al.</i> (2007, p. 399).....	74
Figura 14 – Interior da cabine acústica do LELO, utilizada para as coletas de áudio e ultrassom.....	94
Figura 15 – Exemplo de imagem utilizada na coleta, correspondente à palavra “prato”	98
Figura 16 – Exemplo de imagem utilizada na coleta, correspondente à palavra “pato”	98
Figura 17 – Tipos de sonda. Figura reproduzida de Ferreira-Gonçalves e Brum-de-Paula (2013, p. 90).....	99
Figura 18 – Tipos de sonda compatíveis com o ultrassom <i>Mindray DP 6600</i> , da esquerda para a direita: microconvexa (35C20EA), endocavitária (65EC10EA) e microconvexa (65C15EA). Figura reproduzida de Ferreira-Gonçalves e Brum-de-Paula (2013, p. 101).....	100
Figura 19 – Modelo de capacete estabilizador dos movimentos da cabeça, utilizado para as coletas ultrassonográficas (Fonte: http://www.articulateinstruments.com) .	101

Figura 20 – Equipamentos utilizados para a coleta de áudio e ultrassom: computador (1), aparelho de ultrassom (2), sonda endocavitária (3), sonda microconvexa (4), sincronizador de áudio e vídeo (5), microfone unidirecional (6), gravador de áudio (7) e capacete estabilizador (8)	102
Figura 21 – Tela do <i>software</i> AAA em modo de gravação de vídeo de ultrassom, durante a produção da palavra “boa” por CR	103
Figura 22 – Imagem do teclado do aparelho <i>Mindray</i> DP 6600 com algumas funções sinalizadas: <i>Probe</i> (a); <i>Depth</i> – profundidade (b); <i>Freq.</i> – frequência (c); <i>Gain</i> – ganho (d); <i>IP</i> (e) e <i>F. position</i> – posição do foco (f)	104
Figura 23 – Exemplo de imagem ultrassonográfica da língua no plano sagital, obtida a partir de uma produção de CR. O dorso da língua está localizado à esquerda, enquanto a ponta da língua está localizada à direita	106
Figura 24 – Exemplo de imagem ultrassonográfica da língua no plano coronal (MÉNARD <i>et al.</i> , 2012, <i>apud</i> FRANCISCO, 2015, p. 18).....	107
Figura 25 – Procedimento de segmentação de palavras. Primeiro passo: abrir o áudio da coleta e selecionar um trecho em que ocorra a produção da palavra	112
Figura 26 – Procedimento de segmentação de palavras. Segundo passo: selecionar o botão “sel” ou utilizar o atalho Command+N/Ctrl+N para ampliar a tela na região selecionada	113
Figura 27 – Procedimento de segmentação de palavras. Terceiro passo: selecionar o trecho específico em que a palavra ocorre	114
Figura 28 – Palavra “prato”, produzida por CR, salva a partir do procedimento de segmentação.....	115
Figura 29 – Organização dos arquivos .wav das palavras segmentadas em pastas	116
Figura 30 – Arquivo concatenado contendo todas as produções da palavra “prato” de CR e sua respectiva <i>TextGrid</i>	117

Figura 31 – Arquivo concatenado contendo todas as produções da palavra “prato” de AD1 e sua respectiva <i>TextGrid</i>	117
Figura 32 – Arquivo concatenado contendo todas as produções da palavra “pato” de AD1 e sua respectiva <i>TextGrid</i>	119
Figura 33 – Dado de áudio produzido por CR, descartado por conter forte ruído ambiente.....	120
Figura 34 – Dado de áudio produzido por AD3, descartado por conter sobreposição de vozes.....	121
Figura 35 – Dado de áudio produzido por CR, descartado por sobreposição de barulhos à fala do informante	122
Figura 36 – Dado de áudio produzido por CR, descartado por conter som com características indefinidas na posição de rótico	123
Figura 37 – Exemplo de tela mostrando o processo de sincronização de dados de áudio e vídeo pelo programa AAA, reproduzido de Correa (2017, p. 94).....	126
Figura 38 – Exemplo de tela do programa AAA em modo de análise do vídeo de ultrassom, durante a produção da palavra “prato” por CR	127
Figura 39 – Produção de CR para a palavra “frota”, com ocorrência de rótico vibrante	148
Figura 40 – Produção de CR para a palavra “frita”, com ocorrência de rótico retroflexo	149
Figura 41 – Produção de CR para a palavra “fraca”, com ocorrência de rótico vibrante	149
Figura 42 – Produção de CR para a palavra “praça”, com <i>tap</i> de 51 ms	150
Figura 43 – Produção de CR para a palavra “frita”, com <i>tap</i> de 76 ms	151
Figura 44 – Movimentos de ponta de língua na realização do <i>tap</i> em “prata” (primeira repetição da coleta #6), produzido por CR. Em (a), início do movimento de elevação	

da ponta da língua; em (b), movimento de máxima constrição. Da direita para a esquerda nas imagens, partes anterior e posterior da língua196

Figura 45 – Movimento de máxima constrição de ponta de língua na realização do *tap* em prata (primeira repetição) por AD2. Da direita para a esquerda nas imagens, partes anterior e posterior da língua.....196

Figura 46 – Movimentos de ponta de língua na realização do segmento retroflexo em prata (segunda repetição da coleta #6), produzido por CR. Em (a), início do movimento de elevação da ponta da língua; em (b), movimento de máxima elevação da ponta; em (c), início da retração do dorso; em (d), retração do dorso e encurvamento da ponta. Da direita para a esquerda nas imagens, partes anterior e posterior da língua.....197

Figura 47 – Movimentos reiterados de ponta de língua na realização da vibrante em preso1, quarta coleta, produzido por CR. Em (a), primeiro movimento de elevação da ponta da língua; em (b), leve abaixamento da ponta para a produção do segundo elemento vocálico; em (c), segundo movimento de elevação da ponta. Da direita para a esquerda nas imagens, parte anterior e posterior da língua198

Figura 48 – Imagens acústicas do software AAA do elemento vocálico presente em duas ocorrências de “prata” (primeira e terceira repetições da coleta #6) de CR. Em (a), elemento vocálico mais longo, antecedendo a produção do *tap*; em (b), elemento vocálico mais curto antecedendo a produção do *tap*. Da direita para a esquerda nas imagens, partes anterior e posterior da língua199

Figura 49 – Teste-t do gesto de máxima constrição do elemento vocálico e do gesto de máxima constrição da vogal nuclear nas três ocorrências de “prata” da coleta #6, produzidas por CR. À direita do gráfico, parte anterior do trato vocal; à esquerda, parte posterior do trato vocal.....200

Lista de Gráficos

Gráfico 1 – Médias de duração relativa das obstruintes [p], [b] e [f], em sílabas CCV e CV, produzidas por AD1, AD2 e AD3	132
Gráfico 2 – Médias de duração relativa das vogais nos contextos de [p], [b] e [f], em sílabas CCV e CV, produzidas por AD1, AD2 e AD3	133
Gráfico 3 – Médias de duração relativa, em relação à sílaba, de vogal em sílabas CV, de elemento vocálico + vogal núcleo e de elemento vocálico + rótico + vogal núcleo em sílabas CCV, nos contextos de [p], [b] e [f], produzidas pelas adultas...	140
Gráfico 4 – Médias de duração relativa, em relação à sílaba, de vogal em sílabas CV, de elemento vocálico + vogal núcleo e de elemento vocálico + rótico + vogal núcleo em sílabas CCV, nos contextos de [p], [b] e [f], produzidas por AD1, AD2 e AD3	141
Gráfico 5 – Percentuais de produção da estrutura CCV na fala de CR, dispostos por idade	146
Gráfico 6 – Percentual de segmentos produzidos em C ₂ para alvos CCV	147
Gráfico 7 – Médias de duração absoluta de sílaba em contexto de [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	153
Gráfico 8 – Médias de duração relativa de sílaba em contexto de [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	153
Gráfico 9 – Médias de duração absoluta de sílaba em contexto de [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	155

Gráfico 10 – Médias de duração relativa de sílaba em contexto de [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	155
Gráfico 11 – Médias de duração absoluta de sílaba em contexto de [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	156
Gráfico 12 – Médias de duração relativa de sílaba em contexto de [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	157
Gráfico 13 – Médias de duração absoluta da obstruinte [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	160
Gráfico 14 – Médias de duração relativa da obstruinte [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	160
Gráfico 15 – Médias de duração absoluta da obstruinte [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	161
Gráfico 16 – Médias de duração relativa da obstruinte [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	162
Gráfico 4.15 – Médias de duração absoluta da obstruinte [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	163
Gráfico 18 – Médias de duração relativa da obstruinte [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	164
Gráfico 19 – Médias de duração absoluta de vogal em contexto de [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	169
Gráfico 20 – Médias de duração relativa de vogal em contexto de [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	169

Gráfico 21 – Médias de duração absoluta de vogal em contexto de [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	171
Gráfico 22 – Médias de duração relativa de vogal em contexto de [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	171
Gráfico 23 – Médias de duração absoluta de vogal em contexto de [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	173
Gráfico 24 – Médias de duração relativa de vogal em contexto de [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)	173
Gráfico 25 – Médias de duração absoluta de elemento vocálico, nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)	180
Gráfico 26 – Médias de duração relativa do elemento vocálico em relação à sílaba nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)	181
Gráfico 27 – Médias de duração relativa do elemento vocálico em relação à vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)	182
Gráfico 28 – Médias de duração absoluta do <i>tap</i> nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)	185
Gráfico 29 – Médias de duração relativa do <i>tap</i> nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)	185
Gráfico 30 – Médias de duração absoluta da soma de elemento vocálico e vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)	188
Gráfico 31 – Médias de duração relativa da soma de elemento vocálico e vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)	189

Gráfico 32 – Médias de duração absoluta da soma de elemento vocálico, *tap* e vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)192

Gráfico 33 – Médias de duração relativa da soma de elemento vocálico, *tap* e vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)193

Lista de Quadros

Quadro 1 – Conjunto de variáveis do trato e respectivos articuladores associados, adaptado de Browman e Goldstein (1989, p. 73)	44
Quadro 2 – Conjuntos de articuladores e parâmetros dinâmicos associados, conforme proposta de Browman e Goldstein (1989, p. 74)	45
Quadro 3 – Parâmetros dinâmicos e descritores gestuais associados, conforme proposta de Browman e Goldstein (1989, p. 75)	46
Quadro 4 – Aquisição dos tipos de <i>onset</i> complexo nos dados de G., informante da pesquisa. Quadro adaptado de Bonilha (2005, p. 151)	78
Quadro 5 – Produções incomuns como estratégias de realização de sílabas CCV. Quadro adaptado de Ribas (2002, p. 36)	80
Quadro 6 – Características da criança informante da pesquisa e distribuição de coletas longitudinais realizadas	91
Quadro 7 – Grupo de adultas participantes da pesquisa e coletas transversais realizadas	93
Quadro 8 – Conjuntos de palavras utilizados por Miranda (2007) e Vassoler (2016) para as coletas de dados	95
Quadro 9 – <i>Corpus</i> elaborado para as coletas, com base nos <i>corpora</i> de Miranda (2007) e Vassoler (2016)	97
Quadro 10 – Número de produções por grupo nas coletas de dados acústicos e articulatórios	108
Quadro 11 – Médias de duração da palavra e da sílaba, e média de duração e desvio padrão da obstruinte e da vogal em sílabas CCV e CV, classificadas por tipo de obstruinte em posição de C ₁ . As médias foram calculadas com base em dados de todas as informantes adultas – AD1, AD2 e AD3	130

Quadro 12 – Médias de duração da obstruinte e da vogal em sílabas CCV e CV, classificadas por tipo de obstruinte em posição de C ₁ , calculadas com base em dados da informante AD1.....	131
Quadro 13 – Médias de duração da obstruinte e da vogal em sílabas CCV e CV, classificadas por tipo de obstruinte em posição de C ₁ , calculadas com base em dados da informante AD2.....	131
Quadro 14 – Médias de duração da obstruinte e da vogal em sílabas CCV e CV, classificadas por tipo de obstruinte em posição de C ₁ , calculadas com base em dados da informante AD3.....	131
Quadro 15 – Diferenças significativas das durações – absoluta e relativa – de obstruintes e vogais, quando comparadas as sílabas CCV e CV	134
Quadro 16 – Médias de duração e desvio padrão do elemento vocálico e do rótico em sílabas CCV, classificadas por tipo de obstruinte em posição de primeira consoante da sílaba, calculadas com base em dados de todas as informantes adultas – AD1, AD2, e AD3.....	135
Quadro 17 – Médias de duração do elemento vocálico e do rótico em sílabas CCV, classificadas por tipo de obstruinte em posição de C ₁ , calculadas com base em dados da informante AD1.....	135
Quadro 18 – Médias de duração do elemento vocálico e do rótico em sílabas CCV, classificadas por tipo de obstruinte em posição de C ₁ , calculadas com base em dados da informante AD2.....	136
Quadro 19 – Médias de duração do elemento vocálico e do rótico em sílabas CCV, classificadas por tipo de obstruinte em posição de C ₁ , calculadas com base em dados da informante AD3.....	136
Quadro 20 – Diferenças significativas das durações – absoluta e relativa – do elemento vocálico (EV) e do rótico, quando comparadas em contextos de [p], [b] e [f]	137

Quadro 21 – Médias de duração dos trechos de elemento vocálico + vogal e elemento vocálico + <i>tap</i> + vogal, classificadas por tipo de obstruente em posição de C ₁ , calculadas com base em dados de AD1, AD2 e AD3.....	138
Quadro 22 – Médias de duração dos trechos de elemento vocálico + vogal e elemento vocálico + <i>tap</i> + vogal, classificadas por tipo de obstruente em posição de C ₁ , calculadas com base em dados de AD1	138
Quadro 23 – Médias de duração dos trechos de elemento vocálico + vogal e elemento vocálico + <i>tap</i> + vogal, classificadas por tipo de obstruente em posição de C ₁ , calculadas com base em dados de AD2	138
Quadro 24 – Médias de duração dos trechos de elemento vocálico + vogal e elemento vocálico + <i>tap</i> + vogal, classificadas por tipo de obstruente em posição de C ₁ , calculadas com base em dados de AD3	139
Quadro 25 – Diferenças significativas das durações – absoluta e relativa – das sequências elemento vocálico + vogal nuclear (EV+V) e elemento vocálico + rótico + vogal nuclear (EV+R+V), quando comparadas em contextos de [p], [b] e [f]	140
Quadro 26 – Diferenças significativas das durações – absoluta e relativa – das sequências elemento vocálico + vogal nuclear (EV+V) e elemento vocálico + rótico + vogal nuclear (EV+R+V), quando comparadas à duração da vogal nuclear em sílaba CV, contextos de [p], [b] e [f]	142
Quadro 27 – CR: quantidade de dados coletados, descartados e submetidos à análise acústica.....	144
Quadro 28 – Possibilidades de produção e ocorrências de sílaba CCV dispostas por idade (coletas #1 a #8) e por contexto consonantal de C ₁	145
Quadro 29 – Tipos de segmento produzidos em posição de C ₂ para alvos CCV ...	147
Quadro 30 – Médias de duração de sílaba para as produções da criança (CR), dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs).....	152
Quadro 31 – Médias de duração de obstruente para as produções da criança (CR), dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs).....	159

Quadro 32 – Diferenças significativas das durações – absoluta e relativa – de obstruintes, quando comparadas as duas primeiras coletas da criança com as coletas subsequentes.....	166
Quadro 33 – Médias de duração da vogal para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs).....	168
Quadro 34 – Diferenças significativas das durações – absoluta e relativa – das vogais quando comparados os resultados da criança com os resultados das adultas	175
Quadro 35 – Diferenças significativas das durações – absoluta e relativa – das vogais quando comparados os resultados das duas primeiras coletas da criança com os resultados das adultas.....	176
Quadro 36 – Diferenças significativas das durações – absoluta e relativa – das vogais quando comparados os resultados das coletas 3 a 8 da criança com os resultados das adultas	176
Quadro 37 – Diferenças significativas das durações – absoluta e relativa – das vogais quando comparados os resultados das coletas iniciais e das coletas subsequentes da criança	177
Quadro 38 – Diferenças significativas das durações – absoluta e relativa – de vogais, quando comparadas as sílabas CCV e CV, nas produções das coletas 3 a 8 da criança.....	178
Quadro 39 – Médias de duração do elemento vocálico para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs).....	179
Quadro 40 – Diferenças significativas das durações – absoluta e relativa – do elemento vocálico quando comparados os resultados das adultas e da criança	183
Quadro 41 – Médias de duração do <i>tap</i> para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs).....	184
Quadro 42 – Diferenças significativas da duração do <i>tap</i> quando comparados os resultados das adultas e da criança	186

Quadro 43 – Médias de duração da soma de elemento vocálico e vogal para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs).....	187
Quadro 44 – Diferenças significativas das durações – absoluta e relativa – da sequência elemento vocálico + vogal nuclear, quando comparados os dados da criança e das adultas nos contextos de [p], [b] e [f].....	190
Quadro 45 – Médias de duração da soma de elemento vocálico, <i>tap</i> e vogal para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs).....	191
Quadro 46 – Diferenças significativas das durações – absoluta e relativa – da sequência elemento vocálico + tap + vogal nuclear, quando comparados os dados da criança e das adultas nos contextos de [p], [b] e [f].....	194

Lista de Abreviaturas e Siglas

AAA	<i>Articulate Assistant Advanced</i>
C ₁	primeira consoante do encontro consonantal
C ₂	segunda consoante do encontro consonantal
CCV	consoante + consoante + vogal
CV	consoante + vogal
<i>e.g.</i>	<i>exempli gratia</i> – “por exemplo”
EV	elemento vocálico
FAR	Fonologia Articulatória
FonGest	Fonologia Gestual
fps	<i>frames</i> por segundo
Hz	hertz
<i>i.e.</i>	<i>id est</i> – “isto é”
mm	milímetro(s)
ms	milissegundo(s)
PB	Português Brasileiro
pp	ponto(s) percentual(is)
qtd.	quantidade
TO	Teoria da Otimidade

Sumário

1 Introdução.....	32
2 Referencial teórico	41
2.1 A Fonologia Articulatória	41
2.1.1 O modelo teórico de base	41
2.1.2 A sílaba CCV na perspectiva da Fonologia Articulatória	50
2.1.3 O <i>tap</i> na perspectiva da Fonologia Articulatória	59
2.1.4 Erros de fala: o que podem evidenciar sobre a coordenação gestual	66
2.2 Estudos em aquisição da sílaba CCV	75
2.2.1 Estudos em aquisição da sílaba CCV com análises de outiva	76
2.2.2 Estudos em aquisição da sílaba CCV com análises acústicas.....	81
2.2.3 Estudos em aquisição da sílaba CCV com análises ultrassonográficas	86
3 Metodologia	89
3.1 Sujeitos.....	89
3.1.1 A criança	89
3.1.2 As adultas.....	92
3.2 As coletas de dados	94
3.2.1 O <i>corpus</i> e o instrumento de coleta.....	95
3.2.2 Procedimentos de coleta	98
3.3 Procedimentos de organização e análise dos dados	109
3.3.1 Tratamento e análise dos dados acústicos	111

3.3.2 Tratamento e análise dos dados articulatórios	125
4 Resultados e discussão.....	128
4.1 Análise acústica	128
4.1.1 As adultas: AD1, AD2 e AD3.....	129
4.1.2 A criança: CR	144
4.1.2.1 Emergência da estrutura silábica CCV.....	145
4.1.2.2 Durações de sílaba.....	152
4.1.2.3 Durações de obstruinte	158
4.1.2.4 Durações de vogal.....	167
4.1.2.5 Durações de elemento vocálico	178
4.1.2.6 Durações de rótico	184
4.1.2.7 Durações de elemento vocálico + vogal	187
4.1.2.8 Durações de elemento vocálico + <i>tap</i> + vogal.....	191
4.2 Inspeção articulatória qualitativa	195
4.2.1 Os gestos articulatórios intrínsecos à produção do rótico	195
4.2.2 A coordenação gestual: o rótico e o elemento vocálico.....	199
4.3 Fonologia Articulatória: uma discussão dos dados à luz da teoria	201
5 Considerações finais	206
Referências	211
Apêndices	216
Anexos	232

1 Introdução

Este trabalho investiga o processo de aquisição de sílabas CCV, constituídas por obstruente e *tap*, a partir da análise acústica e articulatória de produções de uma criança em processo de aquisição do português brasileiro (PB). Busca-se, por meio dessa análise, formalizar a gramática da criança no que concerne à emergência da estrutura silábica CCV, com base em pressupostos da Fonologia Articulatória, modelo teórico proposto por Browman e Goldstein (1986; 1989; 1992).

O tema – aquisição de CCV no português – tem seu percurso de investigação e considerável literatura na área. Em pesquisas como Lamprecht (1990) e Bonilha (2005), por exemplo, essa questão se encontra inserida num escopo maior; nesses trabalhos, a aquisição fonológica do português é abordada de maneira ampla, investigando-se a aquisição de várias estruturas e segmentos. Esses estudos partem de dados longitudinais e fornecem resultados sobre a aquisição do *onset* complexo. Essa estrutura, por sua vez, foi o tema específico de trabalhos como os de Ribas (2002; 2004), nos quais se descreve o processo de aquisição do *onset* complexo. Nos trabalhos da autora, dá-se especial atenção à quantificação e descrição de variadas estratégias de reparo adotadas pelas crianças nesse percurso.

Tais estudos são relevantes por terem contribuído para um delineamento do processo de aquisição de sons e estruturas do português, evidenciando produções diferenciadas da criança e os momentos em que a aquisição de determinados segmentos ou estruturas poderia se considerar completa. Entretanto, cabe lembrar que as metodologias empregadas nesses trabalhos envolvem transcrição e análise de dados de fala feitos somente por meio de um julgamento de ouvinte. Os modelos teóricos que sustentam esses trabalhos, por sua vez, adotam traços distintivos e segmentos como unidades representacionais da fonologia. A representação fonológica é concebida, portanto, como categórica. Assim, nesses estudos, o processo de aquisição terá gradualidade, fundamentalmente, no número (ou porcentagem) de produções corretas no transcorrer do processo. Esse número aumenta à medida que a criança se encaminha para o que se chamaria de uma aquisição completa.

Estudos mais recentes, como os de Miranda (2007) e Miranda e Silva (2011), amparam-se em modelos multirrepresentacionais, a saber, a Fonologia de Uso, de

Bybee (2000), e a Teoria de Exemplos, de Pierrehumbert (2001). Miranda (2007) e Miranda e Silva (2011) propõem estágios intermediários de aquisição para a estrutura CCV que revelam a existência de contrastes encobertos nas produções das crianças. Para a descrição desses contrastes, que não se detectam por meio da outiva, a análise acústica foi importante procedimento metodológico, visto que os sons das línguas têm características acústicas particulares. Conhecendo-se as características acústicas dos sons que se pretende investigar, é possível, assim, interpretar as representações visuais da onda sonora, apresentadas em forma de oscilograma e espectrograma. Com esse conhecimento, o pesquisador pode, por exemplo, verificar a presença de um dado som na produção e, assim, extrair seus valores formânticos, calcular sua duração ou detectar mudanças na sua realização.

O aspecto da duração, em especial, foi encontrado nos trabalhos de Miranda (2007) e Miranda e Silva (2011) como um indicativo de uma distinção estabelecida pelas crianças em processo de aquisição da líquida não lateral (ou *tap*). Isso porque, em palavras com sílabas CCV produzidas pelas crianças, não se detectou a realização do *tap*, mas, em contrapartida, houve produção de vogais mais longas em comparação a uma palavra com CV em relação de par mínimo: por exemplo, na palavra “bruxa” ([’bru.ʃa]), a vogal [u] era produzida com maior duração do que na sua contraparte CV, “bucha” ([’bu.ʃa]). A observação desse fenômeno nos dados de crianças que não produziam o encontro consonantal evidenciou que já havia uma distinção sendo feita entre as estruturas CCV e CV, embora esse não fosse, ainda, o contraste normalmente estabelecido na comunidade de fala.

Mezzomo *et al.* (2008) também investigam o uso de uma estratégia de alongamento compensatório no processo de aquisição de encontros consonantais, com base em produções de crianças com desenvolvimento fonológico normal e com desvios. Ao compararem os dados analisados somente por outiva aos mesmos dados analisados acusticamente, ressaltaram a importância de se incluir a acústica na interpretação e descrição de dados, uma vez que os resultados podem diferir consideravelmente. Ainda, apontam a necessidade dessa metodologia para que se obtenham descrições mais fidedignas aos fatos da língua.

Quanto à estratégia de alongamento compensatório, Miranda e Silva (2011) consideram, ainda, que ela pode propiciar condições físicas para que a criança, em algum momento, possa incorporar a produção de um *tap*. Ao produzir a vogal mais longa, a criança criaria, assim, um ambiente para a produção desse segmento e,

eventualmente, com o aprimoramento das rotinas motoras, ela estaria apta a produzir encontros consonantais em início de sílaba. Traz-se, assim, uma interpretação adicional sobre o alongamento: é um fenômeno verificado acusticamente, que pode ser não só um indício de uma distinção estabelecida pela criança, mas uma preparação para rotinas motoras futuras.

A partir dessas considerações, a inspeção acústica foi adotada como procedimento metodológico neste trabalho, a fim de se investigar, dentre outros aspectos, a duração segmental enquanto possível reveladora de contrastes encobertos e de avanços na construção de rotinas articulatórias necessárias para a produção de sequências CCV. Assim, o trabalho vem se somar aos estudos em aquisição fonético-fonológica que utilizam a inspeção acústica como parte da sua metodologia. Também, pretende-se investigar a aquisição de encontros consonantais com o recurso da análise articulatória, a partir de imagens de ultrassom.

A ultrassonografia tem sido utilizada como meio de coleta e análise de dados de fala desde a década de 60. Os trabalhos produzidos no Brasil utilizando essa ferramenta ainda são bastante recentes, mas já trouxeram contribuições importantes para a compreensão de fenômenos acerca da aquisição fonético-fonológica.

O aparelho de ultrassom possibilita a análise de movimentos da língua em tempo real. Em comparação a outros métodos utilizados para uma descrição articulatória, como a ressonância magnética (MRI) e a eletropalatografia (EPG), apresenta as vantagens de ser um procedimento seguro e não invasivo, como apontado por Bressmann (2008). A forma de coleta de dados não é invasiva, pois não exige a inserção de aparelhos no interior do trato vocal para que se obtenham as imagens. Isso faz com que não haja alterações na produção da fala.

A ultrassonografia pode ser aplicada a estudos em aquisição de língua materna, mas também pode ser utilizada em atividades de instrução explícita que envolvam sons da língua estrangeira. Também apresenta contribuições à área da fonoaudiologia, pois a observação dos movimentos da língua auxilia na condução de terapias mais eficazes para diferentes casos de aquisição atípica.

Os estudos em aquisição fonológica de língua materna no Brasil envolvendo a análise da fala com o uso da ultrassonografia são recentes, como apontado por Ferreira-Gonçalves e Brum-de-Paula (2013), mas já é possível encontrarem-se investigações acerca de diferentes classes de segmentos, como plosivas (MELO,

2016), fricativas (FRANCISCO, 2015) e líquidas (BARBERENA, 2016). Também há estudos sobre a aquisição na língua estrangeira, como os de Dias-Cavaleiro (2016), Correa (2017), Correa, Ferreira-Gonçalves e Brum-de-Paula (2017) e Pereira e Ferreira-Gonçalves (2018).

A análise articulatória por ultrassom, nas pesquisas em fonética e fonologia, complementa a análise acústica, na qual se pode apenas inferir o que é produzido ao se estabelecerem correlações entre pistas visuais e aspectos acústicos, como anterioridade, posterioridade, elevação e abaixamento da língua, por exemplo. Nos trabalhos sobre a aquisição de CCV, a acústica tem sido utilizada, fundamentalmente, para medidas de duração compensatória; já com a análise articulatória, podem-se observar os movimentos da língua, o que possibilita estudos sobre a coordenação gestual na produção da sílaba. O ultrassom figura, assim, como uma ferramenta pertinente para os estudos em aquisição fonético-fonológica na língua materna, tanto de ordem típica quanto atípica.

Ainda, com esta pesquisa, busca-se dar continuidade a estudos com a ultrassonografia voltados especificamente para a aquisição de encontros consonantais, os quais são poucos e de recente publicação. Em aquisição de encontros do português brasileiro, há o estudo de Vassoler (2016), no qual se investiga, de forma comparativa, a aquisição da sílaba CCV em dados transversais de crianças com desenvolvimento típico e atípico. Como objetivo geral, a autora busca investigar e descrever, à luz da Fonologia Gestual, os padrões de coordenação gestual envolvidos na produção de sílabas CCV e CV de crianças com essas diferentes condições clínicas.

Por ser um estudo comparativo envolvendo diferentes condições clínicas, os objetivos específicos da tese de Vassoler (2016) se direcionam para dar conta de possíveis diferenças ou semelhanças, tendo em vista esses dois perfis. Esses objetivos são: caracterizar diferenças na coordenação gestual de CCV em função da condição clínica das crianças e verificar em que medida as chamadas reduções de encontros consonantais por crianças com desenvolvimento fonológico atípico podem ser consideradas contrastes encobertos. O segundo objetivo também vai, de certa forma, em direção a uma comparação, já que a redução do encontro consonantal, ou a sua produção sem o segundo elemento, já foi interpretada como evidência de um contraste encoberto na produção de crianças sem desvios (MIRANDA; SILVA, 2011).

A presente pesquisa se baseia na descrição e análise de dados de fala de uma criança em processo de aquisição que foi observado como sendo típico¹. Buscou-se dar ênfase ao processo de construção de uma gramática fonético-fonológica pela criança, propondo-se uma formalização dessa gramática; para tanto, foi importante uma análise com base em dados longitudinais.

Além dos dados de fala infantil, também foram analisados dados de fala de três sujeitos adultos, de caráter transversal. Esses dados não constituem o objeto de estudo desta pesquisa; no entanto, são de especial relevância, pois servem como parâmetro para comparações quanto a fenômenos de duração e coordenação gestual. Isso significa que, para que seja possível aferir um fenômeno como próprio da fala da criança em aquisição, a comparação de dados de fala infantil com dados de fala adulta se faz necessária.

A partir dessas considerações, foram elencadas, para esta pesquisa, as seguintes questões norteadoras:

- i) A duração de vogais e obstruintes tem diferenças significativas na produção de sílabas CCV e CV na fala da criança?
- ii) Se a duração de vogais e obstruintes difere significativamente de acordo com o tipo de sílaba, essa diferença caracteriza apenas um aspecto da fala da criança, constituindo um contraste encoberto, ou também é própria da fala adulta?
- iii) Como se dá a coordenação gestual em sílabas CCV e CV no processo de aquisição fonético-fonológica pela criança?
- iv) Que gestos articulatórios compõem o *tap* e qual o *status* desses gestos, considerando diferentes propostas encontradas na literatura (SILVA, 2002; BARBERENA, 2016; RECASENS, 2016)?

¹ O perfil de aquisição da criança informante desta pesquisa é considerado típico, embora o informante não tenha passado por uma avaliação fonoaudiológica, tendo em vista a ausência desse tipo de serviço na Instituição de Ensino Superior que sedia a presente pesquisa. No entanto, a questão do perfil de aquisição foi pensada e discutida à luz de dados de fala da criança, produzidos antes da coleta e durante a sua realização. Não foram observadas, nesses dados, trocas recorrentes de sonoridade ou de ponto de articulação, por exemplo. A única troca relativa à sonoridade ocorrida na fala da criança foi na palavra “grato”, presente na primeira versão do *corpus* do experimento. Nessa palavra, o [g] era ocasionalmente realizado pela criança como [k]. “Grato” não era uma palavra conhecida pela criança até a realização da primeira coleta, o que pode justificar sua realização sem o domínio completo da sonoridade da consoante inicial. Tendo-se apenas essa exceção, decidiu-se considerar o perfil de aquisição da criança como típico, estando as produções dos sons do português adequadas ao esperado para a idade (LAMPRECHT *et al.*, 2004).

- v) Nas sílabas CCV produzidas pela criança, ocorre sobreposição de gestos do *tap* aos gestos da vogal núcleo da sílaba (SILVA; CLEMENTE; NISHIDA, 2006)?
- vi) Na análise articulatória das produções de sílabas CCV e CV, revelam-se contrastes encobertos outros, não resgatados pela inspeção acústica?

Com base nessas questões, definiu-se o objetivo geral do trabalho: descrever e analisar, acústica e articulatoriamente, a emergência de sílabas CCV constituídas por obstruinte e *tap* na fala de uma criança em processo de aquisição do português brasileiro (PB), com base nos pressupostos da Fonologia Articulatória (FAR). Também foram definidos os seguintes objetivos específicos:

- i) Verificar, por meio de análise acústica, nos dados de fala da criança, se a duração de vogais difere quanto ao tipo de sílaba – CCV ou CV –, havendo, em CCV, um alongamento vocálico que funcionaria como possível mecanismo de distinção entre sílabas na fala dessa criança;
- ii) Verificar, por meio de análise acústica, nos dados de fala da criança, se a duração de obstruintes difere quanto ao tipo de sílaba – CCV ou CV –, havendo, em CCV, uma aspiração de obstruinte que funcionaria como possível mecanismo de distinção entre sílabas na fala dessa criança;
- iii) Descrever e analisar, por meio das imagens de ultrassom dos dados de fala da criança, padrões de coordenação gestual envolvidos na produção de sílabas CCV em comparação a sílabas CV;
- iv) Descrever e analisar, por meio das imagens de ultrassom dos dados de fala da criança, padrões de coordenação gestual envolvidos na produção do segundo elemento consonantal, o *tap*, com base em propostas encontradas na literatura;
- v) Verificar e descrever possíveis complementações entre dados acústicos e articulatórios, no que respeita à presença de contrastes encobertos, com base em diferenças na duração segmental e/ou na sobreposição de gestos.

A partir das questões e objetivos elencados, foram elaboradas as seguintes hipóteses:

- H1) Na fala da criança, a duração, tanto de vogais quanto de obstruintes, terá diferenças significativas conforme o tipo de sílaba produzido, sendo essas durações maiores em CV do que em CCV. Isso se deveria a um processo de ajuste de duração que ocorreria no processo de aquisição até que a criança se aproprie do *tap* e das rotinas motoras para a realização de CCV, ajustando, assim, a obstruinte e a vogal temporalmente.
- H2) A diferença de duração será um aspecto próprio da fala infantil em processo de aquisição, partindo-se da ideia de que a criança passa por um processo de ajuste de durações, como reportado em H1. Assim, no processo de aquisição de CCV, a duração será um aspecto relevante para que a criança marque a estrutura silábica diferente, o que se verificará acusticamente.
- H3) Quando a criança começar a dominar o gesto característico do *tap*, as produções apresentarão gesto de ponta que aumentará em magnitude ao longo das coletas, o que se verificará pela análise articulatória.
- H4) O *tap* não é um segmento complexo, sendo composto apenas por um gesto de ponta. Sua aquisição tardia decorre não da dificuldade de conciliação de dois gestos, mas da sua coordenação temporal.
- H5) Nas sílabas CCV produzidas pela criança, ocorrerá sobreposição de gestos do *tap* e da vogal da sílaba, assim como na fala adulta em relação aos demais gestos que o circundam, especialmente o vocálico.
- H6) Na análise articulatória, serão encontrados contrastes outros, difíceis de serem resgatados ou inferidos pela inspeção acústica.

Tendo-se em vista as questões de pesquisa, entende-se que a aquisição da estrutura CCV seja um fenômeno de aquisição fonético-fonológica que careça de mais descrição, contemplando-se não apenas as presenças, ausências e substituições no nível segmental que se verificam na fala infantil. Os trabalhos de Miranda (2007), Miranda e Silva (2011) e Vassoler (2016) são de grande relevância e fazem considerações inovadoras do ponto de vista do conhecimento fonológico e das rotinas motoras da criança.

De todo modo, ainda há caminho a ser percorrido, no sentido de dar-se conta de uma construção gradual da estrutura CCV no sistema fonético-fonológico da criança. A coleta e a análise de dados longitudinais são procedimentos importantes para se atingir esse objetivo, além de inovarem em relação à análise de dados transversais, encontrada, em geral, nas metodologias de trabalhos sobre aquisição de CCV no português brasileiro. A inspeção acústica e a análise articulatória, por meio de imagens de ultrassom, também serão procedimentos importantes para que se atinja esse objetivo.

Espera-se, com esta pesquisa, contribuir para a compreensão acerca da aquisição de encontros consonantais com o uso da ultrassonografia, ferramenta em crescente utilização nos estudos da fala conduzidos no Brasil. Espera-se, também, contribuir para um aprimoramento da descrição do processo de aquisição típica do português brasileiro a partir de uma análise de caráter longitudinal.

Esta dissertação se organiza em cinco capítulos, a contar com esta introdução. O segundo capítulo, a seguir, é dedicado ao referencial teórico adotado na pesquisa. Nele, é discutido o modelo teórico que embasará o desenvolvimento desta pesquisa e é feita uma revisão de literatura acerca de estudos sobre a estrutura CCV e sobre o *tap*, em termos de sua descrição e aquisição.

O terceiro capítulo é dedicado à metodologia empregada neste trabalho. Nele, o leitor encontrará uma descrição dos sujeitos da pesquisa, dos procedimentos de seleção desses sujeitos, da seleção e composição do *corpus* e dos procedimentos de coleta e tratamento de dados de fala.

O quarto capítulo descreve os resultados encontrados a partir das análises acústica e articulatória e traz uma discussão acerca desses resultados, enquanto o quinto capítulo é reservado às considerações finais deste estudo. Ao final desta

dissertação, estão as referências bibliográficas, apêndices – materiais produzidos para a pesquisa, de autoria da pesquisadora – e anexos – materiais produzidos por outros autores que foram utilizados na pesquisa.

2 Referencial teórico

Este capítulo é dedicado ao aporte teórico que fundamenta a presente pesquisa, dividindo-se em três subseções.

Na subseção 2.1, será feita uma discussão acerca do modelo teórico que serve de base para esta pesquisa: a Fonologia Articulatória. A seguir, em 2.2, faz-se uma revisão de literatura sobre estudos em aquisição fonológica envolvendo a estrutura CCV, seja como um dos vários aspectos fonético-fonológicos investigados ou como tema específico de pesquisa, a partir de diferentes modelos teóricos e metodologias. Por fim, em 2.3, é apresentado e discutido um estudo em aquisição da sílaba CCV no português brasileiro que emprega a análise articulatória por meio de imagens de ultrassom.

2.1 A Fonologia Articulatória

Esta seção é dedicada ao referencial teórico relacionado à Fonologia Articulatória. Na seção 2.1.1, será abordado o modelo teórico de base, com seus principais pressupostos e conceitos. A seção 2.1.2 trata da sílaba CCV na perspectiva da Fonologia Articulatória. A seção 2.1.3 aborda o *tap* e propostas de sua descrição conforme esse modelo teórico e, por fim, a seção 2.1.4 discute o papel dos erros de fala no processo de aquisição, pois, na perspectiva da Fonologia Articulatória, esses podem evidenciar a construção de rotinas motoras.

A seguir, a seção 2.1.1 traz informações principais sobre a Fonologia Articulatória.

2.1.1 O modelo teórico de base

Esta pesquisa tem como modelo teórico de base a Fonologia Articulatória, proposta por Browman e Goldstein (1986; 1987; 1988; 1989; 1992). Enquanto, em modelos de análise fonológica clássicos, as unidades básicas de análise são fonemas ou traços distintivos, na Fonologia Articulatória, o primitivo de análise é o *gesto articulatório*. Este é uma unidade dotada de uma parte física, de ordem motora, mas também de uma parte abstrata, de ordem representacional. Em outras

palavras, o gesto é uma unidade de ação, sendo o resultado do trabalho conjunto de articuladores, mas também é uma unidade de informação, já que a diferenciação de significados em uma língua pode se dar pela presença ou ausência de um gesto, ou também por diferenças quanto às partes do trato vocal envolvidas na produção.

Browman e Goldstein (1989) se perguntavam quais seriam os “átomos” a partir dos quais as estruturas fonológicas se formavam. Na perspectiva teórica elaborada e aperfeiçoada pelos autores, esses átomos seriam os gestos, ações empreendidas pelos articuladores do trato vocal. Na sua instância motora, o gesto seria descrito como a formação (e liberação) de uma constrição em algum dos subsistemas do trato vocal – por exemplo, o subsistema oral, laríngeo ou vélico. No subsistema oral, as constrições poderiam ser formadas pela ação de conjuntos de articuladores relativamente independentes – os lábios, a ponta ou lâmina da língua e o corpo da língua. Quanto ao seu tempo de realização, os gestos têm, como ações, um tempo intrínseco a eles associado. São, para Browman e Goldstein (1989, p. 69, tradução nossa²), “caracterizações de movimentos através do espaço e ao longo do tempo”. As estruturas fonológicas seriam “constelações” estáveis formadas a partir dos gestos; logo, conforme esse ponto de vista, pode-se compreender que os sons da fala são coordenações ou composições de gestos.

Uma característica que diferencia a Fonologia Articulatória de outros modelos teóricos de forma acentuada é o *status* do primitivo de análise da teoria no que diz respeito ao processo de aquisição de uma língua. Para Browman e Goldstein (1989), os gestos são, em princípio, unidades pré-linguísticas, que têm existência e função antes mesmo da emergência de uma gramática. Com base em autores que se referem a esses gestos como “rotinas articulatórias”, assumem que essas unidades emergem no repertório comportamental da criança antes mesmo de qualquer desenvolvimento especificamente linguístico. A partir dessa assunção, propõem que os gestos são unidades de ação, em princípio não linguísticas, que podem ser comandadas por um sistema fonológico em desenvolvimento para se tornarem as unidades básicas de estruturas fonológicas. A evidência de que gestos seriam unidades de ação pré-linguísticas poderia ser vista, por exemplo, no balbúcio de crianças pequenas. A emergência dessas ações poderia ser vista mais como algo

² No original: “characterizations of movements through space and over time”.

que aconteceria em função de desenvolvimentos anatômicos e neurofisiológicos do que como o indício do começo de uma aquisição fonológica em si.

No modelo desenvolvido pelos autores, o gesto é uma caracterização abstrata a partir de movimentos de articuladores do trato vocal, os quais se coordenam de determinada forma de modo a cumprir uma tarefa. Essa tarefa, que seria a formação de uma constrição específica, é cumprida à medida que os movimentos dos articuladores se desdobram no tempo.

Cada gesto é especificado em termos das *variáveis do trato* envolvidas na sua realização. Gestos vélicos, por exemplo, envolveriam apenas a variável de abertura vélica, assim como gestos glotais envolveriam apenas a variável de abertura glotal. Já gestos orais envolvem pares de variáveis do trato que especificam a constrição realizada em termos do seu local e grau.

Gestos são especificados em termos de variáveis do trato e cada variável, por sua vez, está associada a *conjuntos de articuladores* específicos, cujos movimentos determinarão o valor daquela variável. Browman e Goldstein (1989) exemplificam com a variável da abertura labial (*LA – Lip Aperture*), que corresponde à distância vertical entre lábios superior e inferior. Para essa variável, três articuladores são relevantes: mandíbula, lábio inferior – em termos do seu deslocamento vertical em relação à mandíbula – e lábio superior – também em termos de deslocamento vertical. No Quadro 1, tem-se uma descrição do conjunto de variáveis do trato adotadas por Browman e Goldstein (1989) e dos articuladores a elas associados. Já na Figura 1, há uma representação do trato vocal que indica onde as variáveis descritas se localizam.

Variáveis do trato			Articuladores envolvidos
Abreviatura	Nome da variável (inglês)	Nome da variável (português)	
LP	<i>Lip protrusion</i>	Protrusão labial	Lábio inferior Lábio superior Mandíbula
LA	<i>Lip aperture</i>	Abertura labial	Lábio inferior Lábio superior Mandíbula
TTCL	<i>Tongue tip constrict location</i>	Local de constrição de ponta de língua	Ponta de língua Corpo de língua Mandíbula
TTCD	<i>Tongue tip constrict degree</i>	Grau de constrição de ponta de língua	Ponta de língua Corpo de língua Mandíbula
TBCL	<i>Tongue body constrict location</i>	Local de constrição de corpo de língua	Corpo de língua Mandíbula
TBCD	<i>Tongue body constrict degree</i>	Grau de constrição de corpo de língua	Corpo de língua Mandíbula
VEL	<i>Velic aperture</i>	Abertura vélica	Véu palatino
GLO	<i>Glottal aperture</i>	Abertura glotal	Glote

Quadro 1 – Conjunto de variáveis do trato e respectivos articuladores associados, adaptado de Browman e Goldstein (1989, p. 73)

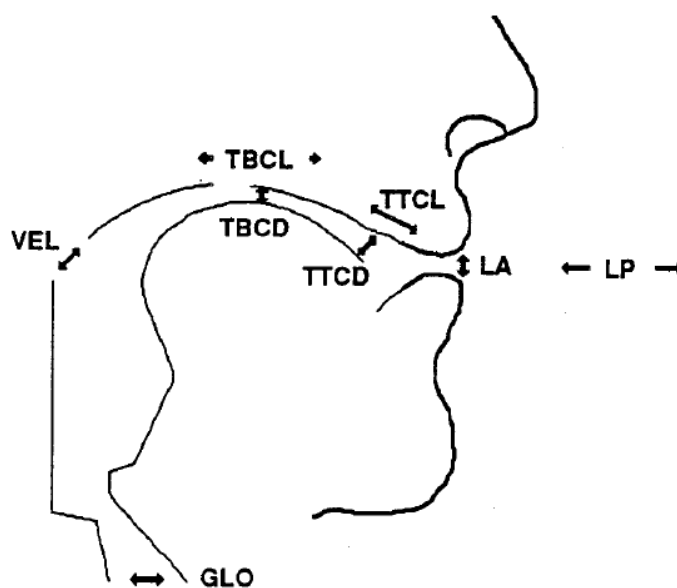


Figura 1 – Representação do trato vocal com a localização das respectivas variáveis, reproduzida de Browman e Goldstein (1989, p. 73)

Há gestos que ocorrem em função das mesmas variáveis do trato. Nesse caso, a distinção entre os gestos se dá por meio dos valores atribuídos aos parâmetros dinâmicos relacionados às variáveis do trato. Esses parâmetros são o *grau de constrição*, o *local de constrição*, o *formato de constrição*³ e a *rigidez*⁴. Os parâmetros dinâmicos relevantes para cada conjunto de articuladores, conforme proposta de Browman e Goldstein, estão listados no Quadro 2, a seguir.

Conjunto de articuladores	Parâmetros dinâmicos associados			
Lábios (LIPS)	Grau de constrição	Local de constrição		Rigidez
Ponta de língua (TT – <i>Tongue tip</i>)	Grau de constrição	Local de constrição	Formato de constrição*	Rigidez
Corpo de língua (TB – <i>Tongue body</i>)	Grau de constrição	Local de constrição	Formato de constrição*	Rigidez
Raiz de língua (TR – <i>Tongue root</i>)	Grau de constrição*	Local de constrição*		Rigidez*
Véu palatino (VEL – <i>Velum</i>)	Grau de constrição			Rigidez
Glote (GLO – <i>Glottis</i>)	Grau de constrição	Local de constrição* ⁵		Rigidez

Quadro 2 – Conjuntos de articuladores e parâmetros dinâmicos associados, conforme proposta de Browman e Goldstein (1989, p. 74)

Esses parâmetros são de natureza contínua, podendo apresentar valores diversos; no entanto, é necessária uma representação desses valores que tenha uma natureza discreta, de forma a se estabelecer um sistema minimamente especificado. Para uma representação discreta desses valores, o modelo emprega um conjunto de *descritores gestuais*, os quais foram elaborados de modo que cada

³ O parâmetro *formato de constrição* (*constriction shape*) encontrava-se, à época, em processo de implementação no modelo. O mesmo se aplica a todos os outros elementos sinalizados com asteriscos no Quadro 2. Logo, o conjunto *raiz de língua*, todos os seus respectivos parâmetros dinâmicos e o parâmetro de *local de constrição* para a glote não se encontravam, então, implementados ao modelo.

⁴ A *rigidez* (*stiffness*) é um parâmetro relevante para todos os conjuntos de articuladores, já que a realização de um gesto é influenciada pelo quão rígidos são os articuladores envolvidos no movimento empreendido.

⁵ O *local de constrição* no conjunto *glote* é destacado, em Browman e Goldstein (1989, p. 74), como um parâmetro a ser implementado no modelo. No entanto, a mesma versão do modelo determina que esse parâmetro é relevante apenas para gestos orais: “*Constriction Location* is relevant only for oral gestures, and refers to the x0 value for the constriction location tract variables (LP, TTCL, or TBCL).” (*op. cit.*)

gesto possa ser especificado a partir de um conjunto de descritores distinto. Os descritores adotados, conforme modelo de Browman e Goldstein (1989), estão no Quadro 3:

Parâmetro dinâmico	Descritores associados (inglês)	Descritores associados (português)
Grau de constrição (CD – <i>Constriction degree</i>)	<i>Closed</i> <i>Critical</i> <i>Narrow</i> <i>Mid</i> <i>Wide</i>	Fechado Crítico Estreito Médio Aberto
Local de constrição (CL – <i>Constriction location</i>)	<i>Protruded</i> <i>Labial</i> <i>Dental</i> <i>Alveolar</i> <i>Post-alveolar</i> <i>Palatal</i> <i>Velar</i> <i>Uvular</i> <i>Pharyngeal</i>	Protruso Labial Dental Alveolar Pós-alveolar Palatal Velar Uvular Faringal

Quadro 3 – Parâmetros dinâmicos e descritores gestuais associados, conforme proposta de Browman e Goldstein (1989, p. 75)

É importante reiterar que os gestos têm dupla função. Eles caracterizam os movimentos dos articuladores que são de fato observados, o que tem implicações metodológicas, como a necessidade de se empregarem, nas pesquisas, tecnologias que possibilitem essa observação de uma forma fidedigna, tais como a ressonância magnética (MRI), a eletropalatografia (EPG) e a ultrassonografia. Ao mesmo tempo, além de caracterizarem os movimentos dos articuladores, os gestos funcionam como unidades contrastivas. Também, a sua dimensão enquanto unidade no sistema difere da de unidades de análise em outros modelos, como o traço e o segmento, representando uma dimensão intermediária em relação a essas duas unidades. Para ilustrar-se a dimensão do gesto enquanto unidade fonológica, os autores estabelecem que

o gesto, como unidade fonológica, difere tanto do traço quanto do segmento (ou nó de raiz, em geometrias de traço atuais). Ele é uma unidade maior que o traço, sendo efetivamente uma ação de constrição unitária, parametrizada conjuntamente por uma estrutura interligada de traços (valores de descritores). Ao mesmo tempo, é uma unidade menor que o segmento: vários gestos ligados entre si são necessários para que se forme

uma unidade no nível segmental ou em níveis superiores. (BROWMAN; GOLDSTEIN, 1989, p. 75, tradução nossa⁶)

Cada gesto possui um aspecto espacial – um alvo a ser atingido, relacionado a uma variável do trato –, como também possui um aspecto temporal intrínseco. Isso permite assumir que os gestos podem se sobrepor, tanto em termos espaciais quanto temporais. Mudanças nos padrões de sobreposição podem produzir formas fonéticas e fonológicas diferentes.

Na produção da fala, os gestos ocorrem em coordenação, e o padrão de organização desses gestos obedece a princípios de faseamento. Esses princípios especificam como se dá a coordenação dos gestos no espaço e no tempo. Na produção de enunciados, os padrões de coordenação observados, assim como os intervalos de ativação dos gestos, resultam em uma organização gestual particular, a qual é mostrada, no modelo, por meio de uma representação gráfica bidimensional, denominada *pauta gestual*.

A pauta gestual é um instrumento importante para a descrição de fenômenos a partir da perspectiva da Fonologia Articulatória, pois resgata tanto aspectos articulatórios quanto temporais na produção de fala. A Figura 2 possui exemplos de pauta gestual fornecidos pelos autores (BROWMAN; GOLDSTEIN, 1989, p. 76), sendo aqui utilizada para ilustrar a sua composição e os aspectos da realização da fala que evidencia.

⁶ No original: “the gesture as a phonological unit differs both from the feature and from the segment (or root node in current feature geometries). It is a larger unit than the feature, being effectively a unitary constriction action, parameterized jointly by a linked structure of features (descriptor values). Yet it is a smaller unit than the segment: several gestures linked together are necessary to form a unit at the segmental, or higher, levels.”

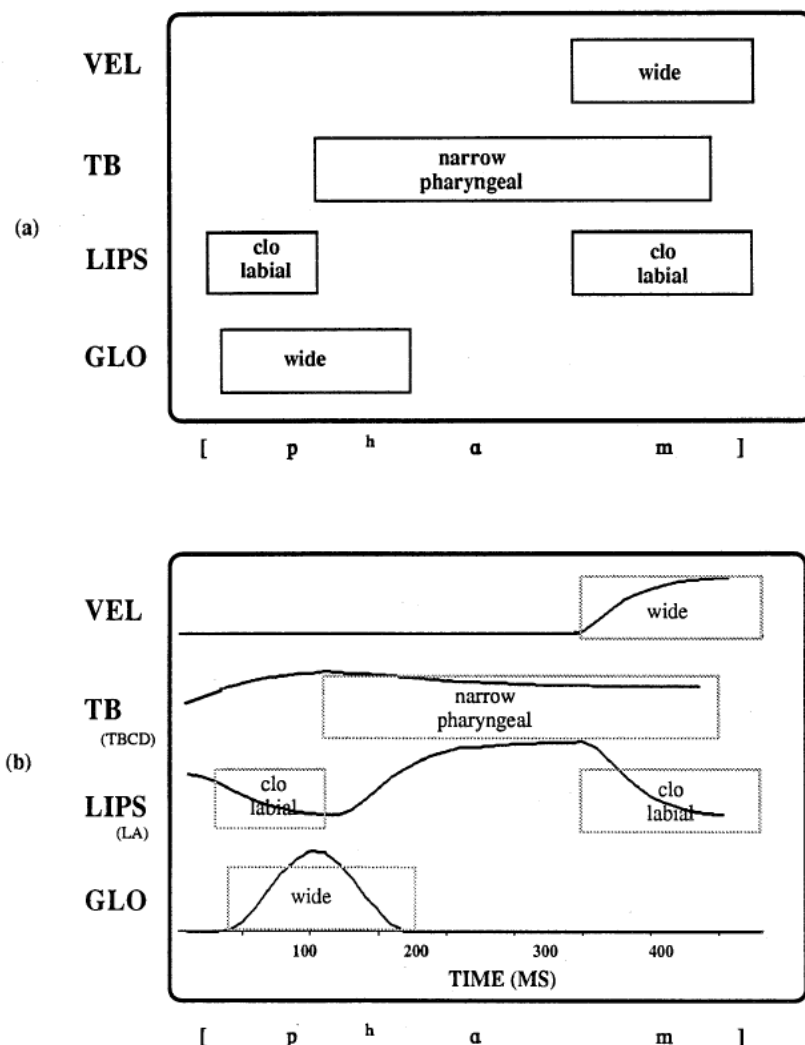


Figura 2 – Exemplos de pautas gestuais com base na palavra *palm*, do inglês, reproduzidos de Browman e Goldstein (1989, p. 76)

Na Figura 2, temos dois exemplos de pautas. Em (a), é mostrada a realização dos gestos por meio de uma estrutura bidimensional. Nela, o eixo vertical representa os conjuntos de articuladores empregados na produção dos gestos, em diferentes *tiers*. Essa representação em *tiers*, um para cada conjunto de articuladores, dá conta das eventuais sobreposições de gestos que empregam diferentes conjuntos de articuladores. A notação dos gestos é em forma de caixas, no interior das quais são especificados, em forma de texto, os descritores, relacionados a valores de parâmetros dinâmicos – grau de constrição e, quando relevante, local de

construção⁷. Para exemplificação, na Figura 2, em (a), pode-se observar que o gesto oral inicial envolve os lábios, pois está representado no *tier LIPS* (lábios). Esse é um gesto de fechamento labial, relacionado à produção do [p] inicial. Seu grau de construção é fechado (representado pelo primeiro descritor, *clo*, abreviação para *closed*) e seu local de construção é labial (representado pelo segundo descritor, *labial*). No eixo horizontal da pauta, é representada a temporalidade, com o intervalo durante o qual cada gesto particular se encontra ativado.

Na Figura 2, em (b), tem-se a mesma representação bidimensional, desta vez com a adição de curvas que mostram a variação nas trajetórias dos articuladores ocorrida em cada gesto. Em (b), pode-se ver que, durante o intervalo de ativação do gesto inicial de fechamento labial de [p], a curva localizada no *tier LIPS*, que representa a abertura labial, apresenta uma queda. A partir das curvas, também se pode ter uma ideia dos intervalos de ativação, os quais capturam aspectos da duração dos gestos.

As informações que são apresentadas na pauta, para Browman e Goldstein (1989), podem identificar uma determinada entrada lexical. Um gesto pode se diferenciar de outro a partir do valor de seus descritores: o gesto de ponta de língua em [s] se diferenciará do gesto envolvendo o mesmo articulador em [t], pois o grau de construção do primeiro será *crítico*, enquanto o do segundo será *fechado*. E, se procedermos ao nível de palavra, pautas gestuais para diferentes itens lexicais se diferenciarão em termos da presença, ausência e/ou valores de descritores de determinados gestos.

Como discutido, cada gesto tem uma extensão temporal particular, o que justifica eventuais sobreposições com outros gestos na produção da fala. Assumindo-se que haja sobreposição gestual, opera-se, dessa forma, uma mudança na observação e explicação de fenômenos fonético-fonológicos. Browman e Goldstein (1987) propuseram que muitos fenômenos da fala casual, tradicionalmente descritos como inserções, apagamentos, assimilações e enfraquecimentos, devem-se a duas variações observáveis em pautas gestuais: i) a redução da magnitude de gestos, tanto no espaço quanto no tempo, e ii) o aumento nas sobreposições

⁷ A necessidade de se especificar o parâmetro *local de construção* para um determinado gesto dependerá da mobilidade dos articuladores nele envolvidos. Gestos de glote e gestos vélicos, por exemplo, não terão esse parâmetro especificado por não possuírem mobilidade significativa para se deslocarem ao longo do trato vocal. Essa é uma mobilidade que os lábios e as partes da língua, por sua vez, possuem, ainda que em diferentes graus. Para esses articuladores, o parâmetro *local de construção* é relevante e, portanto, deve ser especificado nos gestos que os envolvem.

temporais entre gestos. Fenômenos recorrentemente descritos como apagamentos foram caracterizados por Browman e Goldstein (1987) como oclusões (*hidings*) – o segmento considerado apagado, na verdade, teria tido um gesto sobreposto por outro, o qual se revelou mais saliente acusticamente e foi, portanto, o único detectável auditivamente.

Essa é uma leitura que pode ser aplicada também a fenômenos da fala infantil, objeto de grande número de pesquisas em que dados produzidos pelas crianças apresentam fenômenos classificados como apagamentos ou inserções, por exemplo. Por meio da análise da coordenação gestual, é possível que eventualmente se observem movimentos dos articuladores relacionados a segmentos que se presumiriam ausentes ou apagados. Essa compreensão de fenômenos da fala infantil pode evidenciar que a percepção e a aquisição de sons e estruturas têm início antes do que se prevê em trabalhos amparados em teorias fonológicas clássicas. No caso da estrutura CCV, objeto de investigação desta pesquisa, pode-se partir, por exemplo, do pressuposto de que supostos apagamentos de um ou mais elementos da sílaba, em princípio detectados auditivamente em produções da criança, podem, por meio de uma análise articulatória, apresentar, ainda que em menor magnitude, gestos relacionados a um segmento em princípio ausente. Logo, essa interpretação, aferida a partir de princípios da Fonologia Articulatória aqui explicitados, pode oferecer uma visão diferenciada da aquisição fonológica, cujos processos podem se revelar mais precoces do que a literatura apoiada em modelos clássicos aponta.

Nesta seção, buscou-se descrever o modelo teórico de base adotado nesta pesquisa, a Fonologia Articulatória, em termos de seus principais pressupostos, conceitos e instrumentos de descrição e análise. Na seção 2.1.2, a seguir, será discutido em maior detalhe como a sílaba CCV é descrita a partir dos pressupostos desse modelo teórico.

2.1.2 A sílaba CCV na perspectiva da Fonologia Articulatória

Além de terem desenvolvido uma teoria fonológica geral, Browman e Goldstein (1988) também estudaram, de modo mais específico, aspectos como a sílaba e sua organização gestual. Browman e Goldstein (1988) examinaram duas

abordagens para se buscarem padrões estáveis na fala: uma organização local, em que gestos individuais estariam coordenados com outros gestos individuais, e uma organização global, na qual gestos formariam conglomerados maiores.

Com base em evidências articulatórias de palavras do inglês com uma variedade de consoantes iniciais, os autores sugerem que as consoantes em início de sílaba formariam uma organização de ordem *global*. Essa organização seria indexada por uma unidade de medida chamada *centro-C* (*C-center*), um centro de consoante, que se coordena com o gesto de vogal da sílaba. Já nas consoantes em final de sílaba, seria empregada uma organização de ordem *local*, em que o primeiro gesto de consoante pós-vocálico é coordenado com o gesto de vogal.

A partir dessas sugestões, Browman e Goldstein (1988) discutem as implicações desses diferentes estilos de organização para a estrutura perceptual e fonológica da fala. Os autores exploram se há um papel da afiliação na sílaba, isto é, a que parte da sílaba um ou mais gestos consonantais pertencem, no modo de coordenação gestual. Também exploram a possibilidade de existir um tipo de coordenação outro além do de um gesto individual com outro.

Para explorar os efeitos de afiliação na sílaba, os autores compararam padrões gestuais de enunciados em que os gestos consonantais tivessem diferentes afiliações na palavra. Exemplos desses enunciados são *pea splots* ([pi 'sɹplats]⁸) e *piece plots* ([piɹ 'plats]). Nesse par, há sequências de sons iguais em que a diferenciação está na afiliação do [s] – no primeiro enunciado, em início de sílaba e, no segundo, em final de sílaba.

No experimento, também há pares em que a afiliação da consoante é a mesma, mas a diferença está no número de consoantes na sequência, como em [pi 'sɹats] e [pi 'sɹpats]. Essa variação foi incluída para se explorar a medida do *centro-C*, que seria uma medida de coordenação gestual diferente, de ordem mais global, como já discutido.

Os dados examinados são dados articulatórios pertencentes a uma base de dados de microrraio-X. Por meio de eletrodos posicionados em diferentes partes do

⁸ As transcrições fonéticas dos enunciados foram aqui reproduzidas com os símbolos exatos utilizados por Browman e Goldstein (1988). Note-se que a vogal baixa, transcrita como [a], pode ser realizada, no inglês, como [ɑ], vogal baixa posterior, a qual figura em transcrições fonéticas em dicionários de inglês. A transcrição utilizada pelos autores é mais simplificada, omitindo, por exemplo, a aspiração normalmente presente em [p], própria dos segmentos plosivos desvozeados do inglês, realizados como [p^h], [t^h] e [k^h], assim como a duração vocálica de [i], aspecto relevante foneticamente e que habitualmente é notado, nas transcrições, pelos dois pontos – [i:].

aparelho fonador, foi possível acompanhar os deslocamentos horizontal e vertical de articuladores, como a língua, os lábios e a mandíbula. A Figura 3 mostra o enunciado *pea splots*, produzido por falante de inglês, com o deslocamento vertical dos articuladores ao longo do tempo.

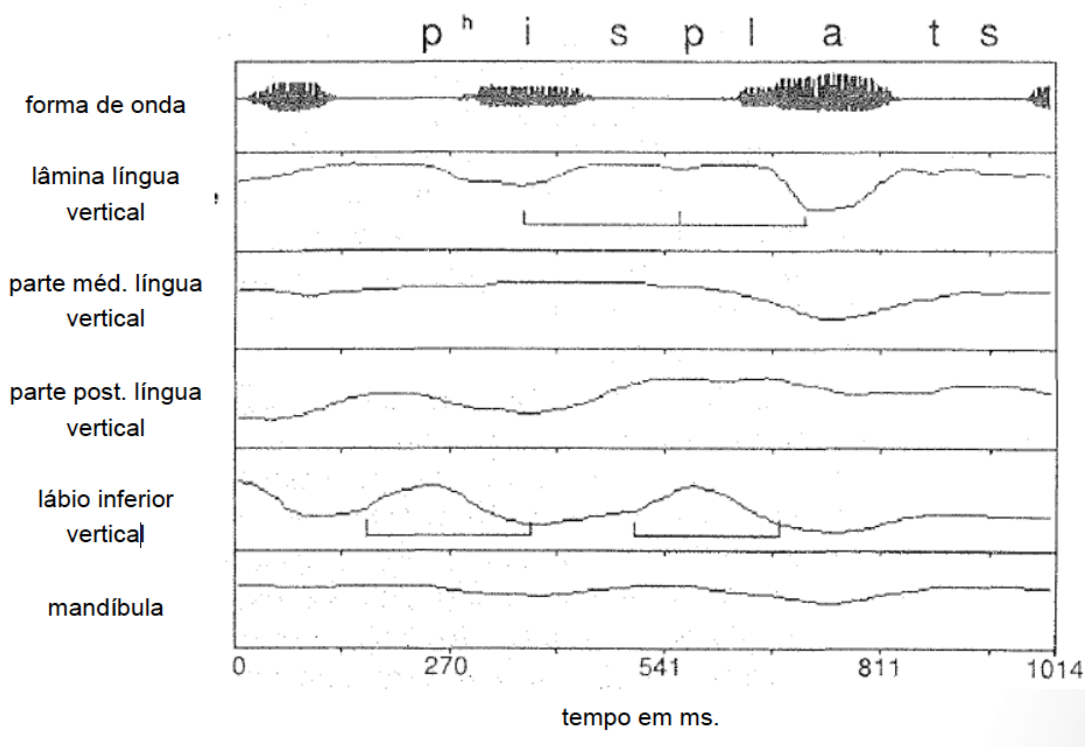


Figura 3 – Enunciado *pea splots*, produzido por falante de inglês, com oscilograma e deslocamento vertical dos articuladores envolvidos. Imagem adaptada de Browman e Goldstein (1988, p. 87)

As análises foram feitas a partir dos movimentos de eletrodos colocados na lâmina da língua (para [s], [l] e [t]) e no lábio inferior (para [p]). Os autores examinaram, primeiramente, a relação dos gestos consonantais intervocálicos com a consoante pós-vocálica, quando os gestos estavam em início de sílaba. Por exemplo, para [pi 'splats], foi examinada a relação entre os gestos associados a [spl] e a [t]. A Figura 4 mostra produções de enunciados, nos moldes [pi#C(C)(C)ats], à esquerda, e [pa#C(C)(C)idz], à direita, com diferentes composições de consoantes em início de sílaba.

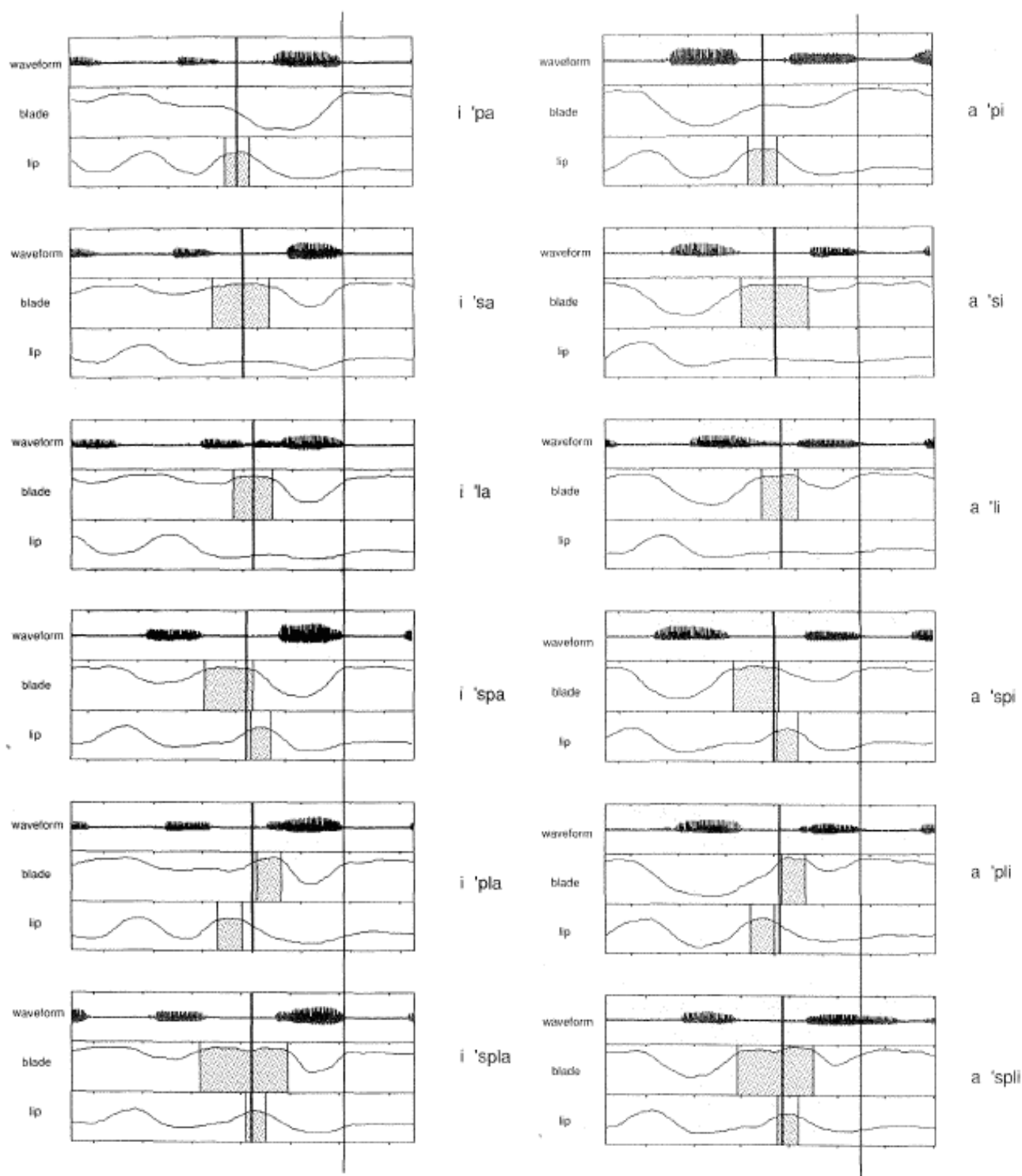


Figura 4 – Oscilograma e deslocamento vertical dos articuladores lâmina de língua e lábio inferior para as produções de sentenças nas formas [pi#C(C)(C)ats], à esquerda, e [pa#C(C)(C)idz], à direita. Imagem adaptada de Browman e Goldstein (1988, p. 90-91)

As porções preenchidas na figura correspondem a platôs do momento de maior deslocamento ocorrido no gesto, e todos os *tokens* foram alinhados no ponto

em que é atingido o alvo da consoante [t] ou [d], com o fechamento alveolar. Essa linha, a qual perpassa todos os *tokens*, localizada mais à direita, é o *ponto de ancoragem*.

As linhas mais espessas, que perpassam a amostra de cada *token*, representam o *centro-C* de cada sequência de gestos consonantais. Ele é o resultado de um cálculo em dois passos: primeiro, para cada gesto consonantal, é computado o ponto médio temporal entre os limites à esquerda e à direita desse gesto. Em seguida, a partir dos pontos médios obtidos, é feito um cálculo de média. O valor de média obtido corresponde ao *centro-C* daquela sequência.

Numa comparação entre enunciados produzidos a partir de um mesmo molde, os autores observam que, nos platôs dos gestos consonantais, os limites à esquerda e à direita não se alinham tão bem. Já o *centro-C* calculado se alinha melhor no tempo em todos os enunciados.

Assim, com base nos dados da Figura 4, os autores sugerem que o *centro-C* parece estar mais relacionado ao ponto de ancoragem do que qualquer outro gesto consonantal individual. O *centro-C* teria uma propriedade global, a qual, aplicada a sequências em início de sílaba, parece fornecer a medida mais estável da relação entre essas sequências e a primeira consoante final daquela mesma sílaba (*i.e.*, o ponto de ancoragem).

A partir desses dados, sugere-se que sequências de gestos consonantais, ao menos em início de sílaba, formam um agrupamento tal que a sua relação com o ponto de ancoragem da mesma sílaba é melhor medida – de modo mais estável – pela métrica global do *centro-C*.

Também são analisados os mesmos enunciados, mas estabelecendo-se um ponto de ancoragem diferente. Na Figura 5, estão dispostas produções de enunciados nos mesmos moldes da figura anterior – [pi#’C(C)(C)ats], à esquerda, e [pa#’C(C)(C)idz], à direita, mas com ponto de ancoragem à esquerda, no fechamento labial de [p]. As palavras estão alinhadas no ponto temporal de fechamento labial, no [p] da primeira palavra, e as linhas verticais que perpassam a amostra de cada *token* representam o limite à esquerda do primeiro gesto consonantal envolvido na sequência.

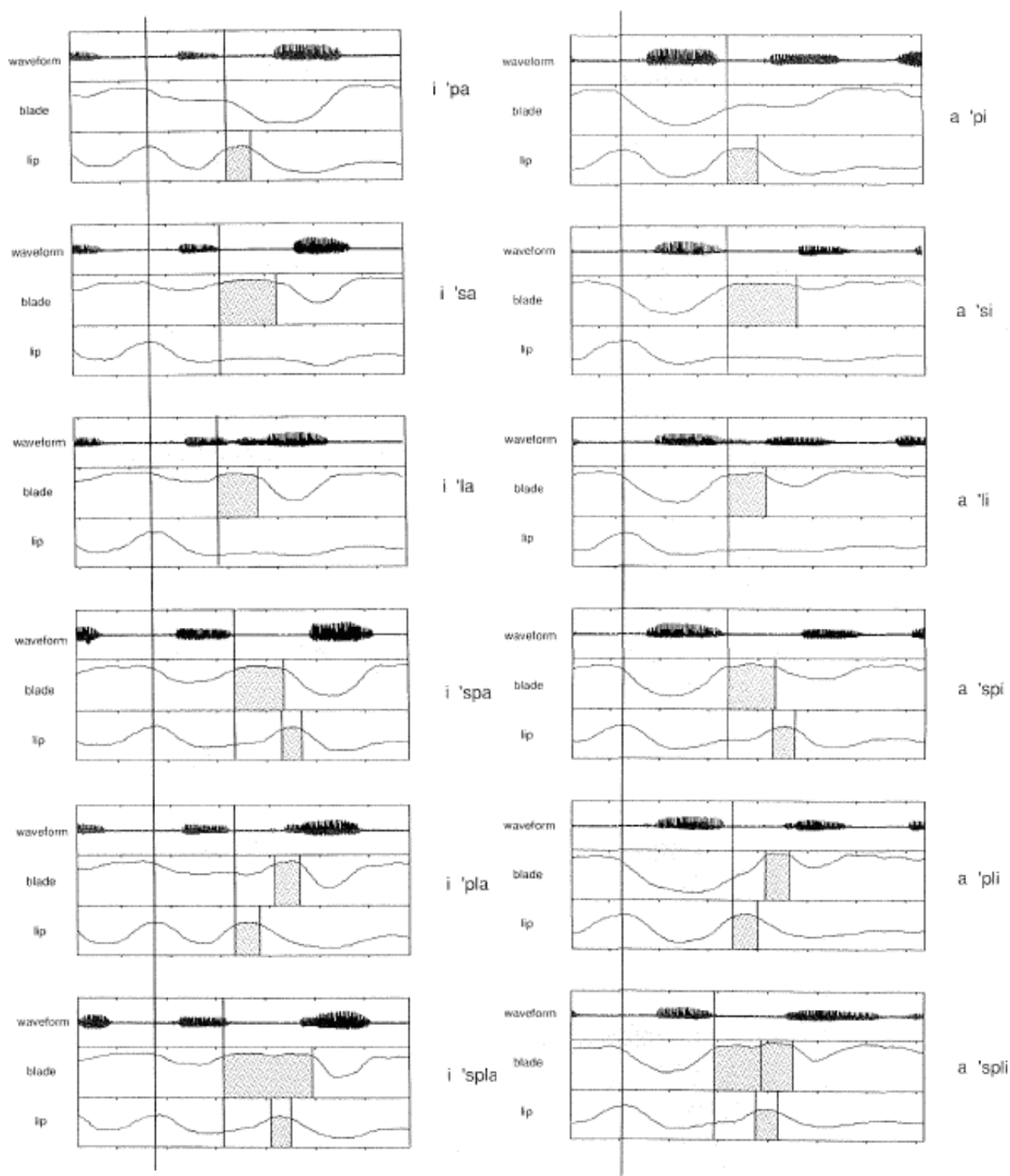


Figura 5 – Oscilograma e deslocamento vertical dos articuladores lâmina de língua e lábio inferior para as produções de sentenças nas formas [pi#C(C)(C)ats], à esquerda, e [pa#C(C)(C)idz], à direita. Imagem adaptada de Browman e Goldstein (1988, p. 92-93)

Na Figura 5, pode-se observar que, desta vez, há uma relação mais estável do ponto de ancoragem com o limite à esquerda da primeira consoante intervocálica,

a qual se alinha, ao longo dos *tokens*, de forma melhor e com menor desvio padrão que o *centro-C*, o qual, embora não demarcado nas produções da Figura 5, é mais variável. Nesses dados, examina-se a relação entre uma sequência consonantal em início de sílaba e uma consoante inicial em palavra anterior. Nesse contexto, a medida mais estável a ser aplicada parece ser uma propriedade local ao invés de global, que seria o atingimento do alvo da consoante mais à esquerda.

Os autores se perguntaram até que ponto essa relação dependeria da afiliação das consoantes a uma sílaba ou uma palavra. Para responderem a essa questão, examinaram enunciados em que a primeira palavra possuía uma sílaba fechada, como [pis 'plats], ao invés de aberta, como [pi 'splats]. Se o princípio de organização fosse a afiliação à sílaba, a métrica mais estável para o [s] final seria o *centro-C*, o qual aparentou ser a melhor métrica para as consoantes afiliadas, conforme os dados na Figura 4. Se, no entanto, o princípio de organização fosse a ordem sequencial, a métrica mais estável seria o limite à esquerda do [s], conforme os dados na Figura 5. Para verificar-se a hipótese correta, foram analisados enunciados produzidos com primeira palavra terminando em consoante. A Figura 6 mostra esses enunciados nos moldes [pip 'adz], [pis 'adz] e [piC#C(C)ats], à esquerda, e [pap 'its], [pas 'its] e [paC#C(C)idz], à direita. As palavras estão alinhadas no ponto temporal do último gesto consonantal da primeira palavra. A linha vertical que perpassa cada *token* representa o limite à esquerda do primeiro gesto consonantal envolvido na sequência.

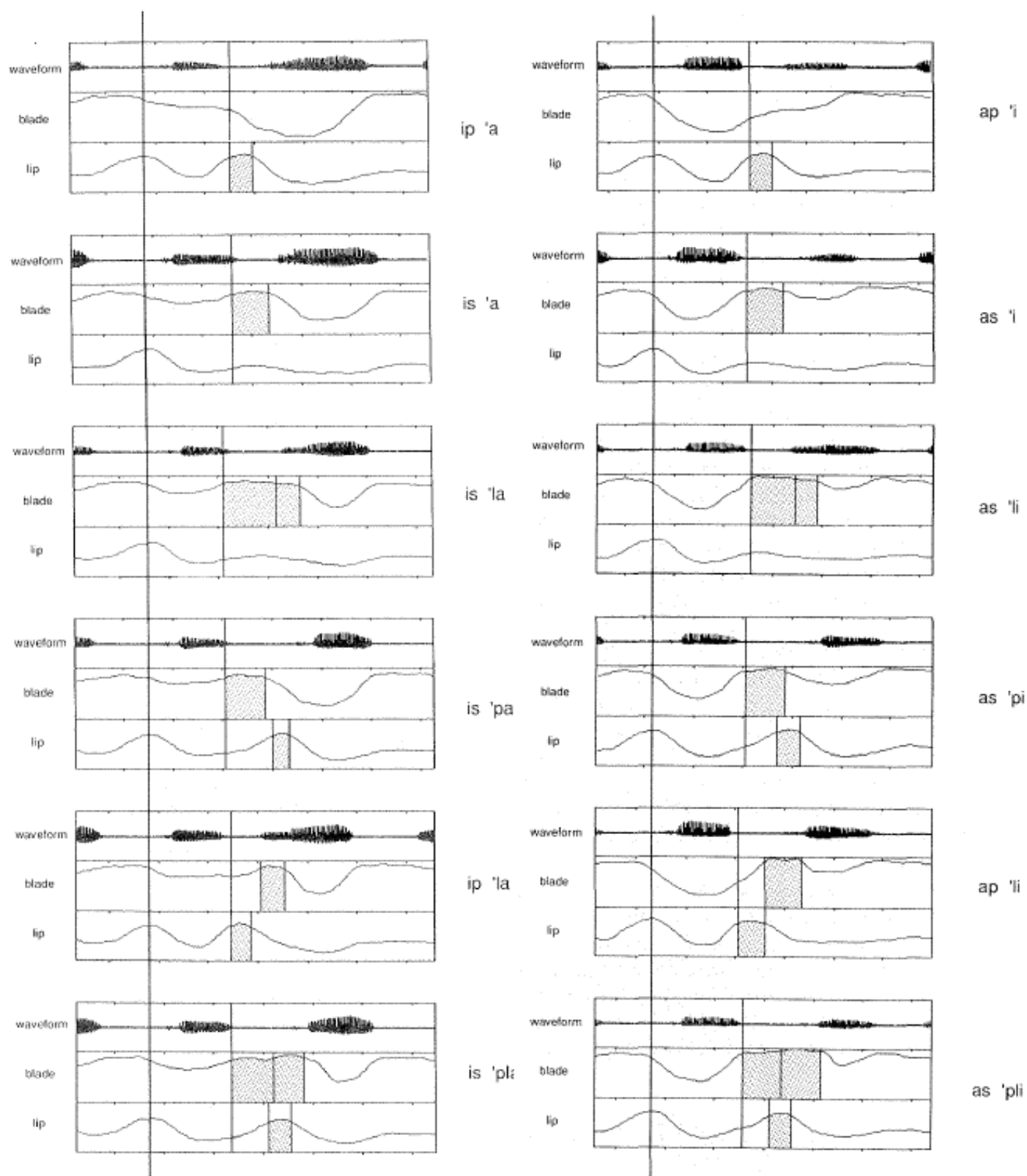


Figura 6 – Oscilograma e deslocamento vertical dos articuladores lâmina de língua e lábio inferior para as produções de sentenças nas formas [pip 'adz], [pis 'adz] e [piC#C(C)ats], à esquerda, e [paC#C(C)idz], à direita. Imagem adaptada de Browman e Goldstein (1988, p. 94-95)

Na Figura 6, os limites à esquerda das consoantes se alinham melhor que os *centros-C*, do mesmo modo que na Figura 5. Parece, assim, que consoantes pós-vocálicas estão organizadas com base na sua relação sequencial com a vogal (limite à esquerda da consoante), e não com base na afiliação à sílaba (*centro-C*). Consoantes pós-vocálicas parecem estar, assim, relacionadas a uma métrica local.

De modo geral, o estudo de Browman e Goldstein sugere que, no interior de palavras ou sílabas, consoantes pós-vocálicas (em final de sílaba) parecem se comportar de forma diferente das pré-vocálicas (em início de sílaba). Consoantes iniciais estão relacionadas a suas palavras em termos de uma métrica global do grupo consonantal, que seria o *centro-C*; já as consoantes finais parecem estar relacionadas a suas palavras em termos da métrica local do atingimento do alvo, que seria o limite à esquerda da consoante. Isso explicaria propriedades fonológicas que diferem sequências consonantais iniciais das finais.

O *centro-C* parece ter relevância na organização articulatória intrassilábica, servindo para coordenar um grupo consonantal inicial com a vogal. E, quanto mais consoantes no grupo consonantal inicial, mais curta será a realização acústica da vogal e mais longa será a sílaba como um todo.

Esse padrão do *centro-C* remete ao padrão do *centro-p*⁹, reportado por autores como Fowler e Tassinary (1981, *apud* BROWMAN; GOLDSTEIN, 1988). Em seu experimento, ouvintes deveriam alinhar palavras com diferentes inícios consonantais, com o auxílio de um metrônomo, em ritmo regular. O alinhamento feito pelos ouvintes não coincidiu com os *onsets* acústicos. Viu-se que, de modo geral, quanto mais consoantes a sequência tivesse, o alinhamento ocorria mais tarde no tempo. Em estudos como o de Marcus (1981), modificar as consoantes em final de sílaba surte efeito muito menor no *centro-p*. Assim, assumindo-se que algo do gesto articulatório de vogal seja importante para o *centro-p*, consoantes finais não interagiriam com o gesto de vogal da mesma forma que consoantes iniciais.

A diferença entre consoantes iniciais e finais quanto à coordenação com a vogal elucida o conceito de peso silábico, sustentando o que já havia sido afirmado por Hyman (1985, *apud* BROWMAN; GOLDSTEIN, 1988), isto é, que *onsets* silábicos não carregariam peso fonológico, estando associados à unidade de peso da vogal, formando uma única mora. Essa diferença entre consoantes finais e

⁹ Em inglês, *p-center*, que significa *perceptual center*.

iniciais se correlaciona com as organizações vogal-consoante e consoante-vogal sugeridas por Browman e Goldstein (1988): as porções alvo de consoantes finais seriam produzidas na sua própria janela de tempo, enquanto as porções alvo de consoantes iniciais sobreporiam a janela de tempo da vogal.

Essa diferença na coordenação também explica, para Browman e Goldstein (1988), o alongamento compensatório, processo, na definição dos autores, em que uma vogal se alonga quando uma consoante em final de sílaba é perdida. Seria esperado que as diferenças na organização articulatória em final de sílaba se correlacionassem com o alongamento compensatório, já que a vogal, antes tendo apenas uma unidade temporal, passaria a ter duas – a da vogal e a da consoante final que se perdeu. Já para Fowler (1983, *apud* BROWMAN; GOLDSTEIN, 1988), isso seria um reflexo da sobreposição entre gestos consonantal e vocálico. A perda do gesto consonantal evidenciaria o gesto vocálico, “alongando-o” de forma que, tanto acústica quando auditivamente, seja detectável uma parte até então oculta, uma “extensão” desse gesto até então inaudível.

A seção a seguir será dedicada a propostas de descrição do *tap*, elemento consonantal presente nos encontros consonantais aqui estudados, na perspectiva da Fonologia Articulatória.

2.1.3 O *tap* na perspectiva da Fonologia Articulatória

Para dar-se base a um estudo da aquisição de encontros consonantais com *tap* na perspectiva da Fonologia Articulatória, é necessário recuperar descrições sobre esse segmento alinhadas com essa mesma perspectiva teórica, visto que os encontros consonantais a serem estudados nesta pesquisa têm esse segmento como segundo elemento consonantal. Assim, nesta seção, serão discutidas propostas de descrição do *tap*. Entender o que está envolvido na sua produção, assim como na sua aquisição, que é relativamente tardia, pode ajudar a explicar como os padrões de coordenação gestual envolvidos em CCV se instalam de modo gradual.

Silva (1996) fez um estudo de caso a partir do qual fornece uma descrição sobre líquidas do português brasileiro. Com base em dados acústicos, propõe que o *tap* é realizado em meio a eventos de natureza vocálica, apoiando-se “sempre em

duas ‘vogais’, independentemente da posição silábica” (SILVA, 1996, p. 68). A abertura oral faria parte da sua realização, de modo que, mesmo em um contexto de encontro consonantal, o *tap* ocuparia “de qualquer forma, uma posição ‘intervocálica’” (*ibid.*, p. 69). Juntamente à sua realização, podem ser observados eventos de natureza vocálica, de duração breve, os quais ocorreriam, em encontros consonantais, antes do *tap* e, em coda, após a realização do *tap*.

Carvalho (2004), por sua vez, fez um estudo comparativo entre vibrantes do português brasileiro e do espanhol colombiano e indicou, no contexto de encontros consonantais, a existência de uma possível semelhança entre o evento vocálico anterior ao *tap* e a vogal nuclear da sílaba em termos de estrutura formântica, tanto no português quanto no espanhol.

Foi com base em trabalhos como os de Silva (1996) e Carvalho (2004), que realizaram descrições visuais do dado acústico, que Silva, Clemente e Nishida (2006) buscaram fazer uma descrição acústica sistemática do evento vocálico adjacente ao *tap* em encontros consonantais¹⁰ e em coda. Para tanto, os autores realizaram medidas das estruturas formânticas presentes no evento vocálico e na vogal nuclear e compararam essas medidas para averiguar a natureza do evento vocálico em relação à vogal nuclear. Também buscaram, a partir desses dados, fornecer uma representação do *tap* em conformidade com modelos dinâmicos de produção da fala.

Silva, Clemente e Nishida (2006) investigaram o comportamento do evento vocálico em encontros consonantais e em codas, para que se observasse se, em ambos os contextos, ele teria a mesma natureza da vogal nuclear. Com base em estudos realizados sobre línguas outras – búlgaro e norueguês –, nas quais o elemento vocálico não apresenta as mesmas características da vogal nuclear, afirmam que a natureza desse elemento é específica da língua ao invés de universal, e, assim, seria necessário representar esse fato na gramática da língua.

Os autores consideram duas possibilidades de descrição desse elemento vocálico: ele poderia ser semelhante a um *schwa*, com características formânticas neutras, ou ser a vogal nuclear em si, “sendo entrecortada pelo *tap*, uma vez que

¹⁰ Ao reportarmos o trabalho de Silva, Clemente e Nishida (2006), referimo-nos às sequências de obstruente e *tap* como encontros consonantais, ainda que, no texto, os autores se refiram a essas sequências como “grupos”.

devido ao seu aspecto descontínuo seja necessário um som contínuo para ele se apoiar” (SILVA; CLEMENTE; NISHIDA, 2006, p. 3).

Para a análise do *tap* em encontros consonantais, foram coletadas palavras dissílabas e paroxítonas, como “prato”, inseridas em sentenças, e.g. “Minha mãe guardou o prato no armário”. Na análise acústica, os *taps* foram identificados visual e auditivamente. Na Figura 7, tem-se um exemplo de palavra submetida à análise acústica, com o *tap* identificado pelo destaque em vermelho.

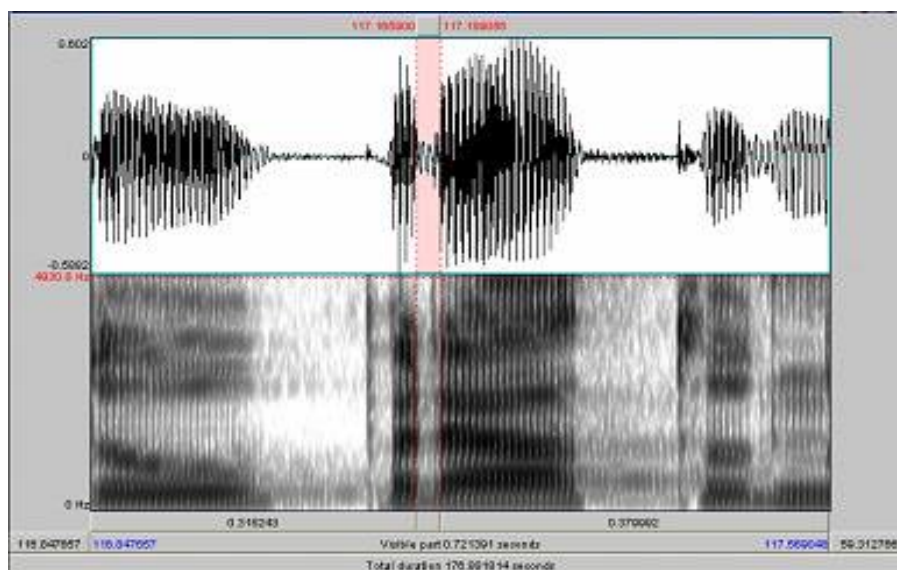


Figura 7 – Oscilograma e espectrograma de uma produção da palavra “prato”. Imagem reproduzida de Silva, Clemente e Nishida (2006, p. 5)

Entre a produção de [p] e o *tap*, destacado em vermelho, é possível observar-se a produção de um segmento contínuo, cuja estrutura formântica se assemelha à da vogal [a] após o *tap*. Esse segmento seria o elemento vocálico. Pela sua duração, consideravelmente menor que a da vogal nuclear, ele pode não ser percebido auditivamente, o que daria a impressão de uma produção de duas consoantes na sequência.

Os dados foram analisados acusticamente, tendo-se extraído os três primeiros formantes, tanto do elemento vocálico quanto da vogal nuclear nos seus pontos médios. A partir dos valores formânticos encontrados para os elementos vocálicos, foi feita uma plotagem dos valores de F1 e F2 e, nas palavras iniciando em oclusivas, os valores dos elementos vocálicos apresentaram disposição

semelhante à do triângulo vocálico que caracteriza as vogais orais do português. Assim, concluiu-se que, nos encontros consonantais iniciados por oclusivas, a estrutura formântica do elemento vocálico é condicionada pela qualidade da vogal em núcleo de sílaba. Em experimento diferenciado, conduzido apenas com palavras cujos encontros consonantais se iniciavam por fricativa, verificou-se a mesma tendência: os elementos vocálicos nesse contexto também formaram, na plotagem, um triângulo semelhante ao das vogais orais.

Silva, Clemente e Nishida (2006), ao constatarem que o elemento vocálico tem qualidade acústica condicionada pela vogal nuclear, levantam a hipótese de que o *tap*, em encontros consonantais, não estaria antecedendo a vogal, mas sim entrecortando-a após alguns milissegundos. Seria produzido, assim, o elemento vocálico com a estrutura formântica mesma da vogal nuclear.

Os autores também investigaram a natureza do elemento vocálico adjacente ao *tap* quando este ocorria em coda silábica. O elemento vocálico também ocorre em sequência com o *tap* nesse ambiente, pois, em coda, o *tap* é seguido de abertura oral, a qual propicia a produção do elemento vocálico. Na Figura 8, é mostrada a produção de uma palavra com *tap* em coda – “expor” – submetida à análise acústica, desta vez destacando-se em vermelho o elemento vocálico:

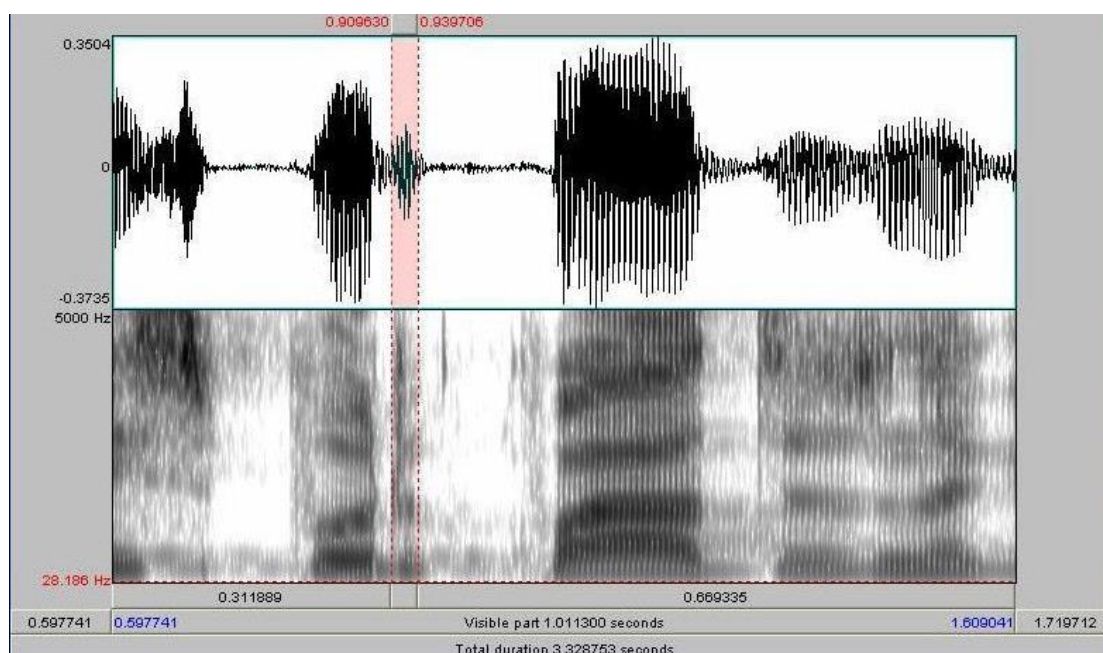


Figura 8 – Oscilograma e espectrograma de uma produção da palavra “expor”. Imagem reproduzida de Silva, Clemente e Nishida (2006, p. 11)

Buscou-se averiguar se o elemento vocálico ocorria de fato no português brasileiro quando o *tap* está posicionado em coda silábica, analisando-se um número maior de dados. As duas possibilidades consideradas sobre a natureza do elemento vocálico foram as mesmas para o *tap* em encontros: suas características seriam semelhantes às de uma vogal neutra, como um *schwa*, ou semelhantes às da vogal tônica, a qual ocorre antes do *tap* em coda.

Para que se investigasse a existência e a natureza desse elemento vocálico em coda, foi elaborado um novo experimento, realizado com quatro sujeitos. Nele, as palavras foram oxítonas dissílabas, pertencentes às classes dos verbos e dos substantivos, produzidas em frase-veículo. E, assim como no estudo do *tap* em encontros consonantais, foram medidos os valores de F1, F2 e F3 nas vogais tônicas e nos elementos vocálicos.

Por meio de uma análise visual dos dados, foi confirmada a existência de elemento vocálico após o *tap* em todas as produções de dois informantes. Para observar-se o comportamento desse elemento vocálico, foram extraídas médias das frequências das vogais e dos elementos vocálicos. Em seguida, foram feitos cálculos de subtração: os valores médios de frequência dos elementos vocálicos eram subtraídos dos valores médios de frequência das vogais. Os valores resultantes mostrariam se o elemento vocálico se aproximava ou não da vogal tônica em termos das suas características acústicas.

Apesar de não terem sido implementados testes estatísticos no estudo, os valores de diferença entre as médias foram interpretados como consideráveis, apontando uma diferença entre as estruturas formânticas da vogal e do elemento vocálico. Os resultados também foram plotados em termos dos valores de F1 e F2 e, enquanto as vogais apresentavam comportamento semelhante ao de um triângulo vocálico, os elementos vocálicos se apresentaram, na plotagem, concentrados ao invés de dispersos, embora estes tivessem sido produzidos com diferentes vogais em posição tônica.

A partir desses resultados, Silva, Clemente e Nishida (2006) concluíram que o elemento vocálico em coda não corresponderia à vogal tônica “entrecortada pelo *tap*” (SILVA; CLEMENTE; NISHIDA, 2006, p. 17), assemelhando-se, de fato, a uma vogal com características neutras. Assim, o elemento vocálico se comportaria de

forma distinta de acordo com a sua posição na sílaba: “em grupos o elemento vocálico é a mesma vogal nuclear, que o *tap* entrecorta. Em codas, por outro lado, o elemento vocálico exhibe [...] natureza distinta daquela da vogal nuclear.” (*op. cit.*)

A partir dessas considerações sobre o elemento vocálico e, por conseguinte, sobre a realização do *tap*, os autores propõem representações desse segmento para modelos dinâmicos de produção da fala. Com apoio em uma representação do *tap* intervocálico, proposta em Silva (2002), fornecem uma descrição do segmento nos dois contextos estudados, ou seja, tanto em encontros consonantais quanto em codas. Essa representação, assim como em Silva (2002), foi proposta na forma de pautas gestuais.

Com base na Fonologia Acústico-Articulatória (FAAR), modelo proposto por Albano (2001), formalizam a realização do *tap* em encontros consonantais por meio de uma pauta gestual, a qual se encontra na Figura 9.

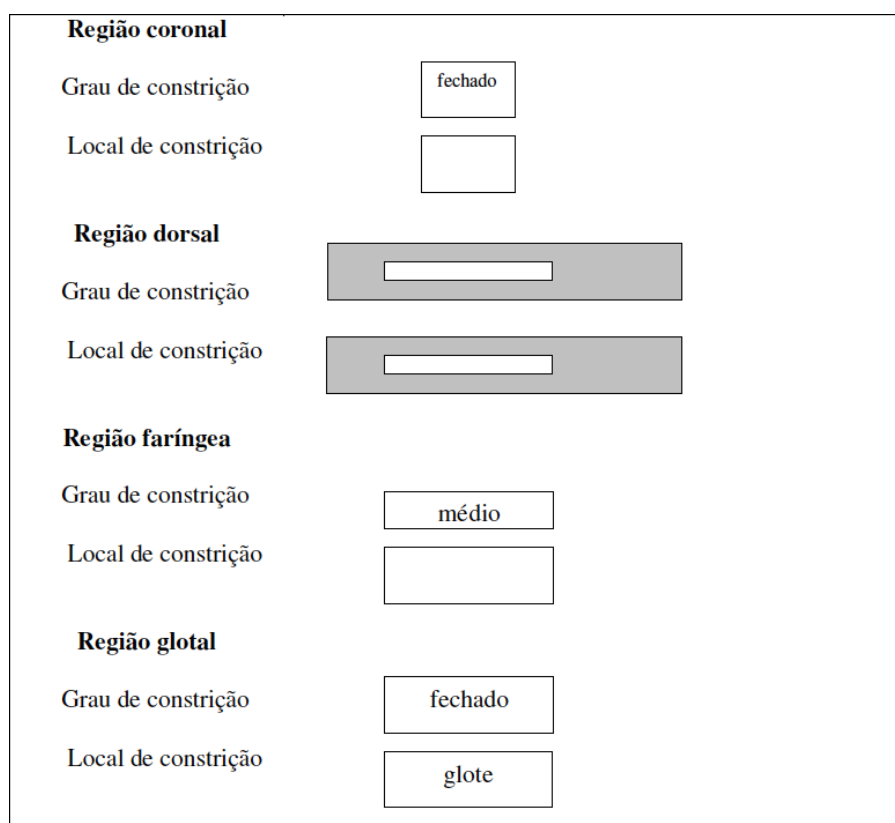


Figura 9 – Proposta de pauta gestual para a sequência de elemento vocálico, *tap* e vogal em grupos do tipo obstruinte + *tap* + vogal. Imagem reproduzida de Silva, Clemente e Nishida (2006, p. 21)

Nessa representação, é previsto um único gesto dorsal, correspondente à produção da vogal, representado nas caixas destacadas em cinza. A esse gesto, sobrepõe-se o gesto de dorso correspondente à produção do *tap*, representado, na imagem, em caixas brancas sobrepostas às caixas em cinza, correspondentes à vogal. Considerando que as pautas gestuais representam o tempo na dimensão horizontal, a imagem mostra que o gesto de dorso relativo ao *tap* se realiza ao mesmo tempo que parte do gesto de dorso da vogal, gesto esse que já havia se iniciado antes. No entanto, o gesto relativo à vogal se sobressai por possuir magnitude e tempo de ativação maiores, parâmetros esses representados, respectivamente, pelas dimensões vertical e horizontal das caixas.

Já na representação do *tap* em coda, observa-se comportamento distinto, já que a vogal tônica e o elemento vocálico não se assemelham acusticamente. Independentemente da vogal em posição tônica, o elemento vocálico após o *tap* será mais centralizado. Tendo-se isso em vista, os autores propõem a pauta disposta na Figura 10.

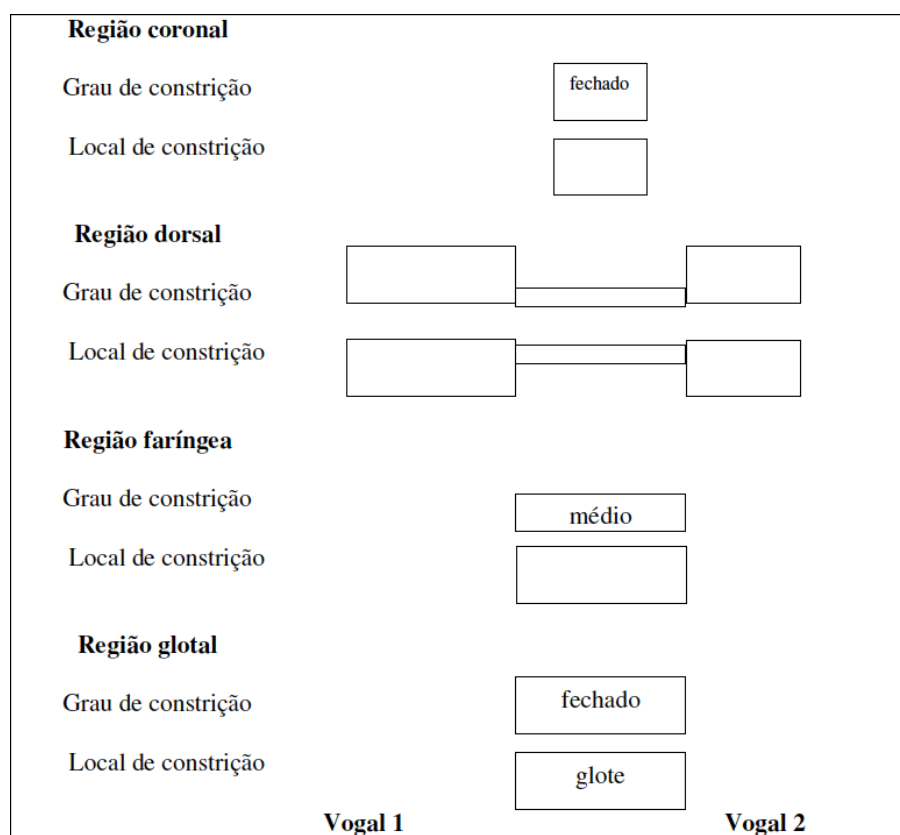


Figura 10 – Proposta de pauta gestual para o *tap* em codas. Imagem reproduzida de Silva, Clemente e Nishida (2006, p. 22)

A pauta gestual da Figura 10 se assemelha à pauta proposta por Silva (2002) para o *tap* intervocálico. No entanto, esta considera que o elemento vocálico à direita será sempre o mesmo, diferindo da vogal à esquerda em termos de qualidade formântica. Esse elemento vocálico também difere da vogal em duração, sendo mais breve.

Silva, Clemente e Nishida (2006) propuseram uma representação dinâmica para os *taps*, considerando que esse segmento está sempre acompanhado de um elemento vocálico, cuja natureza varia conforme a posição na sílaba em que ele se encontra. Também varia entre as línguas: no caso do português, a posição na sílaba é fator relevante na determinação da natureza do elemento vocálico.

Os autores também consideram que, na mesma medida em que o *tap* é uma das razões para se adotarem representações dinâmicas, eles constituem um problema para os modelos dinâmicos, que não operam com a noção de sílaba. Os *taps* indicariam, assim, “a necessidade de um refinamento dessa classe de modelos.” (SILVA; CLEMENTE; NISHIDA, 2006, p. 24)

A representação proposta pelos autores tem maior afinidade com o modelo teórico aqui adotado e também se relaciona a uma das questões norteadoras desta pesquisa – questão iv): que gestos articulatórios compõem o *tap* e qual o *status* desses gestos? O trabalho também é importante para se retomar a questão de como os gestos do *tap* interagem com gestos de outros segmentos em encontros consonantais. Afinal, o *tap* entrecorta ou se sobrepõe a um gesto vocálico? Essa questão tem importância, pois a sobreposição de gestos é uma leitura possível a partir do proposto por Fowler (1983, *apud* BROWMAN; GOLDSTEIN, 1988) e por Browman e Goldstein (1989). Espera-se, com os resultados desta pesquisa, trazer contribuições a essa discussão.

A seção a seguir tratará dos erros de fala e da sua importância para se compreender a coordenação gestual.

2.1.4 Erros de fala: o que podem evidenciar sobre a coordenação gestual

Um exame dos fenômenos da fala que são considerados “erros” pode ser importante na medida em que estes evidenciam que unidades estão envolvidas na

produção da fala. Os avanços na tecnologia possibilitaram o desenvolvimento de estudos sobre os erros de fala por meio de análises acústicas e articulatórias, fornecendo novas descrições para fenômenos antes reportados, de forma categórica, como substituições, inserções e apagamentos de segmentos. Estudos com base em dados articulatórios cinemáticos evidenciam que unidades de dimensões menores que a do segmento atuam em erros de fala.

Um desses estudos é o de Goldstein *et al.* (2007), em que são analisados erros de fala eliciados por meio de tarefas de repetição. São examinados dados articulatórios relacionados ao movimento vertical da língua e dos lábios e os resultados obtidos sustentam a hipótese dos gestos como unidades de produção da fala.

Os autores afirmam que, no passado, a natureza das unidades que compunham a língua era bastante diferenciada daquela de unidades relacionadas a outras tarefas motoras. A fala, em comparação a outras formas de ação coordenada, possuiria propriedades únicas em termos da comunicação e da codificação de informação, o que fez com que ela adquirisse um *status* especial em relação a outras formas de ação coordenada. Goldstein *et al.* (2007), por sua vez, propuseram que é possível a identificação de unidades composicionais de ação, as quais estão sujeitas a princípios dinâmicos que se aplicam à ação de forma geral e que também moldam a atividade da fala. Essas unidades seriam os *gestos* articulatórios e a sua identificação seria possível com base na observação de registros do movimento articulatório, por meio de dados cinemáticos.

Uma evidência dessas unidades – os *gestos* – tem sido observada nos erros de fala. Para Goldstein *et al.* (2007), os erros não são distorções aleatórias; são, pelo contrário, sistemáticos em sua ocorrência e distribuição. A sequência *coffee pot*, por exemplo, poderia ser produzida como *poffee cot* ou *poffee pot*, mas não como *cottee poff* – isso porque as consoantes estão sujeitas a interagirem em erros se compartilham a mesma posição na sílaba ou na palavra. A partir deste tipo de exemplo, supõe-se que os erros obedeçam a leis da fonologia. Logo, as unidades que nele participam são consideradas unidades cognitivas significativas. Resultados de pesquisas a partir de erros de fala têm, assim, desempenhado papel central na construção de modelos de produção da fala.

Estudos sobre os erros de fala conduzidos em laboratório observaram que o erro mais comum era a substituição de uma unidade a nível de segmento por outra.

Também observaram que os erros produzem uma sequência gramaticalmente bem formada na língua, a qual corresponderia a palavras existentes ou possíveis. Os erros obedeceriam, assim, à fonotática da língua em questão.

Os erros, quer fossem coletados em contextos naturais ou em laboratório, tinham a mesma ferramenta básica para seu estudo: a transcrição fonética. No entanto, esse recurso pode ser uma fonte incompleta de evidências da natureza das unidades de produção da fala caso se leve em conta que unidades fundamentais no estabelecimento de uma fonologia sejam, na verdade, menores que segmentos. Então, partindo-se desta ideia, os estudos com base em transcrição não fornecem um registro apropriado dos eventos articulatórios ocorridos, pois não há forma de registrar um gesto quando ele é subsegmental. Do mesmo modo, um movimento de constrição realizado de forma incompleta pode não ser passível de registro por gerar pouco efeito acústico e tornar-se, assim, inaudível ao pesquisador.

Estudos acústicos mostraram que erros de natureza gradiente de fato ocorrem, enquanto estudos de erros com base em dados articulatórios constataram, em produções errôneas, ativações simultâneas de músculos, relacionadas a padrões motores de segmentos distintos e realizadas em diferentes graus de magnitude. Assim, os autores desenvolvem experimentos partindo da ideia de que a observação de movimentos articulatórios, por meio de dados cinemáticos, pode fornecer um registro apropriado dos erros de fala. Consequentemente, também forneceria informação mais adequada quanto à existência e à natureza das unidades de ação na fala.

Algumas suposições dos autores, que orientaram a realização dos experimentos, foram as seguintes:

- a) Erros devem ser interpretáveis em termos de constrições linguisticamente significativas. Desse modo, um gesto deslocado temporalmente não irá resultar em um movimento aleatório.
- b) Erros também devem ser sensíveis a um contexto temporal e dinâmico mais amplo, como, por exemplo, a velocidade da fala.
- c) Os erros poderão envolver gestos individuais ou construções gestuais mais amplas, envolvendo segmentos inteiros.
- d) A fala tem uma base rítmica inerente, o que se verifica por uma métrica e uma alternância entre vogais e consoantes. Dessa forma, aspectos da

organização da fala podem ser elucidados por modelos de sistemas dinâmicos oscilatórios.

O estudo de Goldstein *et al.* (2007) encontrou evidências que sustentaram as quatro suposições. Aqui, são relatados os resultados de um dos dois experimentos desenvolvidos a partir de tarefas de repetição que eliciavam erros na fala. Foram coletados dados articulatórios das produções por meio de um articulômetro eletromagnético médio-sagital, o qual captava movimentos dos articuladores através de transdutores colocados em diferentes partes do aparelho fonador.

O primeiro experimento, a ser aqui detalhado, foi realizado com sete falantes nativos de inglês e estudou os erros na produção de consoantes iniciais. Os sujeitos foram instruídos a repetir construções de duas palavras, com consoantes alternantes em início de sílaba – *cop top*, *top cop*, *tip kip* e *kip tip*. Cada uma dessas construções foi produzida de forma repetida e cada tarefa de repetição, por sua vez, foi realizada em três velocidades, com andamentos entre 76 e 120 bpm. A velocidade de fala foi controlada com um metrônomo visual, colocado acima da tela do computador em que eram mostrados os estímulos. As variáveis controladas no experimento foram a velocidade de fala, o acento (inicial ou final – ênfase na primeira ou na segunda palavra da sequência), a ordem das palavras envolvidas (se *cop top* ou *top cop*, por exemplo) e a vogal envolvida ([a], como em *cop top*, ou [i], como em *kip tip*). Também foram incluídos, no experimento, estímulos-controle, que eram construções com mesma consoante inicial – *cop cop*, *top top*, *kip kip* e *tip tip*.

Nos resultados, o que se pôde observar foi que essas frases-controle, sem alternância de consoantes, mostraram os gestos esperados como em outras produções sem erros. Isso quer dizer que as produções de [t] apresentaram o gesto esperado de elevação de ponta de língua, sem atividade de dorso de língua. Da mesma forma, as produções de [k], em frases-controle, apresentaram o seu gesto de dorso próprio, mas não foi observada atividade de ponta de língua. No entanto, o mesmo não acontece para produções com consoantes alternantes.

Na Figura 11, são apresentados dados de produção de *cop top* por um dos sujeitos. De cima para baixo, as faixas representam: o oscilograma de uma das repetições (áudio); deslocamento vertical de ponta de língua; deslocamento vertical

de dorso de língua. As intrusões gestuais estão indicadas por setas. As 14 repetições estão representadas por curvas, as quais se encontram sobrepostas.

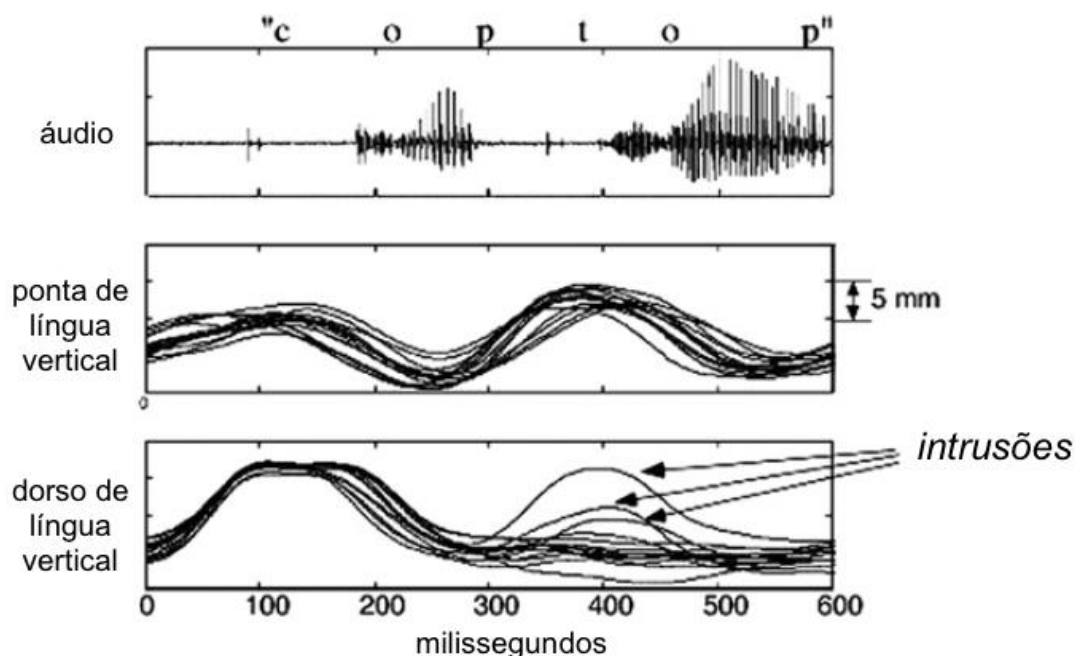


Figura 11 – Sobreposição de 14 repetições de *cop top*, produzidas por um falante de inglês. Imagem adaptada de Goldstein *et al.* (2007, p. 393)

As curvas observadas na imagem representam o deslocamento vertical captado pelos transdutores. Se a frase fosse produzida sem erros, o [t] de *top* não apresentaria movimentos substanciais de dorso de língua. No entanto, em algumas repetições, há uma cópia extra de gesto de dorso, ativada durante o gesto de ponta de [t]. Esses gestos extra estão sinalizados na Figura 11 pelas setas e serão chamados de *erros de intrusão gestual*. As intrusões são caracterizadas pela “adição de um gesto que não é produzido naquela localização temporal em uma produção normal, não-errônea”, na definição de Goldstein *et al.* (2007, p. 392, tradução nossa¹¹).

Ainda, nos dados, foram encontrados os chamados *erros de redução gestual*, que são definidos como “uma redução inapropriada na magnitude de um movimento

¹¹ No original: “addition of a gesture not produced at that temporal location in a normal, non-errorful production”.

articulatório pretendido”, conforme Goldstein *et al.* (*op. cit.*, tradução nossa¹²), dando-se, como exemplo, uma menor elevação do dorso de língua durante a produção de um [k]. Embora os erros de redução tivessem ocorrido com menor frequência, tanto intrusão quanto redução foram observados em [t] e [k] nas repetições com consoantes alternantes, em produções de todos os sujeitos.

Essa forma de descrição dos erros se apóia em dados mais precisos dos movimentos articulatórios ocorridos e é uma alternativa às descrições encontradas nos estudos com base em transcrições fonéticas, nos quais o tipo de erro mais comumente reportado é a *substituição* de um segmento por outro. Para Goldstein *et al.* (2007), uma verdadeira substituição apareceria, no estudo realizado, como a total redução de um gesto, acompanhada por uma total intrusão de um gesto errôneo. Os dados mostraram, no entanto, que o padrão dominante de produção de erros é bastante diferente. Na Figura 12, são mostrados os tipos de erro realizados por sujeito.

¹² No original: “an inappropriate reduction in the magnitude of an intended articulatory movement”.

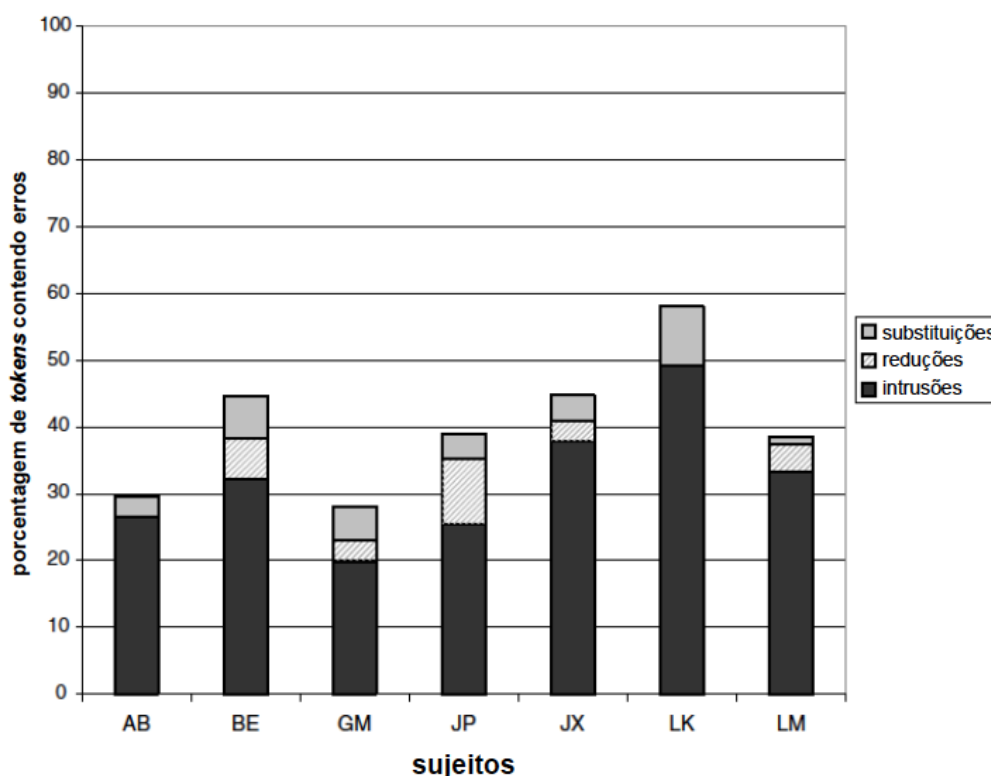


Figura 12 – Distribuição dos tipos de erro ocorridos na produção dos sete sujeitos. Imagem adaptada de Goldstein *et al.* (2007, p. 394)

A partir dos dados, observa-se uma dominância sistemática da intrusão sobre a redução, o que resulta em um padrão de adição de um gesto sem apagamento de outro. Esse fenômeno corrobora a explicação dada pela Fonologia Articulatória de que os casos tradicionalmente descritos como substituições – termo que denota as trocas completas de um segmento por outro – são, na verdade, processos parciais, envolvendo a redução parcial de um gesto e a intrusão de outro.

O estudo investigou o papel do tempo e da velocidade de fala na ocorrência de erros e, a partir de uma análise de variância, viu-se que o número de erros variou de forma significativa em função tanto do tempo (número da repetição na sequência) quanto da velocidade de fala, guiada pelos diferentes andamentos do metrônomo. No estudo, também foram analisados estatisticamente o papel da vogal presente nas construções ([a] ou [i]), do acento (se na primeira ou segunda palavra) e da posição na construção (se a consoante estava na primeira ou segunda palavra). Desses fatores, apenas o da vogal mostrou-se significativo. As frases com [i]

continham significativamente mais erros do que as com [a], o que pode ter a ver com a compatibilidade da constrição realizada para [ɪ] com as constrições de [t] e de [k].

Na discussão dos dados, os autores observaram que os erros não podem ser interpretados como fruto, unicamente, de um processo de substituição de segmentos, mas sim costumam envolver a produção de mais de um gesto – um apropriado e outro intrusivo. A tendência a uma intrusão em detrimento de uma redução poderia parecer, à primeira vista, algo enigmático, pois uma explicação reiterada pela literatura é de que os falantes tendem a minimizar esforço articulatório. No entanto, nos erros de produção observados no experimento de Goldstein *et al.* (2007), observou-se o contrário: foram produzidos mais gestos do que o necessário.

Observado o fenômeno da intrusão gestual, os autores oferecem uma explicação do porquê da sua ocorrência. A tendência à intrusão seria explicada com base no comportamento de osciladores acoplados, no sentido de que a ativação de gestos individuais pode ser orquestrada de acordo com uma dinâmica de planejamento oscilatória intergestual. Por exemplo, em *top top*, o gesto de ponta de [t] e o gesto labial de [p] têm comportamentos oscilatórios e possuem uma frequência 1:1 um em relação ao outro. Isso quer dizer que um ciclo de um constritor está associado a um ciclo de outro.

Esse modo 1:1 é conhecido como o mais estável dentre as possíveis configurações de frequência, estando menos sujeito a erros. E, em certas condições, como uma maior velocidade de fala, modos de frequência mais complexos, como 2:1, por exemplo, exibirão transições para modos mais simples e estáveis. A intrusão pode ser interpretada como um sistema de natureza mais instável sendo capturado pelo modo 1:1, mais estável. Um exemplo dessa transição é fornecido pelos autores e está presente na Figura 13, a seguir.

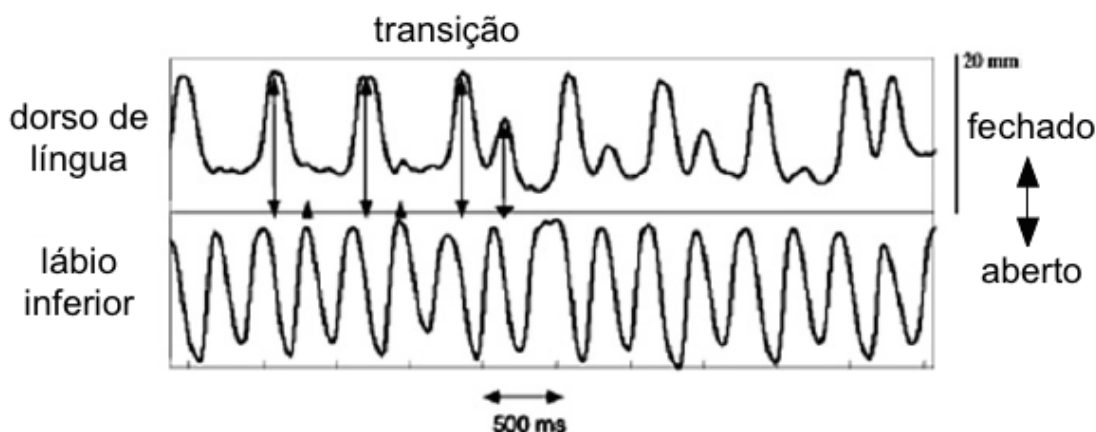


Figura 13 – Repetições de *cop top*, com transição de 2:1 para 1:1 no acoplamento de dorso de língua e lábio inferior. Imagem adaptada de Goldstein *et al.* (2007, p. 399)

Na Figura 13, pode-se observar, nas primeiras produções, que, para cada dois gestos de lábio inferior, é produzido um gesto de dorso de língua. Embora o deslocamento de ponta de língua não esteja representado na imagem, é possível inferir-se a mesma proporção para a frase *cop top*: para cada dois gestos de lábio inferior, é produzido um gesto de ponta de língua.

O que se nota, ao longo das repetições mostradas na Figura 13, é que vão aparecendo curvas extras em trechos em que o dorso não teria atividade. O deslocamento de dorso de língua começa a aparecer com maior frequência ao ponto de se estabelecer uma proporção de um para um: passa-se a observar, no transcorrer das repetições, que é produzido um gesto de dorso para cada gesto labial. Isso se deve a uma transição, a qual ocorre em direção a um atrator, que será um modo dinâmico oscilatório mais estável – neste caso, 1:1.

Assim, de modo geral, os erros de produção de fala podem ser explicados como um resultado da interação entre modos de frequência que são intrinsecamente estáveis e os modos de coordenação lexical próprios da língua que foram aprendidos pelo falante. Esses modos aprendidos têm diferentes padrões de estabilidade e, quando sujeitos a fatores como a velocidade da fala, podem passar por transições para modos mais estáveis.

A partir dos resultados do estudo, concluiu-se que os erros de fala, longe de serem distorções aleatórias, constituem padrões, envolvendo unidades de constrição

linguisticamente significativas. Esses erros têm gradiência, pois podem variar em magnitude, num contínuo, e são sensíveis à velocidade de fala.

É evidente que os erros de fala observados nos dados de adultos são de natureza distinta dos erros que possam ser observados na fala infantil, visto que o adulto já domina o funcionamento da língua e a criança, por sua vez, está em processo de apropriação do sistema e experimentando possibilidades de seu uso. No entanto, alguns fenômenos observados em produções errôneas de adultos também são constatados na produção de crianças em processo de aquisição. Fenômenos interpretados como apagamentos de segmentos podem ser reinterpretados, já que os gestos, apesar de não serem percebidos auditivamente, podem estar presentes.

Dessa forma, é possível pensar em produções da fala infantil com mudanças percebidas no nível segmental que passem, na verdade, por mudanças a nível de gesto, as quais se devem à sobreposição ou intrusão de gestos adjacentes ou à redução de sua magnitude. Para uma estrutura CCV, produções percebidas como errôneas podem evidenciar, a partir de dados articulatórios, gestos relacionados ao segundo elemento consonantal, ainda que em menor magnitude. Tal fato indicia a construção de uma rotina articulatória relacionada à estrutura CCV e, por conseguinte, as evidências de uma aquisição em curso dessa estrutura ocorrendo em etapas mais precoces do que o reportado por uma literatura baseada em descrições de ouvintes. Neste trabalho, a análise de dados longitudinais é empregada justamente com o objetivo de se observar o progresso na aquisição das rotinas articulatórias de CCV, bem como o progresso na coordenação gestual necessária para a realização do *tap*.

Na seção 2.2, a seguir, serão descritos estudos que tematizaram a aquisição da sílaba CCV, conduzidos a partir de diferentes perspectivas teóricas e metodologias.

2.2 Estudos em aquisição da sílaba CCV

Este trabalho se propõe a estudar a aquisição de encontros consonantais com base em pressupostos da Fonologia Articulatória, modelo teórico abordado na seção 2.1. Sendo assim, é importante que se recuperem estudos sobre a aquisição

da estrutura CCV, conduzidos a partir de diferentes propostas teóricas e metodologias.

Esta seção possui três subseções, dedicadas a linhas distintas de estudos relacionados à aquisição da sílaba CCV. A seção 2.2.1 traz estudos envolvendo a aquisição dessa sílaba desenvolvidos a partir de análises de outiva. Na seção 2.2.2, são abordados estudos em aquisição de CCV que utilizam a análise acústica nas suas metodologias, investigando a duração vocálica como indicativo de uma percepção da criança sobre a estrutura silábica, mesmo quando não produzida corretamente. Por fim, a seção 2.2.3 é dedicada a um estudo em aquisição de CCV que utiliza análises articulatórias, por meio da ultrassonografia, como metodologia de pesquisa.

2.2.1 Estudos em aquisição da sílaba CCV com análises de outiva

As primeiras pesquisas em aquisição da sílaba CCV no português brasileiro foram desenvolvidas com análises dos dados de fala por meio do julgamento de outiva. Assim, a qualidade da produção da criança era determinada a partir da percepção auditiva do pesquisador sobre o dado de fala. No caso de produções de palavras com sílaba CCV, por exemplo, ouvir o dado por um número determinado de vezes era o procedimento utilizado para se definir o que, afinal, a criança havia produzido – se realizou a estrutura silábica corretamente, se omitiu a líquida ou mais elementos da sílaba ou se apresentava quaisquer produções diferenciadas, as quais frequentemente se classificavam como *estratégias de reparo*. Logo, o registro e descrição dos fenômenos ocorridos nos dados era feito a partir desse procedimento.

De modo geral, trabalhos de aquisição fonológica que observam o processo de aquisição de um ou mais segmentos e/ou estruturas baseando-se em análises de outiva utilizam o critério da porcentagem para que a aquisição seja considerada completa. Para Lamprecht (2004, p. 23), por exemplo, “é necessário ter-se um critério de proporção de acertos de produção a partir do qual essa afirmação possa ser feita”. A criança não precisaria atingir um percentual de 100% de acertos, pois parte das produções consideradas inadequadas constitui, nas palavras da autora, “resquícios de etapas já superadas ou, até mesmo, simples lapsos de língua” (*op. cit.*). O percentual mínimo estabelecido tem certa variação nas pesquisas:

Lamprecht (1990) adotou 75%; Mezzomo (1999) e Bonilha (2000) consideraram o percentual de 80%; Hernandorena (1990) e Miranda (1996) determinam o percentual de 85%; Azambuja (1998) e Savio (2001) estabeleceram 86%; e Hernandorena e Lamprecht (1997) estabeleceram o percentual de 90%, conforme lista comparativa encontrada em Lamprecht (2004, p. 23). A partir desses e outros estudos elencados pela autora, conclui-se que grande parte dos trabalhos com outiva baseados no critério da porcentagem adotaram a faixa de 80% a 86%.

Os primeiros trabalhos relacionados à aquisição da estrutura CCV versaram sobre a aquisição fonológica do português de forma geral, abordando um número de estruturas silábicas e segmentos de modo a se traçarem perfis de aquisição. É o caso dos estudos de Lamprecht (1990) e Bonilha (2005), entre outros.

O trabalho de Bonilha (2005) é um estudo de caso, realizado a partir de dados longitudinais de uma criança, coletados entre as idades de 1;1 e 3;9 (anos;meses). A autora observa a aquisição do português em relação às unidades do segmento, da sílaba e do acento e busca analisar os fenômenos de aquisição à luz de uma Teoria da Otimidade (TO) baseada em noções conexionistas. E, como parte dos objetivos específicos, busca identificar a ordem de aquisição de segmentos, constituintes silábicos e padrões silábicos no português. Para tanto, a autora observou todas as produções realizadas pelo sujeito e, em seguida, destacou todas as possibilidades de produção em termos de estruturas silábicas e de acento primário, controlando variáveis dependentes e independentes que foram consideradas relevantes para o estudo e interpretando os dados com base na versão da TO proposta no trabalho.

Quanto à aquisição do *onset* complexo, foram controladas duas variáveis dependentes e três variáveis independentes. As variáveis dependentes controladas no estudo foram o constituinte silábico e padrões silábicos. Os padrões silábicos controlados foram vários, constituindo todas as possibilidades no português da realização de uma sílaba com dois elementos consonantais iniciais, como CCV, CCVV, CCVC, entre outros. Já as variáveis independentes, todas de ordem linguística, foram, conforme Bonilha (2005, p. 107):

- a) tipo de segmento vocálico em núcleo (e.g. vogal baixa, vogal média-baixa, etc.);

- b) sequência de segmentos que formam *onset* complexo (e.g. oclusiva + líquida lateral, oclusiva + líquida não-lateral, fricativa + líquida não-lateral, etc.);
- c) classificação do *onset* (inicial, medial ou final).

A autora descreve os resultados referentes à aquisição do *onset* complexo, indicando as principais estratégias adotadas em diferentes etapas da aquisição e as idades em que cada tipo de *onset* complexo tem sua aquisição concluída. Essas informações estão detalhadas no Quadro 4.

Idade	[obstruente + líquida lateral]	[obstruente + líquida não-lateral]	[gw]	[kw]
1:1:22 – 2:1:27	CCV→CV	CCV→CV	CCV→CV	CCV→CV
2:8:16	adquirida	CCV→CV	CCV→CV	CCV→CV
2:8:16 – 2:9:16		CCV→CCV	CCV→CV	CCV→CV
3:0:21		adquirida	adquirida	CCV→CV
3:1:20				adquirida

Quadro 4 – Aquisição dos tipos de *onset* complexo nos dados de G., informante da pesquisa. Quadro adaptado de Bonilha (2005, p. 151)

Para explicar o ordenamento da aquisição dos diferentes tipos de *onset* complexo, a autora considera a escala de sonoridade proposta por Clements (1990). Numa escala de sonoridade, classes de segmentos têm níveis de sonoridade distintos e, na proposta de Clements (1990), a classe das líquidas tem um nível de sonoridade intermediário; acima dessa classe, encontram-se os *glides* e, em seguida as vogais, que detêm nível máximo de sonoridade na escala. Considerando-se a tendência apontada por Clements (1990) de um distanciamento, em sonoridade, entre *onset* e núcleo na composição da sílaba, é possível explicar-se a aquisição de estruturas CCV com líquidas antes das CCV iniciadas com [kw] e [gw].

No entanto, a classe das líquidas, no modelo de Clements (1990), engloba tanto líquida lateral quanto não lateral, o que não explica, nos dados do Quadro 4, por que a aquisição de CCV com líquida não lateral se completa após a de CCV com a líquida lateral. Para explicar esse ordenamento, a autora recorre a uma escala de sonoridade mais especificada, conforme proposta de Bonet e Mascaró (1996). Nessa proposta, o *tap* tem nível de sonoridade 4, o mesmo dos *glides*, estando, assim, atrás apenas das vogais. A lateral, por sua vez, tem nível de sonoridade 3.

Considerando-se que a lateral oferece um contexto de maior distanciamento em relação à vogal, em nível de sonoridade, assume-se, assim, que sílabas CCV com líquida lateral constituam um contexto mais favorável para a produção. Isso não só explica a maior facilidade de aquisição do encontro consonantal com líquida lateral como também justifica as substituições de líquida não lateral por lateral, verificadas nos dados do informante do estudo de Bonilha (2005). Essas substituições de líquida também foram descritas em outros estudos, principalmente aqueles específicos sobre encontros consonantais (RIBAS, 2002; 2004; MIRANDA, 2007; BARBIERI; FERREIRA-GONÇALVES, 2017, entre outros).

Outros estudos que tematizam a aquisição do português de modo mais abrangente referem o processo de aquisição de CCV. No entanto, será dada continuidade a esta seção fazendo-se referência aos trabalhos que abordam a aquisição de encontros consonantais como tema específico de pesquisa.

Dentre os primeiros estudos dedicados à aquisição de CCV no português brasileiro, destacam-se os trabalhos de Ribas (2002; 2004).

Ribas (2002) estuda o processo de aquisição de sílabas CCV constituídas pelas líquidas não-lateral e lateral a partir de dados de fala de crianças com desenvolvimento fonológico normal. O *corpus* de análise são dados transversais de crianças entre 2;0 e 5;3, mas, adicionalmente, também foram consultados dados de crianças entre 1;0 e 1;11, de modo a verificar como essas crianças lidavam com a estrutura em termos de estratégias de produção. O trabalho também utilizou o critério da porcentagem – a autora adota, especificamente, a porcentagem de 85% de produções corretas em duas faixas etárias consecutivas para se considerar completa a aquisição do *onset* complexo.

Os dados são analisados em termos da quantidade de produções corretas de CCV e também são estudadas variáveis nesse processo, com o auxílio de um programa de análise de variáveis. São levados em conta fatores linguísticos, como contexto precedente, contexto seguinte, posição na palavra e tonicidade, e dois fatores extralinguísticos: idade e sexo. A variável dependente foi a produção dos *onsets* complexos, a partir da qual foram determinadas quatro variantes: produção correta, produção não correta (C¹V), produção com substituição de líquida e produções outras, consideradas pouco comuns para o *corpus*.

Eventualmente, as duas últimas variantes foram incorporadas à de produção não correta por terem atingido, na análise feita, “um número muito pequeno de

dados”, segundo Ribas (2002, p. 36). Exemplos dessas produções pouco comuns se encontram no Quadro 5, a seguir:

Produções pouco comuns	Exemplos
Metátese	Livro → [lirvu]
Epêntese	Branco → [barãŋku]
Semivocalização	Prego → [pyɛgu]
Não realização das consoantes do <i>onset</i> complexo	Procurar → [okuya]
Não realização da sílaba com <i>onset</i> complexo	Estrelinha → [liña]
Substituição da obstruinte	Bruxa → [pruša]
Produção C ² V	Placa → [laka]
Coalescência	Blusa → [duza]

Quadro 5 – Produções incomuns como estratégias de realização de sílabas CCV. Quadro adaptado de Ribas (2002, p. 36)

Com base em uma análise de outiva, esses casos foram encontrados em número consideravelmente menor do que os de produção C¹V. No estudo, embora seja mencionada tal variedade de estratégias, o aspecto principal que define o progresso na aquisição é a porcentagem de produções corretas, com a influência de fatores favorecedores ou desfavorecedores. A partir de mais de 2.000 dados de produção de CCV, a autora determina os fatores facilitadores para a produção dos *onsets* complexos, sugerindo, inclusive, palavras-estímulo para terapias fonoaudiológicas envolvendo essa estrutura. No caso das sílabas com a líquida não-lateral, os fatores favorecedores foram a presença de plosiva labial sonora, /b/, como primeiro elemento consonantal do encontro, ser uma sílaba medial, com /o/, /e/ ou /a/ antecedendo o *onset* e ser uma sílaba fraca no pé métrico do acento.

Também, são descritos os aumentos e diminuições nas porcentagens de produção correta conforme as faixas etárias, a partir dos quais se observam regressões e progressos no caminho para a aquisição completa. As estratégias de reparo seriam aplicadas em ambientes específicos, como foi o caso da epêntese para os grupos com consoantes desfavorecedoras; assim, uma palavra como “trem”, que contém uma obstruinte coronal, tida no estudo como desfavorecedora para a produção, sofreria a inserção de um som, uma vogal, que resultaria em uma sequência de sílabas constituídas por *onsets* simples – [te.ˈrẽỹ].

Não apenas essas estratégias são tidas como aplicadas em contextos específicos, como também são consideradas à parte de estágios de aquisição:

Não é possível dizer que as crianças, ao adquirirem o *onset* complexo, passam por etapas intermediárias ao longo do curso do desenvolvimento. As estratégias de reparo são recursos individuais, usados em menos de 5% do total de ocorrências, não podem ser generalizadas a todas as crianças, nem constituir (*sic*) um estágio para a aquisição da sílaba CCV. Pode-se afirmar, então, que as crianças produzem C¹V e depois produzem CCV corretamente. (*ibid.*, p. 143)

Assim, os únicos estágios considerados seriam i) C¹V, produção do encontro consonantal sem a consoante líquida, e ii) CCV, produção correta, correspondente ao alvo. Em Ribas (2004), também são descritas várias estratégias adotadas pelas crianças, mas as etapas de aquisição consideradas também são C¹V e CCV, havendo variabilidade nas estratégias adotadas. O aspecto gradual é encontrado, novamente, nas porcentagens de produções corretas ao longo do processo.

Trabalhos desenvolvidos com o uso de recursos metodológicos como a análise acústica e a análise articulatória, por meio da ultrassonografia, e trabalhos embasados em modelos emergentistas de aquisição propiciaram novas compreensões acerca do processo de aquisição de CCV. Aspectos como a duração segmental passaram a ser observados em maior detalhe e passou-se a considerar a existência de contrastes encobertos, os quais não se detectam por meio da oitiva, mas que constituem estágios na aquisição além dos anteriormente considerados.

As subseções seguintes serão dedicadas a esses estudos mais recentes, os quais propõem novas descrições da aquisição de CCV a partir de outras perspectivas teóricas e metodologias.

2.2.2 Estudos em aquisição da sílaba CCV com análises acústicas

A aquisição de encontros consonantais ganhou novas descrições com trabalhos como os de Mezzomo *et al.* (2008), Miranda (2007) e Miranda e Silva (2011). Mezzomo *et al.* (2008) utilizam a análise acústica para o estudo da estratégia de alongamento compensatório nos casos de simplificação de *onset* complexo, na fala de 28 crianças, entre 1;0 e 8;0, com desenvolvimento fonológico normal e com desvios.

As autoras partem da ideia de aquisição de CCV presente em Ribas (2002; 2004), isto é, embora leve-se em conta a variedade de estratégias de reparo adotadas pelas crianças, considera-se que o processo ocorra, fundamentalmente, em dois momentos: C¹V e CCV. No entanto, investigam uma outra estratégia, a do alongamento compensatório de vogal, a partir da análise acústica, tendo por base estudos como o de Mezzomo *et al.* (2004), em que o alongamento de vogal e o alongamento de fricativa foram percebidos a partir da análise acústica. O grupo de Mezzomo *et al.* (2004) era composto de quatro crianças com desvio fonológico evolutivo; já Mezzomo *et al.* (2008) analisaram a duração de vogal em um grupo de crianças com aquisição normal e outro grupo com desvios, de modo a averiguar se o alongamento era um recurso empregado exclusivamente por crianças em aquisição atípica.

Pode-se compreender que, no trabalho, a análise acústica é considerada um método complementar à análise de ouvida, “fornecendo *insight* naquelas áreas em que a percepção do adulto falha” (MEZZOMO *et al.*, 2008, p. 36); também é apontado pelas autoras que, em estudos sobre aquisição normal e com desvios, “[a] análise espectrográfica tem sido empregada como suplemento à análise perceptual” (*op. cit.*).

As imagens fornecidas pelo espectrograma, por sua vez, são consideradas correlatos diretos do conhecimento linguístico da criança:

A espectrografia indica o conhecimento linguístico da criança, já que releva ensaios, buscas e aproximações, demonstrando que ela sabe que não está produzindo o som desejado, mas procura fazer o possível para “acertar”. (*op. cit.*)

Um espectrograma é, em princípio, uma representação de informações acústicas em imagem, a qual deve ser interpretada conforme parâmetros previamente definidos, tendo-se, a partir dessa interpretação, uma inferência sobre aquilo que é efetivamente produzido. E é essa produção que fornecerá indícios de mudanças no conhecimento linguístico da criança, por meio de comparações, seja ao sistema a ser adquirido, a outras produções daquele mesmo indivíduo e/ou a produções relacionadas à aquisição completa. Sendo assim, a importância dos espectrogramas (e oscilogramas) na análise acústica não seria a de revelar, por si

sós, aspectos do conhecimento linguístico de um indivíduo, mas sim a de tornar acessíveis os dados sonoros em um meio como a imagem, permanente e passível de uma análise mais atenta e detalhada. Essa “tradução” do som em imagem é o que, afinal, possibilitará as inferências e comparações que de fato levam a considerações sobre o conhecimento linguístico do falante.

Isto à parte, o trabalho se prova relevante ao mostrar a diferença considerável entre as análises de outiva e acústica em termos da frequência de detecção do fenômeno em questão em um mesmo *corpus*. Os dados da pesquisa, baseados na produção de palavras que contrastavam em termos da presença de sílaba CCV ou CV, foram analisados por meio de outiva e de acústica e, na comparação entre essas duas formas de análise, houve, nos dados analisados acusticamente, maior ocorrência do alongamento compensatório – o fenômeno foi detectado em 67,70% dos dados por meio da análise acústica e em apenas 4,68% dos dados via julgamento de outiva. Com base nessa comparação de análises, as autoras reforçam a importância da realização da acústica pela sua precisão na análise da fala.

Também comparam a utilização do alongamento por parte das crianças com desenvolvimento normal e com desvios, percebendo comportamentos semelhantes quando o *onset* se inicia por plosiva, contexto já apontado como favorável em Ribas (2002): 100% dos indivíduos com desvios utilizavam o alongamento, em comparação a 92,86% dos indivíduos com desenvolvimento normal. No entanto, apresentaram comportamentos diferentes nos *onsets* iniciados por fricativas: 78,59% das crianças com desvios empregavam o alongamento, contra 7,14% das crianças com desenvolvimento normal. O trabalho não só reforçou o papel do contexto linguístico enquanto favorecedor ou desfavorecedor para a produção como também forneceu o indício de que a aquisição de aspectos prosódicos possa preceder a aquisição segmental. Essa suposição é feita a partir dos dados em que se verifica o alongamento, dados nos quais as crianças parecem manter a posição de *onset* complexo na camada temporal, muito embora ainda não possuam as rotinas motoras para produzirem a forma correspondente ao alvo.

A importância da análise acústica, o potencial desse recurso metodológico nas análises linguísticas e, adicionalmente, a perspectiva teórica de aquisição de linguagem adotada são pontos que carecem de uma definição mais adequada em Mezzomo *et al.* (2008). No entanto, o trabalho é de considerável relevância, sendo

um dos primeiros a investigar o papel da duração segmental na aquisição de CCV a partir de resultados de pesquisa fonoaudiológica. Constitui, assim, parte da literatura que se dedica a estudar fenômenos de aquisição que não se percebem somente a partir da oitiva do pesquisador.

Outros trabalhos em aquisição de encontros consonantais foram, à época, desenvolvidos com a utilização da análise acústica, desta vez claramente fundamentados em modelos emergentistas de aquisição. Miranda (2007) e Miranda e Silva (2011) investigaram o tema com base na Fonologia de Uso e na Teoria de Exemplares.

Miranda e Silva (2011) debatem sobre a natureza multirrepresentacional da linguagem ao analisar a aquisição de encontros consonantais tautossilábicos no português brasileiro, com base em dados de falantes de Belo Horizonte (MIRANDA, 2007). Abordam a natureza do contraste encoberto, “formulado pela criança como estratégia de construção da linguagem” (MIRANDA; SILVA, 2011, p. 14), a relevância de técnicas da Fonologia de Laboratório para estudos em fonologia e a incorporação do detalhe fonético às representações dos sons.

As autoras recuperam estudos que apontam a tendência de uma substituição de sílabas CCV por CV na fala da criança. Com base em tais estudos, salientam que se poderia interpretar que a criança não possui esse contraste, prevendo-se que surgiria abruptamente em algum momento do processo de aquisição. Essa compreensão está relacionada a visões determinísticas, em que um som ou padrão está presente ou ausente das representações.

Miranda e Silva (2011) apresentam, então, com base na Fonologia de Uso e na Teoria de Exemplares, uma proposta não determinística da aquisição de encontros consonantais, sugerindo que “as representações linguísticas contêm detalhes fonéticos finos e que o contraste emerge da capacidade cognitiva do indivíduo em abstrair sobre o conhecimento da língua a que está exposto” (*ibid.*, p. 16). Destacam os principais pressupostos da Fonologia de Uso e da Teoria de Exemplares, dentre eles o de que a experiência e o uso influenciam na organização e no gerenciamento do conhecimento sobre a língua, afetando, assim, as representações. Nos modelos multirrepresentacionais, as categorias não seriam discretas, como fonemas ou traços, mas sim gradientes.

As autoras discutem as características dos contrastes encobertos (SCOBIE *et al.*, 1996), os quais não se percebem auditivamente, mas podem ser detectados

por meio de análise acústica e/ou articulatória. Sendo aspectos gradientes finos, necessitam desses recursos para o seu estudo e compreensão. À luz do conceito de contraste encoberto, as autoras discutem resultados do estudo de Miranda (2007), concentrando-se na duração da vogal que constitui a sílaba CCV. Assim, foi investigado se as crianças que não tinham adquirido o encontro consonantal alongariam a vogal para, de certa forma, compensar a ausência da líquida. A duração, assim, seria uma propriedade fonética fina a serviço de estabelecer o contraste entre CCV e CV na fala em aquisição. Também, segundo as autoras,

a estratégia de alongamento compensatório, neste caso, propicia as condições físicas para que a criança, em algum momento, possa incorporar a produção da consoante líquida. Ou seja, ao produzir a vogal mais longa, a criança cria o ambiente para se introduzir o tepe e, eventualmente, com o aprimoramento das rotinas motoras, será possível produzir encontros consonantais tautossilábicos. (MIRANDA; SILVA, 2011, p. 19).

Amplia-se, desse modo, a compreensão sobre o alongamento, o qual já era definido pela sua propriedade compensatória em Mezzomo *et al.* (2008). Isso porque, em Miranda e Silva (2011), ele passa a ser considerado como uma forma de preparação para rotinas motoras futuras. Esse raciocínio se afina com a perspectiva da Fonologia Articulatória: pode-se interpretar que, embora os gestos necessários para a produção do *tap* não estejam presentes na rotina articulatória da criança, a temporalidade é um aspecto preservado na produção que tem esse alongamento.

A partir de dados de um grupo experimental, que ainda não tinha adquirido os encontros consonantais, e um grupo controle, que já os produzia sistematicamente, Miranda (2007) analisou a duração das vogais em pares mínimos, como “broa”/“boa”, “prato”/“pato” e “pressa”/“peça”. As medidas foram extraídas com o uso de um programa de análise acústica e, com base nas médias de duração de vogal, observaram, nos dados das crianças em processo de aquisição dos encontros consonantais, que as vogais em sílabas CCV pronunciadas como CV tinham duração sistematicamente maior do que as vogais em sílabas CV.

Esse resultado, na compreensão de Miranda e Silva (2011), constitui um caso de contraste encoberto. As crianças produziam de fato o contraste entre CCV e CV, mas ele não era percebido pelos adultos do seu entorno, da sua comunidade de fala. Esse dado relativo à duração também reflete uma categorização dos padrões

silábicos de CCV e CV, de forma diferente daquela que a comunidade reconhece: o alongamento é considerado “o correlato físico utilizado pela criança para expressar o alvo de uma sílaba CCV” (*ibid.*, p. 22). Já nos dados de crianças que produziam CCV corretamente, as durações de vogal foram aproximadas, com diferença na complexidade da sílaba CCV.

Os resultados apresentados em Miranda e Silva (2011) são evidências de que informações a princípio redundantes, expressas por meio de detalhes fonéticos, têm relevância na organização do conhecimento fonético-fonológico da criança. O aspecto da duração, analisado em pares mínimos na fala infantil, em crianças sem e com a produção sistemática da estrutura, é uma dessas informações, evidenciando não apenas a presença de um contraste no repertório fonológico da criança, mas também um possível meio de construção de uma rotina motora relacionada à produção do *tap*. Partindo-se desses pressupostos, a duração é um aspecto abordado neste trabalho, desta vez por meio de uma análise de dados longitudinais de fala infantil. Este trabalho também aborda a duração na aquisição de modo diferenciado por contar com dados de fala adulta, os quais serão importantes para comparações e para se verificar o que pode ou não ser exclusivo da fala em aquisição.

A seção 2.2.3, a seguir, é dedicada a um estudo sobre a aquisição de CCV conduzido com análises articulatórias de dados de ultrassom.

2.2.3 Estudos em aquisição da sílaba CCV com análises ultrassonográficas

Para discorrer sobre a aquisição de CCV no português brasileiro, tendo-se por base a coleta e análise de dados ultrassonográficos, serão aqui reportados alguns resultados do estudo de Vassoler (2016).

O objetivo geral do trabalho foi investigar e descrever o padrão de coordenação gestual envolvido em produções de sílabas CCV de crianças com desenvolvimento típico e atípico, utilizando, para tanto, a coleta e análise de dados ultrassonográficos. Já os objetivos específicos foram caracterizar diferenças na coordenação gestual de sílabas CCV produzidas por crianças típicas e atípicas e verificar em que medida fenômenos considerados como reduções de encontros

consonantais em crianças atípicas poderiam, na verdade, apresentar contrastes encobertos.

O modelo teórico que serviu de base para o trabalho foi a Fonologia Gestual, por meio do qual se fornece uma explicação para a aquisição relativamente tardia de CCV e os “erros” ocorridos no processo. Adotando-se o ponto de vista defendido na FonGest, o “erro” emergiria a partir de um processo dinâmico, podendo ser consequência de deslizos na coordenação de gestos, de desajustes de *timing* ou de mudanças na magnitude de ativação dos gestos. Assim, o que se considera como “erros” são produções que podem apresentar sobreposição de gestos ou ativação parcial de um ou mais gestos. Esses erros, tanto na fala adulta quanto na fala infantil, fornecem indícios de como a coordenação acontece, por exemplo, em diferentes padrões silábicos.

Assim, o estudo de Vassoler (2016) foi desenvolvido a partir de três hipóteses. Na primeira, assumiu-se que a produção de sílabas CCV e CV de crianças típicas apresentaria diferentes medidas, tanto nos dados ultrassonográficos quanto acústicos. Na segunda, assumiu-se que as medidas de dados ultrassonográficos e acústicos diferenciariam a condição clínica das crianças. E, na terceira hipótese, assumiu-se que crianças com produções atípicas poderiam apresentar diferenças nas medidas articulatórias e acústicas, mesmo que sílabas CCV desse grupo de crianças tenham sido classificadas como CV a partir da outiva.

Para a investigação dessas hipóteses, foram realizadas análises acústica e articulatória de dados de fala infantil. A pesquisa contou com 10 sujeitos, dos quais 5 eram crianças com perfil de aquisição típico e 5 eram de perfil atípico. Foram coletados dados de fala, em áudio e ultrassom, de 9 pares mínimos, nos quais a diferença era o tipo de sílaba inicial, se CCV ou CV – e.g. “troco” e “toco”.

A metodologia foi estruturada em três partes: análise de outiva dos dados de fala – realizada por três juízes –, análise acústica e julgamento qualitativo e quantitativo das imagens de ultrassom, este também realizado por juízes. Na análise qualitativa de ultrassom, os juízes deveriam avaliar a natureza do gesto de ponta de língua: se era gesto ausente ou presente (com maior ou menor magnitude) ou se o contorno não era visível. Também deveriam avaliar o gesto de corpo de língua como de maior ou menor magnitude. Para se chegar a um resultado, deveria haver concordância entre pelo menos dois dos três juízes envolvidos.

A análise quantitativa das imagens de ultrassom foi feita em medidas de distância e de área. Na análise quantitativa, foi selecionado um *frame* relativo ao final do *onset* e início da vogal e foram considerados três *frames* antes e três após o *frame* de referência, resultando, assim, em um período de 7 *frames* para análise. A partir desse período de 7 *frames*, foi selecionado aquele que registrasse o momento de máxima constrição do gesto de ponta e, a partir desse *frame* específico, foram realizadas medidas, em mm, das distâncias de ponta, lâmina e dorso de língua até o limite inferior da imagem de ultrassom. Nessa imagem, foram analisados os gestos de ponta e corpo de língua, normalmente envolvidos na produção do *tap*. Também foi feita uma medida de área entre ponta e lâmina de língua.

Após a realização das medidas, os dados passaram por análise estatística. Como resultado, verificou-se que, independentemente da condição clínica, a duração de sílabas CCV é maior que a de sílabas CV. A partir da estatística, o fator palavra foi apontado como significativo – um fato esperado, pois houve variação em contexto vocálico das sílabas e no tipo de consoante em posição de C₁.

Quanto às hipóteses formuladas, a primeira foi corroborada, pois, nas produções de crianças típicas, as sílabas CCV e CV tiveram diferentes medidas ultrassonográficas e acústicas. Já a segunda foi parcialmente confirmada, visto que alguns raios apresentaram valores superiores em crianças atípicas. E a terceira também foi parcialmente confirmada, pois, nas produções de crianças atípicas, algumas medidas foram diferentes, ainda que sílabas CCV tivessem sido julgadas como CV.

De modo geral, os resultados da pesquisa sugerem uma presença de gesto de elevação de ponta de língua na produção do *tap*, verificável pelos dados articulatórios, a não sobreposição de gestos em sílabas CCV, verificável pelos dados acústicos, e a presença de gestos indiferenciados entre ponta e corpo de língua nas produções CCV de crianças atípicas.

Os resultados encontrados pela autora, especificamente de fala infantil típica, foram importantes para o estabelecimento de hipóteses para a presente pesquisa, em especial, as hipóteses H3 e H5 (ver capítulo 1 – *Introdução*).

No capítulo 3, a seguir, será descrita a metodologia deste trabalho. Nesse capítulo, são detalhadas informações sobre os sujeitos, a elaboração do *corpus*, as etapas de coleta de dados e os procedimentos de tratamento e análise desses dados.

3 Metodologia

Neste capítulo, detalha-se o método empregado na pesquisa. Primeiramente, será feita a descrição dos sujeitos participantes – a criança e os adultos. Em seguida, será descrito o modo como se deu a coleta de dados: serão apresentados o *corpus* utilizado e os procedimentos adotados para as coletas de áudio e de ultrassom. Por fim, serão descritos os procedimentos de tratamento dos dados acústicos e dos dados articulatórios.

3.1 Sujeitos

Os sujeitos desta pesquisa estão divididos em dois grupos: i) criança em processo de aquisição dos encontros consonantais e ii) adultas. A responsável pela criança participante e todas as adultas participantes assinaram um Termo de Consentimento Livre e Esclarecido (TCLE) em momento anterior à realização da coleta. O termo apresentado à responsável pela criança participante se encontra no Apêndice A e o termo apresentado às adultas, no Apêndice B. A descrição desses grupos será feita, respectivamente, nas subseções 3.1.1 e 3.1.2.

3.1.1 A criança

Esta pesquisa, caracterizada como um estudo de caso, contou com a participação de uma criança no fornecimento de dados de fala infantil. A criança foi escolhida a partir de três critérios.

O primeiro critério de escolha foi a vivência da criança com outras línguas além da língua materna. Assim, foi escolhida uma criança monolíngue, falante de português brasileiro, cuja prática linguística cotidiana se desse somente no português, sua língua materna, e que ainda não tivesse experiências de aprendizagem formal de língua estrangeira, como, por exemplo, na escola ou em cursos de línguas estrangeiras. Essas informações foram averiguadas com a mãe da criança, por meio de um questionário, presente no Apêndice C.

Satisfeito o primeiro critério, o segundo critério foi a faixa etária inicial para as coletas, que deveria ser entre 4;00 e 5;00. Embora o processo de aquisição de

encontros consonantais comece antes dessa faixa etária, ela foi determinada a partir do que se julgou adequado para coletas experimentais com o ultrassom. É preciso considerar que uma criança abaixo dessa faixa etária pode sofrer desconfortos na situação de coleta caso não tenha a constituição corporal e o desenvolvimento motor adequados para sustentar o capacete estabilizador de cabeça utilizado para as coletas. Do mesmo modo, para o andamento correto da coleta, é necessário que a criança se porte fisicamente de modo adequado, permanecendo sentada, com postura ereta e sem realizar movimentos mais bruscos que possam alterar o posicionamento da sonda ou causar barulhos que prejudiquem a gravação. Assim, buscou-se uma faixa etária em que ainda se pudessem encontrar crianças em processo de aquisição de encontros consonantais, mas com idade inicial mais adiantada em relação a pesquisas com dados naturais¹³. Essa escolha foi feita, em suma, de modo a se manter o bem estar da criança participante e para que as coletas transcorressem com pouca ou nenhuma interrupção.

Satisfeito o segundo critério, o terceiro critério foi a conclusão ou não do processo de aquisição de encontros consonantais. A criança participante deveria apresentar esse processo em andamento, isto é, a criança não estaria produzindo, até então, os encontros corretamente, com os dois elementos consonantais claramente presentes, ou estaria produzindo apenas algumas dessas sequências. Para atestar-se que a criança estava ainda em processo de aquisição, foi realizada a primeira coleta normalmente e, a partir dela, foram analisados dados de áudio, tanto naturais quanto experimentais. Os dados naturais eram palavras com sílabas CCV produzidas pela criança na interação anterior à coleta de ultrassom, por meio de figuras¹⁴. Já os dados experimentais eram palavras com sílabas CCV produzidas durante a coleta com o ultrassom, que também foram gravadas em áudio. A partir desses dados, analisados por meio de outiva e acústica, verificou-se se a criança produzia ou não o encontro consonantal e, se produzia, em que medida essa produção correta acontecia. Caso a maioria das produções constituísse produções incompletas da sílaba, o terceiro critério era considerado também satisfeito e a

¹³ Para exemplificação, no estudo de Bonilha (2005), que trata da aquisição de modo relativamente geral, incluindo a aquisição de estruturas consonantais complexas, analisam-se dados a partir da idade de 1;01. Quanto a estudos sobre aquisição de encontros consonantais, Ribas (2002) analisa dados de crianças a partir dos 2;00, enquanto, em Miranda (2007), a idade inicial considerada é de 3;00.

¹⁴ Para maior detalhamento dos procedimentos de coleta, ver seção 3.2.2.

criança estaria adequada à realização das coletas longitudinais.

A criança participante deste estudo foi a que atendeu a todos os critérios supracitados, procedendo, então, à realização de coletas de dados de fala experimentais, de caráter longitudinal. É necessário esclarecer, também, que a criança não possuía qualquer déficit de ordem cognitiva: nas interações anteriores à coleta e durante a sua realização, a pesquisadora constatou que a criança mantinha a atenção nas atividades propostas e ouvia bem, interagindo e atendendo às instruções dadas sem qualquer dificuldade.

O Quadro 6 apresenta as características do informante e a distribuição das coletas longitudinais realizadas.

Sujeito	Sexo	Naturalidade	Coletas realizadas		
			#	Data de realização	Idade
CR	Masculino	Pelotas/RS	1	31/07/2017	04;01;13
			2	18/09/2017	04;03;00
			3	07/12/2017	04;05;19
			4	08/03/2018	04;08;18
			5	12/04/2018	04;09;25
			6	03/05/2018	04;10;15
			7	15/05/2018	04;10;27
			8	29/05/2018	04;11;11

Quadro 6 – Características da criança informante da pesquisa e distribuição de coletas longitudinais realizadas

Inicialmente, as coletas estavam programadas para ocorrer a cada 30 dias; no entanto, fatores diversos impediram as coletas mensais, como períodos de férias, o que contribuiu para os intervalos irregulares. Sendo assim, há intervalos de doze dias a três meses entre uma coleta e outra. Salienta-se, no entanto, que, a partir da idade de 4;08, momento em que a sequência CCV já era realizada em um maior número de palavras, os dados passaram a ser coletados com um intervalo menor do que 30 dias entre uma coleta e outra.

A seção a seguir descreve o grupo de sujeitos adultos e os critérios de seleção para esse grupo.

3.1.2 As adultas

Nesta pesquisa, participaram três sujeitos adultos no fornecimento de dados de fala. Esses sujeitos foram escolhidos com base em cinco critérios.

O primeiro critério, assim como para o grupo da criança, foi a vivência com outras línguas além da língua materna. Assim, também foram escolhidos sujeitos monolíngues, falantes nativos de português, cujo contato com uma língua estrangeira de forma instrucional tenha ocorrido apenas no ensino fundamental e/ou médio. Nesta pesquisa, os adultos constituem uma referência de produção. Logo, essa escolha metodológica é feita de modo a evitar influências de línguas estrangeiras na produção dos sons ou estruturas alvo. Essas informações foram averiguadas por meio de um questionário aplicado aos informantes, conforme Apêndice D.

A escolha dos sujeitos também foi feita conforme três critérios que favorecem a qualidade da imagem ultrassonográfica obtida, sendo eles o sexo, a idade e o índice de massa corporal. De acordo com Stone (2005), de modo geral, esses fatores podem ter influência na qualidade da imagem do contorno da língua que é gerada pelo ultrassom:

Os sujeitos variam quanto à qualidade de imagem. Sujeitos magros geralmente fornecem melhores imagens do que os sujeitos mais pesados porque há menos gordura na língua¹⁵ para causar refração do som. Sujeitos mais novos geralmente fornecem imagens melhores do que sujeitos mais velhos, talvez por haver uma maior hidratação na boca e menos gordura nos tecidos. Crianças têm excelentes imagens. Mulheres frequentemente fornecem melhores imagens que os homens, especialmente no plano coronal. Não há razão substantiada para isso; possivelmente há uma diferença de gênero quanto ao posicionamento da língua. Outra possibilidade é a de que as línguas tipicamente menores das mulheres tenham uma superfície efetivamente mais regular. Essas generalizações sobre a qualidade de imagem não são absolutas, visto que certas pessoas mais velhas fornecem boas imagens e certas pessoas mais novas, não. (STONE, 2005, p. 465, tradução nossa¹⁶)

¹⁵ O trecho de Stone (2005) evidencia que a gordura presente especificamente na língua pode ser responsável por refrações que comprometam a imagem de ultrassom obtida. No entanto, faz-se necessário apontar a interferência causada pela gordura presente na região submandibular, região essa que fica em contato direto com a sonda. A gordura presente nessa região também pode causar refração considerável da onda ultrassonográfica, gerando imagens menos nítidas. Sujeitos magros fornecem melhores imagens por possuírem menos gordura nessa região.

¹⁶ No original: "Subjects vary in image quality. Thin subjects are generally image (sic) better than heavy ones because there is less fat in the tongue to refract the sound. Younger subjects generally image better than older subjects, perhaps because there is more moisture in the mouth, and less fat in the tissue. Children have excellent images. Women often image better than men, especially in the coronal plane. There is no substantiated reason for this; possibly there is a gender difference in

Tendo-se em vista as generalizações apontadas por Stone (2005) quanto à qualidade de imagem, definiu-se, para esta pesquisa, que os sujeitos escolhidos deveriam ser do sexo feminino, ter entre 18 e 35 anos de idade e possuir baixo índice de massa corporal.

Por fim, o quinto critério foi o local de naturalidade, que deveria ser o mesmo para todos os sujeitos, de modo a minimizar possíveis efeitos nos sons-alvo decorrentes da variação diatópica. Logo, todos os sujeitos escolhidos são naturais da cidade de Pelotas/RS, assim como a criança participante.

Os sujeitos adultos participantes deste estudo atenderam a todos os critérios, estando aptos, assim, à realização de coletas de dados de fala experimentais de caráter transversal. Também é preciso esclarecer que, assim como a criança, os sujeitos adultos participantes também não possuíam qualquer déficit cognitivo. Nas interações anteriores à coleta e durante a sua realização, a pesquisadora observou que as participantes adultas mantinham a atenção nas atividades propostas e ouviam bem, interagindo e atendendo às instruções sem dificuldades.

O Quadro 7 apresenta o grupo de informantes adultas participantes da pesquisa, suas características e informações sobre as coletas realizadas.

Sujeito	Sexo	Naturalidade	Coletas realizadas	
			Data de realização	Idade
AD1	Feminino	Pelotas/RS	12/09/2017	22;00;18
AD2	Feminino	Pelotas/RS	06/09/2017	28;06;02
AD3	Feminino	Pelotas/RS	06/09/2017	29;07;19

Quadro 7 – Grupo de adultas participantes da pesquisa e coletas transversais realizadas

As coletas do grupo das adultas foram transversais, cada uma realizada em um único dia, diferentemente das coletas realizadas com a criança, de caráter longitudinal. A fala adulta é relativamente mais estável, já que o processo de aquisição dos sons e estruturas silábicas da língua materna já se encontra concluído. Assim, foi considerada suficiente a realização de uma coleta por informante, de modo a se constituir um conjunto de dados que fosse uma referência de produção.

Na seção 3.2, a seguir, serão descritos o instrumento e os procedimentos da

tongue positioning. Alternatively, the typically smaller tongues of women may have an effectively smoother surface. These generalizations about image quality are not as absolute as some older people image well and some younger ones do not.”

coleta de dados.

3.2 As coletas de dados

As coletas de dados, realizadas com os dois grupos de informantes, criança e adultas, foram realizadas em cabine com isolamento acústico, no Laboratório Emergência da Linguagem Oral (LELO). O LELO se situa nas dependências do Centro de Letras e Comunicação (CLC), unidade acadêmica da Universidade Federal de Pelotas (UFPel). A Figura 14 mostra o interior da cabine acústica utilizada.

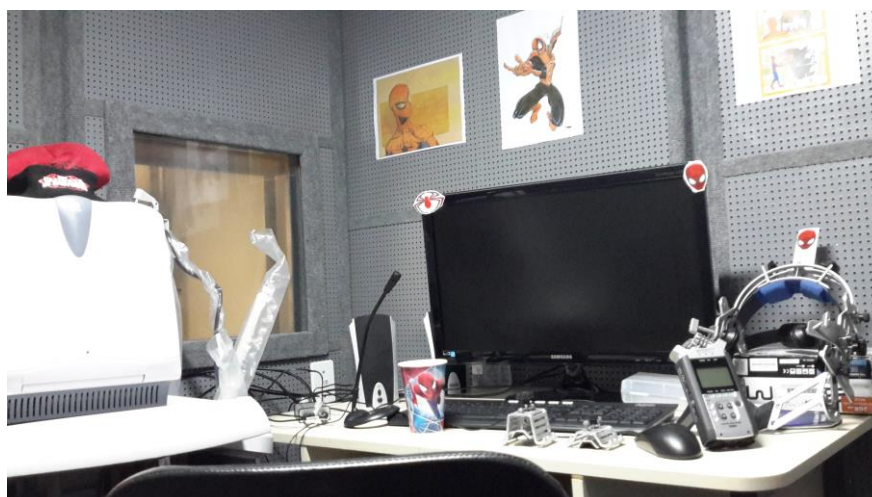


Figura 14 – Interior da cabine acústica do LELO, utilizada para as coletas de áudio e ultrassom

Foi necessário realizar a gravação de dados de fala experimentais em cabine acústica a fim de evitar-se a interferência de ruídos externos, oriundos do próprio laboratório ou dos corredores do *campus*, os quais poderiam interferir na qualidade dos dados acústicos e, também, na análise de dados articulatórios¹⁷.

A seção 3.2.1 descreve o *corpus* escolhido e o instrumento utilizado para as coletas experimentais. Em seguida, na seção 3.2.2, serão expostos os procedimentos de coleta adotados.

¹⁷ Eventuais ruídos nos áudios podem interferir na análise de dados articulatórios, uma vez que o AAA, *software* utilizado para esse fim, utiliza o recorte acústico como base para etiquetamento de *frames*.

3.2.1 O *corpus* e o instrumento de coleta

As coletas de dados desta pesquisa, de natureza experimental, tiveram como finalidade a produção de pares de palavras que se diferenciasssem, fundamentalmente, na estrutura da sílaba inicial – CCV ou CV. Sendo assim, pensou-se em um *corpus* de pares mínimos, de palavras dissílabas e paroxítonas. Esse foi o primeiro critério de escolha do *corpus*, tendo-se por base outros estudos realizados sobre a aquisição de CCV, já citados na seção 2.2.2. Esses estudos são os de Miranda (2007), que utilizou pares mínimos para investigar se a vogal de sílabas CCV era alongada por crianças em processo de aquisição dessa sílaba, e de Vassoler (2016), que estudou como se dava a coordenação gestual na produção de CCVs e CVs por crianças em processo de aquisição típico e atípico, investigando se medidas ultrassonográficas e acústicas eram diferentes para esses dois tipos de sílaba. O Quadro 8 mostra os conjuntos de palavras utilizados para a coleta de dados nessas duas pesquisas.

Miranda (2007)		Vassoler (2016)	
CCV	CV	CCV	CV
broa	boa	broa	boa
prato	pato	prato	pato
prego	pego	prego	pego
pressa	peça	pressa	peça
bruxa	bucha	bruxa	bucha
branco	banco		
frita	fita	frita	fita
grato	gato	grato	gato
troca	toca	troca	toca
troco	toco	troco	toco
20 palavras		18 palavras	

Quadro 8 – Conjuntos de palavras utilizados por Miranda (2007) e Vassoler (2016) para as coletas de dados

Note-se que o estudo primeiro, que constitui a base para essa escolha de *corpus*, é Miranda (2007). O *corpus* de Vassoler (2016), em princípio, é o mesmo proposto por Miranda (2007), com a diferença de que o par “branco”/“banco” foi excluído das análises. Em princípio, haviam sido gravadas as 20 palavras, tal como no conjunto de Miranda (2007), “entretanto o par de palavras, branco e banco, foi excluído [...], uma vez que a nasalidade poderia interferir nas análises” (VASSOLER,

2016, p. 87). Esse par também foi desconsiderado na constituição do *corpus* desta pesquisa, pois as vogais nasais apresentam uma fase de murmúrio que poderia interferir em uma análise acústica em que são medidos valores de duração de vogal e da porção vocálica que se realiza antes do *tap* nas palavras com CCV. Nesta pesquisa, optou-se por partir das 18 palavras presentes em Vassoler (2016), com a realização de mudanças, tendo em vista outros critérios que se mostraram relevantes.

O segundo critério diz respeito à qualidade da primeira consoante nos encontros consonantais das palavras com CCV. Nesta posição, deu-se preferência a sons obstruintes bilabiais ou labiodentais e desvozeados, pois suas características acústicas e articulatórias facilitaram o processo de identificação e segmentação desses sons na análise dos dados. Assim, foram conservadas, a partir do modelo de Miranda (2007), palavras como “prato”, “prego” e “pressa”, que possuem a oclusiva desvozeada [p], e “frita”, que apresenta o som desvozeado [f]. Palavras como “bruxa” e “broa” foram mantidas, ainda que contenham a oclusiva vozeada [b], tendo em vista a importância do contexto bilabial para a análise articulatória em uma sequência do tipo C[r]V. No entanto, foram evitadas, nesta posição, oclusivas alveolares e velares, em função da dificuldade oferecida para a análise articulatória: as alveolares [t] e [d], na sua articulação, envolvem o uso de gestos de ponta, os quais podem ser confundidos com o gesto de ponta relacionado ao *tap* subsequente; já as velares [k] e [g] apresentam gesto de dorso, o que poderia interferir na presença ou ausência de gesto de dorso na articulação do *tap*, fato a ser discutido no presente trabalho, considerando trabalhos como Silva (2002) e Recasens (2016). Logo, as palavras “troca”, “troco” e “grato”, presentes originalmente no *corpus* de Miranda (2007) e conservadas no de Vassoler (2016), foram desconsideradas no conjunto de palavras deste estudo, assim como seus respectivos pares CV, “toca”, “toco” e “gato”.

Por fim, o terceiro critério, no conjunto de palavras deste estudo, diz respeito aos diferentes contextos vocálicos: considerando-se as sete vogais orais do PB, buscou-se ter, no *corpus*, pelo menos um exemplo com cada uma dessas vogais em posição de núcleo, na sílaba CCV/CV em análise. Assim, para serem contemplados os contextos vocálicos com [e] e [ɔ], foram inseridos os pares “preso”/“peso” e “frota”/“foto”, este último um par análogo.

Ainda, para que houvesse uma maior robustez de dados em contexto

considerado ótimo para a análise dos dados articulatórios – consoante labial ou bilabial seguida de vogal baixa –, tendo em vista a sequência C[r]V, foram acrescentadas as palavras “prata”, “praça” e “fraca” ao instrumento, assim como seus respectivos pares CV – “pata”, “passa” e “faca”.

Como resultado, a partir dos critérios estabelecidos para a constituição do *corpus*, elaborou-se o conjunto utilizado nas coletas desta pesquisa. As palavras escolhidas constituem 11 pares, dos quais 10 são pares mínimos e 1 é par análogo, totalizando, assim, 22 palavras. O conjunto resultante de palavras, utilizado para a produção nas coletas, se encontra no Quadro 9, a seguir.

CCV		CV	
prato	['pra.tu]	pato	['pa.tu]
prata	['pra.ta]	pata	['pa.ta]
praça	['pra.sa]	passa	['pa.sa]
fraca	['fra.ka]	faca	['fa.ka]
prego	['prɛ.gu]	pego	['pɛ.gu]
pressa	['prɛ.sa]	peça	['pɛ.sa]
preso	['prɛ.zu]	peso	['pe.zu]
frita	['fri.ta]	fita	['fi.ta]
frota	['fro.ta]	foto	['fo.tu]
broa	['bro.(w)a]	boa	['bo.(w)a]
bruxa	['bru.ʃa]	bucha	['bu.ʃa]
11 palavras		11 palavras	
22 palavras			

Quadro 9 – *Corpus* elaborado para as coletas, com base nos *corpora* de Miranda (2007) e Vassoler (2016)

No Quadro 9, as células hachuradas indicam as palavras adicionadas ao *corpus*. Elas não estão nos *corpora* de estudos anteriores, mas foram pensadas de acordo com os critérios supracitados.

O instrumento de coleta utilizado, por sua vez, foi um conjunto de imagens relacionadas às palavras constantes no Quadro 9 e que foram apresentadas aos informantes por meio do *software* AAA (*Articulate Assistant Advanced*), versão 2.14¹⁸.

Nas Figuras 15 e 16, encontram-se dois exemplos de imagens dentre as vinte e duas utilizadas. O conjunto integral das imagens encontra-se no Anexo A.

¹⁸ Articulate Instruments Ltd 2012. **Articulate Assistant Advanced User Guide**: Version 2.14. Edinburgh, UK: Articulate Instruments Ltd.



Figura 15 – Exemplo de imagem utilizada na coleta, correspondente à palavra “prato”



Figura 16 – Exemplo de imagem utilizada na coleta, correspondente à palavra “pato”

Na seção 3.2.2, a seguir, serão detalhados os procedimentos da coleta.

3.2.2 Procedimentos de coleta

Após a apresentação do Termo de Consentimento e o esclarecimento de eventuais dúvidas, a pesquisadora explicou o funcionamento da coleta de dados aos adultos participantes e ao responsável pela criança participante. Em seguida, como preparação para a gravação dos dados de fala, houve uma etapa de familiarização com as palavras a serem produzidas, na qual foram apresentadas as figuras do instrumento aos informantes. Essas figuras vinculavam-se ao conceito das palavras do *corpus*, a serem produzidas posteriormente na coleta de áudio e ultrassom. Essa etapa se fez necessária não só para eliciar uma primeira produção dessas palavras por parte do adulto ou da criança, mas também para verificar se o informante

compreendeu os procedimentos de coleta.

Para fins de familiarização da criança com as palavras do *corpus*, as imagens foram impressas em pares e apresentadas no formato de jogo da memória. Assim, à medida que as figuras eram mostradas, a pesquisadora fazia perguntas ao informante, tais como “O que é isso?”, “Pra que serve isso aqui?” e “O que essa pessoa tá fazendo?”. O cuidador foi convidado a participar da interação, de modo a facilitar a eliciação das palavras. Essa interação foi gravada em áudio para que se tivesse um banco adicional de dados de produção das palavras. Ressalta-se que a etapa de familiarização se fez necessária também para os adultos, pois, tendo-se em vista o conjunto de imagens escolhidas para o instrumento de coleta, as relações existentes entre o conceito da palavra e sua imagem correspondente não se estabelecem de forma rápida e direta em todos os casos.

Após a etapa de familiarização, procedeu-se à coleta de dados de áudio e ultrassom na cabine. Para a coleta integrada de áudio e ultrassom, o áudio foi capturado com um gravador *Zoom*, modelo H4N, configurado com taxa de amostragem de 44.000 Hz. As imagens ultrassonográficas foram capturadas com um aparelho de ultrassom *Mindray*, modelo DP 6600, com uma sonda transdutora acoplada.

As sondas transdutoras são construídas com materiais piezoelétricos, os quais possibilitam a conversão de energia elétrica em energia mecânica ou acústica. As sondas podem ser de três tipos: linear, microconvexa ou convexa. Por possuírem diferentes tamanhos e formatos, elas produzem imagens diferentes, como pode ser visto na Figura 17:



Figura 17 – Tipos de sonda. Figura reproduzida de Ferreira-Gonçalves e Brum-de-Paula (2013, p. 90)

O modelo de ultrassom utilizado nesta pesquisa aceita três tipos de sonda, apresentados na Figura 18:



Figura 18 – Tipos de sonda compatíveis com o ultrassom *Mindray DP 6600*, da esquerda para a direita: microconvexa (35C20EA), endocavitária (65EC10EA) e microconvexa (65C15EA). Figura reproduzida de Ferreira-Gonçalves e Brum-de-Paula (2013, p. 101)

Para análises linguísticas, o tipo de sonda é escolhido de acordo com o tamanho do trato vocal do informante e, também, de acordo com o fenômeno a ser pesquisado. Para as coletas com a criança, a sonda endocavitária (modelo 65EC10EA, o segundo na Figura 18) foi escolhida. Conforme Ferreira-Gonçalves e Brum-de-Paula (2013), essa sonda possibilita que se visualize a ponta de língua de forma mais clara, já que a sombra da mandíbula é diminuída. Como consequência, o uso dessa sonda possibilita uma melhor captação do gesto de ponta envolvido na produção do *tap*, razão pela qual Barberena (2016) a utilizou para a coleta de dados, em seu estudo sobre aquisição do *tap*.

Para as coletas com adultos, foram utilizadas as sondas endocavitária (modelo 65EC10EA, o mesmo utilizado com a criança) e microconvexa (modelo 65C15EA, o terceiro na Figura 18 da esquerda para a direita). Após a realização de testes com cada informante, foi escolhido o tipo de sonda que fornecesse a melhor imagem de acordo com o trato vocal do indivíduo.

A ligação entre o ultrassom e o computador de mesa utilizados para as coletas é feita por uma placa de vídeo. Para a sincronização de áudio e vídeo na coleta, foi utilizado o sincronizador *SyncBrightUp*, modelo SBU1.0.

Um obstáculo para a obtenção de imagens no ultrassom pode ser a dificuldade de se manter a cabeça do informante imobilizada no momento da coleta. De modo a se aplacar essa dificuldade, foi desenvolvido pela empresa *Articulate Instruments* um capacete¹⁹ para a estabilização de eventuais movimentos da cabeça e da sonda durante a produção. Na Figura 19, é mostrado o capacete, cujo mesmo modelo foi utilizado para as coletas deste pesquisa.

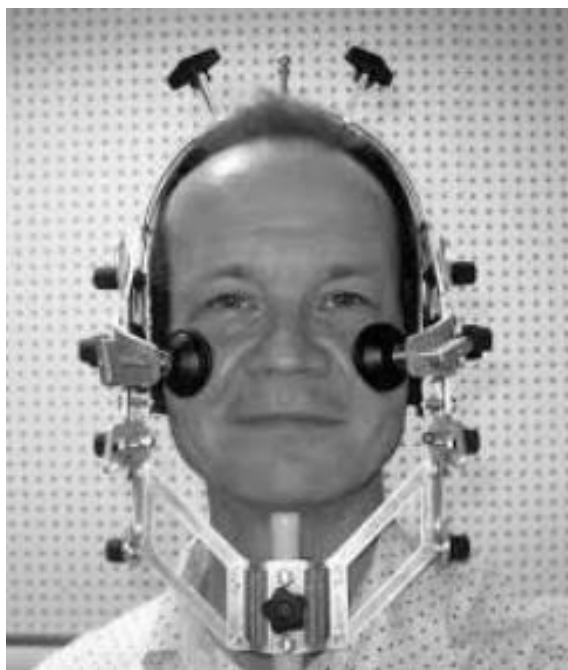


Figura 19 – Modelo de capacete estabilizador dos movimentos da cabeça, utilizado para as coletas ultrassonográficas (Fonte: <http://www.articulateinstruments.com>)

A estabilização dos movimentos de cabeça com o uso do capacete proporciona uma melhor qualidade dos dados articulatórios, os quais poderão, então, ser analisados quantitativamente, caso o pesquisador assim deseje. Logo, todas as coletas desta pesquisa foram realizadas com o uso do capacete nos informantes.

Na Figura 20, pode ser visto o conjunto de equipamentos utilizados para a coleta de áudio e ultrassom, incluindo computador, aparelho de ultrassom, microfone, gravador e capacete estabilizador.

¹⁹ Um maior detalhamento sobre o funcionamento do capacete poderá ser encontrado em Scobbie, Wrench e van der Linden (2008).

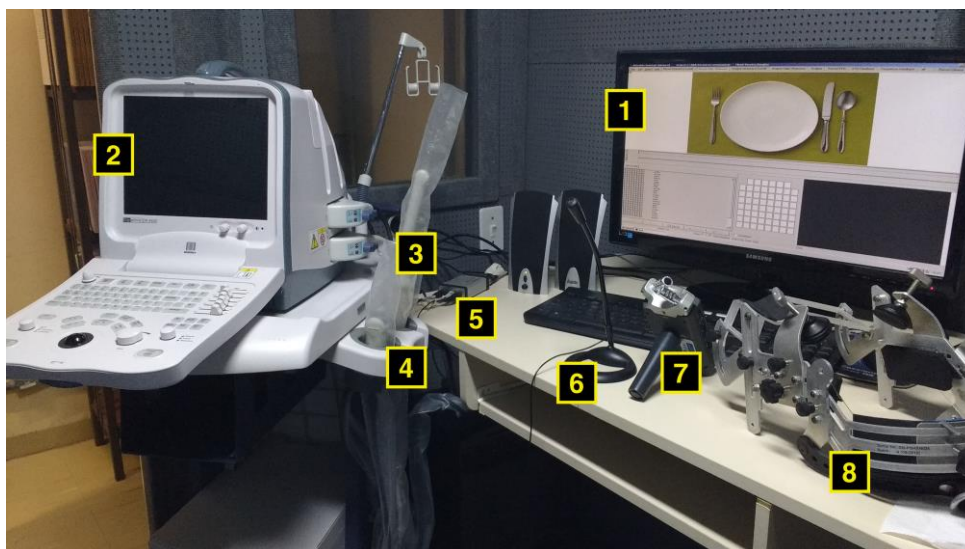


Figura 20 – Equipamentos utilizados para a coleta de áudio e ultrassom: computador (1), aparelho de ultrassom (2), sonda endocavitária (3), sonda microconvexa (4), sincronizador de áudio e vídeo (5), microfone unidirecional (6), gravador de áudio (7) e capacete estabilizador (8)

O *software* utilizado para a coleta de áudio e ultrassom em computador foi o *Articulate Assistant Advanced (AAA)*, versão 2.14, desenvolvido para a coleta e análise de dados de ultrassom, acústicos e eletropalatográficos. Na Figura 21, a seguir, há uma imagem da tela do *software* na função de gravação do vídeo de ultrassom.

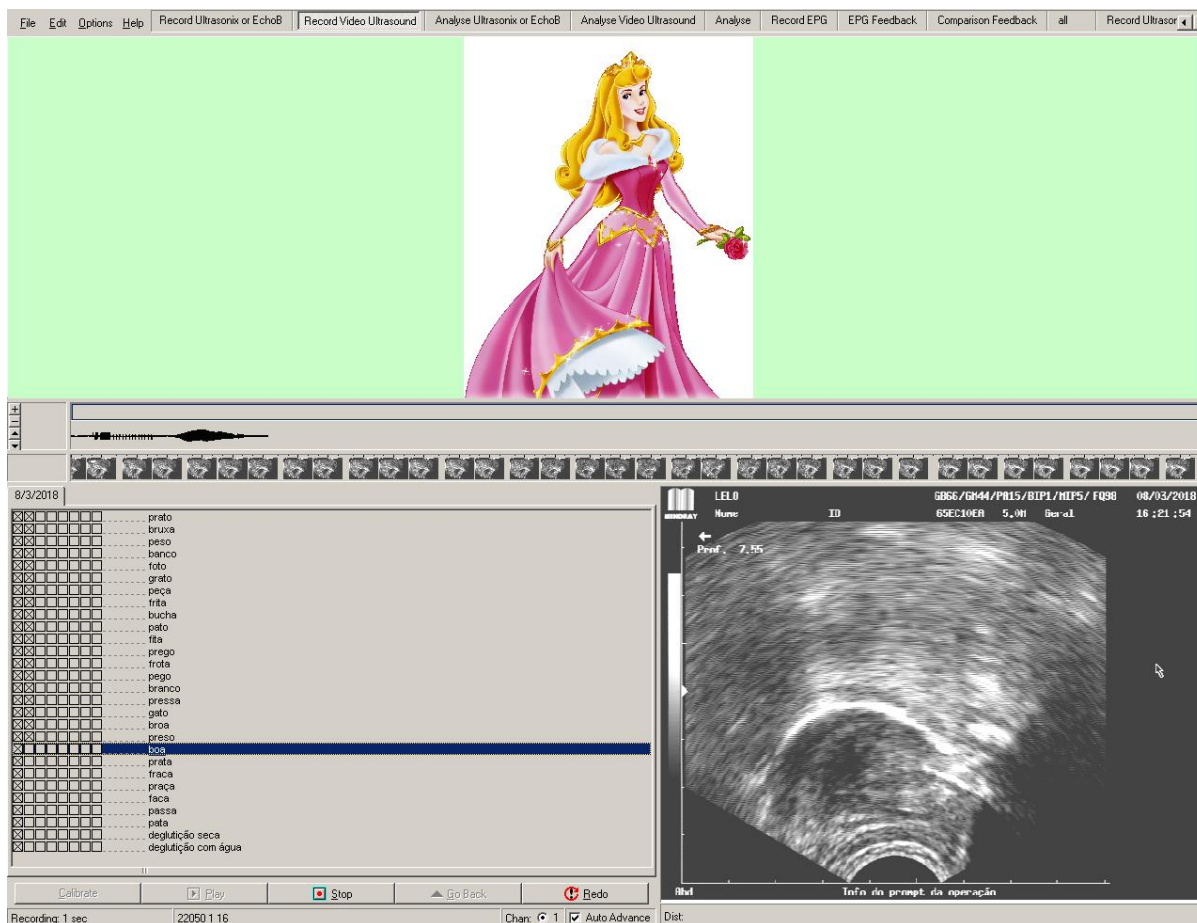


Figura 21 – Tela do software AAA em modo de gravação de vídeo de ultrassom, durante a produção da palavra “boa” por CR

Na parte superior da tela, o programa mostra os estímulos para a produção de fala, os quais podem ser sons, textos ou, como no caso deste trabalho, imagens. No momento da produção, o fundo da tela, que é branco, torna-se verde e um bipe é acionado, sinalizando ao informante que é o momento de produzir a palavra-alvo. Logo abaixo da imagem que serve de estímulo para a produção – na figura, a princesa –, estão dispostos o oscilograma do áudio e os *frames* da imagem de ultrassom. Na parte inferior esquerda, há a lista de sentenças ou palavras a serem produzidas, com o número de repetições a serem feitas, e, na parte direita, há a imagem de vídeo do ultrassom.

Antes da coleta de dados articulatórios, foram feitos testes com a sonda transdutora e o gel para ajuste das configurações do aparelho de ultrassom, de

modo a se obter a melhor imagem possível do contorno da língua para cada informante. O teclado do modelo de ultrassom utilizado, mostrado na Figura 22, apresenta várias opções de configuração.



Figura 22 – Imagem do teclado do aparelho *Mindray DP 6600* com algumas funções sinalizadas: *Probe* (a); *Depth* – profundidade (b); *Freq.* – frequência (c); *Gain* – ganho (d); *IP* (e) e *F. position* – posição do foco (f)

Ajustes nas funções destacadas na Figura 22 foram de maior importância. Abaixo, são detalhadas informações sobre essas funções:

- a) tipo de sonda (*Probe*) – seleciona a sonda conectada ao ultrassom que será utilizada para a coleta;
- b) profundidade (*Depth*) – ajusta a profundidade da imagem de ultrassom de acordo com o tamanho do trato vocal do informante, em termos da distância entre a região submandibular e o palato. Conforme Ferreira-Gonçalves e Brum-de-Paula (2013), essa distância, em geral, não passa de 8 centímetros para adultos; sendo assim, é adequado ajustar a

profundidade entre 7 e 8 cm para esse perfil de informante. Já para crianças, a profundidade deve ser configurada em número menor, já que o trato vocal infantil é menor e, logo, a distância entre a região submandibular e o palato também será menor;

- c) frequência (*Freq.*): ajusta a taxa de varredura da sonda. Os valores de frequência, para as sondas utilizadas, podem variar entre 5.0 e 8.0. Considerando que a frequência é inversamente proporcional à profundidade do feixe de ultrassom, adotou-se o valor de frequência 7.55, mais alto, para a criança. Por ela ter um trato vocal menor, a imagem não precisaria ter um valor alto de profundidade. Regulou-se, assim, a frequência para o valor citado, mais alto, de forma a se ter uma taxa de varredura mais rápida e, portanto, imagens mais nítidas.
- d) ganho (*Gain*) – regula a qualidade de imagem, pois se pode regular a intensidade em diferentes regiões da imagem. Pode-se, por exemplo, aumentar a intensidade na região da língua e diminuir a intensidade em regiões periféricas do trato;
- e) IP (*IP*) – tem papel na otimização da imagem fornecida. Quanto maior for o valor de IP, haverá menor contraste e uma maior suavização da imagem. O valor de IP configurado foi de 5.0;
- f) posição do foco (*F. position*) – define a zona de maior nitidez na imagem ultrassonográfica fornecida. A imagem tem melhor resolução na zona de maior foco.

Além de terem sido feitos ajustes nesses aspectos, a potência acústica foi mantida no nível mais baixo possível, de acordo com o princípio ALARA²⁰.

Após a realização dos testes e a devida configuração do aparelho de ultrassom, o informante foi instruído (com auxílio do cuidador, no caso da criança) a se posicionar de forma adequada para a realização da coleta – sentado, com a coluna ereta e com olhar direcionado à tela do computador. O capacete estabilizador foi colocado e sua altura e largura foram reguladas para que ele se moldasse e

²⁰ Sigla para *As Low As Reasonably Achievable*. De acordo com esse princípio, a exposição de seres humanos, animais ou materiais à radiação deve ser a mais abaixo possível do limite, por razões de segurança.

imobilizasse movimentos da cabeça, com o cuidado de que não ficasse desconfortável para o informante. Em seguida, foi aplicado gel na superfície da sonda a ser utilizada – ele deve ser utilizado nas coletas para um melhor contato entre transdutor e pele, evitando ruídos na imagem que possam ser causados pelo ar. Aplicado o gel, a sonda foi posicionada na região submandibular do informante e fixada na parte inferior do capacete, a qual possui um suporte que mantém a sonda imóvel.

Para este estudo, também foi necessário escolher o tipo de imagem de ultrassom a ser coletada, a qual forneceria o melhor tipo de dado conforme o fenômeno a ser analisado. O ultrassom possibilita a visualização de uma fatia de tecido em duas dimensões. Nos estudos linguísticos, os tipos de imagens mais utilizados são as imagens sagital e coronal: a imagem sagital divide as metades direita e esquerda de uma estrutura, enquanto a imagem coronal divide metades dianteira e traseira. Nas Figuras 23 e 24, são apresentados exemplos de imagens ultrassonográficas nos planos sagital e coronal:



Figura 23 – Exemplo de imagem ultrassonográfica da língua no plano sagital, obtida a partir de uma produção de CR. O dorso da língua está localizado à esquerda, enquanto a ponta da língua está localizada à direita



Figura 24 – Exemplo de imagem ultrassonográfica da língua no plano coronal (MÉNARD *et al.*, 2012, *apud* FRANCISCO, 2015, p. 18)

A escolha por uma coleta de imagens no plano sagital ou coronal dependerá do tipo de segmento ou gesto a ser analisado. Uma pesquisa que investigue a produção de segmentos laterais poderá se beneficiar da coleta de imagens no plano coronal, pois poderá ser observado o comportamento dos músculos laterais da língua. Por sua vez, pesquisas que investiguem segmentos ou gestos em que a altura e o avanço da língua sejam aspectos importantes necessitarão de imagens no plano sagital, em que essas variáveis são visualizadas adequadamente. Esta pesquisa observa o comportamento de partes da língua como a ponta e o dorso em aspectos como a elevação e o avanço; logo, entre os dois tipos de imagem, optou-se pela coleta de imagens no plano sagital.

Após a colocação do capacete estabilizador e o posicionamento da sonda, foi realizada, em cabine acústica, a gravação simultânea de áudio e de imagens ultrassonográficas das produções das 22 palavras do *corpus*. Para os adultos, essas palavras foram produzidas em frase-veículo, inserindo-se a palavra-alvo na estrutura “Digo _____ bem bonito” (exemplo: “Digo ‘prato’ bem bonito.”). Essas produções foram feitas cinco vezes para cada palavra do *corpus*.

Já para a criança, as produções foram realizadas de forma isolada (exemplo: “Prato”), o que facilitou o processo de sincronização entre áudio e vídeo dos dados, realizado no AAA posteriormente. Na coleta com a criança, cada palavra do *corpus* foi produzida três vezes. Esse número de produções é, de fato, reduzido em relação

ao das informantes adultas. Isso se deve a uma particularidade da situação de coleta com a criança em relação a coletas realizadas com adultos, que é o menor tempo pelo qual se conseguia manter o sujeito atento e confortável. Nas primeiras coletas, constatou-se que o número ideal de produções da lista era três, o que resultava em quantidade razoável de dados para análise sem que se desrespeitassem os limites de atenção e conforto da criança. O número menor de produções da lista também não causou escassez de dados de modo geral, já que foram realizadas oito coletas longitudinais.

Assim, foram totalizadas 510 produções pela criança e, pelas adultas, foram totalizadas 330 produções. O Quadro 10 mostra uma relação com o número inicial de produções por grupo.

Grupo de sujeitos	Qtd. de sujeitos	Qtd. de palavras	Qtd. de repetições	Qtd. de coletas	Qtd. de produções
Criança	1 (CR)	x 22 (x 16 na coleta #1)	x 3	x 8	510
Adultos	3 (AD1, AD2 e AD3)	x 22	x 5	x 1	330
Total de produções:					840

Quadro 10 – Número de produções por grupo nas coletas de dados acústicos e articulatórios

Como pode ser visto no Quadro 10, o grupo da criança tem um sujeito, o qual realizou a produção de 16 palavras na primeira coleta, em três repetições ($1 \times 16 \times 3 = 48$), e 22 palavras nas 7 coletas restantes, também em três repetições ($7 \times 22 \times 3 = 462$). Já o grupo das três adultas realizou coletas únicas, produzindo 22 palavras em cinco repetições, em coletas únicas ($3 \times 1 \times 22 \times 5 = 330$). A soma desses cálculos é o número total de dados coletados para esta pesquisa: 840 dados. Salienta-se que o número de repetições foi importante para se manter uma robustez de dados, tendo em vista os critérios, já mencionados na seção 3.2.1, de inclusão e exclusão de palavras com CCV para a constituição do *corpus* desta pesquisa.

Também foi realizada a gravação de deglutições com todos os informantes. O movimento de deglutição envolve elevação da língua até alcançar-se contato com o palato. Esse contato possibilita que se visualize, na imagem ultrassonográfica, o

palato do indivíduo, cujo contorno pode ser extraído e sobreposto a todas as imagens a serem analisadas. O palato serve, assim, como um ponto de referência na imagem de ultrassom que elimina ambiguidades quanto ao posicionamento da superfície da língua, como apontado por Epstein e Stone (2005). Ainda, conforme as autoras, a deglutição pode ser gravada de modos variados: seca, com água ou com refrigerante, por exemplo. Todos esses modos possibilitam a obtenção de imagens do palato.

Nesta pesquisa, as gravações de deglutição foram feitas de dois modos: seca e com água, de modo a se ter mais de uma opção para a obtenção de uma imagem nítida do palato. Essas deglutições foram gravadas no mesmo número de vezes que as gravações de palavras, sempre ao final da lista de estímulos; sendo assim, foram gravadas, da criança, 3 deglutições secas e 3 deglutições com água por coleta; já de cada adulto, foram gravadas 5 deglutições secas e 5 deglutições com água.

Após a coleta em áudio e ultrassom, os dados foram organizados em fichas, segmentados e submetidos a análises acústica e articulatória. Os procedimentos de organização e análise dos dados serão descritos na seção 3.3, a seguir.

3.3 Procedimentos de organização e análise dos dados

Antes de passarem por segmentação e análise acústica, os dados foram organizados em fichas de transcrição, no programa *Word*, versão 14.1.0. Cada áudio de coleta foi escutado e teve todas as produções de palavras com CCV e CV do *corpus* registradas em uma ficha. Uma das fichas preenchidas se encontra no Apêndice E, para exemplificação.

Conforme o Apêndice, pode ser visto que a ficha apresenta três tabelas. A primeira, na parte superior, contém os dados principais para a identificação da coleta específica que foi realizada, quais sejam:

- a) *Arquivo*: nome do arquivo de áudio armazenado no computador que corresponde à transcrição feita (e.g. #5 CR 2018-04-12 – 4;9.wav);
- b) *Duração*: duração, em minutos e segundos, do arquivo de áudio (e.g. 30:41);
- c) *Sujeito*: código do sujeito que teve os dados coletados (e.g. CR, AD1,

AD2, AD3);

d) *Idade*: idade do sujeito no momento da coleta (ex.: 04;09).

A segunda tabela contém um acompanhamento da produção do *tap* em contextos silábicos que não o CCV: em sílaba CV e em coda e em sílaba tônica ou átona. Já a terceira tabela contém o registro de todas as palavras CCV e CV do *corpus* produzidas no áudio, tanto na interação anterior à coleta de ultrassom quanto durante a sua realização. Foram registradas as seguintes informações em colunas:

- a) #: número da produção. As palavras foram numeradas conforme a ordem de ocorrência no áudio;
- b) *mm:ss*: momento no arquivo de áudio, em minutos e segundos, em que a palavra é produzida;
- c) *Palavra*: palavra que foi produzida no momento referido. Além das palavras pertencentes ao *corpus*, palavras produzidas fora da situação de coleta de ultrassom que apresentassem alguma mudança morfológica (e.g. “pecinhas”, ao invés de “peça”) também foram registradas. Do mesmo modo, como poderá ser visto no Apêndice, foram registradas nas fichas palavras com sílabas CCV produzidas durante a situação de coleta que não pertencessem ao *corpus* (e.g. “grande”, “mestre”). Essas produções CCV fora das coletas foram registradas para eventual consulta, de modo a se poder observar padrões de produção ou produções alternativas da criança, se necessário.
- d) *Observações*: eventuais observações quanto a prováveis modificações na produção detectadas perceptualmente, as quais foram averiguadas pela acústica. Também é registrado, nessa coluna, quando um dado é descartado por apresentar qualidade de áudio inadequada para a análise acústica ou quando uma palavra inicialmente presente no *corpus* é excluída do conjunto, de forma que a análise de suas produções fica inviabilizada (caso de “branco/banco e “grato/gato”, excluídos do *corpus*).

Os dados destacados em verde nessa coluna correspondem às palavras

efetivamente coletadas também em ultrassom, pertencentes às situações de coleta experimental. Logo, somente esses dados foram submetidos aos procedimentos de tratamento e análise que serão relatados nas seções a seguir. Do mesmo modo, é a partir desses dados que foram extraídos todos os resultados e elaboradas as conclusões desta pesquisa. Os demais dados foram produzidos em situações de interação anteriores ou simultâneas à coleta de dados experimentais, mas não há nenhum dado de ultrassom a eles correspondente.

Dados destacados em vermelho foram descartados, pois se constatou, na sua escuta e visualização (por oscilograma e espectrograma), que não possuíam a qualidade adequada para a realização de uma análise acústica confiável.

Após o registro dos dados das coletas em fichas de transcrição, as palavras marcadas em verde, coletadas também em ultrassom, foram segmentadas no programa *Praat* (BOERSMA; WEENINK, 2018), versão 6.0.17, e salvas em arquivos .wav separados, para a posterior realização de análise acústica.

Os resultados obtidos por meio da inspeção acústica foram submetidos à análise estatística no programa *SPSS* (*Statistical Package for the Social Sciences*), versão 17.0. Foram utilizados os seguintes testes paramétricos: Teste-T para amostras pareadas e One-Way ANOVA. O valor de significância foi estabelecido para $p < 0,05$. Seguindo Martins (2011), foram considerados ainda valores marginalmente significativos, com p entre 0,05 e 0,10.

Os procedimentos de tratamento e análise dos dados acústicos serão detalhados na seção 3.3.1, a seguir.

3.3.1 Tratamento e análise dos dados acústicos

Após a realização das coletas e o registro das produções em fichas de transcrição, as palavras produzidas na coleta foram analisadas acusticamente com o programa *Praat*. O primeiro procedimento realizado com o programa foi a segmentação manual das palavras produzidas. As Figuras 25 a 28 mostram o procedimento de recorte que foi feito para cada palavra a ser analisada. Na Figura 25, é mostrado o primeiro passo desse procedimento: é preciso abrir, no *Praat*, o áudio da coleta e selecionar, arrastando-se o cursor, um trecho no qual ocorra o dado a ser recortado. Na figura, o trecho selecionado está destacado em azul.

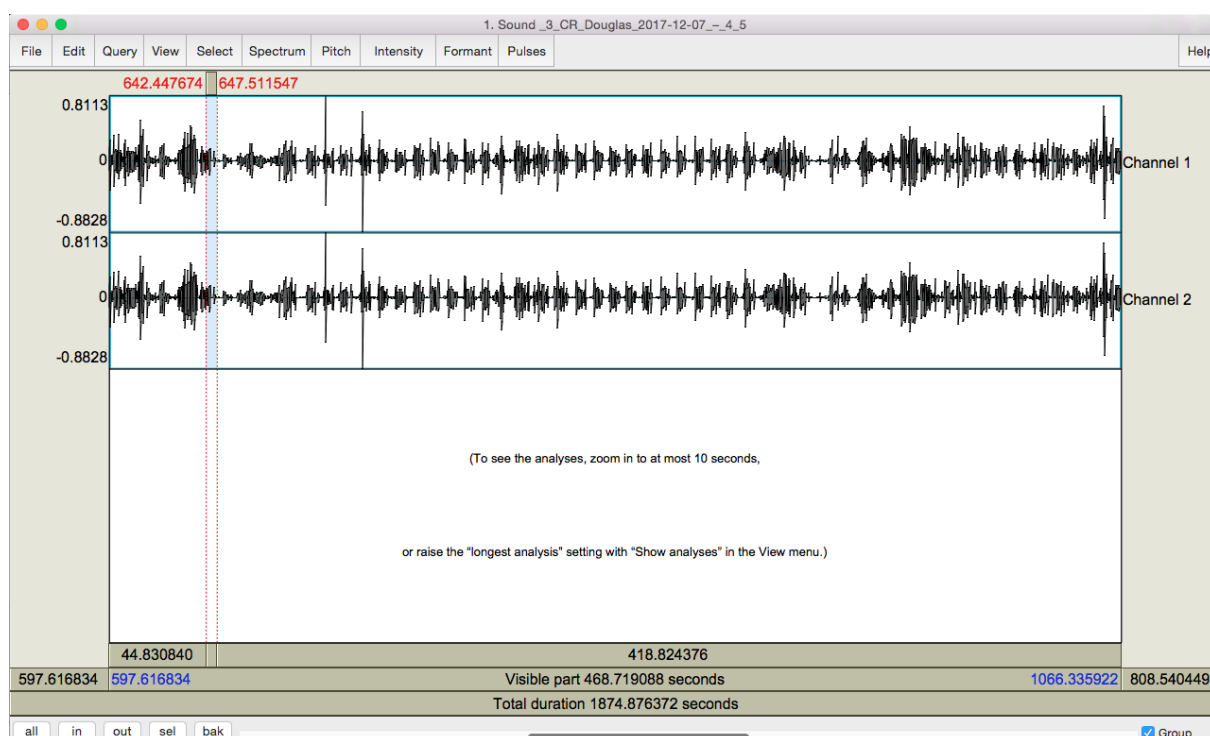


Figura 25 – Procedimento de segmentação de palavras. Primeiro passo: abrir o áudio da coleta e seleccionar um trecho em que ocorra a produção da palavra

Selecionado o trecho, deve-se ampliar o oscilograma na região selecionada, por meio do botão “sel”, no canto inferior esquerdo da tela, ou por meio do atalho Command+N/Ctrl+N. O programa mostrará apenas o trecho selecionado, como se pode ver na Figura 26.

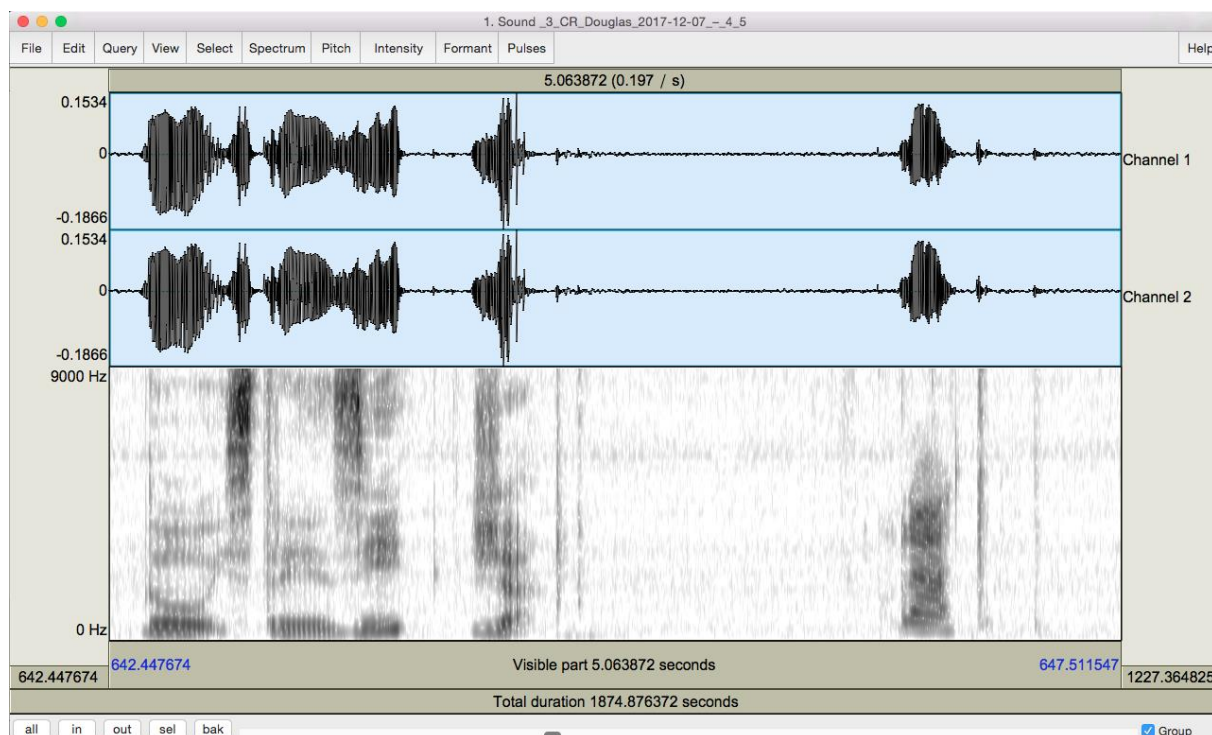


Figura 26 – Procedimento de segmentação de palavras. Segundo passo: selecionar o botão “sel” ou utilizar o atalho Command+N/Ctrl+N para ampliar a tela na região selecionada

No trecho que foi ampliado, deve-se selecionar, com mais precisão, a palavra a ser posteriormente analisada, arrastando-se o cursor sobre o trecho mais específico em que a palavra ocorre. Na seleção desse trecho, é preciso que se certifique de que a palavra foi selecionada em toda a sua realização; logo, devem ser evitados cortes indevidos no oscilograma e no espectrograma que impeçam a medição adequada das durações. Para tanto, nas produções da criança, esse trecho foi selecionado com uma margem de silêncio antes e depois da produção da palavra, já que as suas produções foram feitas de forma isolada. Já para as adultas, o trecho selecionado incluiu parte do final da palavra anterior e o início da palavra seguinte, visto que as produções das palavras para esse grupo foram feitas em frase-veículo.

Após a seleção do trecho específico a partir do áudio da coleta, a palavra foi salva como um arquivo .wav separado, por meio da função “*Save selected sound as WAV file...*”, como mostrado no terceiro passo, na Figura 27.

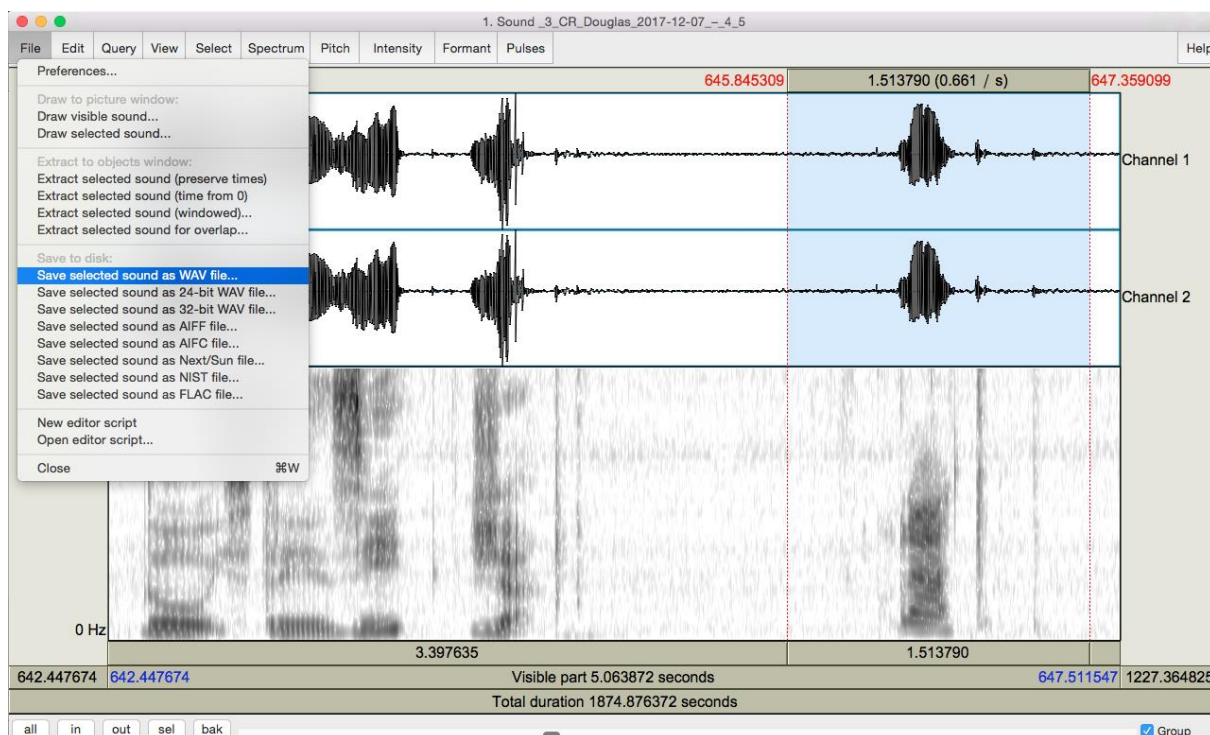


Figura 27 – Procedimento de segmentação de palavras. Terceiro passo: selecionar o trecho específico em que a palavra ocorre

Cada palavra foi salva em um arquivo .wav contendo o código do sujeito, o número da coleta (no caso da criança), o número da produção na ficha de transcrição, a palavra produzida e a repetição específica. O arquivo resultante do procedimento aqui exemplificado foi salvo como “CR #3 019 prato1.wav”. Assim, pelo nome do arquivo, pode-se identificar que foi uma produção realizada pela criança, na sua terceira coleta, de número 019 na ficha de transcrição e que é a primeira repetição da palavra “prato” na coleta experimental. Quando aberto no *Praat*, o arquivo gerado mostra somente o trecho salvo, como mostrado na Figura 28.

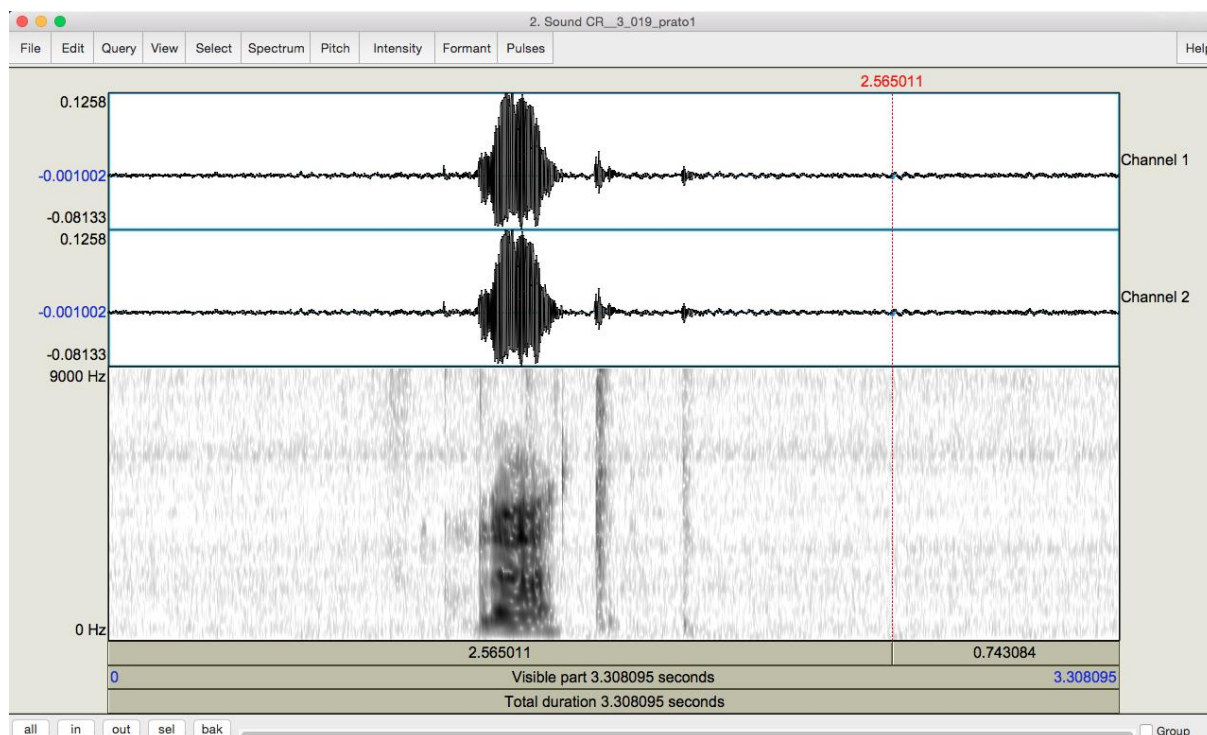


Figura 28 – Palavra “prato”, produzida por CR, salva a partir do procedimento de segmentação

Todos os arquivos .wav das palavras segmentadas foram salvos em pastas distintas, organizadas conforme o sujeito, a data de coleta e a repetição, como pode ser visto na Figura 29. Os arquivos também foram numerados conforme a ordem de aparição no áudio, conforme registrado na respectiva ficha de transcrição.

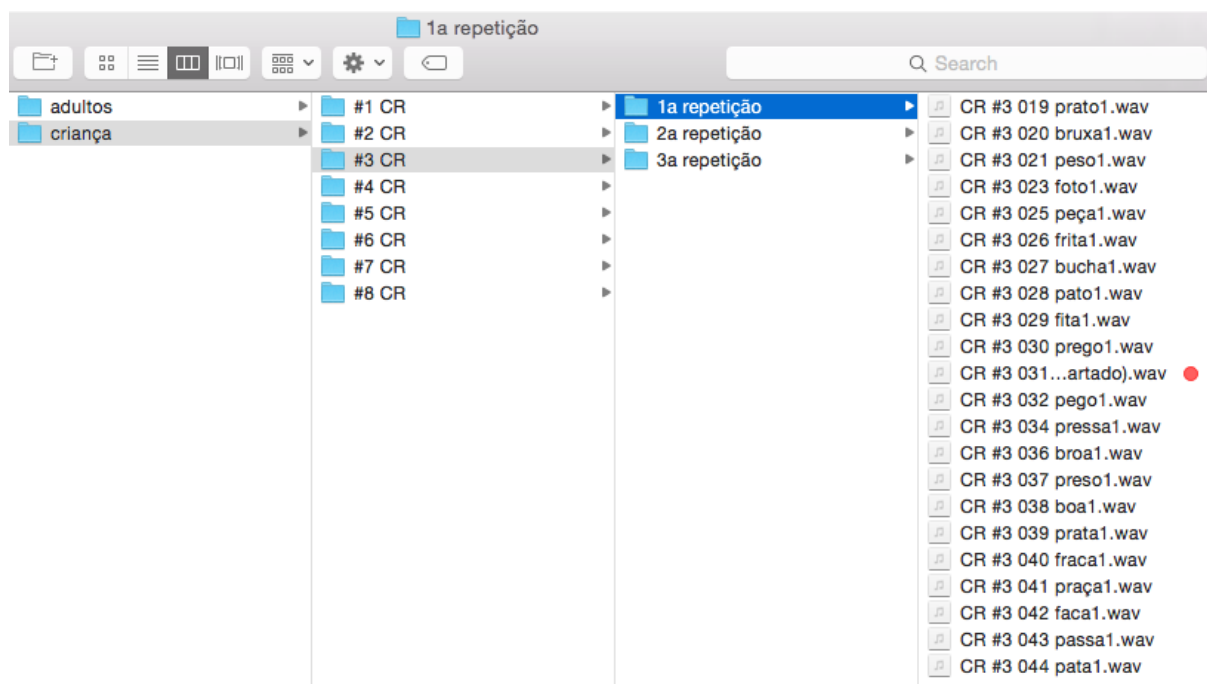


Figura 29 – Organização dos arquivos .wav das palavras segmentadas em pastas

Após essa organização, foi realizado o procedimento de concatenação de dados. No programa *Praat*, todas as produções de uma mesma palavra foram concatenadas num único arquivo .wav, com a função *Concatenate*, a qual justapõe os arquivos de áudio, criando um novo arquivo .wav que é a sequência dos sons selecionados. As Figura 30 e 31 mostram exemplos de arquivos concatenados.

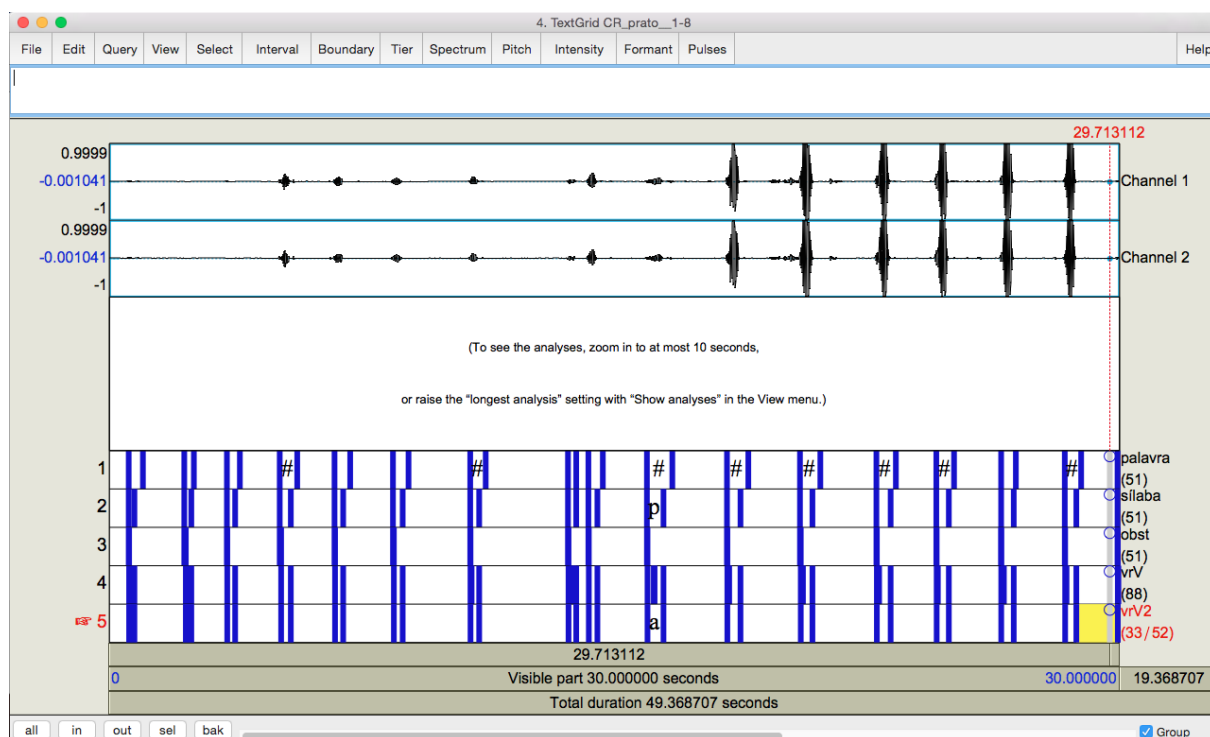


Figura 30 – Arquivo concatenado contendo todas as produções da palavra “prato” de CR e sua respectiva *TextGrid*

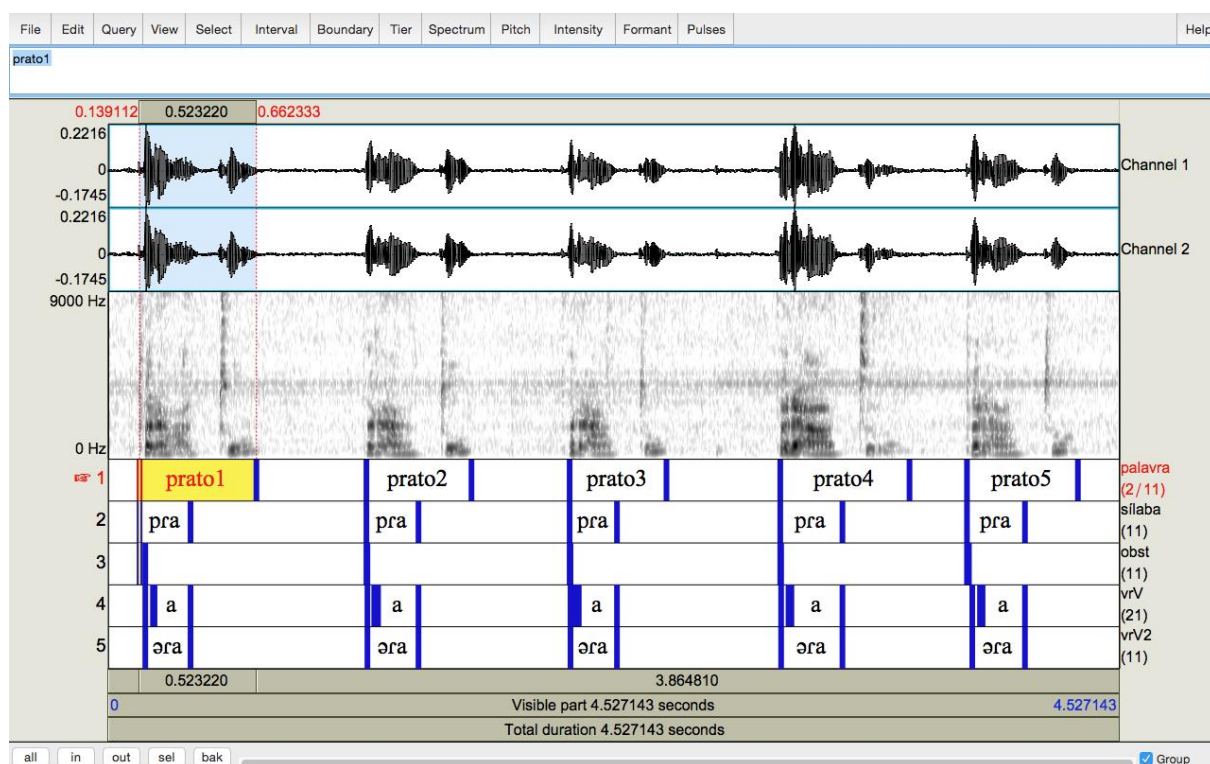


Figura 31 – Arquivo concatenado contendo todas as produções da palavra “prato” de AD1 e sua respectiva *TextGrid*

A concatenação é um procedimento que otimiza o tempo destinado às análises acústicas, pois justapõe várias produções. Ao colocarem-se várias produções de uma mesma palavra em sequência, o pesquisador pode se concentrar em critérios de análise específicos e fazer as medidas de duração de forma mais eficiente, pois os mesmos critérios são aplicados repetidamente.

Para cada arquivo concatenado, foi criada uma *TextGrid*, arquivo de formato específico para trabalho no *Praat* no qual são feitas anotações em relação ao áudio analisado. Essas anotações podem ser feitas em diferentes níveis ou camadas (*tiers*), o que possibilita a medida de durações nos níveis de palavra, de sílaba e de unidades menores. Nas Figuras 30 e 31, mostradas anteriormente, podem ser vistos, de cima para baixo, o oscilograma do áudio, o espectrograma e, na parte inferior, a *TextGrid*, com vários níveis (*tiers*) nomeados (“palavra”, “sílabas”, etc.).

As medidas de duração em palavras com CCV foram feitas para determinadas partes da produção, com a seguinte divisão em *tiers*, presente nas Figuras 30 e 31:

- a) *tier* 1 (“palavra”) – palavra;
- b) *tier* 2 (“sílabas”) – sílaba CCV inicial;
- c) *tier* 3 (“obst”) – obstruinte;
- d) *tier* 4 (“vrV”) – elemento vocálico anterior ao *tap* (quando presente), rótico (*tap*, retroflexo, vibrante ou outras realizações, quando presente) e vogal;
- e) *tier* 5 (“vrV2”) – soma de elemento vocálico (quando presente), rótico (quando presente) e vogal.

Já as *TextGrids* de palavras com CV seguiram um padrão diferente, presente na Figura 32, com os seguintes *tiers* e respectivas medidas realizadas:

- a) *tier* 1 (“palavra”) – palavra;
- b) *tier* 2 (“sílabas”) – sílaba CV inicial;
- c) *tier* 3 (“obst”) – obstruinte;
- d) *tier* 4 (“V”) – vogal.

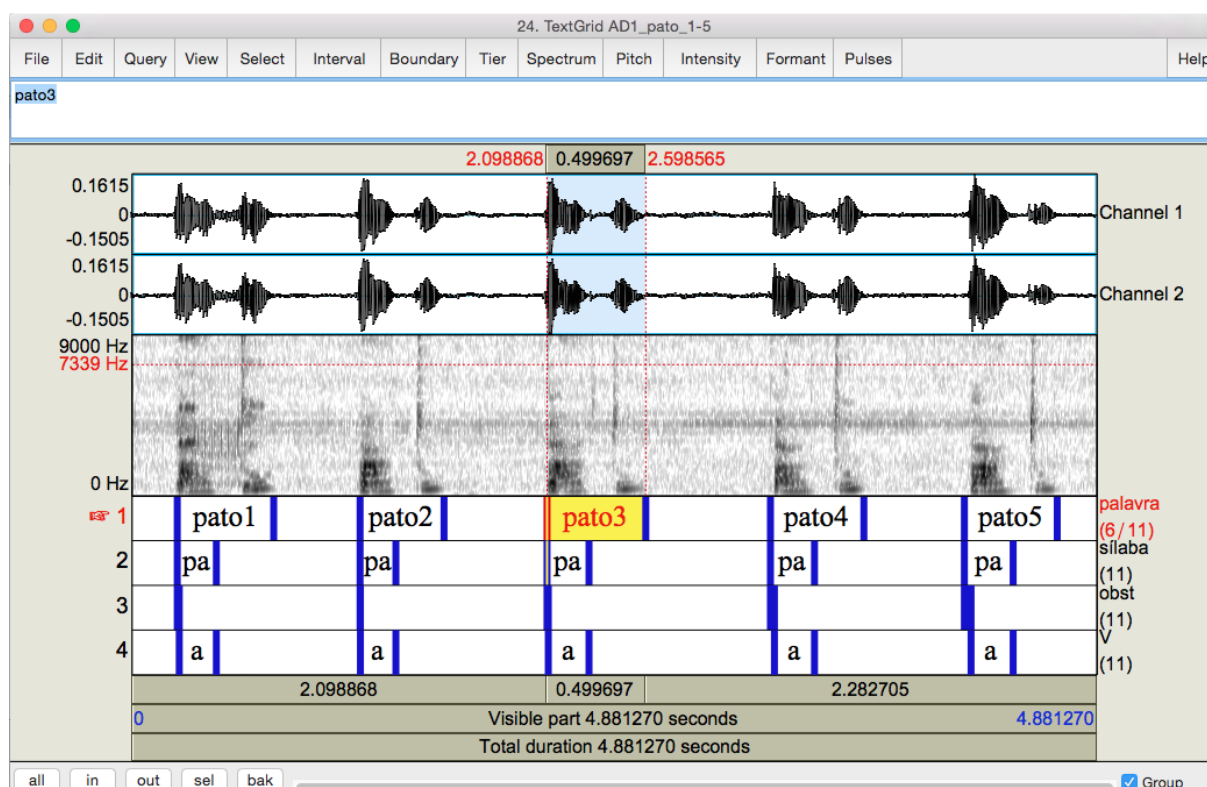


Figura 32 – Arquivo concatenado contendo todas as produções da palavra “pato” de AD1 e sua respectiva *TextGrid*

Nessas *TextGrids*, foram feitas as medidas de duração relativas aos trechos considerados, com anotações em forma de texto. Essas medidas foram realizadas com base na interpretação das imagens de oscilograma e de espectrograma fornecidas pelo *Praat*. Essa interpretação foi feita com apoio em critérios específicos, os quais guiaram a delimitação de cada segmento ou trecho.

Dados que não apresentassem qualidade de áudio suficiente para uma interpretação confiável do oscilograma e/ou do espectrograma foram descartados. Assim, foram excluídos da análise dados que tivessem forte ruído ambiente, uma ou mais vozes se sobrepondo à voz do informante, barulhos ocorrendo ao mesmo tempo da produção da palavra ou sons com características indefinidas ocorrendo na posição de rótico. Esses casos de interferência ou de produção diferenciada dificultaram a identificação dos sons relevantes para a análise no oscilograma e no

espectrograma; logo, não foi possível utilizar dados desse tipo para cálculos de duração e testes estatísticos de modo confiável.

As Figuras 33, 34, 35 e 36, a seguir, mostram exemplos de dados descartados por motivos distintos, os quais estão detalhados nas legendas. Esses dados foram desconsiderados para análise acústica e, conseqüentemente, também para cálculos de duração e testes estatísticos.

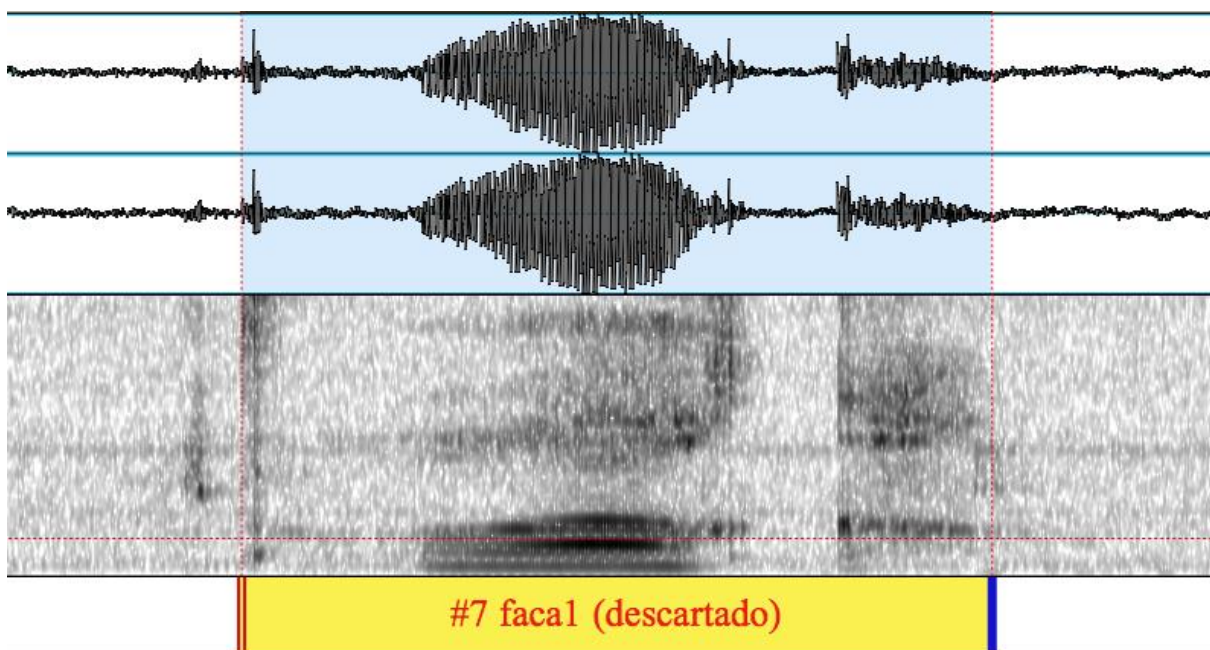


Figura 33 – Dado de áudio produzido por CR, descartado por conter forte ruído ambiente

A Figura 33 mostra um dado de áudio que contém forte ruído ambiente, o que se percebe principalmente pelas partes fora da seleção em azul, nas quais não há produção de fala. Nessas regiões, à esquerda e à direita do trecho da palavra, o oscilograma apresenta uma onda com muitas oscilações e o espectrograma está tomado por um padrão em cinza claro que dá um aspecto “áspero” à imagem. Embora essas características sejam percebidas principalmente nas partes fora da seleção da palavra, o ruído perpassa a produção, modificando o oscilograma e o espectrograma também no trecho em que há fala.

No oscilograma, o ruído se soma à onda sonora da fala da criança, dificultando a percepção de padrões como o aspecto estridente da onda do [f] e a regularidade, ou periodicidade, que caracteriza as ondas das vogais. E, no

espectrograma, o ruído dificulta a percepção de limites dos segmentos na medida em que uma parte mais fraca da fricção que caracteriza um [f] pode se confundir com o ruído ambiente, pela proximidade de tons de cinza. Do mesmo modo, formantes de uma vogal enfraquecida podem se confundir a esse ruído e isso dificultará a delimitação dessa vogal.

O ruído ambiente, somado a um baixo volume de voz, faz com que o dado não possa ser analisado de modo confiável, pois os limites dos segmentos não ficam claros. Portanto, esse tipo de dado teve de ser descartado.

Na Figura 34, a seguir, uma outra voz perpassa a produção de AD3.

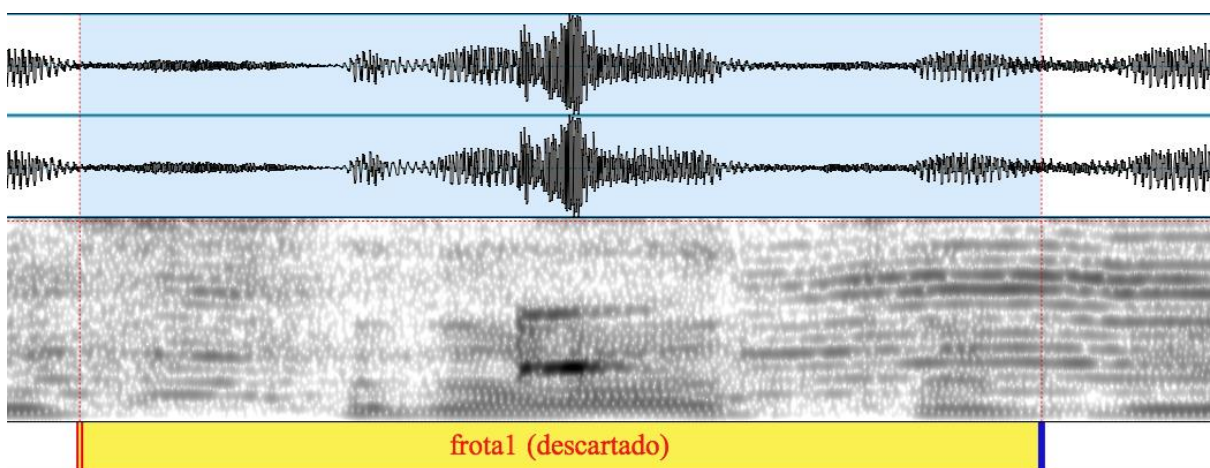


Figura 34 – Dado de áudio produzido por AD3, descartado por conter sobreposição de vozes

Essa sobreposição se percebe pelas ondas ruidosas no oscilograma, mas, principalmente, pelos harmônicos no espectrograma, que são as “manchas” em forma de linhas horizontais. Eles se sobrepõem à fricção da obstruente [f], aos formantes no final da vogal da primeira sílaba e aos formantes da vogal final, o que inviabiliza medidas de duração de palavra, de sílaba, de obstruente e de vogal. Dados desse tipo também foram descartados.

É importante observar que o dado da Figura 34 também apresenta considerável ruído e um desajuste no tempo da produção, pois a informante começou a produzir a palavra antes do bipe²¹. Esse bipe é visível, no oscilograma,

²¹ O bipe referido é o som produzido pelo AAA na gravação de dados de ultrassom. Esse bipe é emitido ao início da gravação de cada dado, indicando ao informante que é o momento de se produzir o dado. Esse sinal sonoro, juntamente com o sinal visual da tela verde, dita o início da produção, que deve ser após o término do som. O bipe também é um ponto de referência para a sincronização

pelas regiões de maior amplitude durante a realização da vogal e, no espectrograma, pelas linhas horizontais mais escuras em relação ao restante da imagem. Esses fatores, por si sós, também podem comprometer as análises e o áudio deverá ser descartado.

Na Figura 35, sobrepõem-se à fala da criança barulhos semelhantes a estalos.

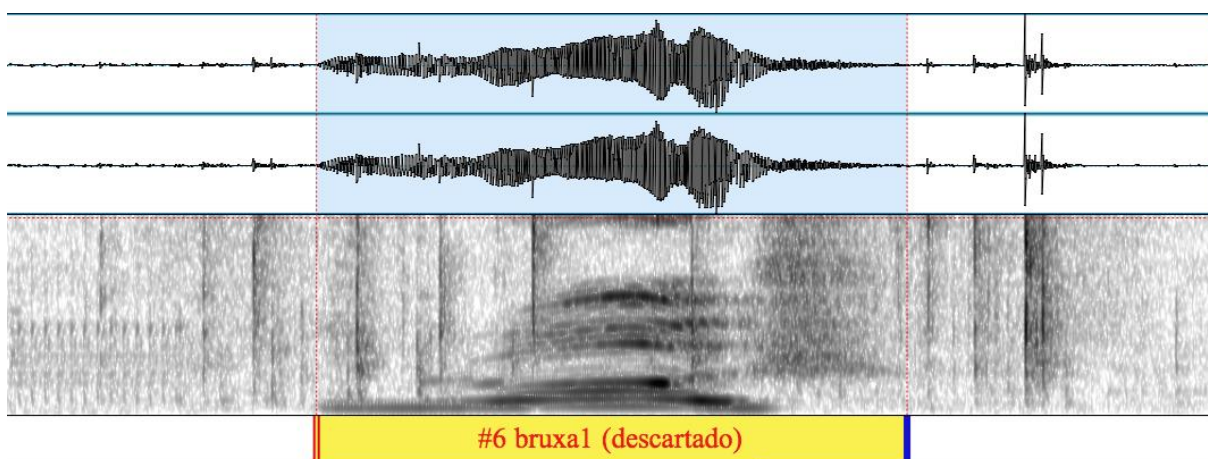


Figura 35 – Dado de áudio produzido por CR, descartado por sobreposição de barulhos à fala do informante

Os barulhos podem ser percebidos, no oscilograma, pelas oscilações intensas e rápidas; já no espectrograma, caracterizam-se pelas linhas que percorrem toda ou quase toda a extensão vertical da imagem. Esses sinais estão presentes em grande número ao longo da produção e se sobrepõem a trechos importantes para a realização da análise, como o início dos formantes de vogal. Também podem se confundir principalmente com sons plosivos como o [b], que têm a linha vertical do *burst* no espectrograma como principal característica. Apesar de o dado também conter ruído, os barulhos constantes foram o principal motivo que inviabilizou a delimitação dos trechos relevantes. Dados que apresentaram características semelhantes foram descartados da mesma forma.

posterior de dados no AAA: como já detalhado na seção 3.3.2, o seu som emite pulsos, os quais aparecem no oscilograma e no espectrograma. Esses pulsos, por sua vez, devem estar alinhados com as imagens geradas na linha dos *frames* para que o dado esteja devidamente sincronizado. Logo, o bipe, além de ser um estímulo sonoro para a produção, é uma referência importante para uma sincronização correta dos dados gravados.

O dado presente na Figura 36 apresenta, na posição de rótico, um espectrograma com características pouco definidas na região em que se localizaria esse rótico. O quadrado vermelho, acrescentado à imagem para melhor explicação, evidencia a provável região do rótico e compreende também o provável início da vogal.

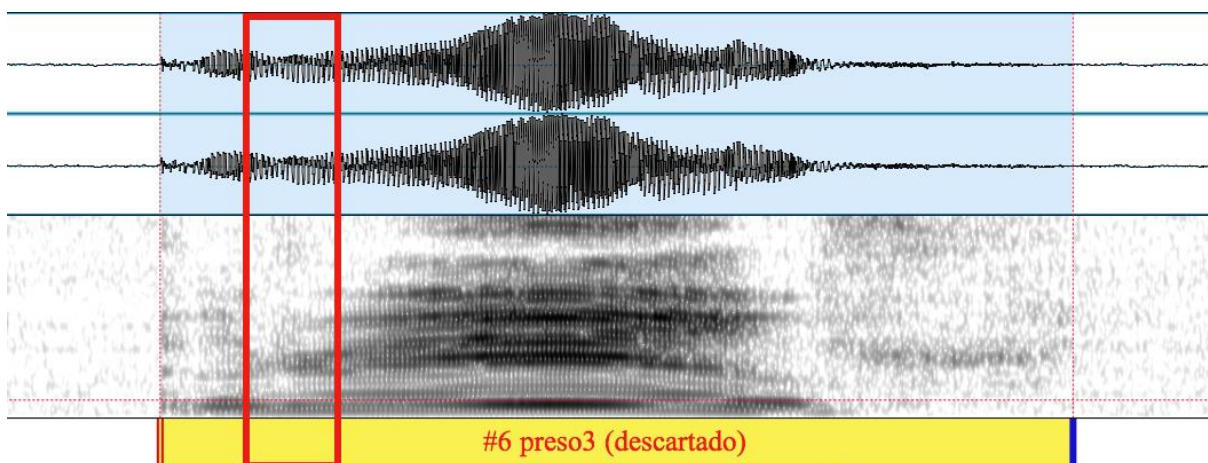


Figura 36 – Dado de áudio produzido por CR, descartado por conter som com características indefinidas na posição de rótico

Nessa imagem, alguns sinais indiciam que o rótico realizado é uma *aproximante*: a porção vocálica de curta duração aparece à esquerda e há alguma mudança na intensidade do tom de cinza dos formantes entre a porção vocálica e a vogal. No entanto, o enfraquecimento dos formantes que caracteriza esse tipo de rótico não se apresenta de forma clara no espectrograma; do mesmo modo, o oscilograma não apresenta o padrão de onda diferente em relação aos da vogal e do elemento vocálico. Observa-se, também, uma transição formântica que pode caracterizar róticos retroflexos. Como a qualidade desse rótico não pôde ser aferida a partir das pistas visuais fornecidas pelo oscilograma e pelo espectrograma, foi considerado inviável medir a sua duração. Logo, esse dado e todos aqueles que apresentaram casos semelhantes foram descartados.

Ao todo, 47 dados foram descartados por se enquadrarem em algum dos casos reportados nas Figuras 33 a 36. Assim, o total resultante para análise foi de 793 dados, a partir dos quais foram feitas as medidas de duração a serem

detalhadas a seguir.

As durações específicas a serem analisadas nesta pesquisa são as de obstruinte, vogal, elemento vocálico e rótico. Para a comparação dessas durações em sílabas CCV e CV, foram consideradas as medidas de duração absoluta (em ms) e relativa (em porcentagem – %). A duração relativa, para as unidades consideradas, foi calculada em relação à duração da sílaba, com exceção da duração relativa de elemento vocálico, calculada também em relação à duração de vogal. A comparação de durações relativa e absoluta em sílabas CCV e CV foi feita para que se verificasse se havia diferença significativa conforme o tipo de sílaba analisado.

Todas as medidas de duração absoluta foram extraídas do *Praat* e organizadas em uma planilha no programa *Excel* (versão 14.1.0) com duas folhas, uma para as produções das adultas e outra para as produções da criança. Após serem lançados os valores de duração absoluta, os cálculos de duração relativa foram feitos com o auxílio das funções do *Excel*. O programa permite a geração e aplicação de equações a uma ou mais células, de modo que um determinado cálculo pode ser aplicado a um intervalo definido de células.

A partir das funções do *Excel*, também foram calculados valores médios de duração. Esses valores foram calculados com base em todas as produções de uma mesma palavra dentro de uma mesma coleta. A partir desses valores médios, procedeu-se a um segundo cálculo de médias considerando todas as produções com mesma obstruinte inicial dentro de uma mesma coleta – e.g. a média de duração de vogal em todas as palavras CCV com obstruinte inicial [f] na coleta #2 de CR.

Após a realização dos cálculos de médias, os valores resultantes de duração da planilha foram lançados no programa *SPSS Statistics* (versão 17.0) e submetidos a testes paramétricos. Os testes específicos utilizados foram Teste-T para amostras pareadas e One-Way ANOVA.

A seção 3.3.2, a seguir, descreve os procedimentos de tratamento e análise dos dados articulatórios.

3.3.2 Tratamento e análise dos dados articulatórios

Para a análise qualitativa das imagens de ultrassom, foi utilizado o programa AAA, empregado também nas coletas de dados articulatórios. Para o tratamento e análise desses dados, foram realizados os seguintes passos:

- a) Sincronização de áudio e vídeo: após a realização das coletas, os dados de áudio e de vídeo fornecidos pelo ultrassom foram sincronizados a partir de instruções de Wrench (2012). Conforme essas instruções, o bipe gerado pelo *software* na gravação de um dado estará presente, no oscilograma, em forma de pulsos. Para que se possa considerar que áudio e vídeo estão sincronizados, esses pulsos devem estar alinhados com as imagens geradas na linha de *frames*. Como pista visual, também deve-se verificar que, no momento do primeiro pulso do bipe, aparecerá um *flash* de cor branca, no canto superior esquerdo da imagem do ultrassom. A taxa de sincronização utilizada para os dados desta pesquisa foi de 59.600 fps. Na Figura 37, há um exemplo de tela que demonstra o processo de sincronização de dados.

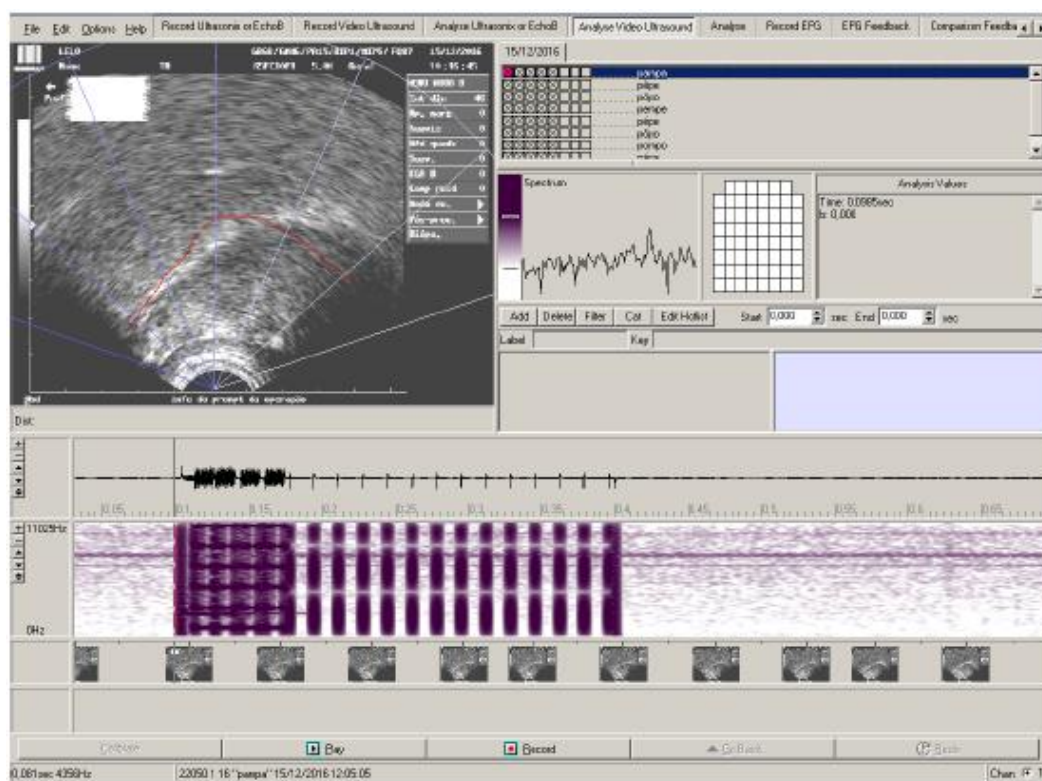


Figura 37 – Exemplo de tela mostrando o processo de sincronização de dados de áudio e vídeo pelo programa AAA, reproduzido de Correa (2017, p. 94)

- b) Criação de etiquetas para anotação dos dados: foram criadas etiquetas para a segmentação dos dados, de modo a controlarem-se momentos específicos de produção dos segmentos analisados. Para a análise qualitativa de imagens ultrassonográficas, foram selecionadas trajetórias articulatórias na produção de sílabas CCV. Uma dessas trajetórias equivale a todos os *frames* que compreendem o início da consoante até o final do *tap*. Outra trajetória a ser analisada consistiu na seleção de todos os *frames* compreendendo o início da produção da vogal – neste caso, o elemento vocálico anterior ao *tap* – até o final da vogal da sílaba. Para a análise da produção do *tap* nos encontros consonantais, foi considerado o *frame* do gesto de ponta de maior magnitude. A Figura 38 mostra uma imagem da tela do programa AAA em modo de análise do vídeo de ultrassom, em que se visualiza o *frame* do gesto de ponta de maior magnitude em uma produção da palavra “prato”.

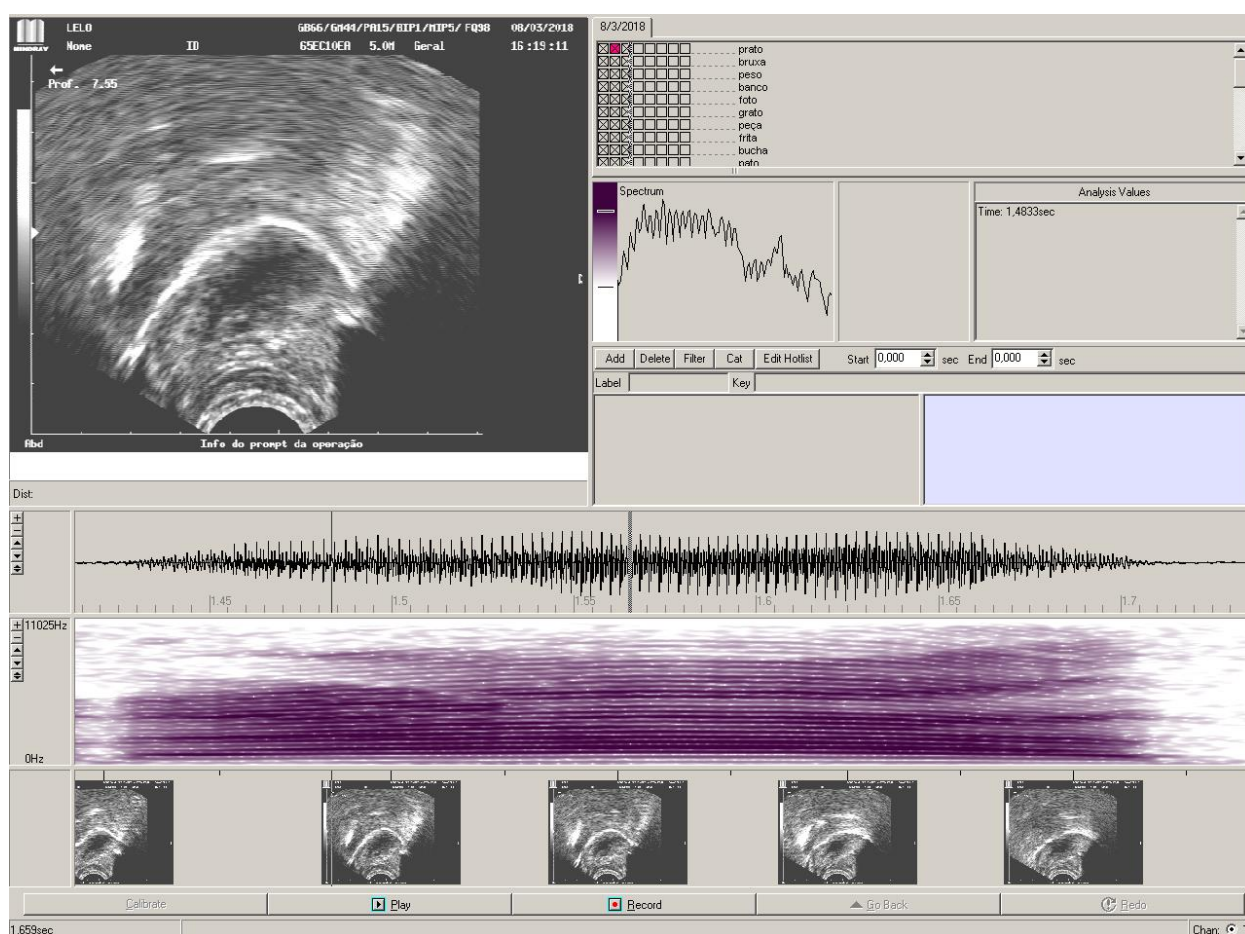


Figura 38 – Exemplo de tela do programa AAA em modo de análise do vídeo de ultrassom, durante a produção da palavra “prato” por CR

- c) Delimitação de um ponto de análise: para a análise qualitativa dos dados, que consiste no traçado de borda da língua, foi considerado como ponto de análise o gesto de maior magnitude na produção do *tap*. Após ser selecionado o ponto de análise, foi feito o desenho sobre o contorno da língua mostrado na imagem de ultrassom. Esse desenho corresponde à configuração que a língua estabelece naquele momento da produção e que é passível de análises pelo programa.
- d) Contorno da borda de língua: para cada produção do informante, foi feito, manualmente, o desenho sobre a imagem do contorno da língua fornecida pelo ultrassom para o momento da produção estabelecido. As imagens geradas pelo aparelho de ultrassom são transmitidas em escala de cinza e, em certos casos, o contorno da língua pode não aparecer de forma

nítida ou com suficiente contraste em relação ao restante da imagem. Nesses casos, o contorno da língua não foi realizado e o dado foi descartado.

- e) Criação de gráficos: na janela *Publisher* do AAA, foi possível gerar gráficos a partir de sobreposições de contornos das línguas, o que viabilizou a comparação de produções de uma mesma coleta.

O capítulo 4, a seguir, apresenta os resultados desta pesquisa, referentes às análises acústica e articulatória.

4 Resultados e discussão

Neste capítulo, serão expostos e discutidos os resultados das análises acústica e articulatória.

Este capítulo contém duas subseções, com suas respectivas subdivisões. Em 4.1, são apresentados e discutidos os resultados da análise acústica e, em 4.2, são apresentados os resultados da análise articulatória qualitativa, realizada com dados selecionados. Ambas as seções apresentam descrição e análise dos dados de fala adulta e infantil.

4.1 Análise acústica

Para as análises realizadas, foram coletados 840 dados de áudio; no entanto, parte desses dados foi descartada por não atender às condições necessárias para a realização de uma análise acústica confiável. Isso resultou em um total de 793 dados com qualidade apropriada para análise.

Esta seção se divide em duas subseções. Em 4.1.1, são apresentados e discutidos os resultados de análise acústica da fala adulta e, em 4.1.2, os resultados de acústica da fala infantil.

4.1.1 As adultas: AD1, AD2 e AD3

A amostra de fala adulta contou, inicialmente, com 330 dados de fala. Destes, 10 foram descartados por conterem forte ruído, sobreposições de vozes ou barulhos interferindo na produção de palavras. Assim, o total resultante de fala adulta para análise foi de 320 dados.

Cabe salientar que, por meio da análise acústica e da outiva dos dados de fala adulta, constatou-se que os sujeitos não tiveram alterações na produção das palavras como, por exemplo, a produção de róticos sem características especificadas ou que fossem diferentes do *tap* em aspectos acústicos.

Primeiramente, serão examinadas, de modo geral, as durações de obstruintes e vogais na fala adulta, de acordo com o tipo de obstruinte em posição de primeira consoante da sílaba (C_1) e com o tipo de sílaba inicial (CCV ou CV). O tipo de consoante em posição de C_1 foi um critério de separação para o cálculo de médias gerais, pois as consoantes iniciais presentes no *corpus* de produção – [p], [b] e [f] – têm diferenças consideráveis no que tange à sua duração e à porção da sílaba que ocupam, aspecto esse verificado por meio do cálculo de duração relativa.

A análise de médias gerais se relaciona a uma das questões norteadoras desta pesquisa, a saber: se a duração de vogais e obstruintes difere de modo significativo nos dados, sendo maior nas sílabas CCV, essa diferença caracteriza apenas um aspecto da fala da criança ou também é própria da fala adulta? Essa é uma das questões levantadas a partir dos resultados de trabalhos como os de Miranda (2007), Mezzomo *et al.* (2008), Miranda e Silva (2011) e Barbieri e Ferreira-Gonçalves (2017). Esses trabalhos, realizados com fala infantil, encontram maiores durações de C_1 e/ou da vogal núcleo em produções de alvos CCV como CV e consideram que essas maiores durações resultam de uma estratégia de alongamento compensatório.

Para se responder à questão, uma análise de como o aspecto da duração se comporta na fala adulta é importante. Isso porque, caso a diferença de duração também apareça na fala adulta, não se poderá afirmar que a diferença de duração é um aspecto próprio apenas da fala em aquisição.

No Quadro 11, a seguir, são mostradas as médias de duração gerais dos sujeitos adultos, separadas por tipo de obstruinte – [p], [b] ou [f] – e por tipo de sílaba – CCV ou CV.

Obst.	Sílabas	Palavra	Sílabas	Obstruinte				Vogal			
		Absoluta (ms)	Absoluta (ms)	Absoluta (ms)	DP	Relativa: sílaba (%)	DP	Absoluta (ms)	DP	Relativa: sílaba (%)	DP
[p]	CCV	427,0	237,0	15,9	6,10	6,8	2,8	170,0	14,0	71,7	2,9
	CV	383,3	198,0	18,6	3,4	8,5	2,5	180,2	13,5	91,1	3,1
[b]	CCV	462,5	283,2	86,6	23,8	29,4	3,8	131,1	20,6	46,2	3,6
	CV	437,6	261,6	102,4	27,4	38,2	2,5	159,3	25,1	61,8	2,5
[f]	CCV	549,4	373,3	159,8	36,1	42,6	2,8	157,4	17,6	42,2	1,9
	CV	513,9	328,0	176,1	33,0	54,0	0,8	151,6	22,2	45,9	0,8

Quadro 11 – Médias de duração da palavra e da sílaba, e média de duração e desvio padrão da obstruinte e da vogal em sílabas CCV e CV, classificadas por tipo de obstruinte em posição de C₁. As médias foram calculadas com base em dados de todas as informantes adultas – AD1, AD2 e AD3

O Quadro 11 demonstra que a duração de palavra e de sílaba é, de modo geral, maior nas palavras com CCV do que nas palavras com CV. Isso é algo esperado, considerando-se que a sílaba CCV apresenta três segmentos e a sílaba CV apenas dois. No entanto, ao examinarem-se as obstruintes e vogais, quase todas as durações, absolutas e relativas, são maiores em sílabas CV do que em CCV, independentemente do tipo de obstruinte. A única exceção, sinalizada em negrito no quadro, é a média de duração absoluta de vogais em sílabas com [f], a qual se mostrou maior em CCV, ao contrário do restante dos dados. No entanto, a duração relativa, que revela a porção que o segmento de fato ocupa na sílaba, se mostra maior em CV, atendendo ao padrão geral.

Com a análise de médias do Quadro 11, pode-se observar que existe, de modo geral, diferença de duração segmental entre CCVs e CVs na fala dos sujeitos adultos. No entanto, foi necessário examinar essas médias de duração também na fala de cada sujeito, para que se verificasse se não havia padrões diferenciados de duração na fala de alguma das informantes. Os Quadros 12, 13 e 14 mostram as médias de duração de obstruinte e vogal, desta vez cada um com base em dados de um sujeito. Os Quadros têm por base os dados de AD1, AD2 e AD3 respectivamente.

Obstruente	Sílabas	Obstruente		Vogal	
		Absoluta (ms)	Relativa: sílaba (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	11,2	4,5	185,5	74,8
	CV	20,1	6,6	194,9	93,3
[b]	CCV	108,3	33,2	131,1	49,0
	CV	132,7	41,0	186,7	59,0
[f]	CCV	197,3	45,2	177,1	40,7
	CV	210,6	55,0	172,8	44,9

Quadro 12 – Médias de duração da obstruente e da vogal em sílabas CCV e CV, classificadas por tipo de obstruente em posição de C₁, calculadas com base em dados da informante AD1

Obstruente	Sílabas	Obstruente		Vogal	
		Absoluta (ms)	Relativa: sílaba (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	13,7	5,9	166,2	71,2
	CV	14,1	7,4	177,6	92,6
[b]	CCV	90,5	29,3	136,0	47,5
	CV	95,3	37,8	154,1	62,3
[f]	CCV	156,9	43,0	152,2	41,6
	CV	173,0	53,6	153,7	46,4

Quadro 13 – Médias de duração da obstruente e da vogal em sílabas CCV e CV, classificadas por tipo de obstruente em posição de C₁, calculadas com base em dados da informante AD2

Obstruente	Sílabas	Obstruente		Vogal	
		Absoluta (ms)	Relativa: sílaba (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	22,8	9,9	158,1	69,0
	CV	20,0	11,3	168,2	87,5
[b]	CCV	61,1	25,6	98,0	42,1
	CV	79,1	36,0	137,2	64,1
[f]	CCV	125,2	39,6	142,9	44,5
	CV	144,7	53,6	128,5	46,3

Quadro 14 – Médias de duração da obstruente e da vogal em sílabas CCV e CV, classificadas por tipo de obstruente em posição de C₁, calculadas com base em dados da informante AD3

Com base nos Quadros 12, 13 e 14, pode-se afirmar que as médias individuais atendem, de modo geral, ao padrão já observado no Quadro 11: durações absoluta e relativa de obstruente e vogais maiores em sílabas CV. São

poucas as medidas em CCV e CV que, comparadas, não atendem a esse padrão, as quais também estão sinalizadas em negrito nos quadros. Essas medidas são: i) em AD1, duração absoluta de vogal, maior em CCV quando a consoante inicial é [f]; ii) em AD3, duração absoluta de obstruente, maior em CCV quando a consoante inicial é [p]; e iii) também em AD3, duração absoluta de vogal, maior em CCV quando a consoante inicial é [f].

Dentre esses casos, chama a atenção que o contexto de [f] propicie, aparentemente, uma maior duração absoluta da vogal em CCVs ao invés de em CVs, de modo que isso ocorre no grupo geral e também nos dados de AD1 e de AD3. No entanto, note-se que, nos casos i) a iii), as respectivas durações relativas atendem ao padrão geral, sendo maiores em CV. Os Gráficos 1 e 2, a seguir, evidenciam uma maior duração relativa da obstruente e da vogal, respectivamente, para cada uma das informantes adultas.

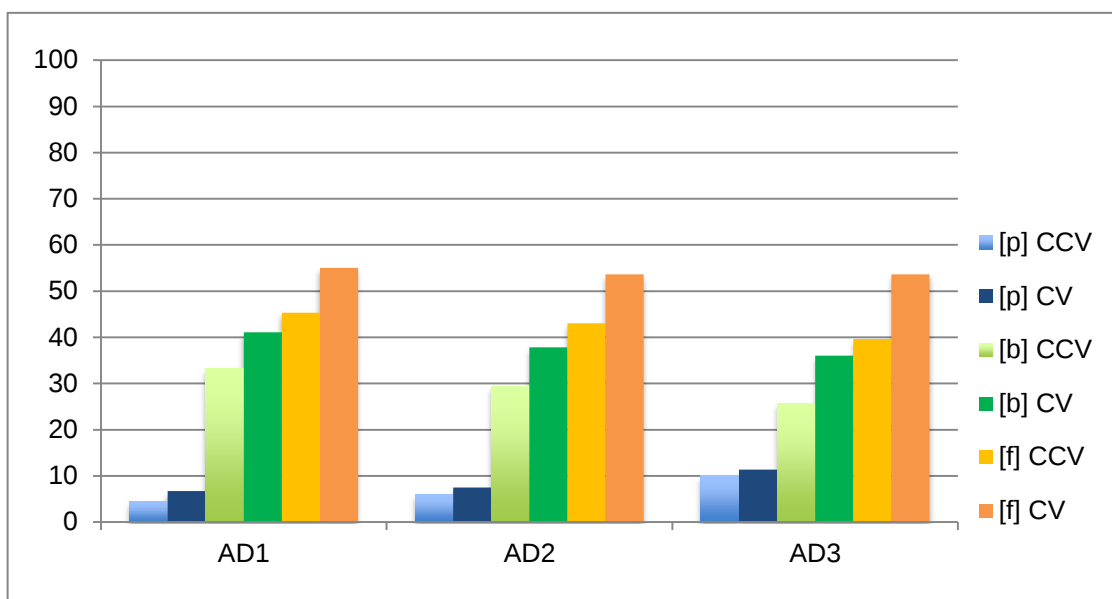


Gráfico 1 – Médias de duração relativa das obstruientes [p], [b] e [f], em sílabas CCV e CV, produzidas por AD1, AD2 e AD3

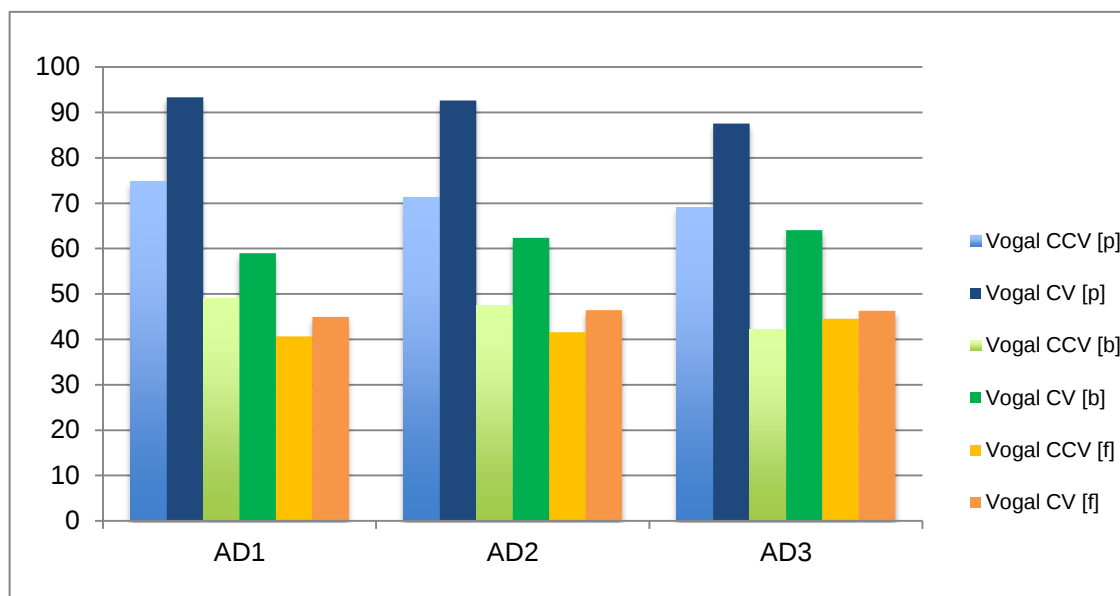


Gráfico 2 – Médias de duração relativa das vogais nos contextos de [p], [b] e [f], em sílabas CCV e CV, produzidas por AD1, AD2 e AD3

As análises de médias dos Gráficos 1 e 2 mostram que há diferença de duração segmental entre sílabas CCV e CV.

A aplicação do teste estatístico paramétrico Teste-T para amostras pareadas evidenciou diferenças significativas quando comparadas as durações – absoluta e relativa – das obstruintes em sílabas CCV e CV. As diferenças significativas constatadas podem ser visualizados no Quadro 15:

Segmento	Variável	Teste-t	Valor de p
[f]	Duração absoluta	-9,093 (2)	0,012
[p]	Duração relativa	-7,625 (2)	0,017
[b]	Duração relativa	-11,458 (2)	0,008
[f]	Duração relativa	-8,906 (2)	0,012
Vogal	Duração absoluta contexto [p]	-17,578 (2)	0,003
Vogal	Duração absoluta contexto [b]	-3,467 (2)	0,074*
Vogal	Duração relativa contexto [p]	-20,138 (2)	0,002
Vogal	Duração relativa contexto [b]	-4,474 (2)	0,047
Vogal	Duração relativa contexto [f]	-3,928 (2)	0,059*

*Diferenças marginalmente significativas

Quadro 15 – Diferenças significativas das durações – absoluta e relativa – de obstruintes e vogais, quando comparadas as sílabas CCV e CV

Por meio da estatística descritiva e inferencial até aqui dispostas por meio dos quadros e gráficos, é possível constatar que a diferença de duração não é um aspecto específico da fala em aquisição, mas também está presente na fala adulta. Tal constatação parece refutar a segunda hipótese deste trabalho.

Os resultados acerca das medidas de duração – relativa e absoluta – das obstruintes e vogais na fala adulta permitem, na verdade, que sejam formuladas duas explicações possíveis para a realização de sílabas CV, para alvos CCV, mais longas pelas crianças, conforme tem sido reportado pela literatura no processo de aquisição de estruturas silábicas complexas, como as constituídas por encontros consonantais:

- i) as diferenças observadas na fala infantil quanto às durações relativas da obstruinte e da vogal, considerando-se alvos CCV produzidos como CV, não evidenciam uma estratégia de alongamento compensatório. Em outras palavras, como o padrão da fala adulta apresenta essa diferença de durações quanto às obstruintes e às vogais, a criança adquire também esse aspecto temporal, produzindo consoantes e vogais mais longas em sílabas CV, mesmo que para alvos CCV;
- ii) a criança, ao alongar consoantes e vogais em uma sílaba CV para um alvo CCV, está de fato realizando um alongamento compensatório, pois o *input*

recebido – constituído por uma sílaba CCV do adulto – apresentaria vogais e consoantes menos longas do que em sílabas CV. A criança, assim, não estaria realizando consoantes e vogais mais longas em acordo com o padrão CV do português, mas como uma estratégia de alongamento compensatório na tentativa de produção de um alvo CCV.

A análise da duração da sequência elemento vocálico + vogal nuclear e elemento vocálico + rótico + vogal nuclear, no entanto, poderá descartar a explicação (ii) aqui formulada.

Nesse sentido, foram examinadas, primeiramente, as durações do elemento vocálico e do rótico, medidas presentes apenas nas palavras com sílabas CCV. Os Quadros 16 a 19 mostram, respectivamente, as médias de duração desses sons para o grupo de adultas e, individualmente, para AD1, AD2 e AD3.

Obst.	Sílabas	Elemento vocálico						Rótico			
		Absoluta (ms)	DP	Relativa: sílaba (%)	DP	Relativa: vogal (%)	DP	Absoluta (ms)	DP	Relativa: sílaba (%)	DP
[p]	CCV	34,5	5,0	14,5	1,6	20,4	1,8	15,9	2,4	6,7	1,5
[b]	CCV	44,8	6,0	16,4	3,5	36,9	10,6	19,9	3,5	7,7	4,8
[f]	CCV	37,5	9,6	10,0	1,2	24,3	3,4	18,3	4,11	5,1	1,9

Quadro 16 – Médias de duração e desvio padrão do elemento vocálico e do rótico em sílabas CCV, classificadas por tipo de obstruente em posição de primeira consoante da sílaba, calculadas com base em dados de todas as informantes adultas – AD1, AD2, e AD3

Obstruente	Sílabas	Elemento vocálico			Rótico	
		Absoluta (ms)	Relativa: sílaba (%)	Relativa: vogal (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	38,3	15,4	20,8	12,2	5,0
[b]	CCV	39,9	12,4	25,6	16,5	5,1
[f]	CCV	45,8	10,5	25,6	15,2	3,5

Quadro 17 – Médias de duração do elemento vocálico e do rótico em sílabas CCV, classificadas por tipo de obstruente em posição de C₁, calculadas com base em dados da informante AD1

Obstruinte	Sílabas	Elemento vocálico			Rótico	
		Absoluta (ms)	Relativa: sílaba (%)	Relativa: vogal (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	36,3	15,5	22,0	16,9	7,3
[b]	CCV	51,5	18,1	38,5	13,1	4,8
[f]	CCV	39,7	10,9	26,6	16,8	4,6

Quadro 18 – Médias de duração do elemento vocálico e do rótico em sílabas CCV, classificadas por tipo de obstruinte em posição de C₁, calculadas com base em dados da informante AD2

Obstruinte	Sílabas	Elemento vocálico			Rótico	
		Absoluta (ms)	Relativa: sílaba (%)	Relativa: vogal (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	28,8	12,6	18,3	15,9	8,0
[b]	CCV	42,9	18,8	46,7	20,1	13,3
[f]	CCV	26,9	8,5	20,6	23,0	7,3

Quadro 19 – Médias de duração do elemento vocálico e do rótico em sílabas CCV, classificadas por tipo de obstruinte em posição de C₁, calculadas com base em dados da informante AD3

De acordo com o Quadro 16, as maiores durações de elemento vocálico e rótico pertencem, de modo geral, às sílabas CCV que se iniciam por [b], sinalizando para o papel do vozeamento da consoante que os antecede. Essas maiores durações são tanto absolutas quanto relativas e estão sinalizadas em negrito no quadro.

Os Quadros 17 a 19 mostram as maiores durações também sinalizadas em negrito. Por se tratarem das médias individuais, é esperado que os padrões de maior duração estejam mais distribuídos, sendo encontrados em mais de um contexto relacionado à obstruinte. É interessante notar, no entanto, que todas as adultas têm a maioria das durações mais longas localizadas no contexto de [b]. No caso de AD2 e AD3, [b] parece propiciar maiores durações de elemento vocálico, porque ambas as informantes apresentam maiores durações de elemento vocálico – absoluta, relativa à sílaba e relativa à vogal – quando a sílaba se inicia por [b].

A aplicação do Teste-T de amostras pareadas revelou diferença significativa na duração relativa do elemento vocálico – com base na sílaba e na vogal – apenas em contexto de [f], quando comparado a [p]. As diferenças de duração de [b], em relação às demais obstruintes, apontadas na estatística descritiva, não se confirmam na estatística inferencial, cujos resultados podem ser observados no Quadro 20:

Segmento	Variável	Teste-t	Valor de p
EV	Duração relativa – sílaba [p] x [f]	19,429 (2)	0,003
EV	Duração relativa – vogal [p] x [f]	-4,193 (2)	0,052*
Rótico	Duração relativa – [p] x [f]	2,810 (2)	0,107*

*Diferenças marginalmente significativas

Quadro 20 – Diferenças significativas das durações – absoluta e relativa – do elemento vocálico (EV) e do rótico, quando comparadas em contextos de [p], [b] e [f]

Quando antecedido por [f], o elemento vocálico apresenta menor duração relativa em relação à sílaba e maior duração relativa em relação à vogal, do que quando antecedido por [p]. Já para o rótico, quando antecedido por [f], apresenta maior duração absoluta e menor duração relativa.

Assim, a estatística inferencial sinaliza para o papel do modo de articulação da consoante antecedente na duração relativa do elemento vocálico – em relação à sílaba e em relação à vogal – e na duração relativa do rótico.

Essas médias tornarão a ser discutidas na seção 4.1.2, quando serão comparadas aos padrões de duração apresentados pela criança.

Os Quadros 21 a 24, a seguir, mostram, então, os valores médios de duração de dois trechos específicos, também analisados nesta pesquisa: a soma das durações de elemento vocálico e vogal e a soma das durações de elemento vocálico, *tap* e vogal.

Tais trechos são aqui considerados para que se possa investigar, de forma mais precisa, a presença de alongamento compensatório em sílabas CV para alvos CCV. Assim, ao se considerar como medida de duração da vogal nuclear não apenas o trecho que a compreende, mas também a duração do elemento vocálico adjacente – que, conforme Silva, Clemente e Nishida (2006), constitui parte da vogal núcleo – o *input* para um alvo CCV poderia apresentar naturalmente vogais mais longas do que em sílabas CV. Ainda, seria possível considerar como medida de duração da vogal nuclear o trecho que compreende do início do elemento vocálico até o final da vogal núcleo, incluindo-se aí medidas de duração do rótico, o qual estaria sobreposto aos gestos vocálicos ainda em curso.

A maior duração de vogais em sílaba CCV, portanto, inviabilizaria a explicação formulada em (ii), pois o *input* CCV recebido do adulto passa a ser

constituído por vogais mais longas do que o *input* CV. A criança, então, quando produz vogais mais longas para alvos CCV realizados como CV, está simplesmente reproduzindo o padrão duracional da vogal (elemento vocálico + vogal nuclear ou elemento vocálico + rótico + vogal nuclear) de um *input* CCV.

Obstruinte	Sílabas	EV + vogal				EV + <i>tap</i> + vogal			
		Absoluta (ms)	DP	Relativa: sílaba (%)	DP	Absoluta (ms)	DP	Relativa: sílaba (%)	DP
[p]	CCV	204,4	18,4	86,3	4,37	220,3	16,5	92,9	3,5
[b]	CCV	175,9	23,6	62,1	5,9	195,8	20,1	69,1	6,5
[f]	CCV	194,9	26,6	52,2	1,11	213,2	22,9	57,1	3,0

Quadro 21 – Médias de duração dos trechos de elemento vocálico + vogal e elemento vocálico + *tap* + vogal, classificadas por tipo de obstruinte em posição de C₁, calculadas com base em dados de AD1, AD2 e AD3

Obstruinte	Sílabas	EV + vogal		EV + <i>tap</i> + vogal	
		Absoluta (ms)	Relativa: sílaba (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	223,7	90,2	235,9	95,1
[b]	CCV	171,0	52,6	187,5	57,7
[f]	CCV	222,9	51,1	238,1	54,6

Quadro 22 – Médias de duração dos trechos de elemento vocálico + vogal e elemento vocálico + *tap* + vogal, classificadas por tipo de obstruinte em posição de C₁, calculadas com base em dados de AD1

Obstruinte	Sílabas	EV + vogal		EV + <i>tap</i> + vogal	
		Absoluta (ms)	Relativa: sílaba (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	202,6	86,7	219,4	93,9
[b]	CCV	187,5	64,2	200,6	68,7
[f]	CCV	191,9	52,5	208,7	57,1

Quadro 23 – Médias de duração dos trechos de elemento vocálico + vogal e elemento vocálico + *tap* + vogal, classificadas por tipo de obstruinte em posição de C₁, calculadas com base em dados de AD2

Obstruinte	Sílaba	EV + vogal		EV + <i>tap</i> + vogal	
		Absoluta (ms)	Relativa: sílaba (%)	Absoluta (ms)	Relativa: sílaba (%)
[p]	CCV	186,9	81,5	202,8	88,5
[b]	CCV	140,9	60,5	161,0	69,2
[f]	CCV	169,8	53,3	192,8	60,6

Quadro 24 – Médias de duração dos trechos de elemento vocálico + vogal e elemento vocálico + *tap* + vogal, classificadas por tipo de obstruinte em posição de C₁, calculadas com base em dados de AD3

Os Quadros 21 a 24 mostram um padrão consideravelmente mais regular para os trechos em análise. De modo geral, todas as maiores durações, absolutas e relativas, se concentram no contexto de [p] como obstruinte inicial. Esse é um resultado esperado, já que essa obstruinte tem, tradicionalmente, as menores durações em relação a [b] e [f]: [p] é um som consonantal que não tem realização contínua e também não exige um pré-vozeamento para a sua realização. Logo, isso explica porque trechos da sílaba que coocorrem com essa obstruinte têm durações maiores, principalmente durações relativas.

O único valor de duração maior que destoa desse padrão, não se localizando no contexto de [p], está no Quadro 22: a duração absoluta de elemento vocálico + *tap* + vogal em contexto de [f], nos dados de AD1. De todo modo, esta duração apresenta menos relevância em relação à duração relativa, a qual normaliza aspectos prosódicos. Também, este valor de duração absoluta apresenta uma diferença de aproximadamente 4 ms em relação à segunda maior duração absoluta, que se localiza justamente no contexto de [p].

O Teste-T de amostras pareadas confirmou a maior duração, absoluta e relativa, das sequências elemento vocálico + vogal nuclear e elemento vocálico + rótico + vogal nuclear no contexto de [p], quando considerados os diferentes contextos antecedentes. As diferenças apontadas como significativas se encontram no Quadro 25.

Sequência	Variável	Teste-t	Valor de p
EV + V	Duração absoluta [p]x[b]	-3,276 (2)	0,082*
EV + V	Duração relativa [p]x[b]	5,100 (2)	0,036
EV + V	Duração relativa [p]x[f]	10,734 (2)	0,009
EV+R+V	Duração absoluta [p]x[b]	4,050 (2)	0,056*
EV+R+V	Duração relativa [p]x[b]	5,122 (2)	0,036
EV+R+V	Duração relativa [p]x[f]	9,378 (2)	0,011
EV+R+V	Duração relativa [b]x[f]	3,121 (2)	0,089*

*Diferenças marginalmente significativas

Quadro 25 – Diferenças significativas das durações – absoluta e relativa – das sequências elemento vocálico + vogal nuclear (EV+V) e elemento vocálico + rótico + vogal nuclear (EV+R+V), quando comparadas em contextos de [p], [b] e [f]

A seguir, os Gráficos 3 e 4 evidenciam, para fins de comparação, as médias de duração relativa – em relação à sílaba – de vogal em sílabas CV, de elemento vocálico + vogal núcleo e de elemento vocálico + rótico + vogal núcleo em sílabas CCV produzidas pelas adultas.

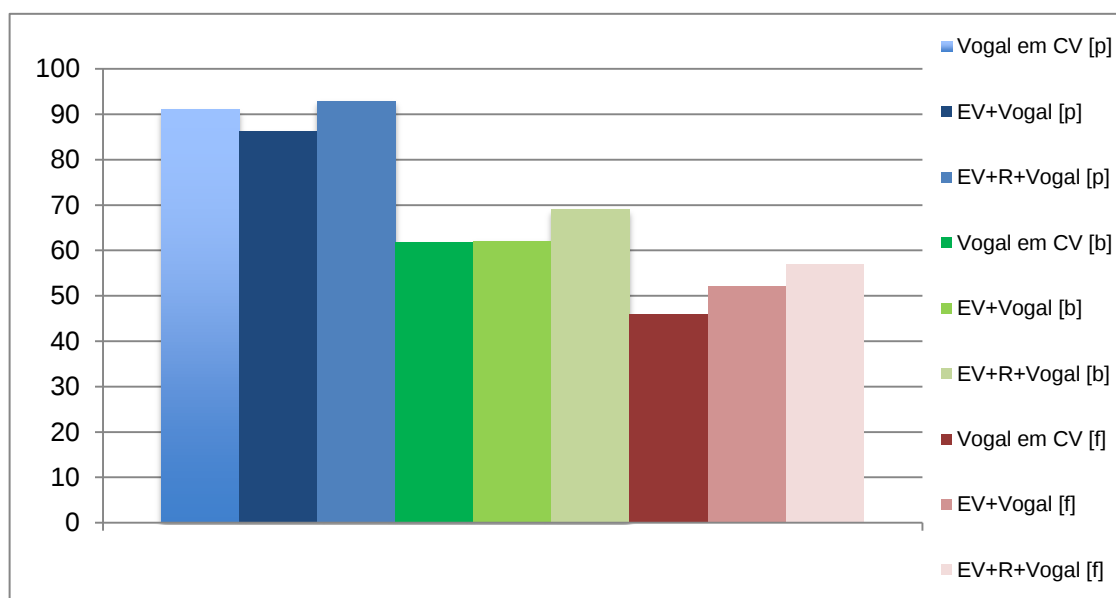


Gráfico 3 – Médias de duração relativa, em relação à sílaba, de vogal em sílabas CV, de elemento vocálico + vogal núcleo e de elemento vocálico + rótico + vogal núcleo em sílabas CCV, nos contextos de [p], [b] e [f], produzidas pelas adultas

De acordo com o Gráfico 3, em termos gerais, as médias de duração relativa da vogal núcleo em sílaba CV são menores dos que as constatadas em sílaba CCV, em contexto de [b] e [f], quando são considerados os trechos de elemento vocálico + vogal nuclear e de elemento vocálico + rótico + vogal nuclear. Apenas em contexto de [p] a duração da vogal nuclear em CV é maior; no entanto sua duração também é menor quando comparada à sequência de elemento vocálico + rótico + vogal nuclear.

No Gráfico 4, os valores de duração podem ser comparados em maior detalhe. Esse gráfico contém as medidas específicas de cada uma das três informantes.

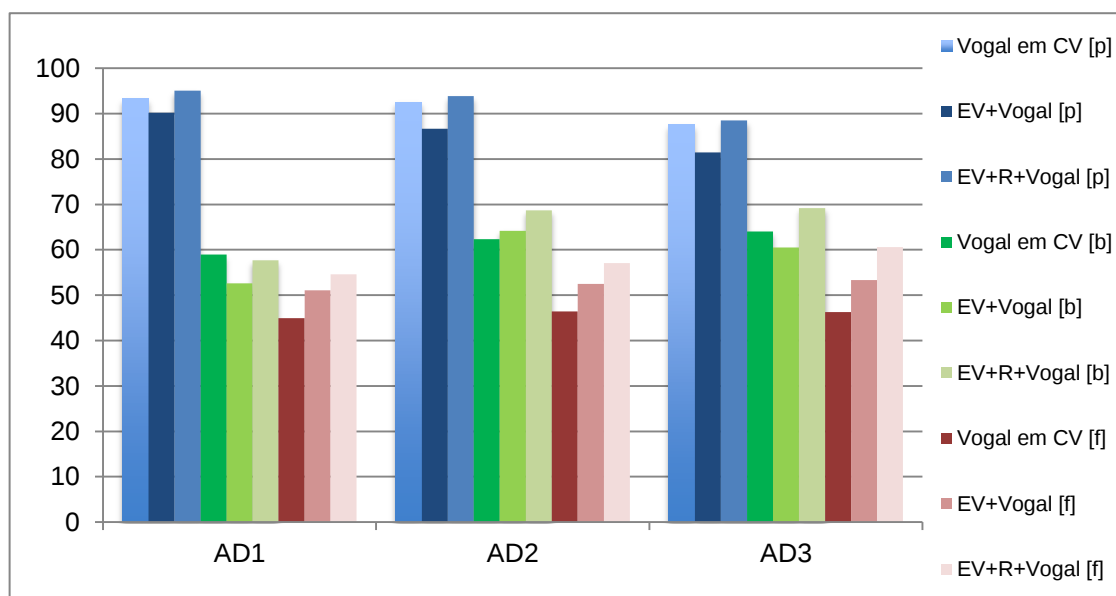


Gráfico 4 – Médias de duração relativa, em relação à sílaba, de vogal em sílabas CV, de elemento vocálico + vogal núcleo e de elemento vocálico + rótico + vogal núcleo em sílabas CCV, nos contextos de [p], [b] e [f], produzidas por AD1, AD2 e AD3

Observa-se, novamente, uma menor duração da vogal em sílaba CV quando comparada à sequência elemento vocálico + rótico + vogal nuclear em todos os contextos, à exceção do contexto de [b], para os dados de AD1. Nos contextos de [p] e de [f], os mesmos padrões da média geral – presente no Gráfico 3 – são constatados para as três informantes.

Com a aplicação do Teste-T para amostras pareadas, foram encontradas significâncias estatísticas quando a duração das sequências elemento vocálico +

vogal nuclear e elemento vocálico + rótico + vogal nuclear foram comparadas ao elemento nuclear em sílaba CV. Os valores significativos encontrados estão dispostos no Quadro 26.

Sequência	Variável	Teste-T	Valor de p
EV + V	Duração absoluta em contexto de [p]	-8,205 (2)	0,015
EV + R+ V	Duração absoluta em contexto de [p]	-17,176 (2)	0,003
EV + V	Duração absoluta em contexto de [f]	12,121 (2)	0,007
EV+R+V	Duração absoluta em contexto de [f]	-18,764 (2)	0,003
EV+V	Duração relativa em contexto de [f]	-22,589	0,002
EV+V	Duração relativa em contexto de [p]	-5,261	0,034
EV+R+V	Duração relativa em contexto de [f]	8,281	0,014
EV+R+V	Duração relativa em contexto de [p]	5,857	0,028

*Diferenças marginalmente significativas

Quadro 26 – Diferenças significativas das durações – absoluta e relativa – das sequências elemento vocálico + vogal nuclear (EV+V) e elemento vocálico + rótico + vogal nuclear (EV+R+V), quando comparadas à duração da vogal nuclear em sílaba CV, contextos de [p], [b] e [f]

No Quadro 26, o Teste-T aplicado aos trechos EV+V e EV+R+V mostrou diferenças significativas, em geral, para os contextos de [p] e [f]. Isso se revelou nas durações absolutas e relativas, tanto de elemento vocálico + vogal nuclear quanto de elemento vocálico + rótico + vogal nuclear, nos contextos das duas obstruintes desvozeadas. A partir dessas significâncias, presume-se que:

- i) no contexto de [f], como, na fala adulta, as durações dos trechos EV+V e EV+R+V são significativamente maiores do que a vogal núcleo de uma sílaba CV, não há, neste caso, uma condição propícia para que se verifique o alongamento compensatório na fala da criança. Assim, ao produzir vogais mais longas em sílabas CV para alvos CCV, a criança está apenas reproduzindo a diferença existente na fala adulta já presente na sequência EV+V;
- ii) no contexto de [p], na fala adulta, apenas a duração do trecho EV+R+V é significativamente maior do que a duração da vogal nuclear em CV; ao

contrário, a duração do trecho EV+V é significativamente menor do que a duração da vogal nuclear. Logo, existe a possibilidade de verificação de alongamento compensatório na fala infantil, pois, ao produzir vogais mais longas em sílabas CV para alvos CCV, a criança estaria reproduzindo diferenças constatadas na fala adulta quando da inclusão do rótico na sequência. O rótico, portanto, não é produzido pela criança, mas o padrão duracional da sequência EV+R+V, sim;

- iii) no contexto de [b], não há significância estatística nas diferenças entre vogais em CV e trechos EV+V e EV+R+V na fala adulta. Logo, seguindo-se essa interpretação, no contexto de [b], as produções CV para alvos CCV realizadas pela criança que apresentarem vogais mais longas do que para alvos CV não estariam reproduzindo os padrões de duração do *input* que são encontrados na fala adulta, podendo, portanto, indiciar outro processo não vinculado ao alongamento compensatório. Salienta-se, no entanto, que os resultados não reportaram diferenças significativas na fala adulta na comparação da duração de vogais em sílabas CV e CCV, nem mesmo ao se considerar a sequência EV+R+V.

Desse modo, quando a duração da média da sequência elemento vocálico + rótico + vogal nuclear é maior em relação à vogal nuclear da sílaba CV, tem-se uma condição que pode sinalizar, portanto, para o que até então é chamado de alongamento compensatório da vogal na fala infantil. A criança, ao alongar a vogal em uma sílaba CV para um alvo CCV, estaria, portanto, tentando realizar o padrão de duração do alvo adulto que envolve não apenas o elemento vocálico e a vogal nuclear, mas uma sequência em que o rótico está presente. Logo, o padrão duracional produzido, maior na sílaba CV para um alvo CCV, estaria de acordo com uma estrutura própria do encontro consonantal. O contexto de [p] parece, pois, mais indicado para essa investigação, pois há diferenças significativas entre a duração da vogal nuclear da sílaba CV e a duração da sequência elemento vocálico + rótico + vogal nuclear – com maior duração da sequência – e a duração da vogal nuclear da sílaba CV e a duração da sequência elemento vocálico + vogal nuclear – com menor

duração da sequência. A produção de formas alongadas pela criança, nesse contexto, poderia, pois, ser considerada um caso de alongamento compensatório.

Essa hipótese de análise será discutida em maior detalhe na próxima seção, na qual os valores médios de duração das adultas serão comparados com os valores correspondentes verificados nas produções da criança.

4.1.2 A criança: CR

A amostra de fala infantil contou inicialmente com 510 dados de fala, dos quais 37 foram descartados por conterem ruídos, sobreposições de vozes, barulhos interferindo na produção das palavras ou sons com características acústicas pouco definidas. Logo, o total resultante de fala infantil para análise foi de 473 dados. A seguir, o Quadro 27 mostra, para cada coleta, a quantidade de dados coletados, a quantidade de dados descartados e a quantidade final de dados para análise, considerados apropriados para a realização de análise acústica.

Coleta	#1	#2	#3	#4	#5	#6	#7	#8	Total
	4;1	4;3	4;5	4;8	4;9	4;10;15	4;10;27	4;11	
Dados coletados	48	66	66	66	66	66	66	66	510
Dados descartados	3	3	3	1	9	8	5	5	37
Dados analisados	45	63	63	65	57	58	61	61	473

Quadro 27 – CR: quantidade de dados coletados, descartados e submetidos à análise acústica

Os dados não descartados foram considerados apropriados para a realização de análise acústica por apresentarem qualidade de oscilograma e espectrograma suficientes para uma análise confiável. Nas análises, foram extraídas medidas de duração dos trechos da produção detalhados na seção 3.3.1.

Da mesma forma que para o adulto, as médias das produções da criança foram extraídas levando-se em conta o tipo de obstruinte em posição de C_1 – [p], [b] ou [f] – e o tipo de sílaba inicial – CCV ou CV. Assim, todas as médias dispostas nesta seção estão separadas por esses critérios. As médias também são examinadas por idade, na ordem em que as coletas ocorreram, para que se

observem as mudanças em duração que ocorrem ao longo da aquisição. As médias gerais dos sujeitos adultos são retomadas para comparação e aparecem sempre ao início das tabelas e gráficos, como referência de aquisição completa.

Os resultados da criança são divididos em seis subseções – 4.1.2.1 a 4.1.2.6. Na primeira seção, serão apresentados os resultados relativos à emergência da estrutura silábica CCV, com base na inspeção acústica dos dados. As seções subsequentes serão dedicadas ao exame das durações por unidade de análise, na ordem que segue: sílaba, obstruinte, vogal, elemento vocálico e rótico.

4.1.2.1 Emergência da estrutura silábica CCV

Para uma análise de como a estrutura silábica CCV emerge ao longo das coletas longitudinais, primeiramente, foi considerada como produção CCV toda realização em sequência de duas consoantes, com ou sem a presença do elemento vocálico à esquerda de C_2 . Esses critérios foram estabelecidos com base na inspeção acústica dos dados.

O Quadro 28 mostra as possibilidades de produção e as ocorrências da sílaba CCV em cada uma das coletas, considerando-se as diferentes obstruintes em posição de C_1 .

Obstruinte	CR: Possibilidades de produção e ocorrências de sílaba CCV							
	#1 (4;01)	#2 (4;03)	#3 (4;05)	#4 (4;08)	#5 (4;09)	#6 (4;10;15)	#7 (4;10;27)	#8 (4;11)
[p]	0/11	0/18	15/17	17/18	13/13	13/15	15/18	14/17
[b]	0/6	0/5	5/6	6/6	3/3	3/3	5/6	6/6
[f]	0/5	0/9	6/7	9/9	9/9	7/7	7/7	7/7
Total	0/22	0/32	26/30	32/33	26/26	23/25	27/31	27/30
%	0	0	86,7	96,9	100	92	87,1	90

Quadro 28 – Possibilidades de produção e ocorrências de sílaba CCV dispostas por idade (coletas #1 a #8) e por contexto consonantal de C_1

Como se pode observar, nas duas primeiras coletas, CR produz apenas sílabas CV para alvos CCV. Nas coletas subsequentes, a partir dos 4;05, a estrutura silábica CCV está presente na maior parte das produções. Os resultados não

indicam que haja um papel do tipo de C_1 na facilitação desse processo²², em termos do ponto de articulação da consoante.

O Gráfico 5 apresenta os percentuais de produção em cada uma das coletas realizadas.

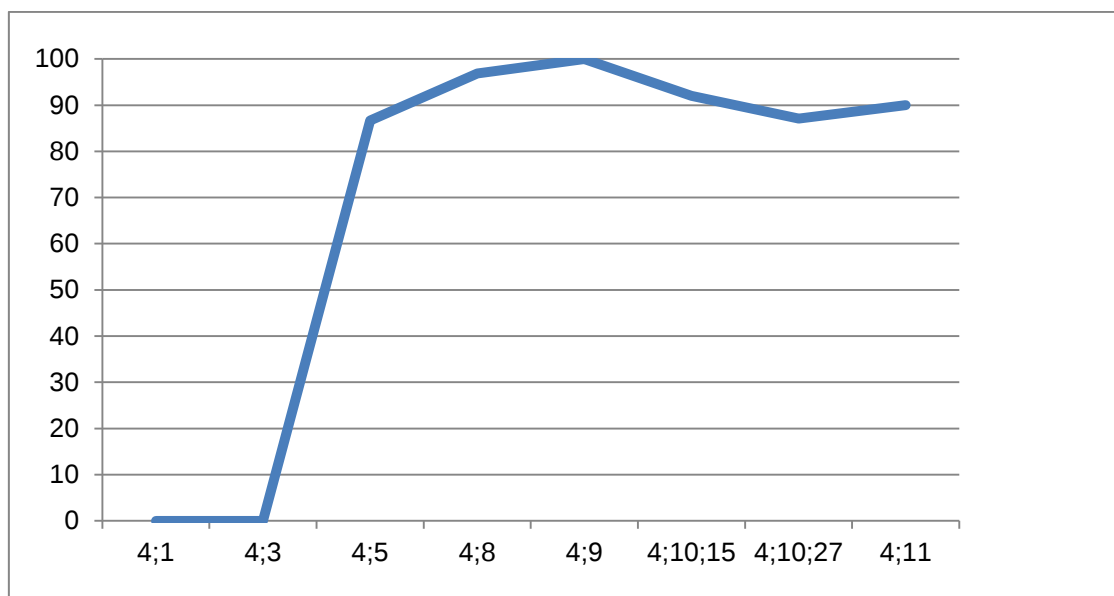


Gráfico 5 – Percentuais de produção da estrutura CCV na fala de CR, dispostos por idade

O Gráfico 5 mostra como o informante parte de uma total ausência de produção CCV, observável nas duas primeiras coletas, para índices que ultrapassam os 85%, atingindo 100% aos 4;09 e mantendo-se próximo aos 90% nas coletas seguintes.

O Quadro 29 e o Gráfico 6 apresentam a distribuição dos segmentos produzidos em C_2 .

²² É importante ressaltar que não se buscou verificar, na constituição do *corpus* deste trabalho, o papel de C_1 na emergência da estrutura silábica CCV. Isso se deve ao fato de o conjunto de palavras conter apenas segmentos labiais nessa posição, conforme discutido na seção 3.2.1.

Tipo de segmento em C ₂	CR: Elementos consonantais produzidos como C ₂								Total
	#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10;15)	#7 (4;10;27)	#8 (4;11)	
<i>tap</i>	0	0	19	21	21	17	26	25	129
vibrante	0	0	7	8	4	2	0	1	22
retroflexo	0	0	0	2	1	4	0	1	7
lateral	0	0	0	0	0	0	1	0	1
rótico nasalizado	0	0	0	1	0	0	0	0	1

Quadro 29 – Tipos de segmento produzidos em posição de C₂ para alvos CCV

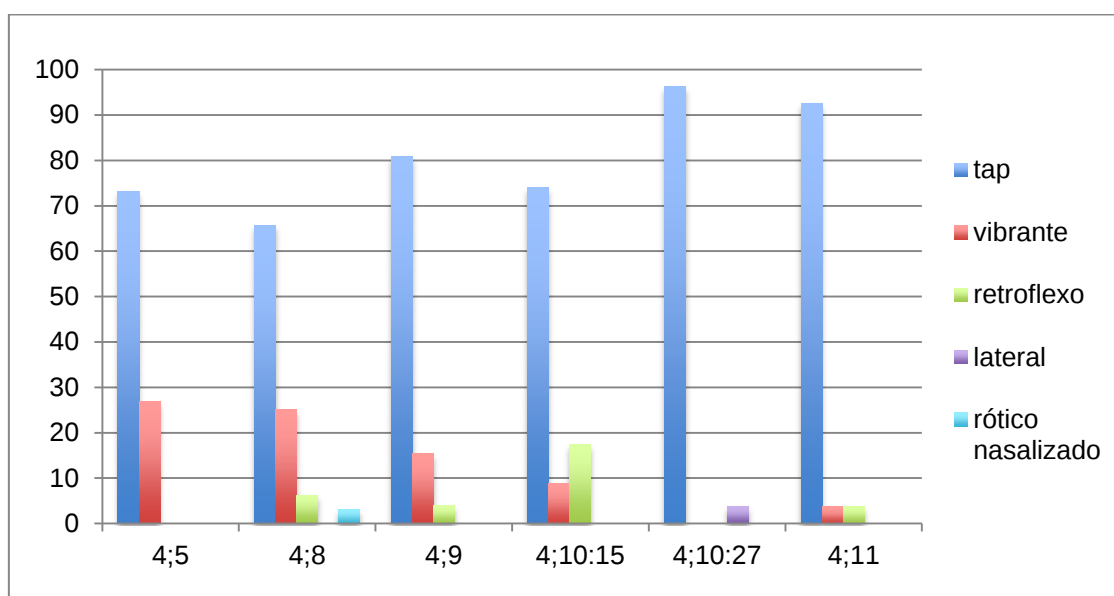


Gráfico 6 – Percentual de segmentos produzidos em C₂ para alvos CCV

Conforme o demonstrado no Quadro 29 e no Gráfico 6, há uma predominância do *tap* em posição de C₂, para alvos CCV. Essa predominância se mostra desde a emergência da estrutura CCV aos 4;05, permanecendo até a coleta final, aos 4;11. Outros segmentos, como vibrantes e retroflexos, estão presentes em quase todas as coletas, ainda que em percentuais baixos. A vibrante, por exemplo, é a segunda forma mais recorrente, sendo produzida em concomitância ao *tap*, de forma mais expressiva, nas três primeiras coletas. É nesse período, dos 4;05 aos 4;09, que estão sendo realizados ajustes temporais relativos à duração do rótico.

Em relação ao modo de articulação, observa-se que a vibrante e o retroflexo são produzidos em todos os contextos, no que diz respeito à obstruinte inicial: são produzidos quando antecidos por plosivas bilabiais e também quando antecidos pela fricativa labiodental. No entanto, a predominância de ocorrência desses outros tipos de rótico é de fato com a consoante fricativa, nas palavras “frita”, “frota” e “fraca”. Exemplos dessas ocorrências estão nas Figuras 39 a 41, a seguir.

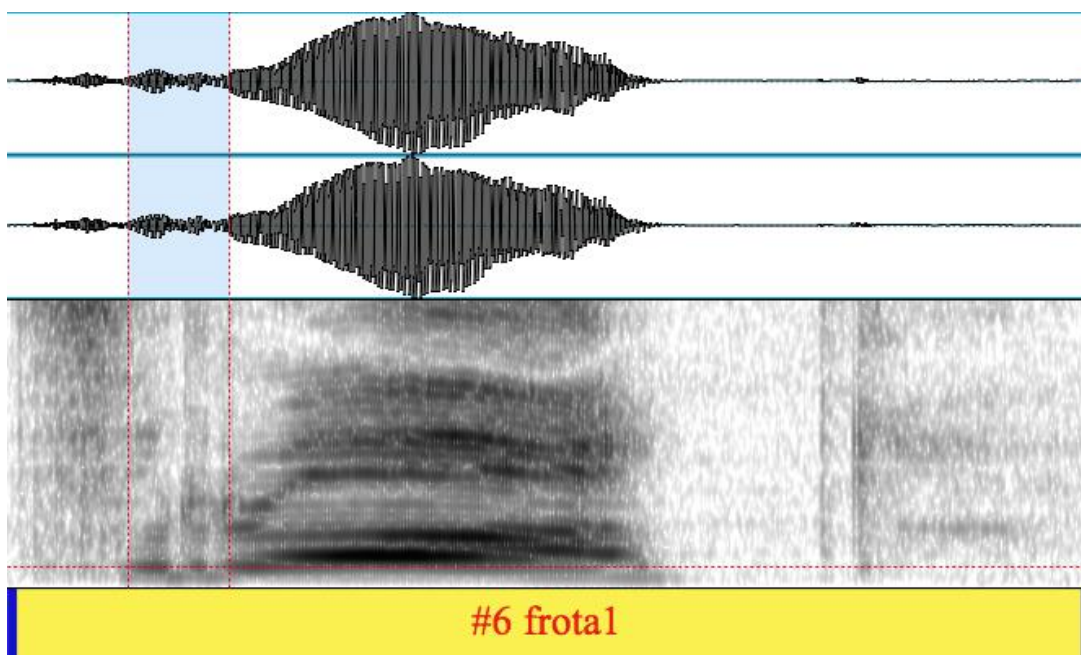


Figura 39 – Produção de CR para a palavra “frota”, com ocorrência de rótico vibrante

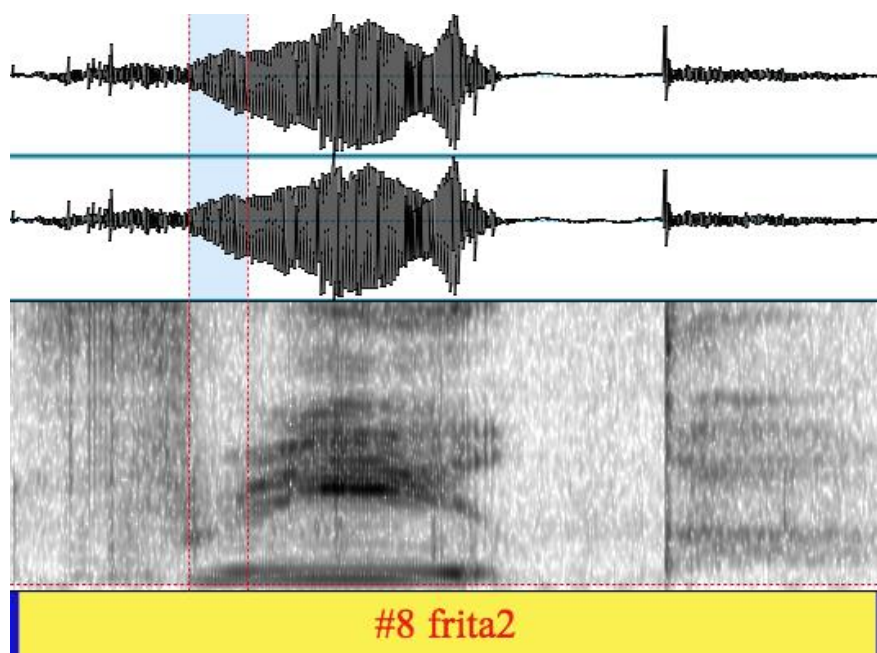


Figura 40 – Produção de CR para a palavra “frita”, com ocorrência de rótico retroflexo

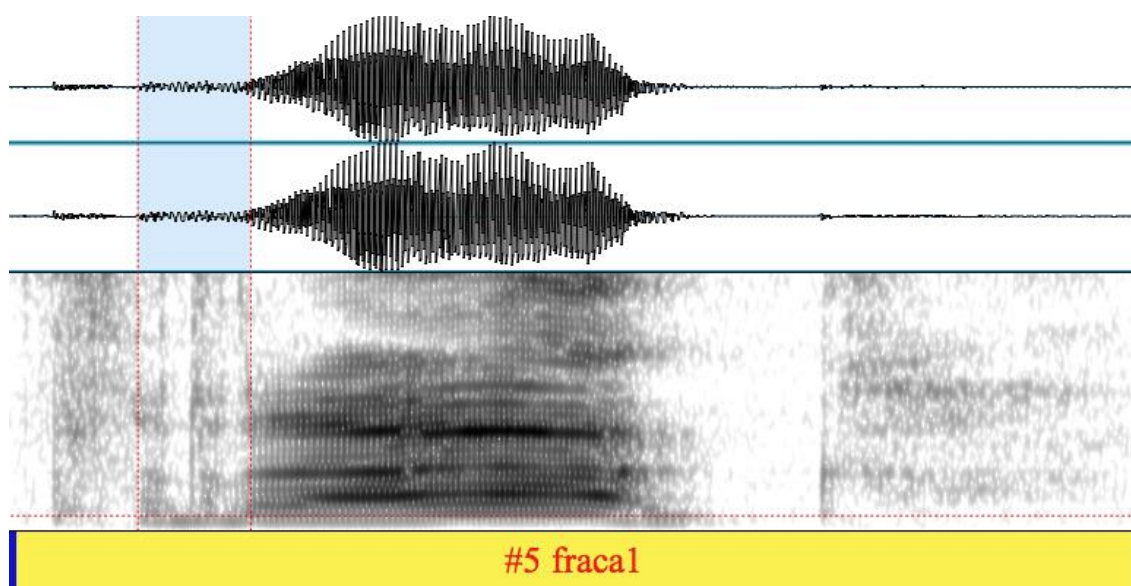


Figura 41 – Produção de CR para a palavra “fraca”, com ocorrência de rótico vibrante

Nas Figuras 39 a 41, é possível observar, nos trechos marcados em azul, produções diferenciadas em relação ao *tap*, segmento alvo. A produção da fricativa surda [f], que é um segmento consonantal de maior duração e com escape contínuo do ar, parece facilitar a repetição do gesto de ponta de língua na realização de

vibrantes, bem como a curvatura da ponta e o recuo do dorso na realização de retroflexos.

Também é importante citar que os *taps* produzidos pela criança não apresentam o mesmo padrão do alvo adulto, fato que será discutido em mais detalhe na seção 4.1.2.6. Esses *taps* são, por vezes, alongados, com durações que excedem os 50 ms. Exemplos dessas produções com maior duração estão nas Figuras 42 e 43.

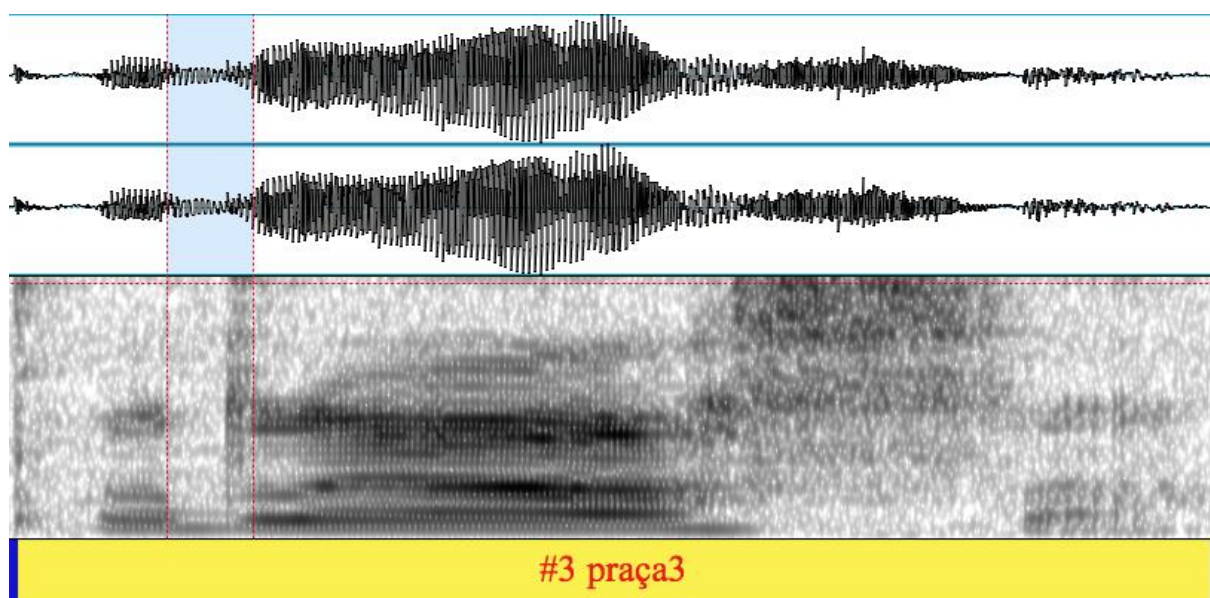


Figura 42 – Produção de CR para a palavra “praça”, com *tap* de 51 ms

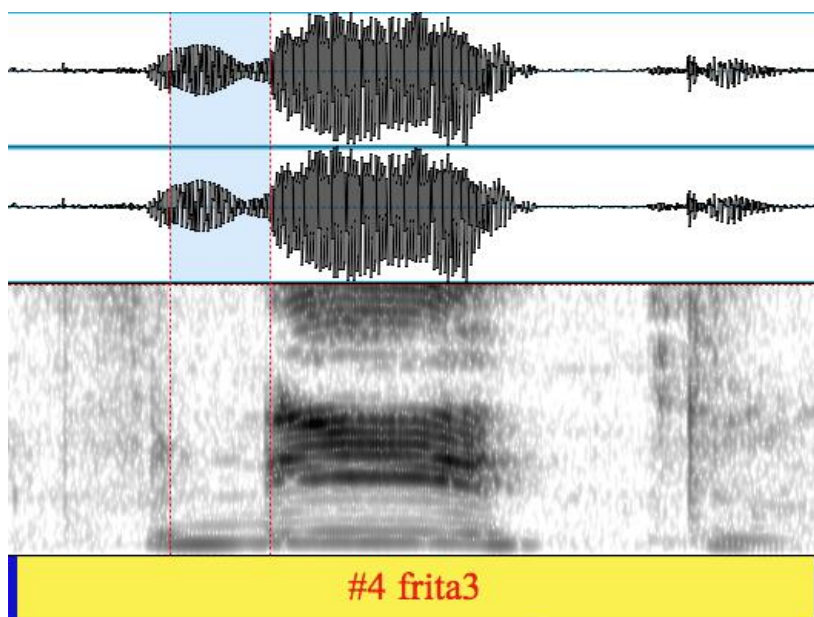


Figura 43 – Produção de CR para a palavra “frita”, com *tap* de 76 ms

É também com a consoante fricativa que são produzidos os *taps* de maior duração: o gesto labial alongado pela fricção de [f] cria o ambiente articulatorio adequado para a produção de um rótico igualmente mais longo, reforçando a sua produção.

Tendo em vista o que foi observado sobre a duração do *tap*, pode-se dizer que CR não busca apenas adequar os gestos necessários para a realização do segmento alvo em C₂. A criança também busca organizar esses gestos em relação ao seu tempo intrínseco e à coordenação aos elementos adjacentes, para, assim, construir o padrão silábico CCV.

Nesse sentido, destaca-se o papel da vogal núcleo da sílaba. Como discutido na seção 2.1.3, o *tap* em encontro consonantal se sobrepõe à vogal núcleo, conforme descrição de Silva, Clemente e Nishida (2006). Ao se sobrepor à vogal, o *tap* a entrecorta, criando, assim, um elemento vocálico adjacente à esquerda. A coordenação gestual necessária para a emergência desse elemento também precisa, portanto, ser dominada pela criança durante o processo de aquisição da estrutura CCV.

Todos os *taps* produzidos por CR são acompanhados por elemento vocálico, o qual ocorre com maior ou menor duração, conforme será discutido na seção 4.1.2.5. O mesmo não ocorre na produção de vibrantes e retroflexos, à exceção de três produções: “broa” (4;08) e “prata” (4;10;15), palavras produzidas com segmento retroflexo, e “frota” (4;11), produzida com segmento vibrante.

As seções 4.1.2.2 a 4.1.2.6, a seguir, são dedicadas à discussão sobre as durações na fala da criança.

4.1.2.2 Durações de sílaba

No Quadro 30, a seguir, estão dispostas as durações de sílaba nas produções de CR, juntamente com as médias gerais da fala adulta.

Sílaba: durações médias										
Obstruinte	Sílaba	Durações absolutas (ms)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	237,0	252,3	276,9	387,6	377,3	370,3	450,2	486,8	504,7
	CV	198,0	245,8	266,6	291,2	305,3	340,3	388,1	398,9	473,0
[b]	CCV	283,2	231,0	303,8	443,2	451,5	480,3	609,0	541,5	515,2
	CV	261,6	235,2	308,3	354,7	385,8	474,3	522,3	459,8	553,8
[f]	CCV	373,3	326,1	424,2	445,2	427,0	434,8	534,8	648,6	642,4
	CV	328,0	336,5	435,7	444,4	477,1	422,2	595,9	577,7	593,0
Obstruinte	Sílaba	Durações relativas (%)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	55,5	59,4	60,9	61,8	56,7	65,1	62,1	62,5	60,1
	CV	51,7	56,6	59,5	53,1	51,6	59,9	57,0	59,3	58,0
[b]	CCV	61,2	56,2	58,6	61,2	61,6	65,6	73,6	63,7	64,3
	CV	59,8	53,1	57,3	55,6	55,7	63,5	62,9	56,0	60,2
[f]	CCV	68,0	60,9	65,2	65,0	61,6	66,1	64,4	66,0	60,8
	CV	63,8	60,5	65,6	64,7	62,7	64,2	67,7	64,6	63,8

Quadro 30 – Médias de duração de sílaba para as produções da criança (CR), dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs)

Os Gráficos 7 a 12, a seguir, mostram as médias de duração conforme o tipo (relativa ou absoluta) e a obstruinte inicial ([p], [b] ou [f]).

Nos Gráficos 7 e 8, serão examinadas as médias de duração de sílaba em contexto de [p] como obstruente inicial.

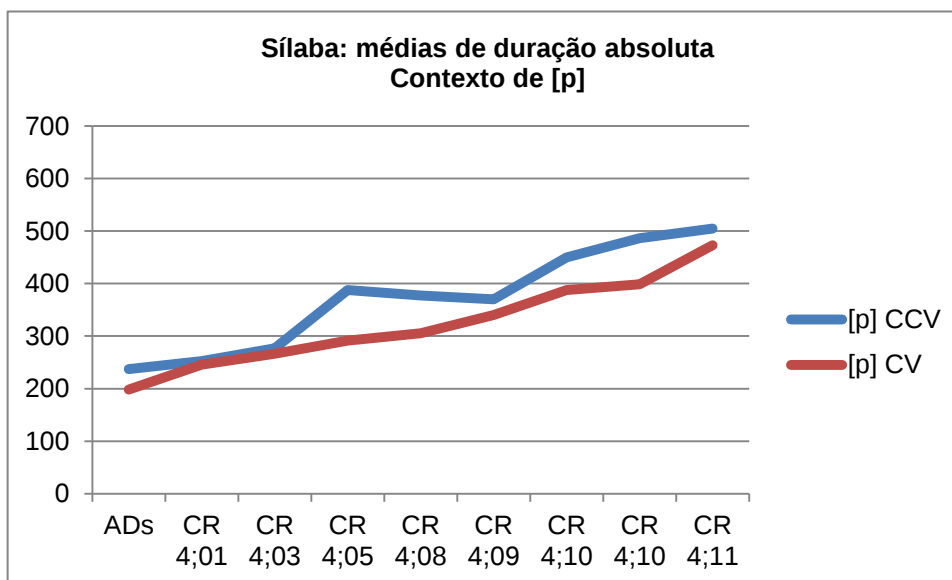


Gráfico 7 – Médias de duração absoluta de sílaba em contexto de [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

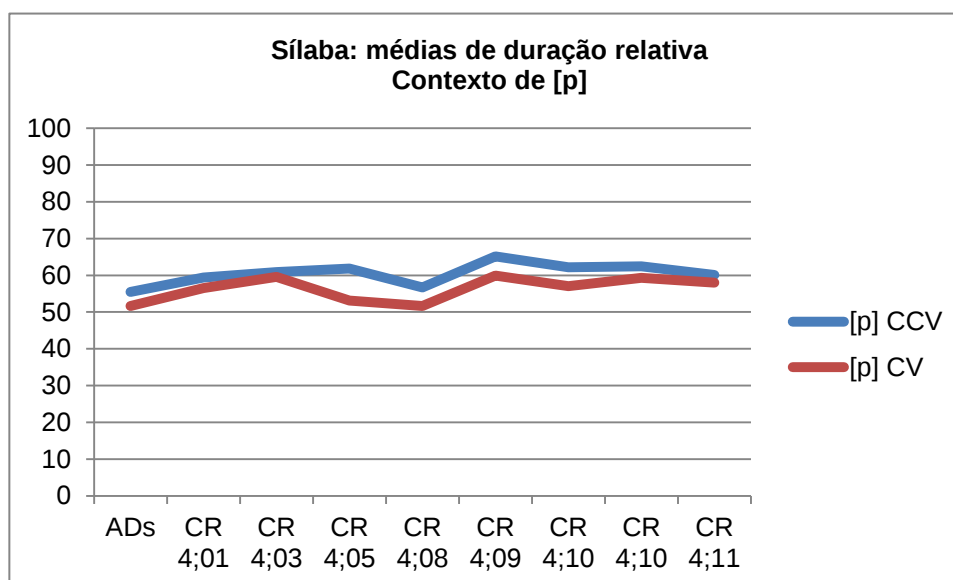


Gráfico 8 – Médias de duração relativa de sílaba em contexto de [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

Os Gráficos 7 e 8 demonstram que a duração de sílaba tem variação conforme o tipo de sílaba, sendo maior em CCV, tanto para as adultas quanto para a criança. As durações absolutas mostram que a criança tem sílabas que duram mais que as das adultas e, apesar de certa oscilação em CCV, a diferença em relação à duração na fala adulta tende a se acentuar. Ao longo da aquisição, as sílabas ficam mais longas. Isso se deve, em parte, ao fato de que, nas duas primeiras coletas, a criança não produzia nenhum encontro consonantal, o que explica, de certo modo, as durações absolutas similares entre CCVs e CVs nas coletas de 4;01 e 4;03. A partir dos 4;05, CR passa a apresentar diferenças mais visíveis entre médias de duração de CCVs e CVs, as quais são maiores em relação à diferença encontrada para as adultas. Isso ocorre principalmente nas coletas entre 4;05 e 4;10, período em que o padrão silábico CCV está sofrendo reajustes temporais mais expressivos.

Considerando-se as durações absolutas e relativas de obstruinte, pode-se presumir que o alongamento compensatório de vogais e consoantes, o qual supriria a ausência da produção do rótico, não se fará presente no processo de aquisição da estrutura CCV nos dados de CR. Como pode ser observado no Gráfico 7, a média de duração absoluta das sílabas CCV e CV de CR são, de modo geral, maiores do que as médias das informantes adultas. Isso ocorre não apenas nas coletas realizadas aos 4;01 e 4;03, em que a criança produz todas as CCVs como CVs, mas também nas coletas subsequentes, quando o rótico já é realizado.

A partir do que foi observado nos Gráficos 7 e 8, o que de fato ocorre é um aumento da duração segmental, tanto de consoantes quanto de vogais. Esse fenômeno continua a ocorrer mesmo quando o rótico já emerge nas produções, o que o configura como um dos ajustes temporais inerentes ao processo de aquisição fonético-fonológica, a partir dos pressupostos da Fonologia Articulatória.

Os Gráficos 9 e 10 mostram as médias de duração de sílaba em contexto de [b] como obstruinte inicial.

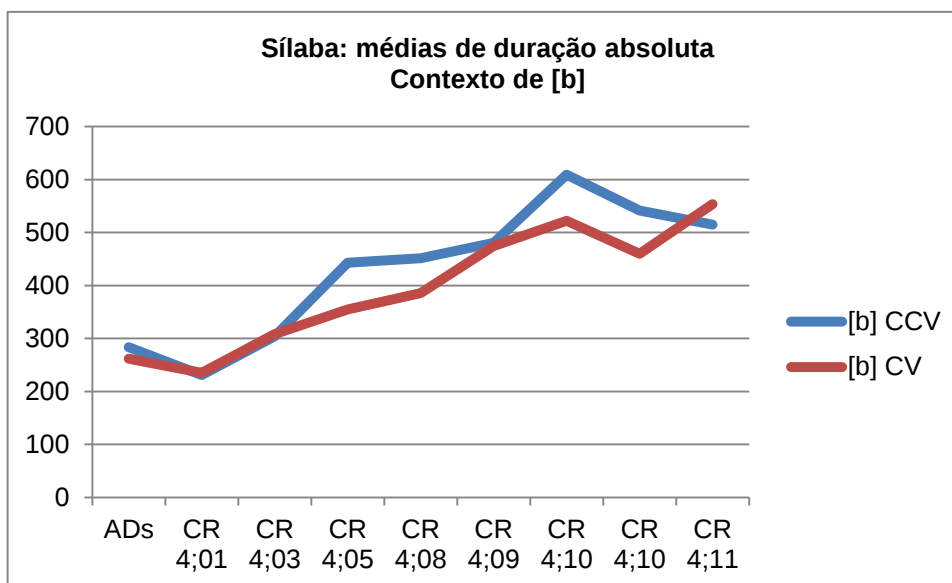


Gráfico 9 – Médias de duração absoluta de sílaba em contexto de [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

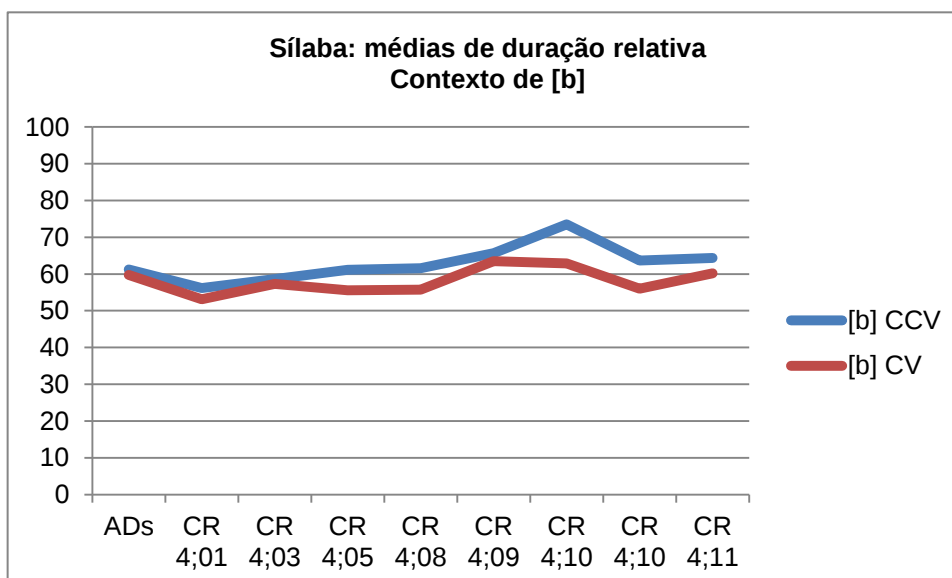


Gráfico 10 – Médias de duração relativa de sílaba em contexto de [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

Os Gráficos 9 e 10 demonstram que também há diferença em duração entre CCVs e CVs na fala da criança. Novamente, as durações absolutas são, de modo geral, maiores do que as das adultas, com exceção da coleta de 4;01. Nas coletas

de 4;01 e 4;03, em que a criança não produz encontros consonantais, as durações são semelhantes, sendo ligeiramente maiores em CVs, mas, nas coletas subsequentes, a duração de CCVs se mantém maior. Apenas aos 4;11 se observa uma mudança, na qual CCV tem duração absoluta inferior a CV, conforme se observa no Gráfico 9.

As durações relativas mostram variações menos expressivas em relação à fala adulta e ao longo do processo de aquisição, com oscilações. Há diferenças mais acentuadas nas duas coletas de 4;10 e uma aproximação dessas durações na última coleta.

De modo geral, o comportamento da sílaba em contexto de [b] é similar ao da sílaba em contexto de [p]: nesses dois contextos, há aumento progressivo da duração absoluta, mesmo quando a produção do rótico já ocorre, e uma diferença observável na duração, em que CCVs são mais longas que CVs.

Os Gráficos 11 e 12 mostram as médias de duração de sílaba em contexto de [f] como obstruente inicial.

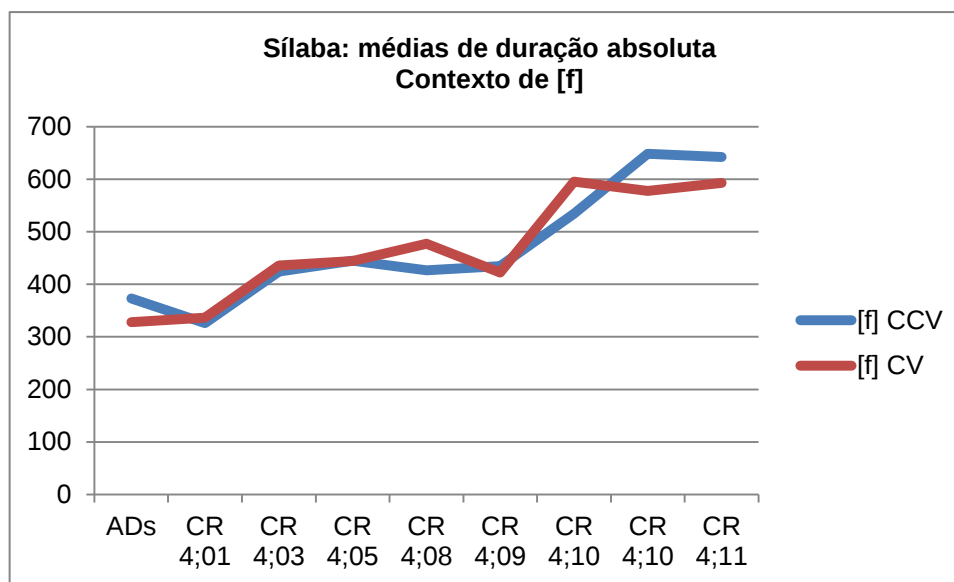


Gráfico 11 – Médias de duração absoluta de sílaba em contexto de [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

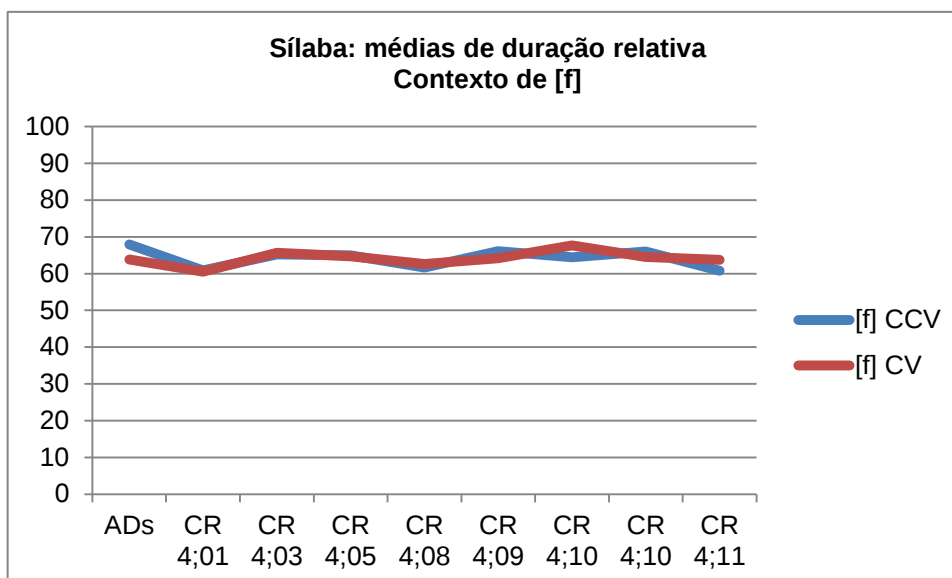


Gráfico 12 – Médias de duração relativa de sílaba em contexto de [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

Os Gráficos 11 e 12 mostram que, de modo similar aos contextos de [p] e [b], as sílabas com [f] produzidas pela criança têm durações absolutas mais longas do que a média das adultas, com exceção da coleta de 4;01. No entanto, os Gráficos não indicam uma diferenciação clara em duração em função do tipo de sílaba: as adultas produzem CCVs mais longas do que CVs, mas não há, para a criança, um claro padrão de um tipo de sílaba sendo mais longo do que o outro. As linhas de duração absoluta se entrecruzam e as durações mais longas são encontradas ora para CCVs, ora para CVs.

Nas durações relativas, as adultas também produzem CCVs mais longas do que CVs, mas, novamente, a criança não apresenta um claro padrão em que um tipo de sílaba é mais longo. Até os 4;08, as durações de CCVs e CVs com [f] são bastante similares, o que se verifica pelas linhas praticamente sobrepostas. Quando começam a aparecer diferenças, as linhas se entrecruzam, de modo que, também mais uma vez, CCVs e CVs se alternam na maior duração.

Considerando-se os Gráficos 7 a 12, as características observadas nas médias de duração de sílaba da criança são as seguintes:

- a) em todos os contextos de obstruinte, a duração absoluta da sílaba da criança aumenta progressivamente em relação à média adulta, com

poucas quedas ao final das coletas. Essa duração continua a aumentar mesmo quando o rótico já aparece nas produções, este com uma porcentagem inicial de ocorrência de mais de 85%. Isso sugere que o aumento na duração de segmentos não tem exatamente uma função compensatória em relação a um rótico ausente;

- b) nas sílabas com plosivas, há diferença na duração conforme o tipo de sílaba, de modo que CCVs são quase sempre mais longas que CVs. Essa diferença se mostra de forma mais sensível nas durações absolutas, mas também aparece nas durações relativas. Ela está presente principalmente a partir dos 4;05, que é justamente quando o rótico já é produzido;
- c) nas sílabas com fricativas, não há um claro padrão de diferença de duração conforme o tipo de sílaba. As médias de duração podem ora ser maiores em CCVs, ora em CVs;
- d) em todos os contextos de obstruinte, as medidas de duração relativa seguem um padrão mais estável, sem aumentos progressivos. Os valores de duração relativa passam por oscilações menores e não apresentam diferenças tão acentuadas em relação à fala adulta.

Tendo em vista as hipóteses deste trabalho, as características observadas não parecem contribuir para que se corroborem H1 e H2. No entanto, para uma real avaliação dessas hipóteses, será importante analisar as durações dos elementos específicos que compõem a sílaba, seja ela CCV ou CV.

Nas seções 4.1.2.3 a 4.1.2.6, a seguir, essas durações serão examinadas para se compreender, de forma mais detalhada, a organização temporal intrínseca de cada um dos segmentos.

4.1.2.3 Durações de obstruinte

O Quadro 31, a seguir, mostra as durações de obstruinte nas produções de CR, ao lado das médias gerais da fala adulta.

Obstruinte: durações médias										
Obst.	Sílaba	Durações absolutas (ms)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	15,88	34,3	25,9	35,4	12,6	12,9	17,7	37,3	31,8
	CV	18,63	29,2	17,2	22,7	8,6	14,6	21,5	46,6	37,2
[b]	CCV	86,63	32,3	85,8	106,7	105,2	240,0	180,3	79,2	67,7
	CV	102,37	54,0	58,4	102,3	152,3	158,5	170,5	94,3	132,1
[f]	CCV	159,81	180,2	183,7	128,9	94,1	109,9	130,4	211,0	194,7
	CV	176,09	141,9	184,6	172,3	190,2	135,8	229,1	230,1	193,6
Obst.	Sílaba	Durações relativas (%)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	6,77	13,3	9,3	9,0	3,3	3,6	3,9	7,6	6,0
	CV	8,46	11,7	6,5	7,9	2,9	4,3	5,7	11,3	8,1
[b]	CCV	29,38	15,2	26,4	22,8	21,8	42,5	29,3	13,7	12,1
	CV	38,24	24,3	17,9	26,6	40,8	32,7	32,6	18,2	22,4
[f]	CCV	42,59	55,7	43,2	29,6	22,0	25,0	24,3	32,6	30,6
	CV	54,04	41,8	42,2	38,2	39,1	31,0	38,5	39,4	32,4

Quadro 31 – Médias de duração de obstruinte para as produções da criança (CR), dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs)

Os Gráficos 13 a 18, a seguir, mostram as médias de duração separadas conforme o tipo (relativa ou absoluta) e a obstruinte inicial ([p], [b] ou [f]).

Nos Gráficos 13 e 14, são mostradas as médias de duração de obstruinte em contexto de [p] como obstruinte inicial.

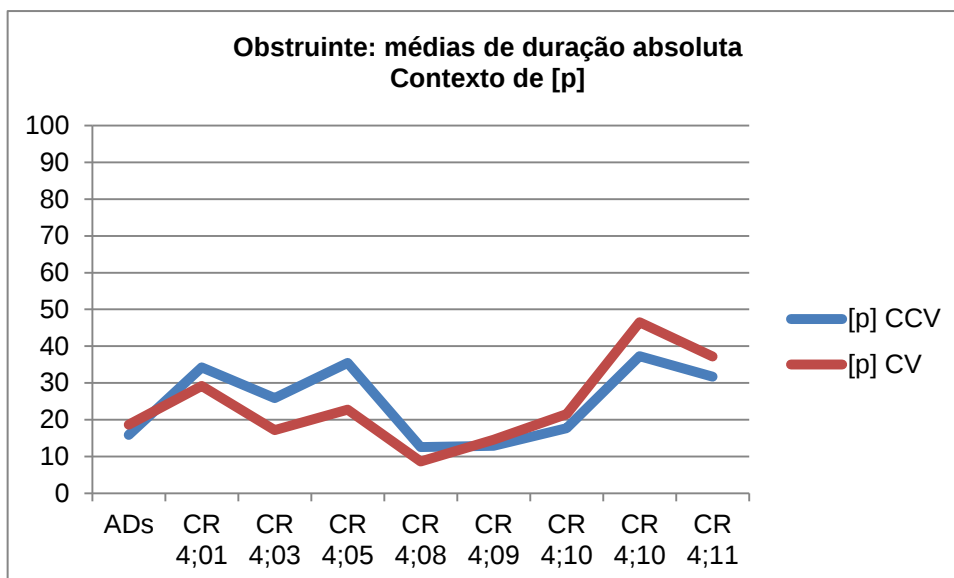


Gráfico 13 – Médias de duração absoluta da obstruinte [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

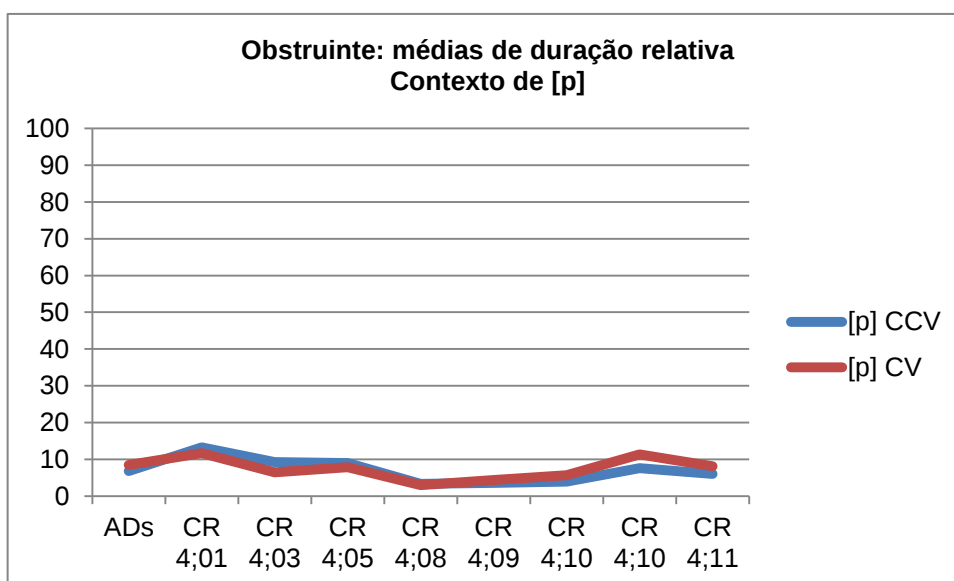


Gráfico 14 – Médias de duração relativa da obstruinte [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

Na fala adulta, as médias de duração absoluta e relativa para [p] basicamente não apresentam diferenças conforme o tipo de sílaba, se CCV ou CV. Os dados de CR evidenciam padrão similar ao se examinar, no Gráfico 14, a duração relativa; CR apresenta, inclusive, valores de duração relativa similares aos das adultas.

As linhas que representam as durações em CCV e CV frequentemente se sobrepõem, apresentando máxima diferença aos 4;05, nas durações absolutas. Essa diferença é de 13,7 ms (35,4 ms em CCV contra 22,7 ms em CV). Esses dados parecem não indicar diferenças de duração conforme o tipo de sílaba.

Quanto às durações absolutas dispostas no Gráfico 13, é interessante observar-se que CR produz a plosiva surda inicialmente com maior duração em sílaba CCV. Esse padrão se altera a partir dos 4;09, quando [p] passa a ser mais longo em sílabas CV, se assemelhando ao padrão apresentado pelas informantes adultas. Esse fato, no entanto, não indica que haja um alongamento compensatório de C_1 em sílabas CV: [p] apresenta inicialmente maior duração em sílabas CCV, não apenas nas duas primeiras coletas, nas quais C_2 não é produzida, mas também na coleta subsequente, em que a estrutura CCV é realizada em mais de 85% das possibilidades de produção. Ainda, as durações de [p] se tornam maiores em CVs justamente aos 4;09, idade em que a produção de róticos apresenta o percentual máximo de 100%, como já exposto anteriormente, no Quadro 28.

Os Gráficos 15 e 16 mostram as médias de obstruente para sílabas com [b].

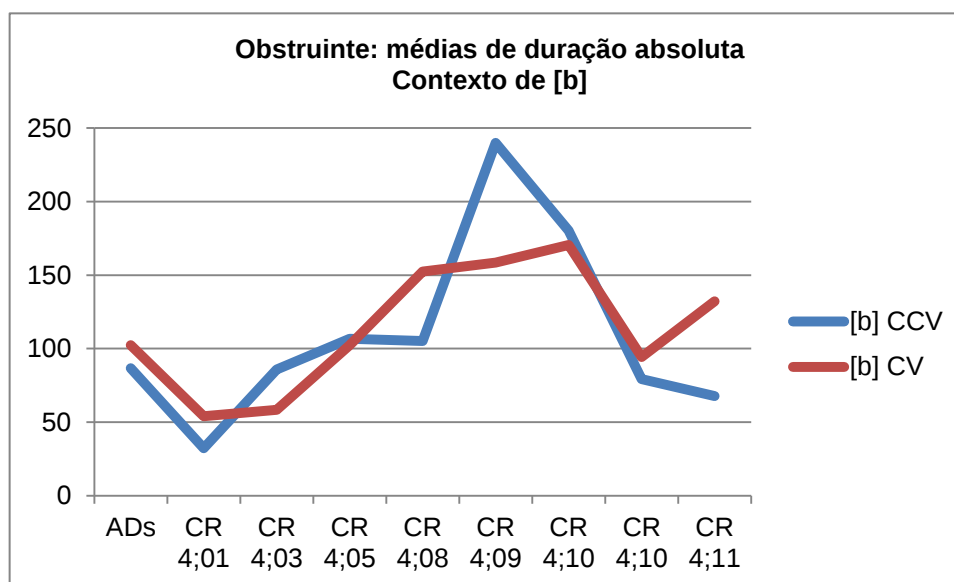


Gráfico 15 – Médias de duração absoluta da obstruente [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

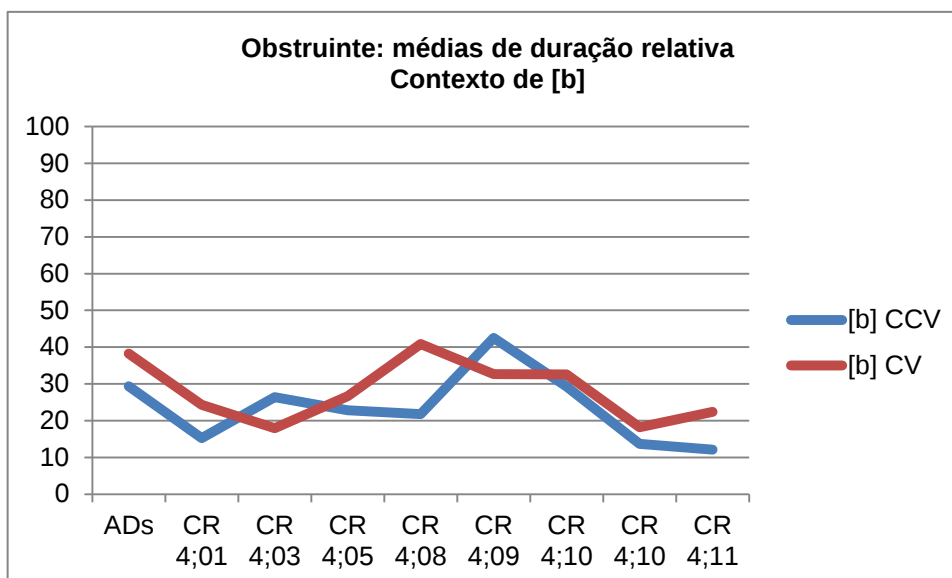


Gráfico 16 – Médias de duração relativa da obstruinte [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

De acordo com o Gráfico 15, as durações absolutas da obstruinte [b] da criança passam por mudanças em relação à mesma obstruinte produzida na fala adulta. Aos 4;01 e 4;03, essa duração começa menor do que a duração do [b] das adultas, provavelmente devido a se encontrar em progresso um domínio do pré-vozeamento que faz parte da produção do [b]. Considerando-se que algumas produções de [b] não tinham esse pré-vozeamento claramente visível na acústica ou não tinham qualquer tipo de vozeamento, sendo detectadas na acústica como produções de [p], foi esperado que as durações dessa obstruinte fossem, em determinado momento, menores do que o observável para um sujeito adulto.

Já entre os 4;05 e 4;10, as durações absolutas são maiores em relação ao padrão das adultas. Isso pode estar relacionado a um momento em que o pré-vozeamento é realizado, mas a sua duração ainda está sendo manipulada e varia consideravelmente, chegando a uma média de 240ms em CCVs produzidas aos 4;09. Na segunda coleta dos 4;10, o aspecto da duração parece se estabilizar, com valores que se aproximam aos das adultas, e, aos 4;11, se abre uma maior diferença de duração novamente. Já as durações relativas, apresentadas no Gráfico 16, têm menor variação em relação ao padrão adulto, de modo que as obstruintes da criança perfazem porcentagens menores da sílaba em relação à porcentagem analisada nas adultas, com poucas exceções.

A diferença de duração entre CCVs e CVs, ao longo da aquisição, também não apresenta um padrão claro, já que as linhas se entrelaçam tanto na duração absoluta quanto na relativa, com obstruintes mais longas ora em CCV, ora em CV, ou durações com valores muito próximos. A partir dessas características, conclui-se que, mais uma vez, não há indícios de que CR realize um alongamento compensatório de obstruinte em produções CV para o alvo CCV.

Os Gráficos 17 e 18 apresentam as médias de duração de [f] em sílabas CCV e CV.

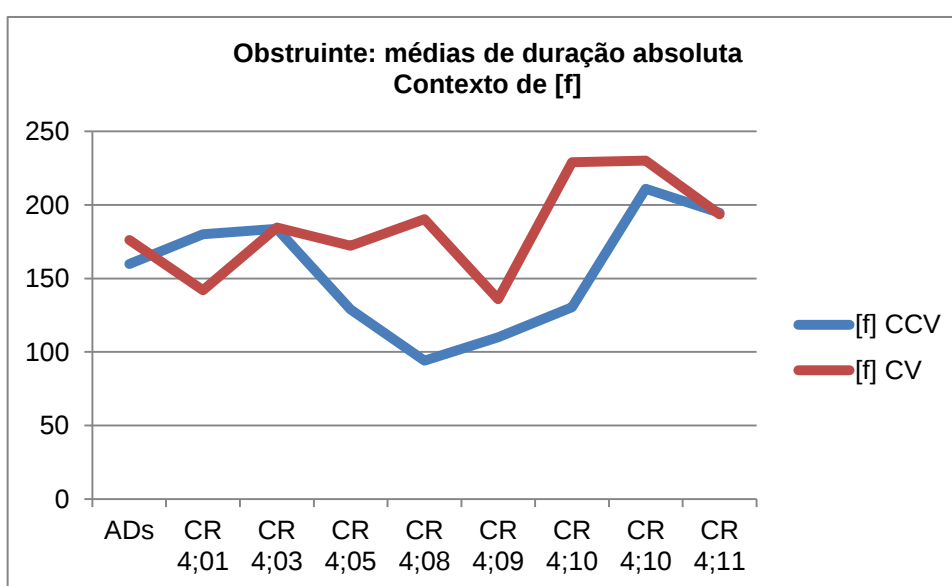


Gráfico 4.15 – Médias de duração absoluta da obstruinte [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

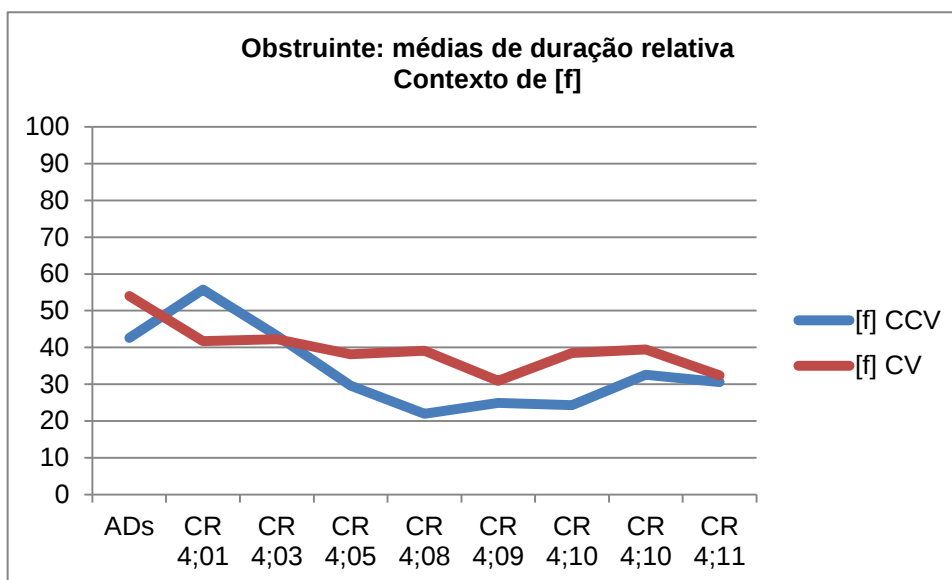


Gráfico 18 – Médias de duração relativa da obstruinte [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

O Gráfico 4.15 demonstra que, em relação à fala adulta, as durações absolutas de [f] da criança passam por oscilações, de modo que são, em alguns momentos, maiores que os das adultas e, em outros momentos, menores. Já no Gráfico 18, as durações relativas apontam pra durações menores em relação às adultas, com exceção das médias de 4;01 para o alvo CCV.

Na comparação entre CCVs e CVs da criança, observa-se que, com exceção da coleta de 4;01, as durações absolutas de [f] são maiores em CV do que em CCV. O mesmo ocorre a partir dos 4;05 nas durações relativas. Também é interessante observar que, aos 4;11, as diferenças de duração se neutralizam tanto nas absolutas quanto nas relativas.

De modo geral, a partir dos Gráficos 13 a 18, as características observadas são as seguintes:

- a) a duração não parece seguir movimentos sistemáticos de aumento ou diminuição no processo de aquisição;
- b) nas sílabas com plosivas, estas são ora mais longas em CCV, ora em CV, com considerável alternância. Essa alternância está expressa, nos gráficos, pelo constante entrecruzamento das linhas;

- c) as sílabas com fricativa, no entanto, parecem apoiar a hipótese do alongamento compensatório. O fato de, aos 4;01, serem constatados valores de duração para [f] maiores em CCV do que em CV pode indicar que a maior duração da fricativa, neste contexto, estaria funcionando como um mecanismo compensatório no processo de aquisição.

A verificação de significância estatística dos resultados aqui apontados foi realizada por meio da aplicação dos testes estatísticos Teste-T e One-Way ANOVA, e incluiu 5 etapas com a comparação das medidas de duração – absoluta e relativa – das obstruintes: (i) resultados de CR x adultas; (ii) resultados das duas primeiras coletas de CR x adultas; (iii) resultados das coletas 3 a 8 de CR x adultas; (iv) resultados das duas primeiras coletas de CR x coletas 3 a 8 de CR e (v) alvo CV x alvo CCV.

Em relação a (i), a aplicação do teste estatístico One-Way ANOVA identificou diferenças significativas entre as produções das adultas e das crianças apenas no que concerne à duração relativa de [b] ($F=5,528$, $p=0,043$) e [f] ($Z=44,287$, $p=0,000$), ambos para alvos CV. A criança apresenta, assim, ao longo das coletas, valores menores de duração relativa do segmento fricativo quando comparado ao padrão constado na fala adulta. Considerando-se que não há diferenças significativas quanto à duração absoluta dos segmentos obstruintes, infere-se, portanto, que [b] e [f] ocupam porções menores da sílaba CV provavelmente em detrimento da duração mais longa dos segmentos vocálicos.

A presença de alongamento compensatório dos segmentos obstruintes para alvos CCV pode ser mais bem investigada se comparadas apenas as durações das produções iniciais – primeira e segunda coletas –, quando o rótico e a estrutura CCV não foram realizados pela criança, com os resultados das adultas. Assim, no que concerne a (ii), para alvos CCV, o teste One-Way ANOVA detectou apenas uma diferença marginalmente significativa para a duração absoluta de [p] ($F=6,610$, $p=0,082$) – maior nas produções de CR –, o que parece frágil para a confirmação da hipótese 1 da presente dissertação. A significância estatística, por outro lado, pode ser constatada na duração relativa de [b] ($F=31,852$, $p=0,011$) e [f] ($F=174,72$, $p=0,000$), mas para alvos CV, com durações menores da obstruinte, conforme já reportado, com base em todas as coletas de CR, no parágrafo anterior.

Na comparação das coletas subsequentes – a partir da produção da estrutura CCV por CR – com os resultados das adultas, o teste One-Way ANOVA constatou apenas diferenças significativas no que concerne a [f]: menor duração relativa para alvos CCV ($F=31,748$, $p=0,001$) e CV ($F=61,834$, $p=0,000$). Uma vez que não foram encontradas diferenças significativas no que se refere à duração absoluta da fricativa, a coordenação temporal do gesto de *abertura labial, crítico, dental* com os demais segmentos da sílaba – seja CV ou CCV – é que, de fato, parece estar sendo reajustada por CR.

Quando comparados os resultados de duração das duas primeiras coletas de CR com os das coletas subsequentes – ausência da estrutura CCV x presença da estrutura CCV –, constataram-se as seguintes diferenças significativas:

Segmento	Variável	F	Valor de p
[b]	Duração absoluta alvo CV	11,509	0,015
[f]	Duração relativa alvo CV	4,032	0,091*
[p]	Duração relativa alvo CCV	8,242	0,028
[f]	Duração relativa alvo CCV	26,679	0,002

*Diferenças marginalmente significativas

Quadro 32 – Diferenças significativas das durações – absoluta e relativa – de obstruintes, quando comparadas as duas primeiras coletas da criança com as coletas subsequentes

A menor duração de [b] e [f] para alvos CV, nas coletas iniciais, justifica-se pelo fato de, nas coletas subsequentes, ocorrerem ajustes no padrão de vozeamento e soltura do ar da fricativa. São esses, inclusive, os dois segmentos apontados pela estatística como apresentando diferenças entre as produções de CR e das adultas. Para o segmento fricativo, inclusive, constatou-se, conforme já reportado, que CR continua apresentando, de forma significativa, uma menor duração relativa do que as adultas, em sílaba CV e CCV, nas coletas subsequentes.

A maior duração relativa de [p] e [f] para alvos CCV nas duas primeiras coletas, em comparação às coletas subsequentes – quando a sílaba CCV é, então realizada –, não parece, no entanto, indiciar a presença de alongamento compensatório, uma vez que, nas produções das adultas, foram igualmente detectadas diferenças entre a duração das obstruintes quando comparadas sílabas

CV e CCV. Por constituir uma sílaba com três elementos, a obstruinte terá, justamente, a tendência de ocupar uma porção menor da estrutura.

Por fim, foram comparados os valores de duração dos segmentos obstruintes para alvos CV e alvos CCV nas duas primeiras coletas de CR e nas coletas subsequentes. Em relação às primeiras coletas, o Teste-T para amostras pareadas não detectou nenhuma diferença significativa nas medidas de duração absoluta e relativa quando comparados alvos CV e CCV de CR. Nas coletas subsequentes, diferenças estatísticas foram constatadas em relação a uma maior duração relativa de [f] em sílaba CV ($t=-3,952$ (5), $p=0,011$).

Os resultados estatísticos até aqui reportados não atestam, portanto, a presença de alongamento compensatório das obstruintes nas produções de CR.

Na seção 4.1.2.4, a seguir, são expostas e discutidas as médias de duração de vogal.

4.1.2.4 Durações de vogal

No Quadro 33, são mostradas as durações da vogal nas produções de CR, ao lado das médias gerais da fala adulta.

Vogal: durações médias										
Obst.	Sílaba	Durações absolutas (ms)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	169,95	218,0	250,9	269,3	303,8	301,3	373,5	372,6	393,7
	CV	180,23	216,7	249,3	268,3	296,6	325,7	366,5	352,3	435,7
[b]	CCV	131,10	198,8	225,0	251,3	335,7	256,3	336,7	371,3	340,0
	CV	159,33	181,3	250,3	252,5	282,5	315,2	351,5	365,2	421,3
[f]	CCV	157,42	151,9	254,7	238,9	267,6	281,4	337,6	356,0	356,8
	CV	151,64	194,7	258,9	281,9	318,3	303,7	385,6	347,3	399,4
Obst.	Sílaba	Durações relativas (%)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	71,65	86,7	90,7	70,7	80,3	81,3	82,9	76,8	78,2
	CV	91,14	88,3	93,5	92,0	97,0	95,6	94,3	88,6	91,9
[b]	CCV	46,17	84,9	75,5	58,2	76,7	56,4	55,4	70,3	66,9
	CV	61,79	75,8	82,2	73,4	73,2	67,2	67,4	81,8	77,5
[f]	CCV	42,24	45,9	60,3	54,1	62,6	64,4	63,2	55,1	56,1
	CV	45,89	58,3	59,6	64,2	67,1	73,5	64,8	60,5	67,6

Quadro 33 – Médias de duração da vogal para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs)

Essas médias de duração estão dispostas nos Gráficos 19 a 24, a seguir, por tipo (relativa ou absoluta) e por obstruente inicial ([p], [b] ou [f]).

Os Gráficos 19 e 20 mostram as médias de duração de vogal em contexto de [p] como obstruente inicial.

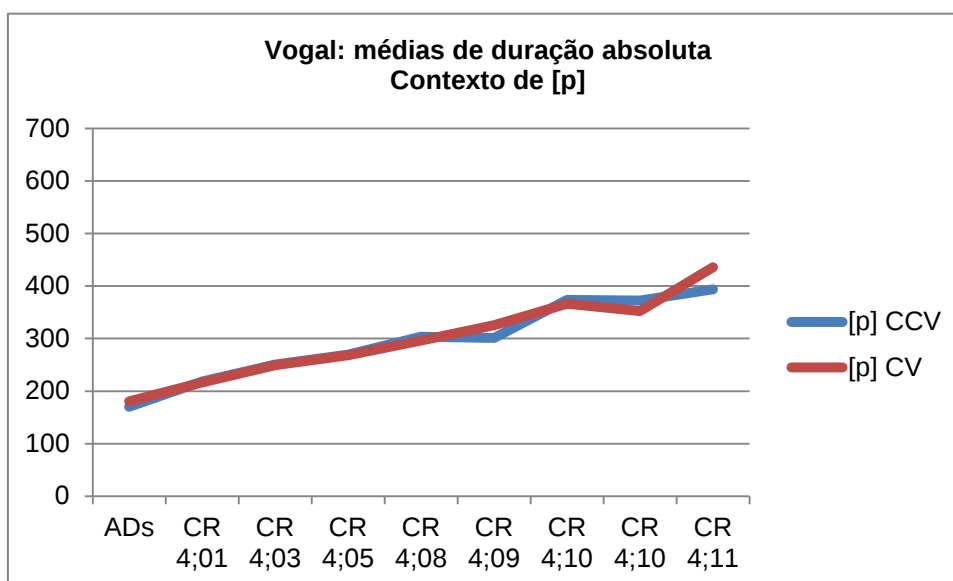


Gráfico 19 – Médias de duração absoluta de vogal em contexto de [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

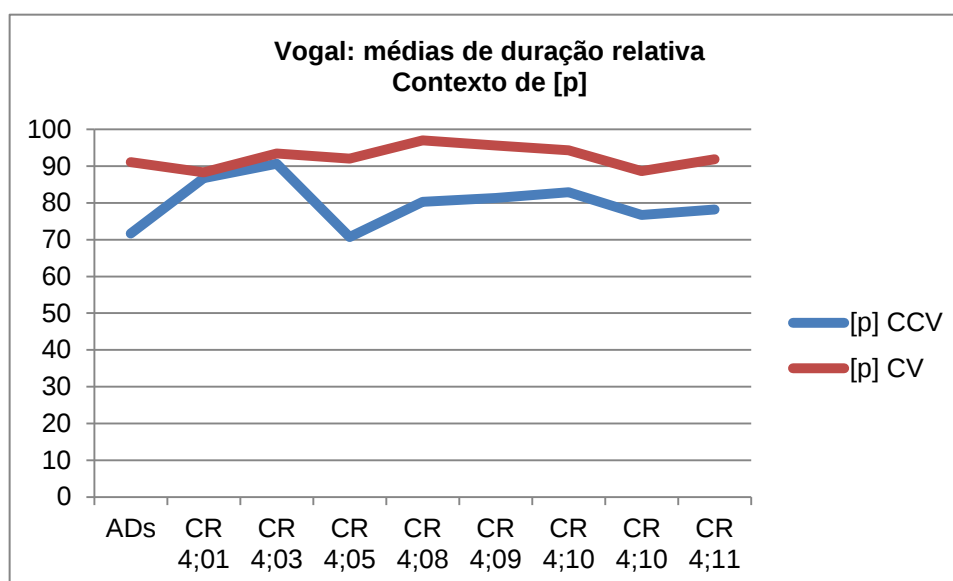


Gráfico 20 – Médias de duração relativa de vogal em contexto de [p] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

No Gráfico 19, as médias de duração absoluta da criança são maiores em relação à média das adultas e vão aumentando progressivamente ao longo dos meses. No entanto, não há diferenças expressivas de duração entre vogais em CCV e CV. As linhas se encontram, em grande parte, sobrepostas, apresentando alguma

diferença em duração a partir dos 4;09, quando o rótico já é realizado em todas as produções CCV. De todo modo, a partir desta idade, pode-se observar que vogais em CCV e vogais em CV se alternam na maior duração, sem que se observe uma dominância de um dos tipos de sílaba sobre o outro. Esse resultado também se mostra contrário à hipótese do alongamento compensatório como estratégia no processo de aquisição de CCV.

A duração relativa, por sua vez, no Gráfico 20, não apresenta um aumento progressivo em relação ao padrão adulto, passando por momentos de maior e menor duração em relação à média adulta. No entanto, a partir dos 4;05, há uma clara diferença entre as durações de vogal em CCV e CV. Nas linhas do gráfico de duração relativa, observa-se que, em todas as idades, a vogal de CV tem maior duração que a de CCV, padrão observado também nas médias das adultas. Esse resultado já era esperado, uma vez que a vogal ocupa um maior percentual em sílabas com dois elementos em comparação a sílabas com três elementos.

Nos dados de CR, essas durações se encontram bastante aproximadas aos 4;01 e 4;03, idade em que a criança não produzia sílabas CCV; no entanto, a partir dos 4;05, as diferenças entre as durações relativas de vogais em CCV e CV passam a variar de 12 a 22 pontos percentuais (pp).

Mais uma vez, os resultados não indicam a presença da estratégia de alongamento compensatório, uma vez que não há uma maior duração da vogal para alvos CCV aos 4;01 e 4;03. O que se observa vai na direção oposta da hipótese do alongamento formulada, pois, nessas idades, as medidas de duração absoluta e relativa nas duas estruturas silábicas têm valores bastante próximos.

Os Gráficos 21 e 22 mostram as médias de duração de vogal em contexto de [b] como obstruinte inicial.

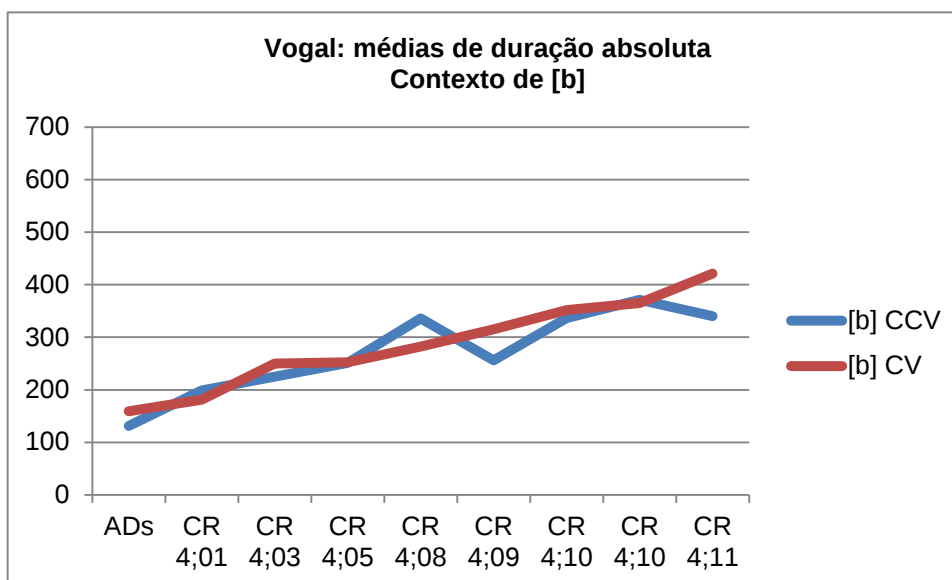


Gráfico 21 – Médias de duração absoluta de vogal em contexto de [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

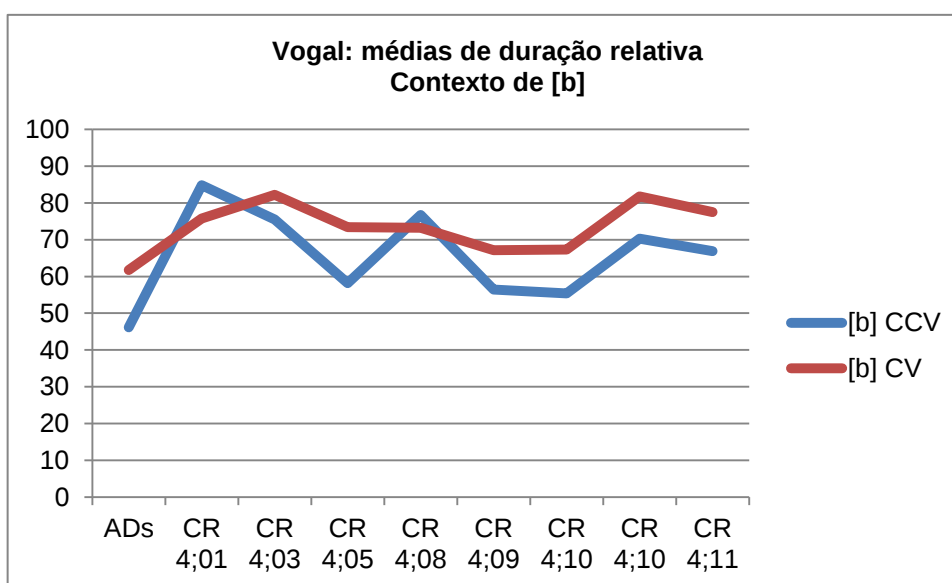


Gráfico 22 – Médias de duração relativa de vogal em contexto de [b] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

As médias de duração absoluta, no Gráfico 21, mostram durações maiores do que o padrão adulto e essa duração aumenta progressivamente, com exceção de quedas do valor em CCV aos 4;09 e 4;11. Seria possível supor que esse aumento progressivo da duração, em relação ao que se observa na fala adulta, pode estar

relacionado com a manipulação do aspecto temporal, a qual se encontraria em andamento, mesmo que a criança estivesse próxima aos seus cinco anos de idade. No entanto, a duração relativa de vogal, no Gráfico 22, apresenta oscilações principalmente nas sílabas CCV. A duração relativa de vogal em CCV começa sofrendo quedas até os 4;05, momento em que a produção do *tap* começa a estar presente na amostra.

Nos dois gráficos, não se observa uma dominância sistemática de duração em algum dos dois tipos silábicos. CCV e CV novamente se alternam na posição de maior duração de vogal; no entanto, a partir dos 4;09, a duração relativa segue um padrão de diferenciação em que as vogais de CVs ocupam cerca de 11 pp a mais da sílaba que as vogais em CCVs, padrão este que se estende até a idade da coleta final, aos 4;11. O padrão inclusive se assemelha ao das informantes adultas, as quais apresentam média em CVs com duração 15 pp maior que a de CCVs. Isso sugere que a criança passou por um período de ajuste de durações entre os 4;01 e os 4;08, estabelecendo, por fim, a partir dos 4;09, uma proporção de durações de vogal em CCVs e CVs semelhante à da fala adulta. Essa é uma interpretação possível dos resultados a partir da FAR. Esse padrão duracional, conquistado após um período de ajuste, segue ocorrendo de forma consistente nas três coletas restantes, o que sugere que houve a sua estabilização.

A maior duração de vogal em CVs observada aos 4;03 poderia contribuir para a confirmação da hipótese do alongamento vocálico; no entanto, essa duração superior de CV continua a ser observada na coleta seguinte, em que mais de 85% das sílabas CCV já possuem rótico, o que contraria a hipótese formulada. Somando-se a essa evidência, pode-se considerar o padrão já descrito em que a duração de vogal se estabiliza com valores superiores em CVs quando o rótico já está presente nas produções. Considere-se, ainda, que as durações de vogal também são superiores em CV nos dados das adultas. Logo, com base nesses indícios, a diferença de duração não parece estar relacionada a um período de aquisição, nem parece desempenhar uma função compensatória.

Os Gráficos 23 e 24 mostram as médias de duração de vogal nas sílabas com obstruente inicial [f].

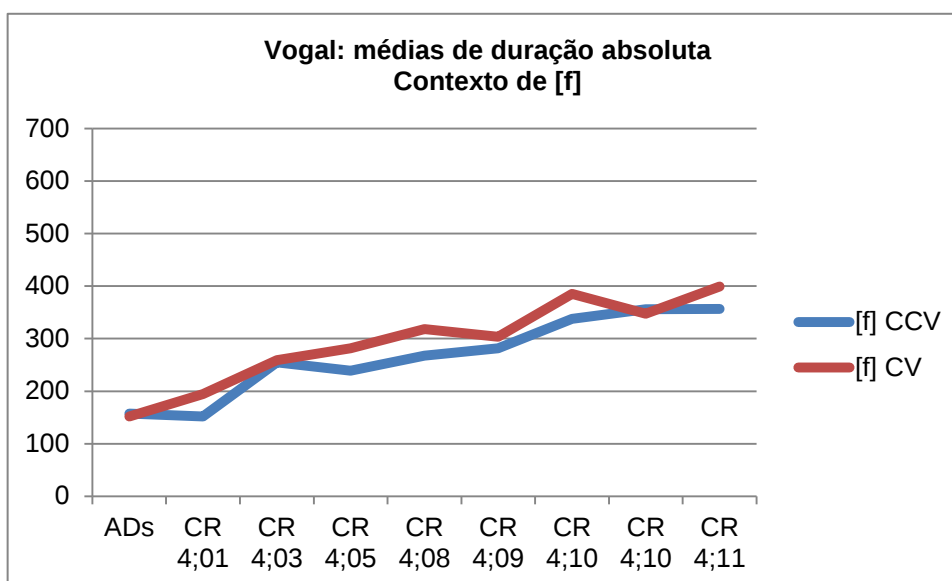


Gráfico 23 – Médias de duração absoluta de vogal em contexto de [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

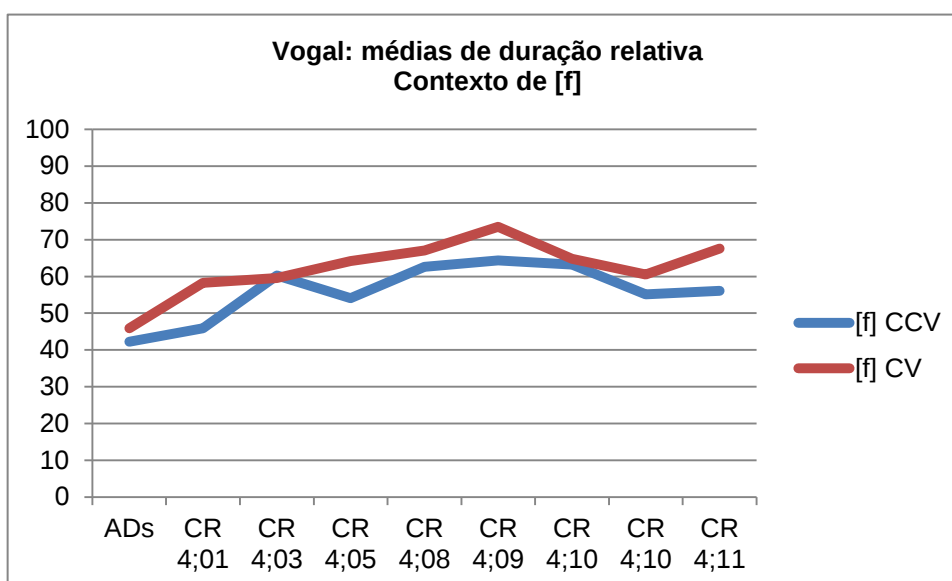


Gráfico 24 – Médias de duração relativa de vogal em contexto de [f] para as produções da criança (CR), dispostas por idade, e para as produções das adultas (ADs)

No Gráfico 23, as médias de duração absoluta de vogal para sílabas com [f] seguem padrão parecido com o de sílabas com [b]: de modo geral, apesar de leves quedas de valores, há um aumento progressivo em relação à média adulta (a qual, por sua vez, pouco diferencia durações em CCV e CV). No entanto, a vogal em CV é, de modo geral, maior em relação à de CCV, tendo duração menor apenas na

segunda coleta dos 4;10. Esse padrão se observa nas produções longitudinais como um todo, não se restringindo às idades em que a criança ainda não produz o encontro consonantal; logo, não é possível afirmar, a partir de tais valores médios, que a maior duração de vogal em CV tenha uma propriedade compensatória.

As médias de duração relativa, no Gráfico 24, seguem comportamento similar às de duração absoluta no sentido de que são superiores à média adulta, apresentando sempre porcentagens maiores. Também seguem o comportamento das durações absolutas porque, apesar da oscilação de valores, existe, de modo geral, um padrão de vogais mais longas em CVs. A única exceção se apresenta aos 4;03, na qual as vogais têm uma duração relativa ligeiramente maior nos alvos CCV, justamente em uma idade na qual a criança ainda não produz encontros consonantais. Isso se soma ao fato de que há um padrão geral de vogais maiores em CVs mesmo quando o rótico já é produzido na maioria das palavras. A partir dessas observações, também não é possível dizer que as vogais mais longas em CV sejam algo específico do processo de aquisição de CCV, nem que tenham uma função compensatória.

Logo, a partir dos Gráficos 19 a 24, as principais características observadas são as seguintes:

- a) nos dados de CR, de modo geral, as durações absolutas de vogal aumentam progressivamente em relação às durações das adultas, enquanto as durações relativas apresentam oscilações;
- b) as durações de vogais de CR, tanto absolutas quanto relativas, são maiores em CVs, com exceções isoladas ao longo das coletas. As adultas apresentam o mesmo padrão de vogais mais longas em CVs, o que sugere que a diferença de duração não seja um aspecto próprio da fala infantil;
- c) CR apresenta um padrão de durações de vogal maiores em CVs, como colocado em b), no entanto, esse padrão não se restringe aos 4;01 e 4;03. Na verdade, esse padrão se repete mesmo quando a criança já produz encontros consonantais, o que não fornece indícios para se considerar que a maior duração tenha função compensatória.

A análise estatística dos resultados relativos à duração da vogal nuclear das sílabas CV e CCV utilizou os mesmos testes e seguiu as mesmas etapas já reportadas para a análise da duração das obstruintes.

O teste One-Way ANOVA identificou diferenças significativas entre as durações – absoluta e relativa – das vogais produzidas pela criança e pelas adultas praticamente em todos os contextos, conforme pode ser visualizado no Quadro 34:

Sílaba	Variável	F	Valor de p
CCV	Duração absoluta contexto [p]	13,321	0,005
CV	Duração absoluta contexto [p]	9,858	0,012
CCV	Duração absoluta contexto [b]	18,777	0,002
CV	Duração absoluta contexto [b]	9,537	0,013
CCV	Duração absoluta contexto [f]	8,638	0,017
CV	Duração absoluta contexto [f]	15,205	0,004
CCV	Duração relativa contexto [p]	6,055	0,036
CCV	Duração relativa contexto [b]	11,151	0,009
CV	Duração relativa contexto [b]	13,713	0,005
CCV	Duração relativa contexto [f]	17,016	0,003
CV	Duração relativa contexto [f]	38,229	0,000

Quadro 34 – Diferenças significativas das durações – absoluta e relativa – das vogais quando comparados os resultados da criança com os resultados das adultas

Confirma-se, assim, na estatística inferencial, o já apontado por meio da estatística descritiva, ou seja, CR produz vogais mais longas do que as adultas de forma recorrente, tanto em sílaba CV quanto em CCV. Tais diferenças são mantidas se comparados os resultados das adultas com as duas primeiras coletas da criança – Quadro 35 – e com as coletas subsequentes – Quadro 36:

Sílaba	Variável	F	Valor de p
CCV	Duração absoluta contexto [p]	15,984	0,028
CV	Duração absoluta contexto [p]	11,159	0,044
CCV	Duração absoluta contexto [b]	24,454	0,016
CV	Duração absoluta contexto [f]	6,667	0,082*
CCV	Duração relativa contexto [p]	41,536	0,008
CCV	Duração relativa contexto [b]	59,013	0,005
CV	Duração relativa contexto [b]	31,454	0,011
CV	Duração relativa contexto [f]	273,675	0,000

*Diferença marginalmente significativa

Quadro 35 – Diferenças significativas das durações – absoluta e relativa – das vogais quando comparados os resultados das duas primeiras coletas da criança com os resultados das adultas

Sílaba	Variável	F	Valor de p
CCV	Duração absoluta contexto [p]	29,231	0,001
CV	Duração absoluta contexto [p]	20,517	0,003
CCV	Duração absoluta contexto [b]	40,170	0,000
CV	Duração absoluta contexto [b]	20,954	0,003
CCV	Duração absoluta contexto [f]	23,389	0,002
CV	Duração absoluta contexto [f]	41,764	0,000
CCV	Duração relativa contexto [p]	5,647	0,049
CCV	Duração relativa contexto [b]	11,039	0,013
CV	Duração relativa contexto [b]	10,782	0,013
CCV	Duração relativa contexto [f]	35,144	0,001
CV	Duração relativa contexto [f]	60,868	0,000

Quadro 36 – Diferenças significativas das durações – absoluta e relativa – das vogais quando comparados os resultados das coletas 3 a 8 da criança com os resultados das adultas

Observa-se que as vogais produzidas pela criança apresentam maior duração do que as produzidas pelas adultas tanto quando a criança não produz a estrutura CCV como quando já a realiza. Neste último caso, os valores de p são ainda menores. Para a confirmação da presença de alongamento compensatório, a maior duração da vogal deveria ocorrer, portanto, nas duas primeiras coletas apenas.

Dessa forma, assim como constatado para a duração das obstruintes, os dados indicam que CR também está ajustando a temporalidade dos gestos

intrínsecos às vogais e a coordenação destes com os demais elementos da sílaba no percurso da aquisição. A comparação entre a duração – absoluta e relativa – das coletas iniciais com as subsequentes indica igualmente diferenças estatísticas na duração da vogal em posição CCV e CV.

Sílaba	Variável	F	Valor de p
CCV	Duração absoluta contexto [p]	6,934	0,039
CV	Duração absoluta contexto [p]	5,892	0,051*
CCV	Duração absoluta contexto [b]	7,661	0,033
CV	Duração absoluta contexto [b]	5,755	0,053*
CCV	Duração absoluta contexto [f]	5,330	0,060*
CV	Duração absoluta contexto [f]	8,852	0,025
CCV	Duração relativa contexto [p]	9,407	0,022
CCV	Duração relativa contexto [b]	5,650	0,055*
CV	Duração relativa contexto [f]	5,078	0,065*

*Diferença marginalmente significativa

Quadro 37 – Diferenças significativas das durações – absoluta e relativa – das vogais quando comparados os resultados das coletas iniciais e das coletas subsequentes da criança

As diferenças significativas e marginalmente significativas dispostas no Quadro 37 corroboram a maior duração absoluta do segmento vocálico em sílabas CV e CCV nas coletas subsequentes. Em relação à duração relativa, no entanto, as diferenças se encontram apenas em sílaba CCV, indicando uma menor duração da vogal, o que era esperado, tendo em vista que, nas coletas iniciais, a criança produzia essa sílaba com dois elementos e, nas coletas subsequentes, com três.

Por fim, foram comparados os valores de duração – absoluta e relativa – da vogal nos pares CV x CCV, com a utilização do Teste-T para amostras pareadas. A aplicação do teste, aos dados das coletas iniciais, evidenciou apenas uma diferença marginalmente significativa, indicando uma maior duração absoluta da vogal em sílaba CCV no contexto de [p] ($t=9,667$ (1), $p=0,66$). As diferenças atestadas quando considerados os dados das coletas subsequentes estão dispostas no Quadro 38:

Variável	Teste-t	Valor de p
Duração absoluta contexto [p]	-3,554 (5)	0,016
Duração relativa contexto [p]	-9,885	0,000
Duração relativa contexto [b]	-3,527	0,017
Duração relativa contexto [f]	-4,533	0,006

Quadro 38 – Diferenças significativas das durações – absoluta e relativa – de vogais, quando comparadas as sílabas CCV e CV, nas produções das coletas 3 a 8 da criança

As diferenças de duração relativa das vogais em todos os contextos, indicando uma maior duração em sílaba CV, é justamente esperada pela presença de um terceiro elemento na sílaba que agora é produzido pela criança.

Para que a hipótese do alongamento compensatório da vogal fosse aqui comprovada, esperar-se-ia encontrar diferenças significativas com uma maior duração da vogal para alvos CCV – produzidos como CV pela criança – em comparação a alvos CV, no entanto, tal diferença surge apenas de forma marginal ($p=0,066$) no contexto de [p]. Por outro lado, neste mesmo contexto, a diferença é significativa ($p=0,016$) quando considerados os dados das coletas subsequentes, ou seja, quando a criança já produz o rótico.

Os resultados aqui expostos não permitem, até aqui, portanto, que se confirme a presença de alongamento compensatório da vogal nos dados da criança.

A seção 4.1.2.4, a seguir, expõe os resultados relativos à duração do elemento vocálico.

4.1.2.5 Durações de elemento vocálico

Para a extração de médias de duração de elemento vocálico, um total de 101 dados foi desconsiderado. Esses dados foram excluídos por não apresentarem produção de *tap* – o que impediria, assim, a produção de uma porção vocálica à parte – ou por apresentarem o *tap*, mas não apresentarem o elemento vocálico em si. O primeiro caso foi observado de forma recorrente na análise acústica, pois alguns dados apresentaram a produção de obstruinte e subsequente produção de um rótico diferente do *tap*, sem a produção dessa porção vocálica entre eles.

No Quadro 39, são mostradas as durações de elemento vocálico nas produções CCV de CR. As células referentes às coletas #1 (4;01) e #2 (4;03) não contêm dados e estão preenchidas em cinza devido ao fato de a criança não ter produzido sílabas CCV com o *tap* nessas coletas, como já relatado na seção 4.1.2.1. Como a criança produziu somente sílabas CV nessas duas coletas, não foi possível extrair medidas em relação ao elemento vocálico, pois este só ocorre junto ao *tap*, ausente nas produções dessas coletas.

Elemento vocálico: durações médias										
Obst.	Sílabas	Durações absolutas (ms)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	34,47			67,8	37,9	33,0	43,3	59,9	66,2
[b]	CCV	44,77			49,3	58,5	32,0	63,0	71,9	72,5
[f]	CCV	37,46			56,0	29,7	28,3	51,5	48,9	72,7
Obst.	Sílabas	Durações relativas: sílaba (%)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	14,49			16,4	10,4	8,8	9,6	12,4	13,2
[b]	CCV	16,44			12,6	20,2	7,0	10,7	12,6	13,7
[f]	CCV	9,96			12,0	6,7	6,8	9,1	7,2	6,8
Obst.	Sílabas	Durações relativas: vogal (%)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	20,37			25,4	13,1	10,7	12,2	17,2	17,5
[b]	CCV	36,93			19,5	18,7	12,4	19,2	19,9	21,0
[f]	CCV	24,28			24,4	10,3	11,3	14,5	14,5	12,5

Quadro 39 – Médias de duração do elemento vocálico para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs)

Essas médias de duração também estão nos Gráficos 25 a 27, a seguir. As durações são divididas por tipo, de forma mais detalhada que nos componentes anteriores. Primeiramente, será examinada a duração absoluta e, em seguida, serão expostos os resultados de dois tipos de duração relativa: um considerando-se a sílaba e outro considerando-se a vogal. Essas durações também são mostradas conforme a obstruinte inicial.

O Gráfico 25 mostra as médias de duração absoluta de elemento vocálico em contexto de [p], [b] e [f] como obstruinte inicial.

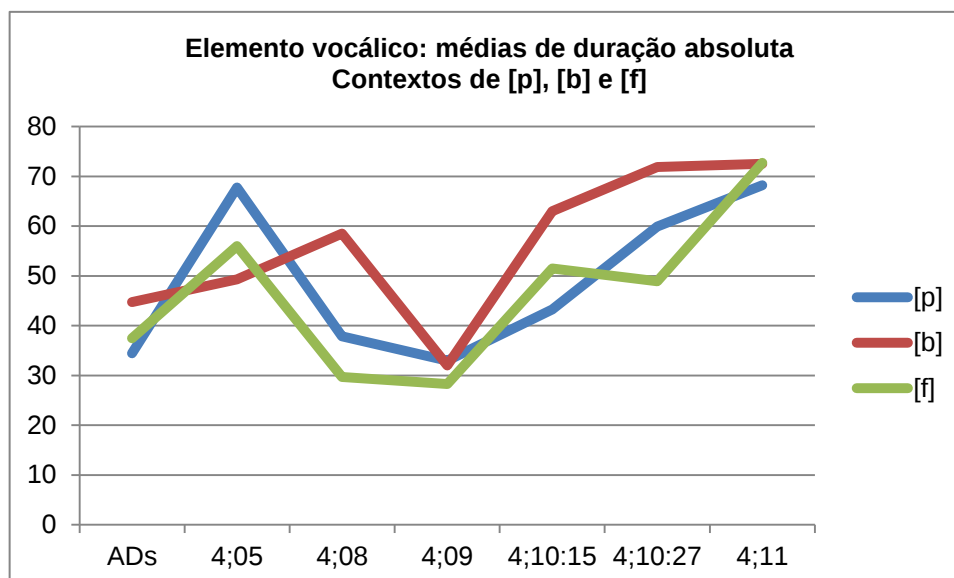


Gráfico 25 – Médias de duração absoluta de elemento vocálico, nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

O Gráfico 25 mostra que, na fala das adultas, as durações de elemento vocálico são aproximadas, com medidas entre 34,5 ms e 44,8 ms. Na fala da criança, essas durações começam superiores às das adultas, passam por quedas e tornam a aumentar a partir dos 4;10;15, atingindo os valores mais altos aos 4;11. Os elementos vocálicos em [p] e [f] têm comportamentos mais similares, com valores mais baixos aos 4;08 e 4;09. Já os elementos vocálicos em [b] têm, na maioria das coletas, durações superiores aos de outros contextos de obstruente, apresentando diferenças de 11 ms a 29 ms a mais em relação a elementos vocálicos com [p] ou [f].

É interessante notar que essas diferenças de duração que são observáveis conforme o contexto de obstruente se normalizam consideravelmente aos 4;11. Nessa idade, CR apresenta durações superiores à média das adultas, mas as diferenças entre as médias com [p], [b] e [f] se tornam pequenas, com uma diferença máxima de 7 ms entre uma duração e outra. Essa proximidade de valores é o que também se observa na média das adultas, na qual a maior diferença entre durações conforme o contexto de obstruente é de 11 ms.

A partir da observação dos dados de duração absoluta somente, o que se sugere é que haja um período de ajuste de durações quando a sílaba CCV já é produzida com os dois elementos consonantais. Uma dessas durações é a do elemento vocálico, a qual parece estar sujeita ao fator da obstruente imediatamente à

esquerda para que se determine a sua duração. A obstruinte [b] parece favorecer produções mais longas do elemento vocálico: ela possui uma duração intermediária entre [p] e [f], sendo mais longa que o [p], e se diferencia de [f] pela sua natureza vozeada. Logo, no íterim do ajuste das durações, o tipo de obstruinte e as suas propriedades podem favorecer uma produção mais longa do elemento vocálico, de modo que, em alguns casos, a sequência de obstruinte e elemento vocálico pode inclusive ser interpretada como uma sílaba à parte. Essa produção mais longa do elemento vocálico, principalmente em certos contextos, será normalizada à medida que a proporção temporal do elemento vocálico em relação a outros elementos da sílaba seja ajustada.

Os Gráficos 26 e 27 mostram as médias de duração relativa do elemento vocálico em contexto de [p], [b] e [f] como obstruinte inicial. O Gráfico 26 apresenta as durações em relação à sílaba, enquanto o Gráfico 27 apresenta as durações em relação à vogal.

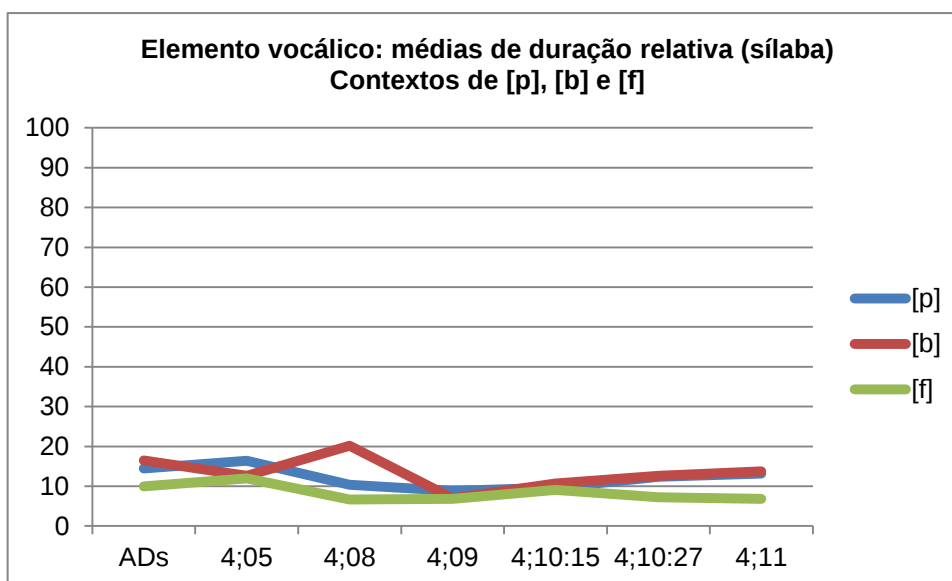


Gráfico 26 – Médias de duração relativa do elemento vocálico em relação à sílaba nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

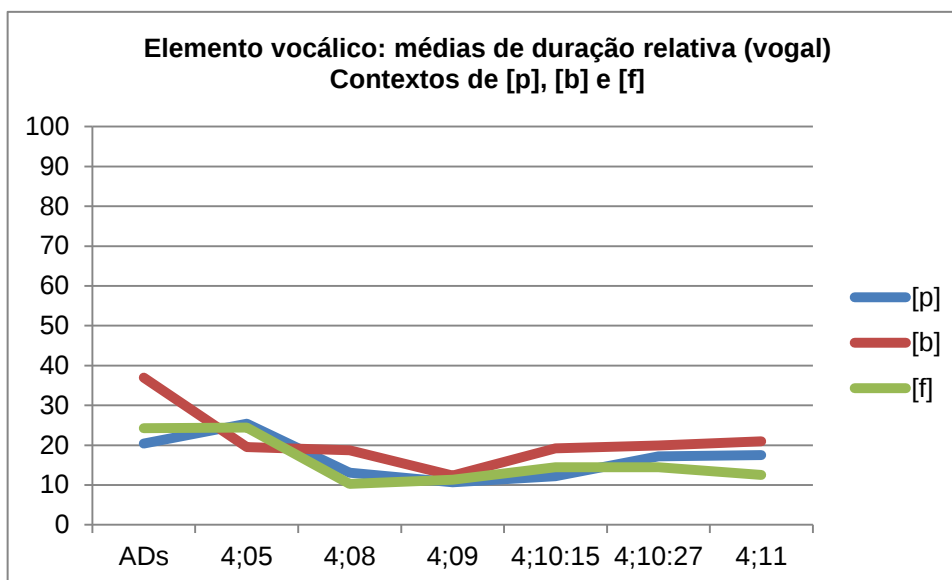


Gráfico 27 – Médias de duração relativa do elemento vocálico em relação à vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

Os Gráficos 26 e 27 mostram que as durações relativas seguem um padrão diferente do das durações absolutas, variando consideravelmente menos em relação às médias das adultas. No Gráfico 26, as durações dos elementos vocálicos de CR em relação à sílaba apresentam menor variação do que as absolutas, com porcentagens entre os 6,7% e 16,4%. A única exceção é a duração do elemento vocálico em contexto de [b] aos 4;08, o qual perfaz 20,2% da sílaba. As durações em relação à sílaba somente se assemelham às absolutas no sentido de que o contexto de [b] detém, de modo geral, as maiores durações. No entanto, essa diferença não é tão expressiva quanto nas durações absolutas.

O Gráfico 27 mostra as durações de elemento vocálico relativas à vogal. As médias adultas, nesse cálculo, se diferenciam das durações absolutas e das relativas à sílaba em que os valores eram bastante aproximados entre si. Nas durações de elemento vocálico relativas à vogal das adultas, [b] representa 12,7 pp a mais da duração da vogal núcleo em relação aos elemento vocálico em contexto de [f]. Curiosamente, o padrão de duração relativa à vogal encontrado nos dados de CR aponta para porcentagens mais aproximadas entre si e um padrão duracional mais estável, mas que retoma o contexto de [b] como aparentemente mais favorável à produção de elementos vocálicos mais longos.

O Gráfico 25 evidencia um período de ajustes na duração do elemento vocálico, com base na análise de durações absolutas que passam por mudanças expressivas. No entanto, as durações relativas apontam para mudanças menos expressivas e um padrão de duração mais estável. De todo modo, chama a atenção que o contexto de [b] propicie, de maneira geral, durações do elemento vocálico mais longas que nos contextos de [p] e [f].

O teste paramétrico One-Way ANOVA identificou diferenças significativas entre a duração absoluta do elemento vocálico quando comparadas as produções das adultas e das crianças. Também não foram identificadas diferenças de duração relativa em relação à sílaba. Assim, apenas duas significâncias estatísticas foram atestadas, sempre no que concerne à duração do elemento vocálico em relação à duração da vogal, conforme valores apresentados no Quadro 40:

Variável	F	Valor de p
Duração relativa em relação à vogal contexto [b]	17,505	0,004
Duração relativa em relação à vogal contexto [f]	9,093	0,020

Quadro 40 – Diferenças significativas das durações – absoluta e relativa – do elemento vocálico quando comparados os resultados das adultas e da criança

Os resultados estatísticos parecem corroborar, portanto, a explicação acerca do fato de que a duração da vogal, em encontros consonantais, é constituída pelo somatório do elemento vocálico, do rótico e da vogal nuclear. No processo de aquisição da estrutura CV, a criança precisa aprender em que ponto da vogal nuclear deve ser realizados os gestos articulatórios necessários para a produção do rótico, que sobrepõem a vogal.

A criança realiza, assim, a produção do elemento vocálico conforme a duração absoluta da forma alvo – já que não foram constatadas diferenças estatísticas em relação à duração absoluta –, mas apresenta diferenças na duração relativa em relação à vogal nuclear – geralmente mais longa na fala da criança –, ou seja, a realização do gesto de ponta de língua do *tap* acaba por sobrepor a vogal nuclear precocemente, gerando durações relativas menores em comparação às adultas.

A seção 4.1.2.6 discute as durações do *tap*, rótico presente nos encontros consonantais analisados nesta pesquisa.

4.1.2.6 Durações de rótico

Do mesmo modo que para as durações de elemento vocálico, parte dos dados foi desconsiderada na extração de médias de duração do rótico. Ao todo, 119 dados foram excluídos das médias por não apresentarem produção de rótico, por apresentarem sons róticos outros que não o *tap* (e.g. vibrante, retroflexa) ou por apresentarem produção da lateral [l] ao invés de um rótico.

O Quadro 41 mostra as durações do *tap* nas produções CCV de CR. Assim como para o quadro referente às durações de elemento vocálico, as células referentes às coletas #1 (4;01) e #2 (4;03) não contêm dados – elas estão preenchidas em cinza porque a criança não produziu sílabas CCV com segundo elemento consonantal nessas coletas, tendo produzido, para esses alvos, sílabas CV. Como o *tap* estava ausente nas produções dessas coletas, não foram extraídas medidas de duração referentes a esse segmento no período mencionado.

Rótico: durações médias										
Obst.	Sílabas	Durações absolutas (ms)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	15,85			25,8	18,6	15,5	22,5	33,6	25,7
[b]	CCV	19,93			38,2	14,8	28,7	28,3	33,9	34,5
[f]	CCV	18,34			31,5	32,3	25,5	21,5	41,7	38,3
Obst.	Sílabas	Durações relativas (%)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	6,73			7,9	5,1	4,3	5,5	6,7	5,2
[b]	CCV	7,73			13,8	3,6	6,9	4,5	6,3	7,2
[f]	CCV	5,13			6,8	5,9	6,9	4,0	6,3	5,8

Quadro 41 – Médias de duração do *tap* para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs)

Essas médias de duração também se encontram nos Gráficos 28 e 29, a seguir, divididas também por absolutas e relativas conforme a obstruinte inicial, contemplando-se os contextos de [p], [b] e [f].

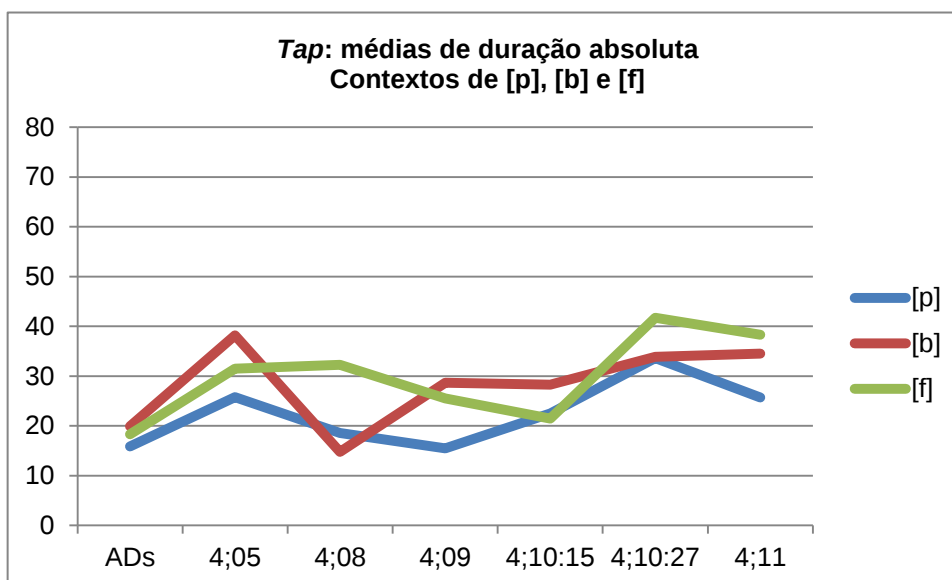


Gráfico 28 – Médias de duração absoluta do *tap* nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

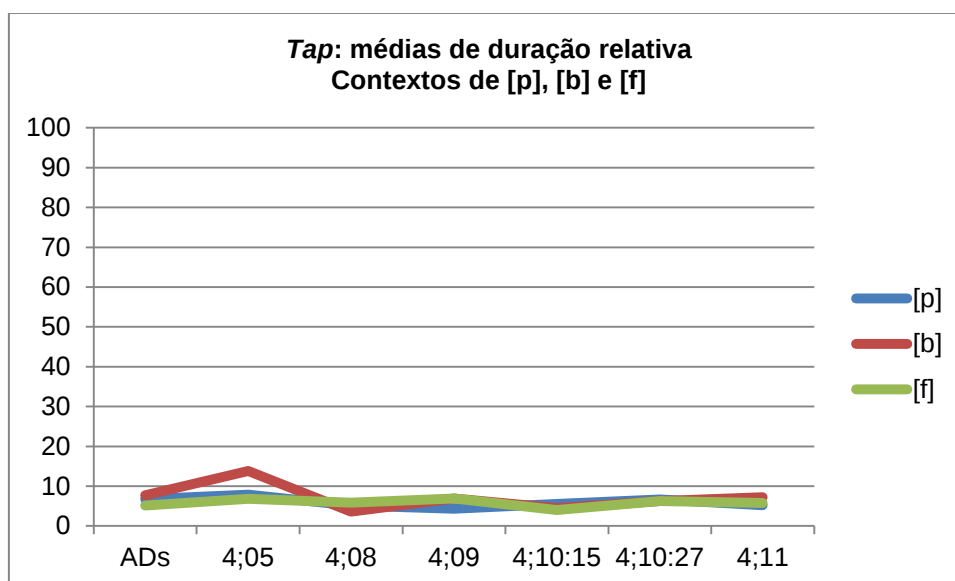


Gráfico 29 – Médias de duração relativa do *tap* nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

De acordo com as durações absolutas apresentadas no Gráfico 28, os *taps* produzidos pela criança em contexto de [p] são, de modo geral, mais longos do que a média das adultas, com exceção das produções aos 4;09. Essa produção, no entanto, passa por oscilações de duração. Já as médias de duração relativa, no

Gráfico 29, indicam porcentagens similares entre as adultas e a criança ao longo da aquisição, num intervalo entre os 5% e 8%.

No contexto de [b], as durações absolutas evidenciam que, de modo geral, os *taps* produzidos pela criança são maiores do que a média adulta, com exceção da média aos 4;08, de 14,7 ms. Já as durações relativas, são, em geral, menores que a das adultas, pois, com exceção da média aos 4;05, todas se mantêm entre os 3% e 7%.

Em contexto de [f], as durações absolutas do *tap* começam menores do que a média adulta, tornando-se maiores a partir dos 4;10;15. No entanto, as durações relativas se mantêm numa faixa de porcentagem de 4 a 7%, incluída a média das adultas.

A aplicação do teste One-Way ANOVA revelou apenas diferenças estatísticas na comparação entre a duração absoluta do *tap* produzido por CR e pelas adultas, conforme disposto no Quadro 42:

Variável	F	Valor de p
Duração absoluta contexto [p]	4,873	0,063*
Duração absoluta contexto [b]	6,697	0,036
Duração absoluta contexto [f]	7,926	0,026

*Diferença marginalmente significativa

Quadro 42 – Diferenças significativas da duração do *tap* quando comparados os resultados das adultas e da criança

Os resultados estatísticos evidenciam que, embora os gestos articulatórios necessários para a realização do *tap* sejam executados por CR, o tempo intrínseco que os constitui ainda é passível de ajustes no transcorrer da aquisição. Por essa razão, os *taps* realizados pela criança apresentam maior duração absoluta quando comparados aos produzidos pelas adultas.

Assim como na análise dos dados acústicos das informantes adultas, também serão examinadas, nos dados da criança, as durações dos seguintes trechos: elemento vocálico + vogal e elemento vocálico + *tap* + vogal. Essas durações serão discutidas, respectivamente, nas seções 4.1.2.7 e 4.1.2.8, a seguir.

4.1.2.7 Durações de elemento vocálico + vogal

O Quadro 43, a seguir, mostra as médias de duração da soma de elemento vocálico e vogal nas produções de CR.

EV + vogal: durações médias										
Obst.	Sílaba	Durações absolutas (ms)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	204,4			337,1	341,7	334,3	416,8	432,5	459,9
[b]	CCV	175,9			300,7	394,2	288,3	399,7	443,3	412,5
[f]	CCV	194,9			294,9	297,2	309,8	389,1	404,9	429,4
Obst.	Sílaba	Durações relativas (%)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	86,3			87,1	90,7	90,1	92,6	89,1	91,4
[b]	CCV	62,1			70,8	96,9	63,4	66,1	82,9	80,6
[f]	CCV	52,2			66,0	69,3	71,2	72,3	62,2	62,8

Quadro 43 – Médias de duração da soma de elemento vocálico e vogal para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs)

Esses valores estão dispostos nos Gráficos 30 e 31, separados por tipo de duração – absoluta ou relativa – e obstruente inicial – [p], [b] ou [f]. O Gráfico 30 mostra as mudanças em duração absoluta.

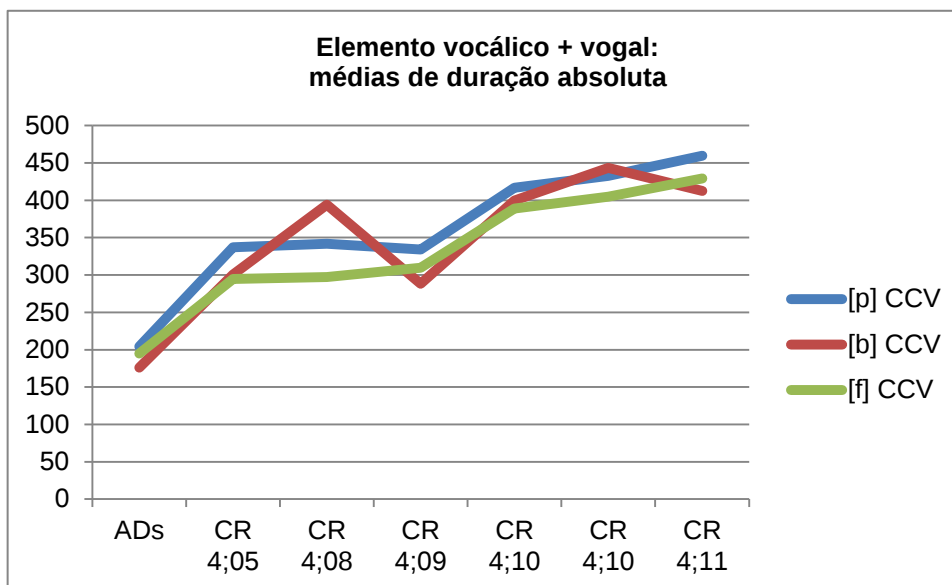


Gráfico 30 – Médias de duração absoluta da soma de elemento vocálico e vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

A partir do Gráfico 30, observa-se que as medidas da soma elemento vocálico + vogal nas produções de CR são sempre maiores que a média das adultas. As médias de duração absoluta, na produção da criança, são pelo menos 100 ms superiores em relação à duração das adultas.

De modo geral, as durações da criança seguem movimento crescente. As médias em contexto de [p] e [f] como obstruintes iniciais seguem comportamento similar, começando relativamente estáveis dos 4;05 aos 4;09 e aumentando a partir da primeira coleta dos 4;10. As médias em contexto de [b] têm comportamento distinto: apesar de o percurso apresentar aumentos na duração na sua maioria, há duas quedas, uma de 106 ms aos 4;09 e outra menos expressiva, de 30 ms, aos 4;11.

O movimento predominante, no que diz respeito aos três contextos, é de aumento na duração do trecho em análise – elemento vocálico mais vogal – ao longo do tempo.

O Gráfico 31 mostra as médias de duração relativa para a soma de elemento vocálico e vogal.

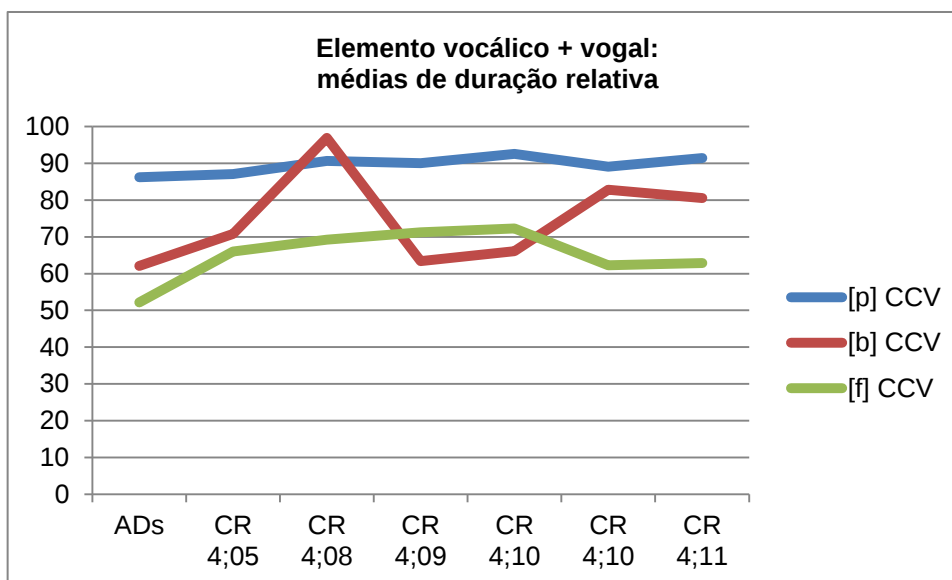


Gráfico 31 – Médias de duração relativa da soma de elemento vocálico e vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

As médias de duração relativa mostradas no Gráfico 31 apresentam comportamentos diferentes entre si. As durações de elemento vocálico e vogal em contexto de [p] nas produções de CR não apresentam mais do que 7 pp de diferença em relação à média adulta. Quando analisadas somente as produções da criança, não há variação de mais de 6 pp nesse contexto de obstruinte. Logo, a porcentagem que elemento vocálico e vogal ocupam nesse contexto é relativamente estável e similar à porcentagem encontrada na fala adulta.

As medidas de duração relativa em contexto de [b], por sua vez, passam por oscilações expressivas, sendo maiores que a média adulta aos 4;05 e 4;08. Nessas idades, são encontradas ocorrências de trocas de [b] por [p]. A ausência do pré-vozeamento normalmente presente no som de [b] diminui a duração da obstruinte, aumentando-se, assim, a duração relativa de outros componentes da sílaba, como o elemento vocálico e a vogal, os quais ocupam um maior espaço na sílaba. Aos 4;09 e na primeira coleta dos 4;10, há quedas de duração nas quais os valores se aproximam das médias das adultas. A partir da segunda coleta dos 4;10, no entanto, os valores tornam a aumentar: isso se deve, provavelmente, ao ajuste de duração do [b], o qual se pôde observar anteriormente na análise das durações relativas de [b], no Quadro 16, presente na seção 4.1.2.3. Naquele quadro, observa-se que [b] passa por um aumento de duração relativa dos 4;08 aos 4;09 e sofre quedas nas

coletas subsequentes. Essas quedas podem estar relacionadas a uma assimilação da duração de [b], provavelmente exagerada nas coletas anteriores devido à aquisição do pré-vozeamento, e que agora, normalizada, dá maior espaço aos outros componentes da sílaba.

As durações relativas em contexto de [f] são maiores que a média adulta, apresentando leve aumento entre os 4;05 e a primeira coleta dos 4;10, com queda na segunda coleta dos 4;10 e outro aumento pouco expressivo, de 1 pp, aos 4;11. Há uma oscilação menos expressiva do que no contexto de [b]: como discutido anteriormente, essa obstruente passou por expressivas oscilações na sua própria duração devido à aquisição do pré-vozeamento e à manipulação da sua duração. A obstruente [f] não exige o domínio desse aspecto: suas durações são mais estáveis dos 4;05 aos 4;11, como pode ser visto no Gráfico 18, presente na seção 4.1.2.3. Logo, os ajustes são menos expressivos e, pelo que se indica no Gráfico 31, caminham, ao final, para uma proporção que se verifica na fala adulta.

A partir das durações relativas examinadas, pode-se concluir que a duração de elemento vocálico mais vogal dependerá consideravelmente da qualidade da obstruente presente. E, numa perspectiva longitudinal, a complexidade da aquisição da obstruente presente influenciará nas mudanças em duração ao longo do tempo.

Tendo em vista as maiores durações – absoluta e relativa – das vogais na fala da criança, a sequência elemento vocálico + vogal apresentou, conforme esperado, diferenças significativas ou marginalmente significativas quando da aplicação do teste One-Way ANOVA. Os resultados podem ser visualizados no Quadro 44:

Variável	F	Valor de p
Duração absoluta em contexto de [p]	28,699	0,001
Duração absoluta em contexto de [b]	28,224	0,001
Duração absoluta em contexto de [f]	18,134	0,004
Duração relativa em contexto de [p]	4,024	0,085*
Duração relativa em contexto de [b]	5,115	0,058*
Duração relativa em contexto de [f]	33,284	0,001

*Diferenças marginalmente significativas

Quadro 44 – Diferenças significativas das durações – absoluta e relativa – da sequência elemento vocálico + vogal nuclear, quando comparados os dados da criança e das adultas nos contextos de [p], [b] e [f]

Quanto à comparação da duração da sequência elemento vocálico + vogal nuclear com a duração da vogal nuclear em sílabas CV, foi encontrada diferença significativa apenas no contexto de [p], para duração absoluta ($t=-4,228$ (5), $p=0,008$) e para duração relativa ($t=2,626$ (5), $p=0,047$).

O resultado aqui reportado se diferencia, portanto, do que foi constatado para as produções das adultas, já que apenas em contexto de [f] foi constatada uma maior diferença da sequência elemento vocálico + vogal nuclear; mas aproxima-se no que tange a [b], pois nem na fala infantil nem na fala adulta a sequência apresentou duração mais elevada do que a vogal núcleo da sílaba CV.

A seção 4.1.2.8, a seguir, trata das durações do trecho que compreende elemento vocálico, *tap* e vogal.

4.1.2.8 Durações de elemento vocálico + *tap* + vogal

O Quadro 45, a seguir, mostra as médias absoluta e relativa de duração do trecho que compreende elemento vocálico, *tap* e vogal.

EV + <i>tap</i> + vogal: durações médias										
Obst.	Sílaba	Durações absolutas (ms)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	220,3			363,0	360,3	349,7	439,3	466,1	485,6
[b]	CCV	195,8			338,8	408,9	317,0	428,0	477,2	447,0
[f]	CCV	213,2			326,4	329,6	335,3	410,6	446,6	467,7
Obst.	Sílaba	Durações relativas (%)								
		Adultas	Criança							
			#1 (4;1)	#2 (4;3)	#3 (4;5)	#4 (4;8)	#5 (4;9)	#6 (4;10)	#7 (4;10)	#8 (4;11)
[p]	CCV	92,9			95,0	95,8	94,4	98,1	95,9	96,6
[b]	CCV	69,1			84,6	78,2	70,3	70,6	89,2	87,8
[f]	CCV	57,1			72,8	75,2	78,1	76,3	68,6	68,7

Quadro 45 – Médias de duração da soma de elemento vocálico, *tap* e vogal para as produções da criança, dispostas por idade (coletas #1 a #8), e para as produções das adultas (ADs)

Esses valores estão dispostos nos Gráficos 32 e 33, separados por tipo de duração – absoluta e relativa – e obstruente inicial da sílaba – [p], [b] ou [f].

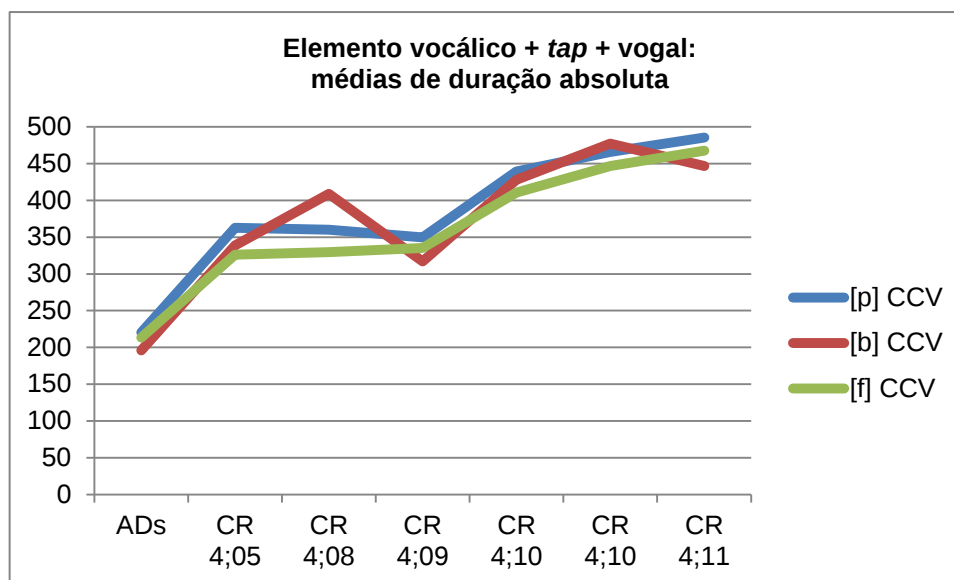


Gráfico 32 – Médias de duração absoluta da soma de elemento vocálico, *tap* e vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

As durações absolutas do trecho elemento vocálico + *tap* + vogal, presentes no Gráfico 32, têm comportamento parecido com o das durações absolutas apresentadas na seção 4.1.2.7, que abrangem o elemento vocálico e a vogal. Em todas as produções da criança, esse trecho tem maior duração do que a média adulta. Os trechos em sílabas com [p] e [f] têm comportamentos ainda mais parecidos, começando de forma relativamente estável e aumentando progressivamente a partir dos 4;09. Os trechos em contexto de [b] como obstruente inicial têm comportamento distinto em relação aos em contexto de [p] e [f]: os trechos em sílabas com [b] apresentam o mesmo padrão da duração absoluta de elemento vocálico + vogal, com queda expressiva de duração (92 ms) aos 4;09 e outra, mais branda (30 ms), aos 4;11.

O Gráfico 33 apresenta as durações relativas para o mesmo trecho.

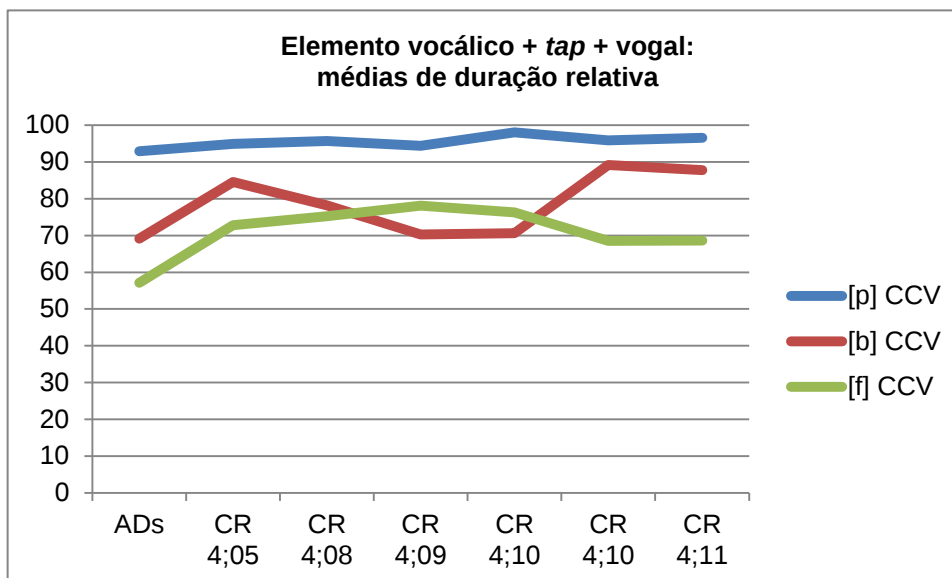


Gráfico 33 – Médias de duração relativa da soma de elemento vocálico, *tap* e vogal nos contextos de [p], [b] e [f], para as produções da criança (CR) e das adultas (ADs)

No Gráfico 33, as médias de duração relativa de elemento vocálico + *tap* + vogal também apresentam comportamento semelhante ao das médias de duração relativa de elemento vocálico + vogal, já apresentadas no Gráfico 31. As semelhanças se observam, principalmente, nas curvas de [p] e de [f].

A curva desse trecho no contexto de [p] é relativamente estável e, embora os trechos produzidos por CR sejam sempre mais longos do que a média das adultas, essa diferença não passa dos 5 pp. E, quando analisadas somente as produções da criança desse trecho no contexto de [p], as variações não ultrapassam os 4 pp. Logo, repete-se o padrão observado nas durações de elemento vocálico + vogal no contexto de [p]: as durações se refletem em altas porcentagens, principalmente devido à duração curta própria do [p], curvas relativamente estáveis e que pouco variam em relação à média adulta.

As durações em contexto de [f], como no caso de elemento vocálico + vogal, são sempre maiores em relação à média adulta e aumentam de forma pouco expressiva. No entanto, o valor de duração sofre queda uma coleta antes, a partir dos 4;09, e, entre a segunda coleta dos 4;10 e os 4;11, há um aumento pouco sensível, de 0,1 ms, o que se traduz, no gráfico, como uma estabilização.

A curva referente à duração do trecho analisado em contexto de [b] tem comportamento semelhante à da duração de elemento vocálico + vogal: os valores de CR são sempre maiores do que a média adulta, mas a medida do trecho incluindo o *tap* mostra oscilações menos expressivas. De todo modo, as curvas complementam o percurso de mudança na duração da obstruinte [b], a qual começa sendo produzida com menores durações, sem o pré-vozeamento característico, passa a ser produzida com duração maior do que a observada na fala adulta, tem sua duração modulada e se estabiliza, com duração menor, após o domínio da fase de pré-vozeamento do [b].

Com base nos dados desta seção, também é possível afirmar que a duração de elemento vocálico + *tap* + vogal dependerá consideravelmente do tipo de obstruinte presente na sílaba. Da mesma forma, o grau de complexidade da aquisição da obstruinte terá influência nas mudanças em duração desse trecho, ao se examinarem dados longitudinais.

A aplicação do teste One-Way ANOVA novamente encontrou diferenças estatísticas em todos os contextos, quando comparados os dados da criança com os das adultas no que concerne à duração da sequência elemento vocálico + rótico + vogal nuclear. Tal resultado não poderia ser diferente, considerando-se os resultados já apontados acerca da maior duração da vogal nuclear e do *tap* produzido por CR.

Variável	F	Valor de p
Duração absoluta em contexto de [p]	27,573	0,001
Duração absoluta em contexto de [b]	33,097	0,001
Duração absoluta em contexto de [f]	19,622	0,003
Duração relativa em contexto de [p]	5,084	0,059*
Duração relativa em contexto de [b]	7,132	0,032
Duração relativa em contexto de [f]	28,330	0,001

*Diferença marginalmente significativa

Quadro 46 – Diferenças significativas das durações – absoluta e relativa – da sequência elemento vocálico + *tap* + vogal nuclear, quando comparados os dados da criança e das adultas nos contextos de [p], [b] e [f]

Quanto à comparação da duração absoluta da sequência elemento vocálico + rótico + vogal nuclear com a vogal núcleo da sílaba CV, foram constatadas

diferenças significativas para todos os contextos: [p] ($t=2,626$ (5), $p=0,003$); [b] ($t=3,594$ (5), $p=0,016$); [f] ($t=3,548$ (5), $p=0,016$). Os resultados se aproximam dos constatados para as informantes adultas, à exceção da maior duração da sequência no contexto de [b].

A seguir, a seção 4.2 apresenta os resultados referentes à inspeção articulatória qualitativa.

4.2 Inspeção articulatória qualitativa

Tendo em vista questões relativas ao curto período de tempo para o desenvolvimento da descrição e análise dos dados da presente dissertação, bem como as tarefas relacionadas à análise acústica qualitativa, quantitativa e estatística, esta seção apresenta uma breve inspeção articulatória com o objetivo de mais bem detalhar alguns aspectos das produções da criança já apontados na seção 4.1.

4.2.1 Os gestos articulatórios intrínsecos à produção do rótico

Conforme já evidenciado, CR apresentou, de forma preponderante, a partir da terceira coleta – na qual há a emergência da estrutura CCV –, a produção de *taps*, ainda que com durações maiores do que as constatadas nas produções das informantes adultas. Formas outras de produção, como segmentos retroflexos e vibrantes, foram igualmente atestadas, indiciando a presença de produções variáveis.

Nas produções com realização do *tap*, a inspeção articulatória dos dados identificou a produção de róticos formados por apenas um gesto de ponta de língua, conforme expresso na Figura 44 – aproximando-se do padrão do adulto –, o qual tem início ao final do elemento vocálico que o antecede.

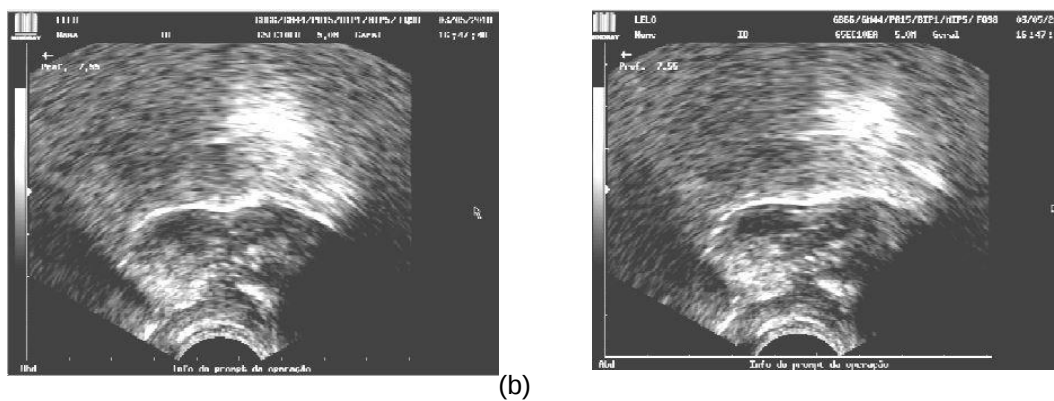


Figura 44 – Movimentos de ponta de língua na realização do *tap* em “prata” (primeira repetição da coleta #6), produzido por CR. Em (a), início do movimento de elevação da ponta da língua; em (b), movimento de máxima constrição. Da direita para a esquerda nas imagens, partes anterior e posterior da língua

Chama a atenção, nas imagens, o sulco formado na região do corpo da língua, característico de produções retroflexas. Na produção da informante AD2, o gesto de ponta de língua presente na produção do *tap* não apresenta o sulco constatado na configuração articulatória de CR, como pode ser visto na Figura 45.



Figura 45 – Movimento de máxima constrição de ponta de língua na realização do *tap* em prata (primeira repetição) por AD2. Da direita para a esquerda nas imagens, partes anterior e posterior da língua

A sequência do movimento de CR, no entanto, como pode-se constatar de (a) para (b) na Figura 44, apresenta apenas a elevação da ponta, sem encurvamento e sem retração do corpo.

Outras formas atestadas nas produções de CR são a produção de retroflexas e vibrantes, como pode ser constatado nas imagens dispostas nas Figuras 46 e 47:

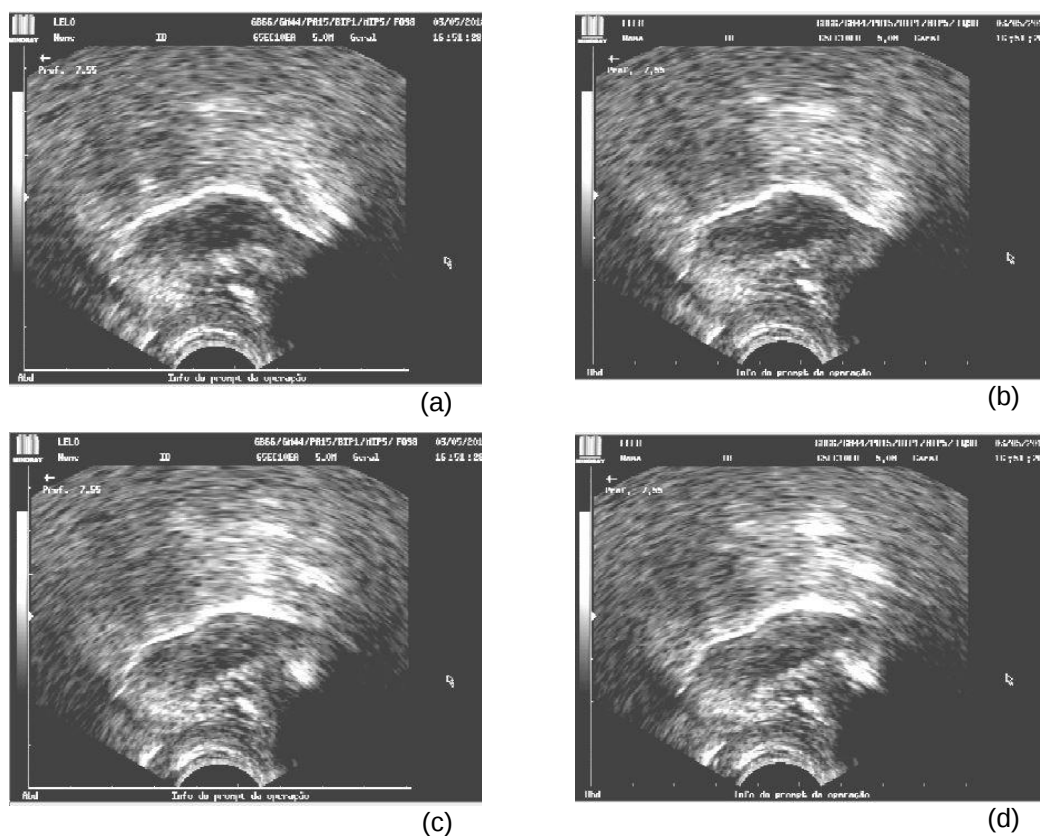


Figura 46 – Movimentos de ponta de língua na realização do segmento retroflexo em prata (segunda repetição da coleta #6), produzido por CR. Em (a), início do movimento de elevação da ponta da língua; em (b), movimento de máxima elevação da ponta; em (c), início da retração do dorso; em (d), retração do dorso e encurvamento da ponta. Da direita para a esquerda nas imagens, partes anterior e posterior da língua

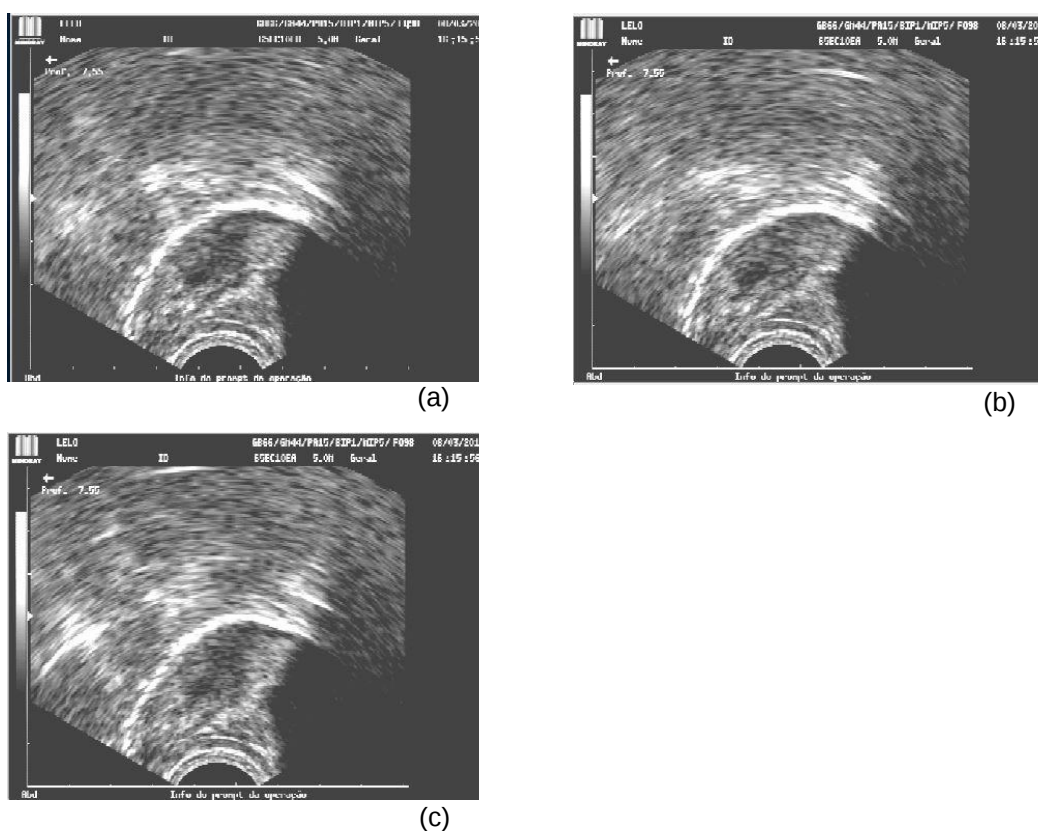


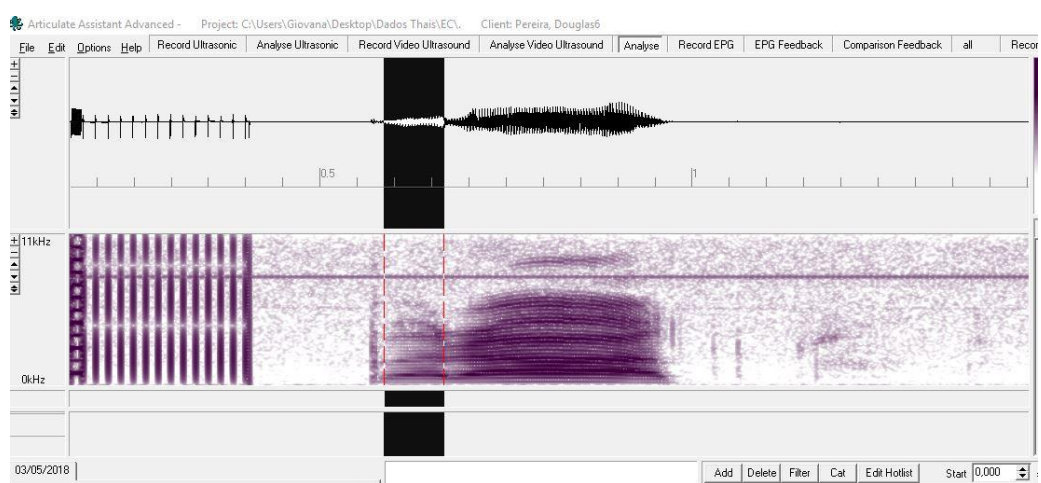
Figura 47 – Movimentos reiterados de ponta de língua na realização da vibrante em preso1, quarta coleta, produzido por CR. Em (a), primeiro movimento de elevação da ponta da língua; em (b), leve abaixamento da ponta para a produção do segundo elemento vocálico; em (c), segundo movimento de elevação da ponta. Da direita para a esquerda nas imagens, parte anterior e posterior da língua

A inspeção articulatória dos dados confirma, portanto, os resultados da análise acústica no que concerne à presença de diferentes formas de róticos a constituir a estrutura CCV. O padrão do *tap* evidenciado na Figura 44, com a clara presença de um sulco na região do corpo, sinaliza, pois, para a instabilidade na coordenação do articulador. O movimento único, preciso, balístico de ponta esperado – conforme visualizado na produção de AD2 – não ocorre – caso da produção do retroflexo – ou ocorre “arrastado”, quase por dar sequência à constituição articulatória necessária à produção do retroflexo, ou ocorre de forma repetida, constituindo os segmentos vibrantes.

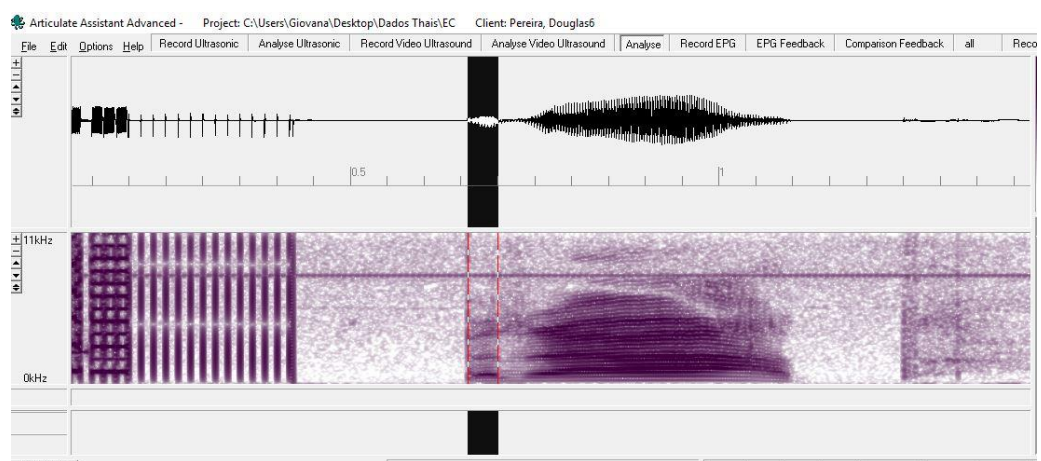
4.2.2 A coordenação gestual: o rótico e o elemento vocálico

Os resultados já evidenciados na seção 4.1 indicaram o papel central do elemento vocálico na coordenação gestual que constitui a sílaba CCV. Nos dados de CR, o elemento vocálico se apresenta com maior duração em comparação às produções das adultas, mas essa duração é variável.

A Figura 48 evidencia a forma variável com a qual a duração do elemento vocálico se apresenta nos dados de CR.



(a)



(b)

Figura 48 – Imagens acústicas do software AAA do elemento vocálico presente em duas ocorrências de “prata” (primeira e terceira repetições da coleta #6) de CR. Em (a), elemento vocálico mais longo, antecedendo a produção do tap; em (b), elemento vocálico mais curto antecedendo a produção do tap. Da direita para a esquerda nas imagens, partes anterior e posterior da língua

A criança está, na verdade, adquirindo o padrão do português que prevê a realização do gesto de ponta de língua do rótico após o desdobramento inicial do gesto vocálico, desta forma, surgem formas mais longas e menos longas.

Confirmando o proposto por Silva, Clemente e Nishida (2006), o *tap* se sobrepõe ao gesto vocálico – ainda que não o entrecorte –, deixando-o eclipsado por alguns poucos milissegundos.

A discussão acerca da qualidade do elemento vocálico, discutida pelos autores com base em dados acústicos, pode ser aqui aprofundada com base na inspeção articulatória.

Na Figura 49, observa-se a comparação dos traçados da borda da língua de CR, para o gesto de máxima de constrição do elemento vocálico e da vogal nuclear nas três repetições de “prata” da coleta #6, produzidas por CR.

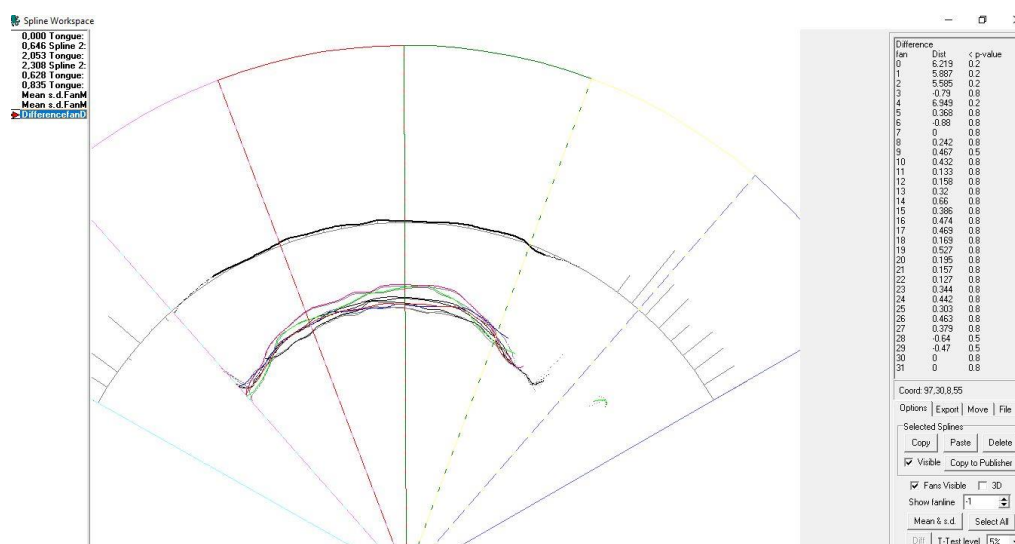


Figura 49 – Teste-t do gesto de máxima constrição do elemento vocálico e do gesto de máxima constrição da vogal nuclear nas três ocorrências de “prata” da coleta #6, produzidas por CR. À direita do gráfico, parte anterior do trato vocal; à esquerda, parte posterior do trato vocal.

A análise dos traçados das bordas da língua, por meio do software AAA, não evidenciou diferenças significativas quando comparada a média do gesto de maior constrição do elemento vocálico com a média do gesto de maior constrição da vogal nuclear. Tal resultado indicia, pois, que elemento vocálico e vogal nuclear são parte da vogal que constitui a sílaba CCV.

A breve inspeção articulatória aqui realizada evidencia que, para adquirir a forma CCV alvo realizada pelo adulto, CR tenta aprimorar a produção do gesto de ponta de língua no que concerne a seus aspectos intrínsecos e extrínsecos. A temporalidade, tem pois, papel central nesse sentido.

4.3 Fonologia Articulatória: uma discussão dos dados à luz da teoria

Esta seção é dedicada a uma discussão dos resultados já apresentados nas seções 4.1 e 4.2 com ênfase no modelo teórico adotado nesta pesquisa. Assim, pretende-se relacionar os principais achados deste estudo a pressupostos da Fonologia Articulatória (BROWMAN; GOLDSTEIN, 1989).

A seção 4.1.1 apresentou os resultados de análise acústica com base em dados de fala adulta. O padrão de fala desse grupo apresentou diferenças de duração no âmbito da sílaba, de modo que as sílabas CCV são mais longas que as CV. Já no nível segmental, as obstruintes e vogais – quando consideradas apenas enquanto vogal nuclear, sem o elemento vocálico – analisadas se mostraram com duração maior nas sílabas CV. Nas hipóteses deste trabalho, presumiu-se que essa diferença se mostraria na fala infantil e seria exclusiva da fala em aquisição, com base em trabalhos que investigaram a duração como forma de estratégia compensatória. No entanto, a análise, que se iniciou pelos dados das adultas, revelou que a fala adulta também mostra essa diferença.

Partindo-se do pressuposto da FAR de que as palavras são constelações de gestos, as sílabas apresentam as suas “subconstelações”, os seus ramos de gestos. Esses gestos, como já discutido na seção 2.1.1, têm uma dimensão espacial e outra temporal. As diferenças observadas quanto à duração de obstruintes e vogais mostram com mais clareza a organização gestual no âmbito temporal. Sílabas CCV têm mais elementos – logo, a duração relativa destes será menor do que em sílabas CV. Nestas, não há um segundo elemento consonantal – logo, a obstruinte e a vogal ocuparão um maior espaço da sílaba. As durações possibilitam observar como determinados gestos se desdobram no tempo e como interagem entre si. A própria análise acústica apresenta sinais que permitem a observação de sobreposições ou justaposições de gestos – não por acaso, os valores formânticos das vogais, por exemplo, são preferencialmente medidos fora das regiões de coarticulação. Isso

remete à afirmação de que os gestos estão em coordenação, e essa coordenação se dá no espaço e no tempo. Quando há coocorrência entre os gestos, seja durante toda a sua realização ou durante parte destes, é porque há sobreposição, seja ela total ou parcial.

Os dados das adultas também corresponderam à descrição do *tap* em encontros consonantais do português brasileiro, proposta por Silva, Clemente e Nishida (2006). Como já discutido na seção 2.1.3, os autores propuseram uma pauta gestual para a ocorrência do *tap* em encontros consonantais, na qual existe um evento de natureza vocálica entre a obstruinte e o *tap*. Nos dados acústicos das adultas desta pesquisa, esse elemento vocálico, uma “pequena vogal”, não só se mostrou presente como também apresentou características formânticas da vogal núcleo. Logo, tanto os dados de fala adulta do estudo de Silva, Clemente e Nishida (2006) quanto os deste estudo apontam uma relação entre elemento vocálico e vogal núcleo em que ambos fazem parte do mesmo segmento, que é realizado de forma contínua. O rótico, na visão dos autores, entrecorta a realização da vogal. A partir dos dados de Silva, Clemente e Nishida (*op. cit.*), dos dados desta pesquisa e da teoria de base que a sustenta, ou seja, a Fonologia Articulatória, pode-se dizer que não há exatamente um entrecortamento, pois isso pressuporia uma interrupção do gesto em curso. O que há, de fato, é uma sobreposição dos gestos do *tap* aos da vogal, e essa sobreposição é algo previsto de acordo com a FAR: é uma forma de organização temporal dos gestos envolvidos na produção de uma sílaba ou palavra. Logo, seria possível definir esse fenômeno como um eclipsamento²³ da vogal: o *tap* é produzido com gesto de ponta de maior magnitude, o que o torna mais saliente acusticamente, ocultando, assim, os gestos vocálicos. Ainda, a análise articulatória com base nos traçados das bordas de língua no AAA não apontou diferenças significativas comparando-se a maior constrição de elemento vocálico e vogal nuclear.

Isso à parte, outro resultado que retoma pressupostos da FAR é o da comparação entre vogais em CV e os trechos de elemento vocálico + vogal e elemento vocálico + rótico + vogal. Nos resultados descritivos, as vogais em CV se

²³ O termo *eclipsamento* foi sugerido pela Profa. Mirian Rose Brum de Paula, em reunião de estudos ocorrida em 2017. A professora utilizou esse termo para explicar a presença do elemento vocálico que acompanha o *tap* em estruturas silábicas CCV e CVC. Esse elemento seria observável a partir de uma sobreposição do *tap* à vogal, em que parte desta é eclipsada. O trecho anterior à realização do *tap*, que não se encontra sobreposto e é distinguível acusticamente, seria o elemento vocálico.

mostraram menores do que o trecho EV+R+V e, na maioria dos contextos, menores do que o trecho EV+V. Na estatística, não foi apontada diferença significativa entre essas medidas no contexto de [b]. Esse resultado corresponde em parte à proposta de descrição em que o rótico ocorre em sobreposição a uma vogal, após seu início e antes de seu fim. As diferenças significativas ocorreram nos contextos de [p] e [f]. A partir dos dados de estatística descritiva e inferencial, a real duração de vogal em CCV parece ser o trecho de elemento vocálico, rótico e vogal.

Assim, considerando as diferenças significativas nos contextos de [p] e [f], seria justamente nesses contextos que poderia ocorrer o fenômeno chamado de alongamento compensatório. A criança, ao alongar a vogal em uma sílaba CV com alvo CCV, estaria realizando uma produção em que o rótico está marcado temporalmente, apesar de seus gestos não estarem presentes. Nessa perspectiva, casos reportados pela literatura como alongamento compensatório, são, na verdade, a realização de uma vogal com seu padrão duracional usual. Esta vogal, no entanto, não tem o *tap* sobrepondo sua produção, estando em sua plena duração. É uma vogal exposta, à qual não há outro som se sobrepondo, somente.

Tendo em vista que os padrões de duração apontam para uma sobreposição gestual, isso possibilita, por exemplo, uma nova leitura acerca das produções anteriormente definidas, na literatura, como epênteses. É importante retomar os resultados de estudos como o de Ribas (2002), apoiados fundamentalmente na porcentagem de produções corretas como parâmetro para que se considerasse a aquisição como completa. Produções outras, que não C₁V e CCV, foram consideradas estratégias de reparo aplicadas em contextos específicos, classificados como desfavorecedores. É necessário, também, retomar-se os exemplos de epêntese citados na seção 2.2.1. A palavra “trem” era produzida como [te.ˈrẽỹ] devido a conter uma obstruinte coronal considerada desfavorecedora para a produção. A palavra “branco”, produzida como [ba.ˈrãn.ku], foi citada, no estudo de Ribas (2002), como outro exemplo de epêntese e classificada como uma das produções consideradas pouco comuns (*ibid.*, p. 36). Essas palavras, na ótica do estudo realizado, sofreram a inserção de um som, uma vogal, resultando em duas sílabas com *onsets* simples.

É interessante notar que, nas produções exemplificadas a partir de Ribas (2002), a vogal “inserida” é transcrita com características no mínimo similares à da vogal núcleo: são versões não nasalizadas da primeira parte de um ditongo ou da

vogal nuclear. Pode-se retornar à proposta de Silva, Clemente e Nishida (2006), em que se afirma que há um evento de natureza vocálica que apresenta as mesmas características da vogal núcleo. Comparando ambos os estudos, o que se chama de epêntese, no primeiro, será considerado, no segundo, um elemento vocálico. Numa proposta apoiada em um modelo gestual, não seria, de fato, uma coincidência essa vogal apresentar as mesmas características. Essa “nova vogal” não seria uma inserção, mas sim o elemento vocálico, que, em uma leitura conforme a FAR, emerge em demasia por um ajuste temporal inadequado. Esse ajuste continuará a ser realizado no processo de aquisição. Isso porque a duração, na perspectiva da FAR, é um aspecto intrínseco ao gesto, mas esse aspecto não emerge pronto – na fala infantil, essa duração intrínseca passará por ajustes. Do mesmo modo, a coordenação intergestual também será ajustada ao longo do tempo, considerando-se que as chamadas epênteses sejam produções em que a sobreposição de um gesto de ponta está deslocada no tempo, ocorrendo mais tarde e deixando descoberta uma parte maior da vogal antes da sua realização. Essa duração da parte descoberta da vogal à esquerda se torna, assim, maior, mais saliente auditivamente e considerada uma vogal à parte por adquirir um padrão duracional semelhante ao da porção vocálica após o *tap*.

Tendo essas observações em vista, é preciso rever afirmações como a de que não é possível dizer que as crianças passem por estágios intermediários. Do mesmo modo, os fenômenos considerados como estratégias de reparo não parecem ocorrer em pequenas escalas. Nos dados da criança da presente pesquisa, o elemento vocálico apresentou durações maiores que o padrão adulto com grande frequência, em várias coletas. Cabe, também, lembrar que muitas dessas produções da criança seriam consideradas, em princípio, erros. No entanto, esses erros devem ser investigados detalhadamente, visto que as produções assim consideradas na fala não são distorções aleatórias. Os erros de fala, como definido por Goldstein *et al.* (2007), constituem padrões e devem ser interpretáveis em termos de unidades linguisticamente significativas – neste caso, os gestos. Assim, um gesto deslocado no tempo não irá resultar em um movimento aleatório, mas sim, refletir uma solução encontrada quando o padrão ainda não foi devidamente assimilado.

As produções da criança, de modo geral, passam por oscilações ou aumentos progressivos na duração ao longo das coletas, o que só pôde ser observado por meio de análises longitudinais. Os ajustes observados na fala da criança apenas

reforçam que os gestos, apesar de terem sua duração intrínseca, passam por ajustes de duração e coordenação temporal ao longo do desenvolvimento.

Estudos futuros, realizados com amostras maiores, poderão questionar afirmações como a de que os ajustes na duração, definidos como estratégias de reparo em estudos anteriores, sejam recursos individuais. De todo modo, essas afirmações podem ser melhor revistas a partir de teorias como a Fonologia Articulatória e de métodos que utilizem análises acústica e articulatória. É a partir desse aparato teórico e metodológico que aspectos finos da linguagem em desenvolvimento podem ser identificados e redefinidos.

O capítulo 5, a seguir, apresenta as considerações finais deste trabalho.

5 Considerações finais

A partir dos resultados obtidos, foi possível avaliar hipóteses formuladas no início deste trabalho. Essas hipóteses serão, aqui, retomadas e comentadas.

A primeira hipótese (H1) coloca que, na fala da criança, a duração, tanto de vogais quanto de obstruintes, teria diferenças significativas conforme o tipo de sílaba produzido, sendo essas durações maiores em CV do que em CCV. Isso se deveria a um processo de ajuste de duração que ocorreria no processo de aquisição até que a criança se apropriasse do *tap* e das rotinas motoras para a realização de CCV, ajustando, assim, a obstruinte e a vogal temporalmente.

Essa hipótese não pôde ser confirmada, devido aos poucos resultados que a favoreciam. Quanto às obstruintes, a comparação entre as duas primeiras coletas de CR e as produções das adultas, considerando-se apenas alvos CCV, apontou apenas uma diferença, de ordem marginalmente significativa. Ao se considerar as coletas seguintes para comparação com as produções das adultas, há apenas um contexto em que a diferença de duração é significativa. Também foi feita a comparação entre obstruintes em CV e CCV nas duas primeiras coletas de CR e nas coletas subsequentes: embora a diferença aparecesse na duração relativa de [p] e [f], as produções das adultas apresentaram também diferenças entre CCVs e CVs nesta medida. Logo, nas obstruintes, os ajustes de fato ocorrem; no entanto, as diferenças nem sempre são significativas.

Quanto às vogais, os resultados de estatística descritiva mostram um aumento progressivo da duração absoluta de vogal da criança em relação à das adultas, mas as relativas apresentam oscilações. Ainda que haja essas oscilações, a criança tem maiores durações em CCVs; no entanto, esse padrão se repete mesmo após as coletas em que a criança só produzia CVs para alvos CCV, o que não fornece indícios para se considerar essa duração algo compensatório.

De acordo com a segunda hipótese (H2), a diferença em duração seria um aspecto próprio da fala infantil em processo de aquisição, partindo-se da ideia de que a criança passaria por um processo de ajuste de durações. Desse modo, no processo de aquisição de CCV, a duração seria um aspecto relevante para que a criança marcasse a estrutura silábica diferente, o que se verificaria acusticamente.

Essa hipótese também não pôde ser confirmada. No caso das obstruintes, a criança apresentou diferenças significativas em segmentos que as adultas também

produziam durações significativas. As diferenças de duração se devem a uma organização temporal, a qual causa durações diferentes na fala adulta, ao invés de uma tentativa de se marcar a estrutura CCV.

Quanto à duração de vogal, as vogais de CR, de modo geral, apresentam vogais maiores em CCV; no entanto, as adultas apresentam esse mesmo padrão, o que sugere que essa diferença não seja própria da fala infantil. Isso aponta a importância de se incluir um parâmetro que represente a fala adulta nas pesquisas, como forma de descartar a ideia de que fenômenos sejam exclusivos da fala em aquisição.

As análises estatísticas, por sua vez, apontam durações das vogais da criança que são significativamente maiores do que as adultas em vários contextos. No entanto, a significância acontece em todos os contextos, sem exceção, a partir da coleta 3, na qual a criança já produz encontros consonantais na maioria dos contextos. A duração, conforme H2, seria considerada um aspecto relevante para a marcação de estrutura silábica diferente na ausência do rótico; no entanto, para que isso fosse atestado, a maior duração de vogal em relação às adultas deveria ocorrer apenas nas duas primeiras coletas. Logo, a estatística indica que as diferenças em duração estão mais relacionadas a ajustes em duração e coordenação gestual do que a um recurso de compensação.

A investigação das medidas de elemento vocálico + vogal e de elemento vocálico + rótico + vogal mostraram resultados que indicam diferenças significativas de duração, especificamente no caso de vogais em CV quando comparadas ao trecho de elemento vocálico + rótico + vogal. Esta última medida parece apresentar a real duração das vogais em CCV, a partir dos resultados já apresentados.

A terceira hipótese (H3) afirma que, quando a criança começa a dominar o gesto característico do tap, as produções apresentarão gesto de ponta que aumentará em magnitude ao longo das coletas. Essa hipótese está a ser averiguada, visto que a análise articulatória empreendida até a conclusão desta dissertação não contemplou todos os dados de todas as coletas.

Na quarta hipótese (H4), afirma-se que o *tap* não é um segmento complexo, sendo composto apenas por um gesto de ponta, e que sua aquisição tardia decorre da sua coordenação temporal particular. Essa hipótese foi parcialmente confirmada a partir dos dados articulatórios analisados na seção 4.2, em que se observa o gesto de ponta sem retração sensível do corpo de língua.

Para a quinta hipótese (H5), as sílabas CCV produzidas pela criança apresentarão sobreposição de gestos do *tap* e da vogal núcleo. Isso foi confirmado tanto nas análises acústicas, em que elemento vocálico e vogal núcleo apresentavam as mesmas características no espectrograma, quanto na inspeção articulatória.

Por fim, a sexta hipótese (H6) afirmou que, na análise articulatória, seriam encontrados contrastes outros, difíceis de serem resgatados ou inferidos pela inspeção acústica. De fato, dados inspecionados e apresentados na seção 4.2 confirmaram pressupostos formulados na etapa de análise acústica, em especial quanto à qualidade dos róticos produzidos.

O presente estudo demonstra a importância de se considerarem, nos estudos em aquisição, os dados de fala adulta como parâmetro para comparações e as coletas longitudinais, as quais revelam as mudanças e tendências que de fato acontecem no processo de aquisição de um sujeito específico. Ao mesmo tempo, as hipóteses relacionadas à inspeção acústica (H1 e H2) ainda poderiam ter sido melhor investigadas a partir de dados de fala em idades mais jovens do que os 4;01, idade inicial das coletas. Logo, caso se deseje investigar a duração segmental e a emergência de CCV por meio de análise acústica exclusivamente, um estudo a partir de dados de fala naturais abrangendo faixas etárias menores poderá revelar interessantes resultados, quer na direção dos achados desta pesquisa ou não.

Conclusões sobre os dados à parte, faz-se necessário refletir, talvez imprimindo-se aqui uma impressão de ordem mais pessoal, sobre a real importância do alto grau de detalhe metodológico para a condução de pesquisas em fonologia.

Com os avanços na tecnologia e na pesquisa, novas metodologias vão se incorporando ao trabalho do pesquisador em ciências humanas, e isso não é diferente no caso do pesquisador em fonologia. No domínio das pesquisas mais recentes, por exemplo, pode-se pensar nos instrumentos de coleta experimentais, na análise acústica, na análise articulatória, na ultrassonografia e na análise estatística. Com a demanda por rigor metodológico, esses e outros recursos vão se somando à metodologia de pesquisa em aquisição da linguagem. O uso desses recursos implica uma carga de atenção e trabalho que faz com que o pesquisador, por vezes, chegue ao ponto de esquecer daquilo que trata. É recorrente que se abstraia do fato de que o objeto de estudo é a linguagem da criança, seja lendo trabalhos alheios, seja desenvolvendo um trabalho próprio.

No entanto, é essa mesma sistematicidade e complexidade metodológica que faz com que se averiguem hipóteses, as quais são formuladas parte com base em leituras, parte com base em pressentimentos, expectativas, desejos. Em outras palavras, é esse rigor que faz com que o investigador, por meio de números, comparações e testes, alcance descobertas de modo objetivo. A partir das análises de duração empreendidas neste trabalho, são muitos os indícios de que uma criança pode ter percepções de aspectos extremamente finos da língua do seu entorno. E não apenas percebê-los, mas tentar reproduzi-los, ajustá-los, tentar reproduzi-los de novo e novamente ajustá-los num caminho para uma estabilização. De todo modo, nesse caminho ela poderá demonstrar, na sua produção, que já aprendeu alguns aspectos do que recebe do seu entorno e que encontra soluções temporárias para o que ainda não domina, com aparente maestria. O papel de modelos teóricos como a Fonologia Articulatória e de recursos metodológicos como as análises acústica e articulatória tem importância nesse sentido: investigar aspectos finos da descrição e da aquisição da linguagem de modo objetivo e preciso.

Browman e Goldstein, ao discutirem os pressupostos da Fonologia Articulatória, pontuaram que os gestos são, em princípio, unidades pré-linguísticas, as quais, mais tarde, são comandadas para servirem a propósitos linguísticos, adquirindo um papel no sistema. Segundo essa visão, movimentos articulatorios podem ser, em princípio, como qualquer outro movimento que uma criança experimenta. Nesse sentido, engatinhar, andar ou pegar um objeto, por exemplo, são funções que vão sendo experimentadas e ajustadas. Os gestos articulatorios emergem não sendo diferentes nesse sentido, mas têm o potencial para desempenharem funções simbólicas, o que de fato ocorre.

Pode-se pensar que a linguagem verbal é, nas mãos da criança, um brinquedo. É um artefato que vai sendo experimentado, manipulado, brincado. No entanto, parece ser um brinquedo que a criança não abandona quando cresce. É um brinquedo que se transforma num artefato outro, à medida que a criança aprende suas outras funções. É por meio dela que brinca, mas também que pede, que demonstra seus sentimentos, que interage, que conquista oportunidades. É um brinquedo querido que ganha novas formas, significados e usos para o adulto, mas que continua o acompanhando por toda a vida.

Mães, pais, cuidadores, professores e fonoaudiólogos, assim como outros indivíduos que convivam constantemente com a criança, observam o sistema se

criar, pouco a pouco, mesmo quando não parece haver sistema. Os familiares e profissionais podem perceber e sentir essas questões, assim como o linguista que investiga a fala infantil.

Assim como a linguagem verbal no seu completo desenvolvimento, a linguagem verbal em processo de aquisição é complexa, fascinante e infinita nas suas possibilidades de observação. Espera-se, com este trabalho, contribuir para uma descrição do processo de aquisição do português brasileiro e para o conhecimento em áreas afins, como a fonoaudiologia, mas também contribuir para que mais pesquisadores investiguem aspectos da aquisição de língua materna.

Referências

ALBANO, E. C. **O gesto e suas bordas**: esboço de fonologia acústico-articulatória do português brasileiro. Campinas: Mercado de Letras, 2001.

AZAMBUJA, E. J. **A aquisição das líquidas laterais do português**. Dissertação (Mestrado em Letras). Instituto de Letras e Artes, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1998.

BARBERENA, L. da S. **Caracterização de aspectos da produção articulatória do [r] TAP por análise instrumental e resultados de intervenção**. 2016. Tese (Doutorado em Distúrbios da Comunicação Humana). Centro de Ciências da Saúde, Universidade Federal de Santa Maria.

BARBIERI, T. T.; FERREIRA-GONÇALVES, G. Aquisição de encontros consonantais: uma análise de dados naturais e experimentais. **Letrônica**, Porto Alegre, v. 10, n. 2, p. 624-633, 2017.

BOERSMA, P.; WEENINK, D. **Praat**: doing Phonetics by Computer. Disponível em: <http://www.fon.hum.uva.nl/praat/>. Acesso em: 13 mar. 2018.

BONET, E.; MASCARÓ, J. **On the representation of contrasting rhotics**. Universidade Autônoma de Barcelona, 1996.

BONILHA, G. F. G. **Aquisição dos ditongos orais decrescentes**: uma análise à luz da Teoria da Otimidade. Dissertação (Mestrado em Letras). Universidade Católica de Pelotas, 2000.

_____. **Aquisição fonológica do português brasileiro: uma abordagem conexional da Teoria da Otimidade**. 2005. Tese (Doutorado em Letras). Instituto de Letras e Artes, Pontifícia Universidade Católica do Rio Grande do Sul.

BRESSMANN, T. Quantitative assessment of tongue shape and movement using ultrasound imaging. In: COLANTONI, L.; STEELE, J. (eds.). **Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology**. Somerville: Cascadia Proceedings Project, 2008.

BROWMAN, C., GOLDSTEIN, L. Towards an articulatory phonology. **Phonology Yearbook**, 3, 219-252, 1986.

_____. Tiers in articulatory phonology, with some implications for casual speech. **Haskins Laboratories Status Report on Speech Research**, SR-92, 1-30, 1987.

_____. Some notes on syllable structure in articulatory phonology. **Haskins Laboratories Status Report on Speech Research**, SR-93/94, 85-102, 1988.

_____. Articulatory gestures as phonological units. **Phonology**, 6 (2), 201-251, 1989.

_____. Articulatory phonology: an overview. **Phonetica: International Journal of Speech Science**, 49, 155-180, 1992.

BYBEE, J. The phonology of the lexicon: Evidence from lexical diffusion. In: M. BARLOW, M.; Kemmer, S. (eds.). **Usage-based models of language**. Stanford: CSLI, 2000.

CARVALHO, K. C. H. P. **Descrição fonético-acústica das vibrantes no português e no espanhol**. Tese (Faculdade de Ciências e Letras de Assis). Universidade Estadual Paulista (Assis), 2004.

CLEMENTS, N. The role of the sonority cycle in core syllabification. In: KINGSTON, J.; BECKMAN, M. (eds.). **Papers in laboratory phonology I: between the grammar and physics of speech**. New York: Cambridge University Press, 1990.

CORREA, B. T. **Aquisição das vogais nasais francesas [ɛ̃], [ã] e [õ] por aprendizes brasileiros**: aspectos acústico-articulatórios. Dissertação (Mestrado em Letras). Programa de Pós-Graduação em Letras, Universidade Federal de Pelotas, 2017.

CORREA, B. T.; FERREIRA-GONÇALVES, G.; BRUM-DE-PAULA, M. R. Aquisição das vogais nasais francesas [ɛ̃], [ã] e [õ] por aprendizes brasileiros: aspectos articulatórios. **Ilha do Desterro**, Florianópolis, v. 70, n. 3, p. 131-149, 2017.

DIAS-CAVALHEIRO, B. S. **Aquisição da vogal [a] espanhola por falantes de Português Brasileiro**. Dissertação (Mestrado em Letras). Programa de Pós-Graduação em Letras, Universidade Federal de Pelotas, 2016.

EPSTEIN, M. A.; STONE, M. The tongue stops here: ultrasound imaging of the palate. **Journal of the Acoustical Society of America**, 118, 2128-2131, 2005.

FERREIRA-GONÇALVES, G.; BRUM-DE-PAULA, M. R. A ultrassonografia em pesquisas linguísticas. In: FERREIRA-GONÇALVES, Giovana; BRUM-DE-PAULA, Mirian Rose (orgs.). **Dinâmica dos Movimentos Articulatórios**: sons, gestos, imagens. Pelotas: Editora UFPel, 2013.

FRANCISCO, D. T. **Contorno de língua na produção do /s/ e /ʃ/ na fala de adultos e crianças com e sem transtorno fonológico**. 2015. Dissertação (Mestrado em Ciências da Reabilitação). Faculdade de Medicina, Universidade de São Paulo.

GOLDSTEIN, L.; POUPLIER, M.; CHEN, L.; SALTZMAN, E.; BYRD, D. Dynamic action units slip in speech production errors. **Cognition**, n. 103, p. 386-412, 2007.

HERNANDORENA, C. L. M. **A aquisição da fonologia do português**: estabelecimento de padrões com base em traços distintivos. Tese (Doutorado em Letras). Instituto de Letras e Artes, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1990.

HERNANDORENA, C. L. M.; LAMPRECHT, R. R. A aquisição das consoantes líquidas do português. **Letras de Hoje**, Porto Alegre, v. 32, n. 4, p. 7-22, 1997.

LAMPRECHT, R. R. **Perfil da aquisição normal da fonologia do Português. Descrição longitudinal de 12 crianças: 2:9 a 5:5**. 1990. Tese (Doutorado em Letras). Instituto de Letras e Artes, Pontifícia Universidade Católica do Rio Grande do Sul.

_____. Antes de mais nada. In: _____ (org). **Aquisição fonológica do português: perfil de desenvolvimento e subsídios para terapia**. Porto Alegre: Artmed, 2004.

MARCUS, S. M. Acoustic determinants of perceptual centers (P-center) location. **Perception & Psychophysics**, v. 30, p. 247-256, 1981.

MARTINS, C. Manual de análise de dados quantitativos com recurso ao IBM SPSS: saber, decidir, fazer, interpretar e redigir. Braga: Psiquilibrios Edições, 2011.

MELO, R. M. **O contraste entre oclusivas alveolares e velares: estados gradientes mediados por análise acústica e ultrassonográfica**. 2016. Tese (Doutorado em Distúrbios da Comunicação Humana). Centro de Ciências da Saúde, Universidade Federal de Santa Maria.

MEZZOMO, C. L. Aquisição dos fonemas na posição de coda medial do português brasileiro em crianças com desenvolvimento fonológico normal. Dissertação (Mestrado em Letras). Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1999.

MEZZOMO, C. L.; KESKE-SOARES, M.; MOTA, H. B. Análise acústica como instrumento de auxílio na descrição do sistema fonológico infantil. In: **Anais do XII Congresso Brasileiro de Fonoaudiologia e II Congresso Sulbrasileiro de Fonoaudiologia**. Foz do Iguaçu, 2004.

MEZZOMO, C. L.; MOTA, H. B.; DIAS, R. F.; GIACCHINI, V. O uso da estratégia de alongamento compensatório em crianças com desenvolvimento fonológico normal e desviante. **Letras de Hoje**, Porto Alegre, v. 43, n. 3, p. 35-41, 2008.

MIRANDA, A. R. M. **A aquisição do “r”**: uma contribuição à discussão sobre seu status fonológico. Dissertação (Mestrado em Letras). Instituto de Letras e Artes, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1996.

MIRANDA, I. C. C. **Aquisição e variação estruturada de encontros consonantais tautossilábicos**. 2007. Tese (Doutorado em Linguística). Faculdade de Letras, Universidade Federal de Minas Gerais.

MIRANDA, I. C. C.; SILVA, T. C. Aquisição de encontros consonantais tautossilábicos: uma abordagem multirrepresentacional. **Linguística**, Rio de Janeiro, v. 7, n. 1, 2011.

PEREIRA, O. T. A. ; FERREIRA-GONÇALVES, G. A ultrassonografia e a aquisição da retroflexa do inglês. Disponível em: http://cti.ufpel.edu.br/siepe/arquivos/2015/LA_04725.pdf. Acesso em: 27 dez. 2018.

PIERREHUMBERT, J. Exemplar dynamics: Word frequency, lenition, and contrast. In: BYBEE, J; HOPPER, P. (eds.). **Frequency effects and the emergence of lexical structure**. Amsterdam: John Benjamins, 2001.

RECASENS, D. What is and what is not an articulatory gesture in speech production: The case of lateral, rhotic and (alveolo)palatal consonants. **Gradus: Revista Brasileira de Fonologia de Laboratório**, Curitiba, v. 1, n. 1, p. 23-42, 2016.

RIBAS, L. **Aquisição do onset complexo no português brasileiro**. 2002. Dissertação (Mestrado em Letras). Instituto de Letras e Artes, Pontifícia Universidade Católica do Rio Grande do Sul.

_____. Sobre a aquisição do onset complexo. In: LAMPRECHT, R. R. (org). **Aquisição fonológica do português: perfil de desenvolvimento e subsídios para terapia**. Porto Alegre: Artmed, 2004.

SAVIO, C. B. Aquisição das fricativas /s/ e /z/ do Português Brasileiro. 2001. Dissertação (Mestrado em Letras). Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2001.

SCOBIE, J. M.; GIBBON, F.; HARDCASTLE, W. J.; FLETCHER, P. Covert contrast as a stage in the acquisition of phonetics and phonology. In: SCOBIE, J. M. (ed.). **QMC Working Papers in Speech and Language Sciences**, v. 1, p. 43-62, 1996.

SCOBIE, J. M.; WRENCH, A.; VAN DER LINDEN, M. **Head-Probe Stabilisation in Ultrasound Tongue Imaging Using a Headset to Permit Natural Head Movement**. Disponível em: <http://issp2008.loria.fr/Proceedings/PDF/issp2008-87.pdf>. Acesso em: 4 mai. 2016.

SILVA, A. H. P. **Para a descrição fonético-acústica das líquidas no português brasileiro**: dados de um informante paulistano. 1996. Dissertação (Mestrado em Lingüística). Instituto de Estudos da Linguagem, Universidade Federal de Campinas.

SILVA, A. H. P. **As fronteiras entre Fonética e Fonologia e a alofonia dos róticos iniciais em PB**: dados de dois informantes do sul do país. 2002. Tese (Doutorado em Lingüística). Instituto de Estudos da Linguagem, Universidade Federal de Campinas.

SILVA, A. H. P.; CLEMENTE, F. C; NISHIDA, G;. Para a representação dinâmica do tap em grupos e codas: evidências acústicas. **Revista Virtual de Estudos da Linguagem – ReVEL**. v. 4, n. 7, p.1-26, 2006.

STONE, M. A guide to analyzing tongue motion from ultrasound images. **Clinical Linguistics and Phonetics**, v. 19, p. 455-501, 2005.

VASSOLER, A. M. de O. **Coordenação gestual na produção de encontros consonantais em crianças com desenvolvimento típico e atípico**. 2016. Tese (Doutorado em Estudos Linguísticos). Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista.

WRENCH, A. **Articulate Assistant Advanced User Guide**: Version 2.14. Edinburgh, UK: Articulate Instruments Ltd., 2012.

Apêndices

Apêndice A – Termo de Consentimento Livre e Esclarecido assinado pelo responsável pela criança participante



UNIVERSIDADE FEDERAL DE PELOTAS
CENTRO DE LETRAS E COMUNICAÇÃO
LABORATÓRIO EMERGÊNCIA DA LINGUAGEM ORAL



TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Pesquisa:

Aquisição fonológica: análise acústica e articulatória

Pesquisadoras responsáveis:

Prof.^a Dr.^a Giovana Ferreira Gonçalves

Thais Telles Barbieri

Senhores pais ou responsáveis,

convidamos seu filho a participar da pesquisa *Aquisição fonológica: análise acústica e articulatória*, sob a responsabilidade da Prof.^a Dr.^a Giovana Ferreira Gonçalves e da mestranda Thais Telles Barbieri. Por meio dessa pesquisa, pretende-se investigar o processo de aquisição e a produção de sons na fala infantil e na fala adulta. Se você autorizar a participação de seu filho, estará contribuindo para o desenvolvimento dos estudos em fonologia do português como língua materna.

Para que a participação de seu filho seja possível, você deverá estar ciente das seguintes informações:

- 1)** A participação consiste em uma coleta de dados, a qual ocorrerá no Laboratório Emergência da Linguagem Oral (LELO). Esse laboratório se situa nas dependências do Centro de Letras e Comunicação (CLC) da Universidade Federal de Pelotas (UFPEL), *campus* Porto (situado à rua Gomes Carneiro, nº 1, em Pelotas/RS). Essa coleta será feita em duas etapas:
 - a)** gravação, em áudio, de interação guiada por figuras. Nessa etapa, o informante identificará figuras e passará por uma familiarização com palavras novas;
 - b)** coleta de dados de fala em áudio e vídeo, por meio de gravador, microfone e aparelho de ultrassom, em cabine com isolamento acústico. Nessa etapa, o

informante visualizará imagens em uma tela de computador e será condicionado a produzir palavras relacionadas a essas imagens.

- 2) A participação de seu filho nesta pesquisa é voluntária, podendo ser interrompida a qualquer momento sem qualquer prejuízo aos envolvidos. Pelo mesmo motivo, não haverá qualquer tipo de despesa ou remuneração para o participante ou para o seu responsável.
- 3) Os dados fornecidos pelos informantes serão analisados e poderão ser utilizados em eventos científicos, publicações e estudos futuros. Nesses casos, será preservada a identidade dos envolvidos – informantes e seus responsáveis.

Caso haja qualquer dúvida em relação a esta pesquisa, sinta-se livre para entrar em contato conosco por e-mail:

giovanaferreiragoncalves@gmail.com (Giovana Ferreira Gonçalves)
ttbarbieri@gmail.com (Thais Telles Barbieri)

Eu, _____, RG _____, responsável por _____, fui informado sobre as intenções desta pesquisa e entendo por que a participação de meu filho é necessária. Estou ciente de que não terei despesas e não serei remunerado por essa participação, bem como posso interrompê-la quando julgar necessário. Assino este documento por livre e espontânea vontade, certificando que estou de acordo com a realização desta pesquisa e autorizando a participação de meu filho.

Assinatura do responsável

Eu, Thais Telles Barbieri, RG 5091716273, confirmo que entreguei este documento ao responsável nele nomeado. Confirmo, também, que esclareci eventuais dúvidas acerca da pesquisa e do conteúdo deste documento ao responsável nomeado.

Thais Telles Barbieri
Mestranda em Letras

Prof.^a Dr.^a Giovana Ferreira Gonçalves
Pesquisadora responsável

Pelotas, _____ de _____ de 201__.

Apêndice B – Termo de Consentimento Livre e Esclarecido assinado pelos adultos participantes



UNIVERSIDADE FEDERAL DE PELOTAS
CENTRO DE LETRAS E COMUNICAÇÃO
LABORATÓRIO EMERGÊNCIA DA LINGUAGEM ORAL



TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Pesquisa:

Aquisição fonológica: análise acústica e articulatória

Pesquisadoras responsáveis:

Prof.^a Dr.^a Giovana Ferreira Gonçalves

Thais Telles Barbieri

Você está convidado a participar da pesquisa *Aquisição fonológica: análise acústica e articulatória*, sob a responsabilidade da Prof.^a Dr.^a Giovana Ferreira Gonçalves e da mestranda Thais Telles Barbieri. Por meio dessa pesquisa, pretende-se investigar o processo de aquisição e a produção de sons na fala infantil e na fala adulta. Se você aceitar participar, estará contribuindo para o desenvolvimento dos estudos em fonologia do português como língua materna.

Para que a sua participação seja possível, você deverá estar ciente das seguintes informações:

- 1) A participação consiste em uma coleta de dados, a qual ocorrerá no Laboratório Emergência da Linguagem Oral (LELO). Esse laboratório se situa nas dependências do Centro de Letras e Comunicação (CLC) da Universidade Federal de Pelotas (UFPEL), *campus* Porto (situado à rua Gomes Carneiro, nº 1, em Pelotas/RS). Essa coleta será feita em duas etapas:
 - a) gravação, em áudio, de interação guiada por figuras. Nessa etapa, o informante identificará figuras e produzirá palavras relacionadas a elas;
 - b) coleta de dados de fala em áudio e vídeo, por meio de gravador, microfone e aparelho de ultrassom, em cabine com isolamento acústico. Nessa etapa, o informante visualizará imagens em uma tela de computador e será condicionado a produzir palavras relacionadas a essas imagens.

- 2) A sua participação nesta pesquisa é voluntária, podendo ser interrompida a qualquer momento sem qualquer prejuízo aos envolvidos. Pelo mesmo motivo, não haverá qualquer tipo de despesa ou remuneração para o participante.
- 3) Os dados fornecidos por você serão analisados e poderão ser utilizados em eventos científicos, publicações e estudos futuros. Nesses casos, é assegurada a preservação da sua identidade.

Caso haja qualquer dúvida em relação a esta pesquisa, sinta-se livre para entrar em contato conosco por e-mail:

giovanaferreiragoncalves@gmail.com (Giovana Ferreira Gonçalves)
ttbarbieri@gmail.com (Thais Telles Barbieri)

Eu, _____, RG _____, fui informado sobre as intenções desta pesquisa e entendo por que a minha participação é necessária. Estou ciente de que não terei despesas e não serei remunerado por essa participação, bem como posso interrompê-la quando julgar necessário. Assino este documento por livre e espontânea vontade, certificando que estou de acordo com a realização desta pesquisa e autorizando a minha participação.

Assinatura do participante

Eu, Thais Telles Barbieri, RG 5091716273, confirmo que entreguei este documento ao participante nele nomeado. Confirmo, também, que esclareci eventuais dúvidas acerca da pesquisa e do conteúdo deste documento ao participante nomeado.

Thais Telles Barbieri
Mestranda em Letras

Prof.^a Dr.^a Giovana Ferreira Gonçalves
Pesquisadora responsável

Pelotas, _____ de _____ de 201__.

Apêndice C – Questionário apresentado ao responsável pela criança participante



UNIVERSIDADE FEDERAL DE PELOTAS
CENTRO DE LETRAS E COMUNICAÇÃO
LABORATÓRIO EMERGÊNCIA DA LINGUAGEM ORAL



FICHA DE CADASTRAMENTO E QUESTIONÁRIO

Pesquisa:

Aquisição fonológica: análise acústica e articulatória

Pesquisadoras responsáveis:

Prof.^a Dr.^a Giovana Ferreira Gonçalves

Thais Telles Barbieri

I) Ficha de cadastramento	
Dados da criança	
Nome:	
Idade:	
Data de nascimento:	
E-mail do(a) responsável:	
Telefone do(a) responsável:	
Nacionalidade:	☐ Brasileira ☐ Outra: _____
Naturalidade:	

II) Questionário	
Escolaridade e domínio de outras línguas	
Nível de escolaridade:	☐ Pré-escola ☐ Ensino Fundamental (Série: _____) ☐ Outro: _____
A criança utiliza outra língua além do português (por exemplo, quando está em casa ou quando fala com a família)?	☐ Sim (Qual/quais? _____) ☐ Não
A criança aprende outra língua na	☐ Sim (Qual/quais? _____)

escola?	 () Não
---------	---------------------------------

Pelotas, _____ de _____ de 201__.

Thais Telles Barbieri
Mestranda em Letras

Prof.^a Dr.^a Giovana Ferreira Gonçalves
Pesquisadora responsável

Apêndice D – Questionário apresentado aos adultos participantes



UNIVERSIDADE FEDERAL DE PELOTAS
CENTRO DE LETRAS E COMUNICAÇÃO
LABORATÓRIO EMERGÊNCIA DA LINGUAGEM ORAL



FICHA DE CADASTRAMENTO E QUESTIONÁRIO

Pesquisa:

Aquisição fonológica: análise acústica e articulatória

Pesquisadoras responsáveis:

Prof.^a Dr.^a Giovana Ferreira Gonçalves

Thais Telles Barbieri

I) Ficha de cadastramento	
Dados pessoais	
Nome:	
Idade:	
Data de nascimento:	
E-mail:	
Telefone:	
Nacionalidade:	() Brasileira () Outra: _____
Naturalidade:	

II) Questionário	
Escolaridade e domínio de outras línguas	
Nível de escolaridade:	() Ensino Fundamental () Ensino Médio () Graduação () Pós-graduação
Situação:	() Em curso () Concluído(a) () Interrompido(a)
Curso:	

(responda apenas se você marcou “Graduação” ou “Pós-Graduação” na seção “Nível de Escolaridade”)	
Tem domínio de outra língua além do português?	() Sim () Não
a. Inglês	() Produção oral () Produção escrita () Compreensão oral () Compreensão escrita
b. Espanhol	() Produção oral () Produção escrita () Compreensão oral () Compreensão escrita
c. Francês	() Produção oral () Produção escrita () Compreensão oral () Compreensão escrita
d. Alemão	() Produção oral () Produção escrita () Compreensão oral () Compreensão escrita
e. Outra língua: _____	() Produção oral () Produção escrita () Compreensão oral () Compreensão escrita
f. Outra língua: _____	() Produção oral () Produção escrita () Compreensão oral () Compreensão escrita

Pelotas, _____ de _____ de 201__.

Thais Telles Barbieri
Mestranda em Letras

Prof.^a Dr.^a Giovana Ferreira Gonçalves
Pesquisadora responsável

Apêndice E – Exemplo de ficha de transcrição utilizada para registro das produções CCV e CV

Arquivo	#5 CR 2018-04-12 – 4;9.wav
Duração	30:41
Sujeito	CR
Idade	04;09

Produção do <i>tap</i> [ɾ]: acompanhamento			
em onset simples (CV)		em coda	
tônico	átono	tônico	átono
00:40 guri	09:53 agora	01:38 turma	02:14 apertar
01:14 guri	11:28 tiro	['tru.ma]	07:38 apertado
18:11 bateria	11:32 adora	01:43 turma ø	
22:44 tirar	11:48 tira	01:45 turma	
	11:50 tira	['tru.ma]	
	18:01 jogadores		
	20:09 seguro		
	20:15 seguro		
	27:28 coceira		
Observações: –			

1ª repetição			
#	mm:ss	Palavra	Observações
1	01:41	outra	
2	01:51	outra	
3	01:55	outra	
4	07:49	sempre	
5	08:03	criança	
6	08:52	cobra	
7	09:02	grande	
8	09:06	grande	
9	10:35	mestre	
10	12:01	prato	
11	12:03	prato	
12	12:46	prato	
13	12:55	bruxa	
14	13:02	peso	
15	13:30	banco	A palavra não faz mais parte do <i>corpus</i> .
16	13:44	foto	
17	13:59	grato	A palavra não faz mais parte do <i>corpus</i> .
18	14:06	peça	
19	14:15	frita	
20	14:26	bucha	

21	14:34	pato	
22	14:43	fita	
23	14:56	prego	
24	15:12	prego	
25	15:23	frota	
26	15:23	frota	
27	15:29	frota	
28	15:32	frota	
29	15:33	frota	
30	15:43	pego	
31	15:53	preta	
32	15:53	preta	
33	15:54	preta	
34	15:58	branco	
35	15:59	branco	
36	15:00	branco	
37	15:00	branco	
38	16:01	branco	
39	16:07	branco	A palavra não faz mais parte do <i>corpus</i> .
40	16:21	pressa	
41	16:22	pressa	
42	16:50	gato	A palavra não faz mais parte do <i>corpus</i> .
43	16:57	broa	
44	17:17	preso	
45	17:45	boa	
46	17:51	prata	
47	17:55	prata	
48	18:02	prata	
49	18:10	fraca	
50	18:11	fraca	
51	18:25	praça	
52	18:35	faca	
53	18:50	passa	
54	19:02	pata	
2ª repetição			
55	20:45	prato	
56	20:54	bruxa	
57	21:00	peso	
58	21:06	banco	A palavra não faz mais parte do <i>corpus</i> .
59	21:12	foto	
60	21:37	grato	A palavra não faz mais parte do <i>corpus</i> .
61	21:42	peça	
62	21:57	frita	
63	22:04	frita	
64	22:16	bucha	
65	22:22	pato	
66	22:29	fita	
67	22:36	prego	

68	23:07	frita	
69	23:10	frota	
70	23:16	frota	
71	23:25	pego	
72	23:45	branco	
73	23:50	branco	A palavra não faz mais parte do <i>corpus</i> .
74	23:56	pressa	
75	23:58	pressa	
76	24:04	pressa	
77	24:30	gato	A palavra não faz mais parte do <i>corpus</i> .
78	24:38	broa	
79	24:45	preso	
80	24:51	boa	
81	25:00	prata	
82	25:05	fraca	
83	25:09	fraca	
84	25:20	praça	
85	25:30	faca	
86	25:38	passa	
87	25:45	pata	
3ª repetição			
88	26:47	prato	
89	27:10	bruxa	
90	27:15	peso	
91	27:24	banco	A palavra não faz mais parte do <i>corpus</i> .
92	27:36	foto	
93	27:45	grato	['ga.tru] A palavra não faz mais parte do <i>corpus</i> .
94	27:52	peça	
95	28:01	frita	
96	28:09	bucha	
97	28:15	pato	
98	28:20	fita	
99	28:24	prego	
100	28:25	prego	
101	28:31	frota	
102	28:40	pego	
103	28:46	branco	A palavra não faz mais parte do <i>corpus</i> .
104	28:51	pressa	
105	28:57	gato	A palavra não faz mais parte do <i>corpus</i> .
106	29:10	broa	
107	29:15	preso	
108	29:17	preso	
109	29:22	boa	
110	29:27	prata	
111	29:32	prato	
112	29:37	fraca	
113	29:38	fraca	
114	29:43	praça	

115	29:48	faca	
116	29:54	passa	
117	30:00	pata	
Total de dados para análise: 66			

Anexos

Anexo A – Imagens utilizadas para a coleta de áudio e ultrassom**prato****pato****prata****pata****praça****passa**



fraca



faca



prego



pego



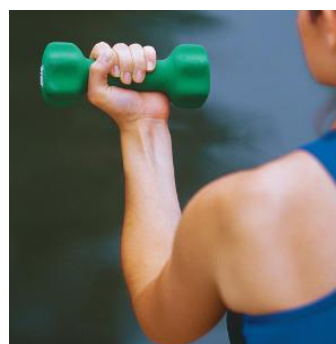
pressa



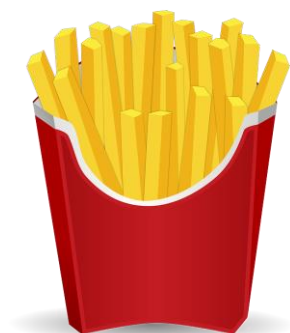
peça



preso



peso



frita



fita



frota



foto



broa



boa



bruxa



bucha