Aglael Juliana Aparecida Gama Rossi

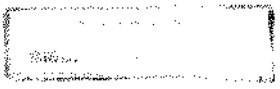
RELAÇÕES ENTRE DESENVOLVIMENTO LINGÜÍSTICO E NEUROMOTOR: A AQUISIÇÃO DA DURAÇÃO NO PORTUGUÊS BRASILEIRO

Tese apresentada ao Curso de Lingüística do Instituto de Estudos da Linguagem da Universidade Estadual de Campinas como requisito parcial para a obtenção do título de Doutor em Lingüística.

Orientadora:
Profa. Dra. Eleonora Cavalcante Albano

991643

UNICAMP
Instituto de Estudos da Linguagem
1999



UNIDADE BC
N.º CHAMADA:
Ex.
TOMBO 30/38543
PROC. 229/99
C D X
PREÇO R$ 11,00
DATA 27/08/99
N.º CPD

CM-00125660-0

este exemplar é a redação final da tese defendida por Aglael Juliana Aparecida Gama Rossi e aprovada pela Comissão Julgadora em 28/06/99.

Dra. Eleonora Cavalcante Albano

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA IEL - UNICAMP

Rossi, Aglael Juliana Aparecida Gama

R735r Relações entre desenvolvimento lingüístico e neuromotor: a aquisição da duração no português brasileiro / Aglael Juliana Aparecida Gama Rossi. - - Campinas, SP: [s.n.], 1999.

Orientador: Eleonora Cavalcante Albano

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Aquisição de linguagem. 2. Acústica. 3. Língua portuguesa - fonética. I. Albano, Eleonora Cavalcante. II. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. III. Título.

AGRADECIMENTOS

Ao fazer este trabalho, tive a sorte de contar com a atenção, disponibilidade, conhecimento e amizade de muitas pessoas. Quando paro e revejo todos esses anos, acho difícil colocar em poucas palavras aquilo que recebi de cada uma delas. O meu desejo mais profundo é que o drama eterno se encarregue de retribuir-lhes tudo aquilo que me deram. Para tentar expressar o que essas pessoas, *que me são muito queridas*, significaram, ofereço algumas estrofes da *Bhagavad Gita*:

1
Árjuna disse:
Qual o yogui mais perfeito,
o que está sempre ocupado
em Teu serviço amoroso
ou o que adora o Supremo
impessoal e sem forma?
13-14
O Senhor Supremo disse:
Aquele que não inveja;
que é amigo sincero
de todos os seres vivos;
que não tem senso de posse;
que está livre do egoísmo;
que tem a mesma atitude
na tristeza ou na alegria;
que está sempre satisfeito,
servindo com devoção;
que é sempre determinado
tendo a mente e o intelecto
harmonizados Comigo;
é muito querido a Mim.
15
Quem nunca perturba os outros
nem se deixa perturbar,
além da dualidade
do sofrimento e prazer,
livre do medo e da angústia,
também é muito querido.

16
Quem se livrou da esperança
e afeição materiais;
quem é esperto mas puro
e em todos os seus esforços
renuncia aos resultados
é muito querido a Mim.
17
Aquele que não se apega
nem ao prazer nem à dor,
que não rejeita ou deseja,
renunciando igualmente
ao que agrada ou aborrece
é muito querido a Mim.
18-19
Quem age do mesmo modo
com amigos e inimigos,
e não muda de atitude
no ostracismo ou na glória,
no sucesso ou no fracasso;
quem nunca se contamina;
quem está sempre contente,
com tudo que lhe oferecem;
quem, silente e satisfeito,
não está preocupado
em ter uma moradia;
quem está sempre ocupado
em servir com devoção
- este Me é muito querido.

(***Bhagavad Gita* Canção do Divino Mestre**.
Tradução, Introdução e Notas de Rogério
Duarte. Companhia das Letras, São Paulo,
1998, Capítulo 12, p.149-53)

Muito obrigada a:

No LAFAPE:
Eleonora Cavalcante Albano (minha orientadora), Edson Françozo, Plínio Almeida Barbosa, Sandra Madureira (PUC-SP), Paulo Rehder (responsável pela análise estatística), Mário Fontes (aquisição, instalação e uso do *CSRE 4.5*), Adelaide Hercília Pescatori Silva, Agnaldo A. Moreira, Ana Luiza Navas, Beatriz Raposo de Medeiros, Bruno Dallari, Cláudio Luís Marques Sampaio, Elizabeth M. G. de Souza, Evelin Fortes de Albuquerque, Guilherme Rios, Leonardo Couto Franco de Oliveira, Lígia Formico Paoletti,

Luciana Garcia Roriz, Luiz Arthur Pagani, Marcia Rafaela Arnold,
Patrícia Aparecida de Aquino, Paula Lenz Costa Lima, Regis K. Kakinohana,
Sandro Oliveira Simas, Vera Pacheco e Wilmar D'Angelis.

Professores visitantes do LAFAPE:
Magnus Petturson e Osamu Fujimura.

No IEL:
Ester M. Scarpa e Maria Bernadete M. Abaurre.

Nas Secretarias de Pós-Graduação e de Projetos do IEL:
A todos os funcionários e, em especial, a: Rose, Rogério, Bete, Ilda, Graziela e Cláudio.

Na Biblioteca do IEL:
Aos funcionários e, em especial, a: Haroldo, Cidinha e Ana Maria.

Na Limpeza do IEL:
A todas as *tias*, porque tornaram minha vida melhor.

No Xerox do IEL:
Helton e demais rapazes.

Ao CNPq:
Pela Bolsa de Doutorado, processo 140564/94-2.

Na EMEI Marcílio Dias:
Ligia Scomparini (diretora), Nadir Fernandes Bresson (coordenadora pedagógica) e
a querida professora Sônia e suas crianças, do primeiro estágio de 1996,
em especial, M e E.

No Estúdio EGG:
Eduardo, Guto e Gabriel.

No Centro Universitário São Camilo:
Maria de Jesus Gonçalves (a amiga de sempre), Sílvia Tavares (pela compreensão), colegas
e alunos.

Na Brahma Kumaris:
A todos os irmãos e irmãs.

Em casa:
Márcia, Paulo, Iná (pela força para conseguir a escola) e Idinha.

Nas estrelas e no coração:
Graça, Geraldo, Mariano, Norberto e Margarida.

Porque me fazem dar os melhores sorrisos:
Vivi, Khelf, Tuca, Kiko, Flor, Pituca, Bat, Berri, Leão, Francis, Neguinha, Penha, outra Neguinha, Nenê, Fofinho, Macaquinho, Batman, Kelly, Nina (Pipoca), Átila e Aretha.

Para Mauro,
por todo amor e respeito
que nos unem.

Banca Examinadora:

Profa. Dra. Eleonora Cavalcante Albano - Orientadora

Profa. Dra. Claudia Thereza Guimarães de Lemos

Profa. Dra. Leonor Scliar-Cabral

Prof. Dr. Plínio Almeida Barbosa

Profa. Dra. Regina Ritter Lamprecht

Suplentes:

Prof. Dr. Edson Françozo

Profa. Dra. Maria Cecília Marconi Pinheiro Lima

SUMÁRIO

RESUMO

SUMMARY

INTRODUÇÃO 01

1. EXPLICITANDO FALHAS DA LITERATURA E ELABORANDO UMA NOVA PROPOSTA SOBRE A INTERAÇÃO ENTRE ASPECTOS LINGÜÍSTICOS E NEUROMOTORES.....06

1.1 Histórico da relação entre duração e variabilidade e as hipóteses da maturação neuromotora e do artefato estatístico 06

1.2 Falhas metodológicas da literatura 29

1.3 Resultados de um estudo piloto e uma hipótese alternativa às hipóteses da maturação neuromotora e do artefato estatístico 35

1.4 Sujeitos e situação de gravação do novo experimento 39

1.5 Material de gravação 43

1.6 Critérios adotados na realização das medidas de duração 44

1.7 Critérios adotados na elaboração do *corpus*.....48

1.8 O subconjunto de dados analisados e os critérios lingüísticos adotados na análise estatística 52

2. NOVOS DADOS E REFLEXÕES SOBRE A INTERAÇÃO ENTRE DURAÇÃO E VARIABILIDADE NA AQUISIÇÃO DA DURAÇÃO NO PORTUGUÊS BRASILEIRO.....71

2.1 Comparações entre os sujeitos de cada par para o teste t 75

2.1.1 Duração e variabilidade de segmentos acústicos vocálicos 75

2.1.2 Duração e variabilidade de segmentos acústicos consonantais.....84

2.1.3 Duração e variabilidade de segmentos acústicos do tamanho de sílabas.....94

2.1.4 Duração e variabilidade de segmentos acústicos do tamanho de palavras.....106

2.2 Comparações entre os sujeitos para o teste estatístico de Duncan 113

2.3 Os novos fatos sobre a interação entre duração e variabilidade a partir do estudo da aquisição da duração no português brasileiro 122

3. MODELOS DINÂMICOS NO ESTUDO DA AQUISIÇÃO DA LINGUAGEM.....130

3.1 O problema das unidades de análise no estudo da fala infantil 130

3.2 Uma unidade com tempo intrínseco na explicação de processos fônicos da fala corrente: a Fonologia Articulatória 134

3.3 Necessidade do tempo intrínseco na explicação de processos fônicos da fala corrente 142

3.4 A unidade com tempo intrínseco e a abstração das representações subjacentes 145

3.5 A falta de uma unidade de controle auditivo e de um controle extrínseco de tempo na Fonologia Articulatória 150

3.6 Um modelo dinâmico do ritmo 153

3.7 Modelos dinâmicos e os fatos sobre as relações entre desenvolvimento lingüístico e neuromotor na aquisição da duração no português brasileiro.....158

ANEXO A.....171

ANEXO B.....173

ANEXO C.....175

ANEXO D.....176

ANEXO E.....177

ANEXO F.....179

ANEXO G.....180

ANEXO H.....181

REFERÊNCIAS BIBLIOGRÁFICAS.....182

APÊNDICE.....190

RESUMO

O trabalho apresenta uma discussão sobre as relações entre desenvolvimento lingüístico e neuromotor a partir do estudo da aquisição da duração no português brasileiro (PB). Ele parte de uma literatura que compara grupos de crianças de diferentes faixas etárias a um grupo controle de adultos para medidas de duração. As hipóteses encontradas para explicar os maiores valores de duração e a maior variabilidade na fala infantil referem-se a fatores neuromotores e estatísticos. Faz-se uma crítica aos experimentos dos estudos resenhados e persegue-se a hipótese de que, no PB, um fator lingüístico, o grau de acentuação do segmento acústico em função da posição por ele ocupada na palavra e na sentença, influencia a diferença encontrada entre adultos e crianças, para medidas de duração, isso porque, as crianças não reduzem os segmentos em posições não-acentuadas. Realizou-se um experimento no qual sentenças pronunciadas pela pesquisadora eram repetidas três vezes pelos sujeitos, tendo-se três pares formados por: criança de 4 anos e 1 mês (M) e pesquisadora; criança de 4 anos e 9 meses (E) e pesquisadora; e a professora delas (S) e a pesquisadora. Segmentos acústicos vocálicos, consonantais e do tamanho de sílabas foram demarcados para suas posições de acento nas palavras (pré-tônica, tônica e pós-tônica) e na sentença (início e final absolutos). Por meio de uma análise via teste t, os sujeitos dentro de cada par foram comparados para a média e o desvio-padrão das diferenças de duração de cada tipo de segmento acústico, em cada posição de acento. A hipótese a ser testada era se as médias das diferenças de duração entre os sujeitos de cada par seria estatisticamente igual a zero. Encontrou-se que a criança mais nova (M) já adquiriu a implementação do parâmetro de duração para a realização do acento, embora ainda não o tenha adquirido para os elementos não-acentuados que participam da construção das alternâncias rítmicas do PB. Já a criança E está muito próxima do padrão adulto para a duração de segmentos vocálicos, mas não para a duração de consoantes, sílabas e palavras. Quanto à variabilidade, embora M e E apresentem valores maiores de desvios-padrão, seus contornos de duração para os vários tipos de segmentos acústicos, nas várias posições de acento, não são díspares em relação àquele obtido para o par formado pelos adultos. Nos segmentos acústicos do tamanho de sílabas e palavras, há uma maior coordenação com sobreposição de gestos na fala adulta, levando, aparentemente, a uma maior coarticulação que na fala infantil. A partir da adoção de modelos dinâmicos, a maior variabilidade da fala infantil, na produção de segmentos acústicos vocálicos e consonantais, é vista como uma menor coordenação entre gestos articulatórios e uma menor freqüência de oscilação dos articuladores, que dificulta a produção de segmentos reduzidos e estáveis nas posições não-acentuadas. Propõe-se que as crianças estudadas possuem um oscilador para variações macrorrítmicas, provavelmente ao nível da sentença, do contorno duracional do PB, mas não para variações ao nível da sílaba, como é demonstrado pela menor coordenação e coarticulação entre os gestos articulatórios envolvidos em sua produção.

Palavras-chave: Aquisição da Linguagem, Acústica, Língua Portuguesa (Fonética)

SUMMARY

This work offers a discussion on the relationship between linguistic and neuromotor development, based on the study of the acquisition of duration in Brazilian Portuguese (BP). Many studies in the literature have shown that acoustic segment durations are more variable in young children's speech than in adults' speech. The hypotheses that have been advanced to explain this fact refer to neuromotor and statistical factors. Our work criticizes the experiments in the literature and proposes that, in BP, a linguistic factor, which combines stress status and position of the acoustic segment in the word and sentence, can influence the difference between adults' and children's duration measures, since children do not reduce acoustic segments in stressless positions as adults do. We ran an experiment in which sentences produced by the author were repeated three times by the subjects. The analyses made comparisons within and among three pairs of subjects, made up by the author and: one child of 4 years and 1 month (M), one child of 4 years and 9 months (E), and their teacher. The acoustic segments of each sentence, corresponding to vowels, consonants, syllables and words, were demarcated for its stress position in the word (pre-stressed, stressed and post-stressed) and in the sentence (beginning and final). A statistical analysis via t test compared the subjects in each pair as to the mean and standard deviation of the differences in duration for each acoustic segment type in each stress position. The hypothesis to be tested was whether the mean of the differences in duration would be statistically equal to zero. We found that the younger child (M) had already acquired the duration parameter for the production of stress, but not for the production of acoustic segments in the stressless position. Child E's production is very similar to the adult's in relation to vowels, but not to consonants, syllables and words. Concerning variability, although the children present higher standard deviation values, their variability contours for the different kinds of acoustic segments in different stress positions are similar to the ones obtained for the adults. In syllables and words, a coordination with superimposed gestures was found to be higher in the adults' speech than in the children's speech. We adopt dynamic models to explain the higher variability of the children's speech, mainly in the production of isolated acoustic segments, corresponding to vowels and consonants, in view of the lesser coordination among articulatory gestures and the lower oscillation frequency of the articulators, which prevents the production of reduced and stable acoustic segments in stressless positions. Also in line with dynamic models, we propose that the child production mechanisms studied here incorporate an oscillator which accounts for the greater variation in the acoustic duration contour of BP, probably at the level of the sentence, but not at lower level of the syllable, as shown by the lesser coordination and coarticulation among the articulatory gestures involved in their production.

Key words: Language Acquisition, Acoustic, Portuguese Language (Phonetics)

INTRODUÇÃO

Nas últimas quatro décadas, o desenvolvimento de uma tecnologia de fala, representada, num primeiro momento, pelo espectrógrafo e, posteriormente, pelo microcomputador, permitiu o surgimento e a consolidação da Fonética Acústica. Ou seja, de um espaço onde é possível pensar a relação entre o acústico e o articulatório e, portanto, onde se pode validar ou não os diversos modelos de produção e percepção de fala, explorando-se ao máximo os limites dessa relação (KENT e READ, 1992).

No que concerne à fala infantil, nas décadas de 50 e 60, poucos pesquisadores aventuraram-se na descrição de seus aspectos acústicos. Contudo, a partir do início da década de 80, os estudos acústicos da fala infantil adquiriram volume e importância significativos, porque deles se tentou depreender implicações sobre o desenvolvimento motor, que subjaz à aquisição e desenvolvimento da fala, e seus possíveis desvios.

Até pouco antes dos anos 80, os dados acústicos sobre crianças falantes do inglês americano referiam-se basicamente às mudanças, em função da idade, nos valores de freqüência fundamental, nos formantes de sons vocálicos e em algumas propriedades

temporais da fala (KENT, 1976). A partir de 1980, com a publicação de um estudo de KENT e FORNER sobre a duração na fala de crianças e adultos, surge uma extensa série de trabalhos empenhados, durante uma década e meia, em explicar as diferenças de *timing* entre a fala adulta e infantil. Logo em seguida, começam a aparecer também estudos sobre a coarticulação na fala adulta e infantil (entre eles, TURNBAUGH, HOFFMAN, DANILOFF e ABSHER, 1985, REPP, 1986, SERENO, BAUM, MAREAN e LIEBERMAN, 1987, SERENO e LIEBERMAN, 1987, FLEGE, 1988, NITTROUER, STUDDERT-KENNEDY e McGOWAN, 1989). A partir da década de 90, adultos e crianças passam a ser comparados para medidas acústicas e articulatórias (STATHOPOULOS, 1995, SMITH e KENNEY, 1998).

No português brasileiro (doravante, PB), não há ainda uma linha de pesquisa que utilize sistematicamente as várias medidas acústicas no estudo da fala infantil. O primeiro trabalho de que temos conhecimento é o de LEVY (1993), que realizou uma inspeção acústica na fala de crianças com dificuldade na produção do vozeamento, quanto à distinção de obstruintes.

Partimos do trabalho de KENT e FORNER (1980) e do conjunto de trabalhos que a eles se seguiram empregando medidas de duração para comparar grupos de crianças de diferentes faixas etárias a um grupo controle de adultos. A discussão por eles levantada tem-se atido à questão de até onde fatores maturacionais e/ou estatísticos constituem as causas da maior variabilidade do *timing* da fala infantil, entre repetições de um mesmo enunciado.

Ao contrário da literatura, procuramos mostrar que fatores somente maturacionais e estatísticos são insuficientes para explicar as diferenças entre a fala de adultos e crianças: é

necessário contemplar a interação entre aspectos lingüísticos e neuromotores na aquisição e realização da língua. Para isso, apresentamos uma comparação entre três pares de sujeitos, falantes do PB, formados pela pesquisadora, de um lado, e por duas crianças da faixa etária de 4 anos e um adulto, de outro, para medidas de duração, principal parâmetro acústico na realização do acento em PB (MORAES, 1984, MASSINI, 1991). Os sujeitos de cada par são comparados entre si para a duração de segmentos acústicos correspondentes a vogais, consoantes, sílabas e palavras, medidos dentro de sentenças, enunciadas pela pesquisadora e repetidas pelos sujeitos.

Para discutir a interação entre aspectos lingüísticos e neuromotores na aquisição da duração no PB, adotamos modelos dinâmicos, nos quais o controle motor da fala é visto como decorrente de uma dinâmica abstrata que controla o movimento de objetos físicos reais (língua, mandíbula, etc.) e é expressa por meio de equações dinâmicas utilizadas na descrição da motricidade em geral (BROWMAN e GOLDSTEIN, 1985, 1986, 1989, 1990, 1992, PORT, CUMMINS e GASSER, 1995).

No LAFAPE[1], a adoção e discussão de modelos dinâmicos, a partir da análise de dados acústicos da fala adulta (ALBANO, 1999, AQUINO, 1997, BARBOSA, 1994, 1996, SILVA, 1996, SOUZA, 1994) e infantil (GAMA-ROSSI e ALBANO, 1996, 1998), têm levado ao debate da tese da comensurabilidade entre a Fonética e a Fonologia (ALBANO, BARBOSA, GAMA-ROSSI, MADUREIRA e SILVA, 1998), que, de acordo com ALBANO (1998), remete-nos à relação corpo-mente, na medida em que nos obriga a questionar se um dado processo fônico se enraiza na mecânica da produção da fala ou é

desencadeado por categorias fônicas ou gramaticais de ordem superior ou, ainda, se a própria mecânica da produção da fala obedece exclusivamente a princípios físicos ou se ajusta às categorias da língua, tornando impossível falar de automatismos fonéticos universais.

No que se refere à aquisição da linguagem, há muito (GONÇALVES, 1989, GAMA, 1989, ALBANO, 1990) as questões teóricas e metodológicas sobre a relação entre o controle físico da fala e a emergência de categorias lingüísticas têm nos tocado. Assim, por meio de uma investigação acústica, este trabalho tentará apontar para possíveis relações entre o desenvolvimento de categorias lingüísticas e o controle neuromotor da fala infantil.

Na seqüência, o Capítulo 1 (Explicitando Falhas da Literatura e Elaborando uma Nova Proposta sobre a Interação entre Aspectos Lingüísticos e Neuromotores) apresenta uma resenha das hipóteses encontradas na literatura da qual partimos, a respeito da relação entre duração e variabilidade, na fala adulta e infantil, e a crítica às falhas metodológicas, dos experimentos dos estudos resenhados, falhas essas difíceis de enxergar para não-lingüistas. O capítulo traz também uma nova hipótese advinda de um estudo piloto, e descreve os sujeitos, materiais e critérios de elaboração do *corpus* utilizados no experimento que serviu de base a este trabalho. Além disso, são apresentados os critérios empregados na análise estatística dos dados, os quais fizeram com que o trabalho estabelecesse definitivamente uma diferença em relação à literatura anterior, por cruzar parâmetros de medidas de variabilidade de fala com parâmetros de análise lingüística. No Capítulo 2 (Novos Dados e Reflexões sobre a Interação entre Duração e Variabilidade na Aquisição da Duração no Português

---

[1] Laboratório de Fonética Acústica e Psicolingüística Experimental, do Instituto de Estudos da Linguagem.

Brasileiro), são apresentados os resultados da análise estatística que comparou os sujeitos dentro de cada par (criança de 4 anos e 1 mês e pesquisadora; criança de 4 anos e 9 meses e pesquisadora; e professora das crianças e pesquisadora), por meio do teste t, para a média e o desvio-padrão das diferenças de duração de segmentos acústicos vocálicos, consonantais, e do tamanho de sílabas e palavras. Analisam-se também os resultados da comparação entre: (1) os modelos da pesquisadora oferecidos aos três sujeitos; (2) as crianças e a professora; (3) as duas crianças e os dois adultos; (4) as crianças e (5) os adultos entre si; para a duração dos segmentos acústicos vocálicos e consonantais e do tamanho das sílabas de cada sentença e de todas as sentenças, a partir do teste estatístico de Duncan. No Capítulo 3 (Modelos Dinâmicos no Estudo da Aquisição da Linguagem), são expostos, discutidos, criticados e adotados os modelos dinâmicos da Fonologia Articulatória (BROWMAN e GOLDSTEIN, 1985, 1986, 1989, 1990, 1992) e de PORT, CUMMINS e GASSER (1995). Propomos, ainda, um diálogo entre esses modelos, num esforço de oferecer uma explicação para as diferenças encontradas entre os dois adultos e as duas crianças aqui estudados.

---

da Universidade Estadual de Campinas.

1. EXPLICITANDO FALHAS DA LITERATURA E ELABORANDO UMA NOVA PROPOSTA SOBRE A INTERAÇÃO ENTRE ASPECTOS LINGÜÍSTICOS E NEUROMOTORES

Neste capítulo, mostra-se que a literatura precedente cometeu um sério equívoco ao desconsiderar a interação entre aspectos lingüísticos e neuromotores, na análise das diferenças entre fala adulta e infantil, para medidas de duração. A partir da discussão de suas hipóteses e falhas metodológicas, que consistem sobretudo em pressupor análises lingüísticas ingênuas ou, no mínimo, superficiais, apresentamos nossa tentativa de superá-las, tanto na elaboração do experimento proposto, como nos critérios empregados na análise estatística, visando, com isso, destrinçar a intricada e pouco conhecida relação entre o lingüístico e o neuromotor, a partir da aquisição do parâmetro de duração no PB.

1.1 Histórico da relação entre duração e variabilidade e as hipóteses da maturação neuromotora e do artefato estatístico

Na literatura, produzida em geral em Departamentos de Psicologia ou Distúrbios da Comunicação, são encontradas duas hipóteses para explicar a diferença de *timing* entre fala

adulta e infantil. A hipótese da maturação neuromotora prediz que a relação entre duração e variabilidade é uma função da idade e independe da duração (SMITH, SUGARMAN e LONG, 1983). O que está por trás dessa hipótese é que a imaturidade do sistema nervoso central das crianças pequenas faz com que elas apresentem uma produção motora mais variável que os adultos e as crianças mais velhas, o que, obviamente, não se reflete apenas na fala, mas abrange toda a realização de tarefas que envolvem habilidades motoras específicas ou especializadas (TINGLEY e ALLEN, 1975 apud KENT, 1976). A realização de atividades motoras pelas crianças pequenas, principalmente aquelas abaixo de 6 anos (KENT e FORNER, 1980), mostra indícios de imaturidade neurológica tanto no que diz respeito ao *timing* dos movimentos, como à sua precisão[1] (KENT, 1976). Assim, devido a um controle neural já estabelecido e estável, o adulto é capaz de, durante repetições de um mesmo enunciado, mover seus articuladores na direção dos mesmos alvos, ao passo que as crianças pequenas apresentam uma maior variação.

A variabilidade da produção da fala é aferida a partir de transformações dos valores das medidas acústicas do *corpus* a ser estudado em parâmetros estatísticos, sendo que os mais usados pelos estudos que enfocam a variabilidade temporal da fala infantil são a média e o desvio-padrão, calculados para as durações obtidas a partir das várias repetições de um mesmo segmento acústico[2], por um sujeito. O desvio-padrão, também designado por

---

[1] Segundo KENT (1976), *precisão* é entendida como a capacidade de o sistema nervoso central controlar as forças musculares necessárias para a realização de movimentos finos, por exemplo, aqueles envolvidos na fala.

[2] De acordo com FANT (1973), as fronteiras de segmentos acústicos (*sound segments*) não deveriam ser confundidas com as fronteiras de fonemas, uma vez que sons adjacentes, na fala corrente, podem carregar informação de um único fonema, enquanto a sobreposição entre segmentos acústicos faz com que um único segmento carregue informação de vários fonemas adjacentes. A análise espectrográfica da fala mostra, segundo ele, fronteiras distintas entre partes sucessivas no eixo do tempo e essas fronteiras estão

variabilidade absoluta, fornece uma medida do quanto cada uma das repetições de um sujeito distancia-se da média de suas repetições. Na comparação entre sujeitos de diferentes grupos etários ou de um mesmo grupo etário para diferentes taxas de elocução, emprega-se ainda o coeficiente de variação (SMITH et al., 1983), também designado por variabilidade relativa, o qual constitui a razão entre o desvio-padrão e a média das durações de um dado segmento acústico. Assim, a maior precisão da fala adulta é expressa por meio de valores geralmente mais baixos tanto do desvio-padrão como do coeficiente de variação, na produção repetida de um segmento acústico.

A segunda hipótese, ou hipótese do artefato estatístico, decorre de uma crítica à hipótese da maturação neuromotora e prediz que a maior variabilidade temporal da fala infantil é um viés da análise estatística das medidas de duração. Por essa hipótese, a relação entre duração média e variabilidade depende da duração e independe da idade (SMITH et al. 1983). Ela surgiu da questão levantada por KENT e FORNER (1980) de que a maior variabilidade da fala infantil poderia ser uma conseqüência de sua taxa de elocução

---

relacionadas a eventos que mudam em função de mudanças no mecanismo da fala, ou seja, em função da fonte sonora primária (voz ou ruído) ou do tipo de abertura ou fechamento formado dentro das cavidades (oral, nasal e lateral). Fronteiras de sons menos distintas são definidas a partir de mudanças no padrão de formantes, sendo que a variação mais ou menos contínua de alguns formantes depende do local de suas freqüências e seus padrões podem variar dentro e entre fronteiras de segmentos acústicos.
FANT explica que o número de segmentos acústicos sucessivos dentro de um enunciado é maior que o número de fonemas. Ele dá como exemplo uma oclusiva surda, que contém pelo menos dois segmentos acústicos, a saber: a oclusão e o estouro, este último podendo ainda ser dividido em uma explosão transitória e uma fricativa curta. Da mesma forma, a primeira parte da vogal que se segue a um estouro assimila seu desvozeamento e é uma questão de convenção, segundo ele, estipular se o segmento acústico correspondente ao estouro é atribuído à vogal ou à consoante precedente.
Por fim, o autor nos fala da relação entre segmentos acústicos e traços. Segundo ele, fronteiras entre segmentos acústicos e traços são devidas ao início ou fim de pelo menos um dos traços, mas um único traço pode se estender sobre vários segmentos acústicos. Esse seria o caso da nasalização da vogal adjacente a uma consoante nasal. FANT diz que o exemplo mais comum seria aquele da continuidade de vibrações das pregas vocais ao longo de uma série de sons vocálicos.

caracteristicamente mais lenta, o que faria da maior variabilidade das crianças pequenas um artefato estatístico de suas durações no geral mais longas.

Apesar de a literatura ter-se debatido entre as duas hipóteses durante uma década e meia aproximadamente, tentaremos mostrar, ao longo dessa resenha, que houve pouco avanço em estabelecer a superioridade de uma hipótese em oposição à outra, uma vez que os esforços no sentido de destrinçar a complexa relação entre duração e variabilidade acabaram produzindo evidências esparsas, ora a favor de uma, ora a favor da outra hipótese.

A hipótese da maturação neuromotora aparece esboçada numa resenha (KENT, 1976) que arrola os primeiros estudos acústicos da fala infantil, realizados para uma variedade de medidas acústicas que não somente as de duração. Nessa resenha, o autor mostra que os valores das medidas acústicas tendem a diminuir com a idade e que uma variação dentro dos padrões adultos é atingida por volta dos 12 anos.

EGUCHI e HIRSH (1969 apud KENT, 1976) encontraram que os desvios-padrão dos valores de f0 (freqüência fundamental), F1 e F2 (correspondentes aos picos de resposta do primeiro e segundo formantes) diminuem progressivamente com a idade, até que um mínimo é alcançado por volta de 10-12 anos. Entretanto, os resultados foram questionados por LINDBLOM (1972 apud KENT, 1976), que mostrou que a estimativa dos formantes a partir de espectrogramas é limitada pela freqüência fundamental da voz do falante, ou seja, quanto mais alto o valor de f0, maior o erro associado ao cálculo dos formantes. Esse é o caso das vozes infantis e femininas que apresentam freqüências fundamentais altas, devido a características das configurações anatômicas semelhantes entre os tratos vocais de crianças e

de mulheres, e que as tornam distintas das vozes masculinas (FANT, 1966 apud KENT, 1976).

Para as distribuições de VOT (ou *Voice Onset Time*, definido como o intervalo de tempo entre a soltura da oclusiva e o início do vozeamento do elemento vocálico seguinte), KENT (1976) reporta dados de diversos autores, traçando uma linha de desenvolvimento, segundo a qual a distribuição do VOT das primeiras palavras de crianças falantes do inglês americano muda de unimodal para bimodal e a distribuição similar àquela dos adultos é observada por volta dos 8 anos de idade, quando as crianças apresentam dois valores de VOT, não sobrepostos, para oclusivas surdas e sonoras.

No que concerne aos aspectos temporais da fala, KENT (1976) cita os trabalhos de DISIMONI (1974a,b,c, apud KENT, 1976), que comparou grupos de crianças de 3, 6 e 9 anos, para as durações de vogais e consoantes produzidas dentro de enunciados CVC/VCV e encontrou que a variabilidade das durações tende a diminuir com a idade, com influências do contexto e da extensão do enunciado sobre as durações segmentais surgindo entre 3 e 6 anos. Esse dado de DISIMONI (1974a,b,c apud KENT, 1976) serviu de base para a afirmação de KENT (1976) de que talvez a aquisição de programas motores esteja condicionada ao aprendizado de regularidades fonológicas.

O dados acústicos disponíveis sobre a fala infantil apontavam então para uma diminuição dos valores das medidas acústicas e de sua variação em função do aumento da idade, o que fez com que a hipótese da maturação neuromotora fosse, inicialmente, tomada como base para o trabalho inaugural (KENT e FORNER, 1980) da linha de pesquisa que, tentando entender a maior variabilidade temporal da fala infantil, ocupou-se exclusivamente

das medidas de duração e, por conseguinte, da relação de dependência entre duração e variabilidade.

KENT e FORNER (1980) compararam grupos, com dez sujeitos cada, de crianças de 4, 6 e 12 anos, a um grupo controle de adultos, para medidas de duração, realizadas dentro de três sentenças (*We saw you hit the cat, The box is blue and red* e *I took a spoon and a knife*), repetidas quatro vezes pelos sujeitos, a partir de modelo fornecido por um dos pesquisadores. Eles optaram por utilizar sentenças porque os poucos dados sobre durações segmentais medidas na fala corrente de crianças não deixavam claro se o alongamento de segmentos era uma propriedade uniforme da fala infantil.

Uma parte dos resultados do estudo parece confirmar os achados anteriores (DISIMONI, 1974a,b,c apud KENT, 1976) de que os valores da duração média, do desvio-padrão (variabilidade absoluta) e do coeficiente de variação (variabilidade relativa) diminuem em função do aumento da idade, e, na maior parte das vezes, não são encontradas diferenças substanciais entre as crianças acima de 6 anos e os adultos.

As comparações realizadas para segmentos acústicos das sentenças (VOT de [k] em *cat*, de [t] em *took* e de [b] em *blue*, duração da vogal [a] e do fechamento da oclusiva [k] em *box*) mostraram que a duração média e a variabilidade diminuíam em função da idade e que havia uma maior dispersão dos valores de desvios-padrão dos sujeitos dos grupos mais novos em relação aos grupos mais velhos.

Quando os grupos etários foram comparados para trechos[3] delimitados dentro de cada uma das três sentenças do *corpus*, KENT e FORNER (1980) encontraram novamente uma redução nas durações dos trechos em função do aumento da idade dos sujeitos e as crianças de 4 anos apresentaram os maiores valores tanto no desvio-padrão como no coeficiente de variação do grupo, ou seja, na variabilidade absoluta e relativa, respectivamente. Além disso, não foram observadas diferenças consistentes na variabilidade entre os três grupos mais velhos, embora os adultos tendessem a ser menos variáveis que as crianças de 6 e de 12 anos.

A comparação entre os sujeitos para a variação relativa, ou coeficientes de variação, de alguns segmentos acústicos (VOT de [k] em *cat*, duração da fricativa [s] em *box*, e duração da fricativa [s] em *spoon*) mostrou uma diminuição de seus valores conforme a idade aumentava. Os adultos, enquanto grupo, apresentaram os menores valores, seguidos pelas crianças de 12 anos, ao passo que as crianças de 4 e 6 anos apresentaram os maiores valores para os coeficientes de variação. KENT e FORNER (op. cit.) ressaltam que, no grupo, os coeficientes de variação das crianças de 12 anos e dos adultos apresentaram ainda uma menor dispersão que aqueles das crianças de 4 e 6 anos, o que os fez supor que os sujeitos mais velhos são mais uniformes com relação à precisão individual.

Porém, nos três segmentos acústicos medidos, havia duas ou três crianças de 4 anos, num total de dez, que possuíam coeficientes de variação dentro do espectro de variação dos adultos. Isso fora observado nas comparações realizadas anteriormente, para os demais

---

[3] Vale esclarecer que aquilo designado por "trecho" (*phrase length*, no original) não corresponde a uma unidade lingüística, por exemplo, do tamanho de um sintagma, como poder-se-ia imaginar. Ao contrário, o

segmentos acústicos e para os trechos delimitados dentro das sentenças. Os autores presumiram, então, que crianças de 4 anos eram capazes de um controle bem mais seguro da produção de fala, o que os fez afirmar que não era possível inferir que elas tivessem consistentemente segmentos acústicos mais longos e mais variáveis, mas que elas teriam segmentos mais longos e variáveis que os adultos sob certas circunstâncias. KENT e FORNER (1980) apelaram então para a necessidade de observar a influência de fatores fonéticos ou lingüísticos, que estariam relacionados às diferenças de desenvolvimento encontradas nas durações de segmentos acústicos. A não ser por uma tentativa muitíssimo modesta (SMITH, 1994), como veremos a seguir, essa sugestão será ignorada por todo o conjunto de estudos que se sucedeu a eles.

O embrião da hipótese do artefato estatístico, a qual será perseguida juntamente com a hipótese da maturação neuromotora pelos estudos posteriores, encontra-se também em KENT e FORNER (1980) e decorre do achado de que, quando as crianças de 4 anos apresentaram durações de segmentos acústicos similares àquelas dos adultos e crianças mais velhas, os valores de suas medidas de variabilidade assemelharam-se aos dos sujeitos mais velhos. KENT e FORNER (1980) afirmaram então que pelo menos parte da variação das medidas de duração poderia estar relacionada à taxa de elocução, como demonstrado para a fala adulta. Seguindo KLATT (1974, apud KENT e FORNER, 1980) e LEHISTE (1972, apud KENT e FORNER, 1980) eles presumiram que se a precisão do *timing* da fala de um indivíduo pode variar com fatores que influenciam a duração segmental, então é importante reconhecer a possibilidade de que uma maior variabilidade esteja relacionada simplesmente a

---

trecho delimitado dentro da sentença corresponde a uma seqüência de segmentos acústicos para os quais foi

uma taxa de elocução mais lenta (e, portanto, a segmentos acústicos mais longos), e não necessariamente a uma imaturidade neurológica, na fala em desenvolvimento, ou a uma lesão neurológica, na fala com desvio.

KENT e FORNER (1980) deixaram em aberto a discussão sobre a variabilidade do *timing* da fala infantil e a duração de segmentos acústicos. Os estudos que se seguiram passaram a procurar correlações entre duração e variabilidade, sem que a influência de fatores fonéticos e lingüísticos na relação entre ambas fosse sequer considerada como objeto de investigação. Os estudos focalizaram em como mudanças na duração afetavam a variabilidade, e, para isso, empregaram modificações das condições experimentais nas quais as amostras de fala eram coletadas. Essas modificações incluíam mudanças na taxa de elocução, fala com *bite block* e estímulos de fala produzidos a partir de leitura.

Segundo SMITH, SUGARMAN e LONG (1983), se as durações diminuíssem na taxa de elocução rápida e aumentassem na taxa de elocução lenta, na comparação entre sujeitos de um mesmo grupo etário, ou se sujeitos de diferentes grupos etários variassem igualmente entre as três taxas de elocução, haveria evidência a favor da hipótese do artefato estatístico. Em contraposição, se a variabilidade relativa permanecesse constante entre as diferentes taxas de elocução, para sujeitos de um mesmo grupo etário, ou se sujeitos de diferentes grupos etários variassem diferentemente entre as três taxas de elocução, a hipótese da maturação neuromotora seria favorecida. Como será visto, os resultados de SMITH et al. (1983) indicam que nenhuma das duas hipóteses explica de modo exclusivo as diferenças na

---

possível comparar todos os sujeitos.

variabilidade temporal da fala infantil, mas que tanto vieses estatísticos como fatores maturacionais parecem contribuir para sua maior variabilidade.

Os autores utilizaram três grupos de crianças, cada um deles formado por 5 sujeitos, com idade média de 5, 7 e 9 anos, e um grupo de adultos. Os sujeitos foram gravados para dez repetições da sentença *My puppy ate the cookie today*, em três taxas de elocução, normal, rápida e lenta. Foram medidas as durações de dois trechos delimitados dentro das sílabas acentuadas de *puppy* e *cookie*, e a duração de um trecho delimitado dentro da sentença (da soltura do /p/ inicial de *puppy* à soltura do /t/ de *today*).

As relações observadas entre duração e variabilidade, a partir da variação da taxa de elocução, não foram simples de interpretar em relação às hipóteses em jogo. Na comparação entre grupos, uma análise de variância mostrou que a taxa de elocução teve um efeito significativo sobre as medidas de duração dos trechos delimitados dentro das sílabas e dentro da sentença e não houve interações significativas entre idade e taxa de elocução. Contudo, quando a atuação de cada grupo foi considerada isoladamente, para cada taxa de elocução, constatou-se que as durações dos trechos delimitados dentro das sílabas e dentro da sentença eram maiores para as crianças de 5 anos em comparação aos adultos e crianças mais velhas. Em outras palavras, nas três taxas de elocução, as crianças pequenas apresentaram as maiores durações, o que foi confirmado por uma análise de variância que mostrou que o fator idade teve um efeito significativo sobre as medidas de duração. Portanto, até aqui, segundo SMITH et al. (1983), os resultados mostram que a taxa de elocução influencia a variabilidade das medidas de duração, o que vai na direção da hipótese do artefato estatístico, mas também que as medidas de duração, entre as três taxas de

elocução, diminuem em função da idade, o que vai de encontro à hipótese da maturação neuromotora.

Na comparação entre os sujeitos de um mesmo grupo etário, para a variabilidade relativa, aferida por meio dos coeficientes de variação[4], foram encontradas evidências favoráveis e contrárias à hipótese do artefato estatístico, dependendo do tamanho da unidade que serviu de base para a comparação entre os sujeitos. SMITH et al. (1983) constataram que a variabilidade do trecho delimitado dentro da sentença aumentava conforme suas durações aumentavam da taxa de elocução rápida para a taxa lenta. Essa relação entre duração e variabilidade seria predita pela forma mais simples da hipótese do artefato estatístico. Entretanto, quando os sujeitos de cada grupo etário foram comparados entre si para os coeficientes de variação dos dois trechos delimitados dentro das sílabas, os valores eram menores, para os quatro grupos, na taxa de elocução normal, e muito mais variáveis nas taxas de elocução lenta e rápida. Os autores limitaram-se a dizer que esse achado contradiz as predições da hipótese do artefato estatístico, uma vez que, na taxa de elocução rápida, seriam esperados valores menores tanto de duração quanto de variabilidade, devido a uma maior precisão da produção da fala. Não questionaram sequer a possibilidade de os sujeitos, principalmente as crianças, terem tido dificuldade em realizar as diferentes taxas de elocução e muito menos levantaram hipóteses que explicassem esses resultados.

---

[4] Note que o coeficiente de variação, ou variação relativa, é usado porque, apesar de tratar-se de uma comparação de sujeitos dentro de um mesmo grupo, ela é feita para suas produções em diferentes taxas de elocução, donde a necessidade de utilizar uma medida de normalização.

Na tentativa de explicar as diferenças entre a variabilidade das durações do trecho delimitado dentro da sentença e dos dois trechos delimitados dentro das sílabas, em função da taxa de elocução, SMITH et al. (1983) apelaram superficialmente para ALLEN[5] (1973 apud SMITH et al.,1983), propondo que dois relógios afetam o controle do *timing* da fala, um deles responsável pela especificação das durações segmentais e, o outro, pela taxa de elocução ou pelo *timing* de unidades maiores. No máximo, os autores sugerem que uma interação complexa entre fatores maturacionais e vieses estatísticos pode ser responsável pela maior variabilidade normalmente observada na fala infantil.

Uma outra tentativa de verificar de que forma mudanças na duração afetam a variabilidade do *timing* da fala infantil pode ser encontrada em CHERMAK e SCHNEIDERMAN (1986), que contrastaram duração e variabilidade para palavras e sentenças, produzidas por crianças de 7 anos, adolescentes de 13 anos e adultos, cada grupo formado por cinco sujeitos, para quinze repetições de uma sentença (*Bob hit the big dog*), enunciada dentro de um conjunto de três outras sentenças (*Bob caught a bad cold, Bob lost the green kit* e *Bob fed the black cat*), em duas taxas de elocução, definidas por conversacional e rápida. As sentenças foram lidas ainda sob cinco condições de acento enfático, correspondentes à colocação do acento da sentença sobre cada uma de suas quatro palavras de conteúdo. Entretanto, apenas a sentença produzida no estilo de fala designado

[5] ALLEN (1973) propõe que, na fala, há dois tipos de controle de tempo. O controle de tempo global é aquilo que normalmente é referido como taxa de elocução ou tempo, ou seja, uma taxa média na qual sílabas, palavras e frases *(phrases, no original)* são enunciadas. O controle de tempo local refere-se à especificação de durações segmentais dentro de sílabas e, segundo o autor, possivelmente à especificação de durações silábicas dentro de frases (*phrases*) rítmicas. Esses dois níveis de controle de tempo não são, de acordo com ALLEN, totalmente independentes, mas há determinadas condições nas quais ocorre uma aproximação da condição de independência entre eles, quando então eles podem ser considerados separadamente e pode-se determinar aspectos dos processos de controle subjacentes a cada um.

por neutro, no qual não era usado acento enfático sobre nenhuma palavra, serviu ao propósito dos autores de avaliar a relação entre taxa de elocução (duração) e variabilidade.

A análise dos dados (durações médias em valores absolutos, desvios-padrão e coeficientes de variação, para cada grupo) revelou que as crianças falaram mais lentamente e apresentaram maior variabilidade que os adultos e adolescentes, tanto para cada uma das quatro palavras da sentença, como para a sentença como um todo, nas duas taxas de elocução. A variabilidade relativa para as durações das palavras e da sentença diferiram significativamente em função da idade. A variabilidade das crianças foi maior que a dos adolescentes e adultos, tanto na análise entre os sujeitos de cada grupo quanto na análise entre grupos.

As comparações dentro de cada grupo, para as durações das palavras, nas duas taxas de elocução, revelaram um aumento na variabilidade absoluta (desvios-padrão) e relativa (coeficientes de variação), conforme as durações das palavras e da sentença aumentavam. Foram encontradas correlações significativas altas ($r = 0.90$) entre a variabilidade e a duração de palavras e sentenças, para crianças e adolescentes, o que explicou 80% da variância. Poucas correlações foram significativas para os adultos.

Os resultados de CHERMAK e SCHNEIDERMAN (1986) coadunam-se com os de SMITH et al. (1983) ao indicar que nenhuma das duas hipóteses, maturação neuromotora ou artefato estatístico, pode explicar com exclusividade as diferenças quanto à variabilidade do *timing* na produção da fala de crianças e adultos, uma vez que, também para CHERMAK e SCHNEIDERMAN (op. cit.), vieses estatísticos e fatores maturacionais contribuem para a maior variabilidade nas crianças. Forneceram sustentação à hipótese do artefato estatístico:

(1) as comparações dentro de cada grupo para as durações das palavras e da sentença, nas duas taxas de elocução, que revelaram um aumento na variabilidade conforme a duração aumentava[6] e (2) as fortes correlações positivas entre variabilidade e duração, explicando aproximadamente 80% da variação. Forneceram sustentação à hipótese da maturação neuromotora: (1) as comparações entre grupos quando a duração era comparável para todos (palavra *dog*), as quais mostraram que havia uma variabilidade significativamente maior para as crianças que para os adultos ou adolescentes[7] e (2) a constatação de que 20% da variação não foram explicados pela correlação entre variabilidade e duração.

SMITH (1992) deu prosseguimento à tentativa de compreender se duração e variabilidade poderiam ser vistas como indicadores separados do desenvolvimento neuromotor da fala ou se a variabilidade era basicamente um epifenômeno, ou seja, uma função secundária da duração. Segundo o autor, o objetivo era verificar se duração e variabilidade poderiam ser utilizadas como indicadores diferentes do controle motor da fala pelas crianças. Note-se uma preocupação prematura com a aplicação, sem uma reflexão maior sobre os dados.

Para isso, ele refez a análise de dados apresentados anteriormente em dois outros estudos. Do primeiro deles (SMITH, 1978 apud SMITH, 1992), foram utilizados os resultados da comparação entre dois grupos de crianças, das faixas etárias de 2 e 4 anos, com dez sujeitos em cada grupo, para as durações dos monossílabos *Bob* e *tot* (para *taught*) e de suas vogais, repetidos dez vezes por cada um dos sujeitos. Do segundo estudo (SMITH

[6] A hipótese da maturação neuromotora prevê um valor constante da variabilidade relativa entre as diferentes taxas de elocução (CHERMACK e SCHNEIDERMAN, 1986).

et al., 1983 apud SMITH, 1992), resenhado acima, o autor utilizou os resultados da comparação entre grupos de crianças de 5, 7 e 9 anos, com cinco sujeitos em cada grupo, para a sentença *My puppy ate the cookie today*, repetida dez vezes por cada sujeito, numa taxa de elocução normal. Da sentença foram extraídas as durações das sílabas acentuadas, de *puppy* e de *cookie*, e de um trecho, da soltura do /p/ inicial de *puppy* à soltura do /t/ de *today*.

SMITH (1992) buscou correlações entre as durações médias para as várias unidades (segmentos, sílabas e palavras), extraídas dos *corpora* dos dois estudos considerados em conjunto, e a variabilidade a elas associada, aferida por meio do desvio-padrão. Para as vinte crianças dos grupos de 2 e 4 anos, comparadas primeiramente para as durações das vogais de *Bob* e *tot*, não foram encontradas correlações significativas entre as durações das vogais e sua variabilidade. Para as quinze crianças dos grupos etários de 5, 7 e 9 anos, comparadas para as durações das duas sílabas acentuadas de *puppy* e de *cookie* e para o trecho definido dentro da sentença (*My puppy ate the cookie today*), a única correlação significativa entre duração e variabilidade foi obtida para a sílaba acentuada de *cookie*. SMITH (op. cit.) concluiu então que a relação entre duração e variabilidade na fala infantil deve ser relativamente modesta. Não houve modificação dos resultados quando se utilizou a medida de variabilidade relativa, o coeficiente de variação.

Segundo o autor, uma explicação para a falta de correlações entre duração e variabilidade pode estar nas diferenças do controle motor de fala pelas crianças, o que se torna evidente quando suas atuações são observadas individualmente, dentro de cada grupo.

---

[7] A hipótese do artefato estatístico prevê uma variabilidade comparável em durações comparáveis

No grupo das crianças de 2 anos, a criança que apresentou a menor duração média para a vogal de *Bob* obteve um dos maiores valores de desvio-padrão, enquanto uma outra criança que apresentou uma duração média quase tão baixa obteve o menor valor de desvio-padrão de todo o grupo. O mesmo tipo de discrepância entre duração e variabilidade foi observado no grupo das crianças de 4 anos, no qual as duas crianças que apresentaram respectivamente a menor e a maior duração média para a vogal de *Bob* apresentaram também os menores valores de desvio-padrão de todo o grupo. Achados similares foram ainda obtidos para as produções da vogal de *tot*, por esses dois grupos etários, e também para a sílaba acentuada de *puppy* e para o trecho medido dentro da sentença, para as crianças de 5, 7 e 9 anos. O autor ressalta então que as novas análises das atuações dos sujeitos de cada grupo, considerados individualmente, mostraram que a relação entre duração e variabilidade não é nem tão forte, nem tão clara como havia sido suposto.

Fiel ao seu objetivo de buscar um indicador do desenvolvimento, SMITH (op. cit.) tentou determinar para qual medida, duração ou variabilidade, as crianças estavam mais próximas da produção adulta, utilizando para cada conjunto de dados a comparação com o grupo controle de adultos do experimento original. Os valores dos desvios-padrão das crianças de 2 e de 4 anos, cujas durações estavam dentro do espectro de variação dos sujeitos adultos, eram bem maiores. Para as crianças de 2 e 4 anos, com durações médias da vogal de *Bob* dentro da variação dos adultos, os desvios-padrão médios eram, respectivamente, cerca de 96% e 57% maiores que o desvio-padrão médio para os dez sujeitos adultos do grupo controle. Porém, entre essas crianças, nos dois grupos etários,

---

(CHERMACK e SCHNEIDERMAN, 1986).

havia sempre uma ou outra cujos desvios-padrão médios estavam dentro do espectro de variação dos adultos. Na análise das durações da vogal da palavra *tot*, apesar de dezesseis, das vinte crianças do conjunto de 2 e 4 anos, apresentarem durações dentro do espectro de variação dos adultos, o desvio-padrão médio era cerca de 65% maior que o dos adultos. Entre as dezesseis crianças, apenas cinco, pertencentes ao grupo de 4 anos, apresentaram desvios-padrão dentro da variação dos adultos.

A comparação entre as quinze crianças combinadas dos grupos de 5, 7 e 9 anos e o grupo controle de adultos, para as durações da sílaba acentuada de *puppy*, mostrou que sete delas, cinco de 9 anos e duas de 7 anos, apresentaram durações dentro da variação dos adultos, apesar de apenas uma delas, de 9 anos, ter apresentado variabilidade dentro do padrão adulto. Como o experimento original havia também utilizado modificações da taxa de elocução, as mesmas comparações foram feitas para a taxa de elocução rápida, quando se encontrou que treze dos quinze sujeitos, exceção para duas crianças de 5 anos, produziram durações para a sílaba acentuada de *puppy* semelhantes ou mais curtas que os adultos. Contudo, apenas uma criança de 9 anos, a qual apresentou uma duração média menor que a dos adultos enquanto grupo, alcançou o limite superior do espectro de variabilidade deles.

Por fim, SMITH (1992) comparou os valores de duração, desvio-padrão e coeficiente de variação para os grupos de crianças e cada um dos respectivos grupos de controle (crianças de 2 e 4 anos e adultos, para a duração da vogal de *Bob*, e crianças de 5, 7, 9 anos e adultos, para a duração da sílaba acentuada de *puppy*), e observou que, em todas as idades, os valores de duração estavam sempre mais próximos aos dos adultos, o que não

acontecia com os valores de desvio-padrão (variabilidade absoluta) e coeficiente de variação (variabilidade relativa).

SMITH (1992) concluiu então que as medidas de duração e variabilidade não alcançam simultaneamente o nível de produção da fala adulta. Devido ao padrão típico encontrado na literatura, segundo o qual os sujeitos mais velhos apresentam valores menores de duração e variabilidade na comparação aos sujeitos mais novos, as medidas, segundo o autor, têm sido entendidas como tendo desenvolvimentos paralelos e fornecendo informações comparáveis sobre o desenvolvimento motor de fala das crianças. Enquanto a análise original dos dados mostrou uma tendência geral para a diminuição dos valores de duração e variabilidade em função da idade, a nova análise apontou não apenas para uma modesta correlação entre duração e variabilidade, mas também para divergências entre ambas, o que pode ser visto, de acordo com SMITH (op.cit.), na atuação individual das crianças em cada grupo, sugerindo que é necessária uma certa cautela em considerar duração e variabilidade como indicadores equivalentes da maturação do controle motor da fala pelas crianças e como medidas intimamente relacionadas.

No que concerne ao controle motor da fala pelas crianças, SMITH (1992) questiona se uma das duas medidas poderia ser considerada um melhor indicador para avaliar o desenvolvimento, uma vez que duração e variabilidade talvez reflitam aspectos diferentes daquele. Para o autor, a confirmação de que a duração atinge o padrão adulto de variação antes da variabilidade implica verificar que fator explica cronogramas de maturação aparentemente separados para essas duas variáveis e o que a duração, em comparação com a variabilidade, revela sobre o desenvolvimento neuromotor da fala pela criança. SMITH (op.

cit.) especula até que ponto a duração não seria um reflexo mais direto de habilidades de um nível mais baixo da produção da fala e da taxa com que os gestos articulatórios são produzidos, enquanto possivelmente a variabilidade estaria mais intimamente relacionada a habilidades organizacionais de níveis mais altos, as quais o falante possui para executar os gestos.

Partindo da possibilidade de que duração e variabilidade poderiam ser parâmetros relativamente independentes, que fornecem informações diferentes sobre o desenvolvimento do controle motor de fala, SMITH (1994) empreendeu uma tentativa de dar sustentação à hipótese da maturação neuromotora.

Para isso, ele propôs novamente modificações nas condições de produção de fala, sob a alegação de que essa seria uma forma de aumentar a variabilidade da fala dos sujeitos em relação a seus padrões normais. O experimento utilizou grupos de crianças de 5, 8 e 11 anos e um grupo de adultos, com cinco sujeitos em cada grupo. A partir de estímulos visuais, os sujeitos produziram oito palavras CVC (*peak*, *pack*, *beack*, *back*, *fin*, *fan*, *mean*, *man*), dentro de uma frase-veículo (*He saw ___ again*), em três condições experimentais, fala nas taxas de elocução normal e rápida, e fala com *bite block* (ou seja, segurando um "bloquinho" entre os molares), e cinco repetições dos oito estímulos em cada condição. A opção por usar pares mínimos visava permitir ainda a comparação entre os sujeitos para diferenças de duração intrínseca (por exemplo, vogais altas x vogais baixas) e a variabilidade a elas associadas. Entre os estudos examinados até aqui, essa foi a primeira e única tentativa, ainda que muito acanhada, de incorporar uma variável fonética a um *corpus* que visa estudar a relação entre duração e variabilidade na fala infantil.

Para cada sujeito, foram calculadas, em cada uma das três condições de fala, correlações entre as médias das medidas de duração e os desvios-padrão obtidos para cada segmento acústico medido. Num total de sessenta correlações, apenas seis foram estatisticamente significativas. Cinco dessas correlações foram obtidas para as crianças de 5 anos, duas em cada uma das taxas de elocução normal e rápida e uma na condição de fala com *bite block*. Apenas uma correlação significativa entre duração e variabilidade foi encontrada para um adulto e isso ocorreu na condição de *bite block*. Essas seis correlações explicaram menos da metade da variância dos dados. SMITH (op. cit.) concluiu que as poucas correlações significativas não manifestaram uma forte relação entre duração média e desvio-padrão (variabilidade).

Foram feitas então comparações entre os grupos etários para os seguintes segmentos de *He saw peak again*: [s], [a], [p], a aspiração ([^{h}]) de [p], [i] e [k], utilizando a sentença produzida na taxa de elocução normal. Dentro de cada grupo, a unidade mais curta foi a aspiração ([^{h}]) de [p], a qual variou de 53 a 71 ms, e os segmentos mais longos foram o [s], para os adultos e crianças mais velhas, e o [a], para as crianças de 5 anos, os quais variaram, respectivamente, de 133 a 155 ms. Contudo, em nenhum grupo, a menor unidade foi a menos variável, e para três dos quatro grupos, a maior unidade não foi a mais variável. De fato, em dois grupos, a unidade mais longa foi a menos variável. O mesmo foi observado na comparação entre os grupos para os segmentos [s], [a], [f], [ae] e [n] de *He saw fan again*. As durações variaram de 55 a 70 ms para [n] e de 155 a 197 ms para [f] ou [ae], sendo que novamente o menor segmento acústico não foi o menos variável, para três dos quatro grupos, nem o maior segmento foi o mais variável para dois grupos.

Quando os valores de duração e variabilidade foram comparados entre as taxas de elocução normal e rápida, houve uma diminuição das durações dos segmentos acústicos [s] e [a], pertencentes à palavra *saw* da frase-veículo, em todos os grupos, para todos os sujeitos, da taxa de elocução normal para a rápida. Os desvios-padrão também diminuíram, de 3 a 4 ms, para as crianças de 5 anos e para os adultos, na condição de taxa rápida, mas não apresentaram mudanças para os grupos de 8 e 11 anos, apesar de neles ter havido uma diminuição das durações de 25 a 30%, ou seja, as mudanças foram menores para a variabilidade que para a duração. As comparações estatísticas entre essas duas condições, para os quatro grupos de sujeitos, não foram significativas para [s] ou [a]. Para os quatro grupos de sujeitos combinados, a variabilidade dos segmentos [s] e [a] diminuiu cerca de 15% da taxa de elocução normal para a rápida, enquanto a duração relativa diminuiu quase duas vezes mais, cerca de 27%. As durações diminuíram cerca de 15% para as crianças de 5 anos, 25% para as de 8 anos, 30% para as de 11 anos e 35% para os adultos. Na fala com *bite block*, não foram encontradas diferenças sistemáticas entre os grupos nem para as durações médias nem para os desvios-padrão de [s] e [a], em comparação com a fala produzida sem *bite block* e na taxa de elocução normal (condição controle).

Duração e variabilidade foram ainda examinadas em relação às diferenças de duração intrínseca das vogais do *corpus*. As durações médias das vogais baixas [a] e [ae] foram divididas pelas durações médias da vogal alta [i], em cada grupo etário, quando se observou que as vogais baixas foram de 15% a 75% mais longas que a vogal alta [i], nos diferentes grupos. A diferença de variabilidade entre as vogais baixas em relação à vogal alta não foi tão grande quanto a diferença de duração e nenhum grupo apresentou mais que 20% de

variação para as vogais baixas em comparação com a vogal alta, com exceção do grupo das crianças de 5 anos, que apresentou 48% a mais de variabilidade para a vogal baixa [a] que para a vogal alta [i]. Apesar de, na média, entre os quatro grupos de sujeitos, a vogal [i] ter sido cerca de 43% mais curta que as vogais baixas, sua variabilidade foi apenas 13% menor que a daquelas vogais. Para SMITH (op. cit.), houve uma disparidade substancial entre os efeitos observados para duração e desvio-padrão a partir das comparações para as durações intrínsecas, uma vez que às maiores durações das vogais baixas não estavam associados aumentos iguais de variabilidade.

SMITH (1994) afirma que apesar de as diferenças entre duração e variabilidade não estarem claras, no processo de desenvolvimento, a variabilidade parece alcançar os níveis adultos mais tarde que a duração. Ele usou uma medida relativa, na qual foram divididos os valores das durações médias e dos desvios-padrão das quarenta repetições de [s] e [a], de cada grupo de crianças, pelas respectivas durações e desvios-padrão dos adultos, nas taxas de elocução normal e rápida, e fala com *bite block*. Para os três grupos de crianças e nas três condições de fala, as durações relativas de [s] e [a] estavam sempre mais próximas dos níveis do grupo adulto que a variabilidade. Para todos os três grupos de crianças e para os dois segmentos combinados nas três condições de fala, a duração foi aproximadamente 26% maior que o valor médio dos adultos, enquanto a variabilidade foi cerca de 64%.

Para SMITH (1994), de um modo geral, os achados do estudo enfraqueceram de modo substancial as afirmações de que existe uma forte ligação estatística entre duração e variabilidade e o fato de duração e variabilidade serem basicamente independentes uma da outra forneceu sustentação para a hipótese da maturação neuromotora. O autor coloca que

os impulsos nervosos são transmitidos mais lentamente por meio de neurônios de menor diâmetro, assim como por meio de um sistema nervoso ainda não totalmente mielinizado, e que o processo de mielinização não se completa antes da puberdade. Entretanto, segundo ele, o que não está claro é se o sistema nervoso menos maduro é necessariamente mais variável no processamento das mensagens sensoriomotoras simplesmente por ser mais lento. Ele diz que duração e variabilidade talvez não estejam intimamente relacionadas para funções neurofisiológicas básicas, tais como a velocidade de condução do nervo.

Um ponto por ele levantado, e que a nosso ver é importante na compreensão das diferenças entre as duas crianças que fazem parte de nosso estudo, refere-se ao fato de que, enquanto grupo, as crianças pequenas geralmente tendem a ser menos hábeis que os adultos (e que as crianças mais velhas) no controle de fatores necessários para produzir a fala de uma forma consistente, quanto a seu aspecto temporal. Porém, na análise intra-grupo, é possível encontrar crianças pequenas que apresentam valores das medidas de duração e variabilidade equiparáveis àqueles dos adultos. Segundo o autor, isso pode ser devido, pelo menos em parte, ao fato de que as habilidades motoras desenvolvem-se de modo exponencial. Assim, duas crianças da mesma idade podem diferir de modo considerável em relação às suas atuações motoras porque uma delas começou seu processo de mudança exponencial um pouco antes da outra. Dessa forma, algumas crianças podem ser mais hábeis na produção da fala que outras da mesma idade, simplesmente devido a diferenças fisiológicas no desenvolvimento, ou seja, porque individualmente as crianças amadurecem em diferentes taxas.

SMITH (op. cit.) recorre a SHARKEY e FOLKINS (1985 apud SMITH, 1994) que propuseram que a maior variabilidade nas crianças pequenas poderia ainda ser resultado de sua maior flexibilidade ou de uma atuação menos automatizada que a apresentada pelos adultos. Na medida em que fatores desse tipo podem também ser responsáveis pela variabilidade individual, SMITH (1994) coloca que duas crianças da mesma idade e com sistemas nervosos comparáveis poderiam ainda manifestar diferentes graus de variabilidade se uma fosse mais inclinada, que a outra, a explorar as capacidades de seu trato vocal.

1.2 Falhas metodológicas da literatura

Um *corpus* elaborado para o estudo da duração não pode perder de vista que ela constitui um parâmetro acústico bastante suscetível a uma variedade de fatores lingüísticos, uma vez que ela serve para exprimir e demarcar acento (ALBANO, 1996), lexical ou nas fronteiras inicial (FOUGERON e KEATING, 1997) e final (KLATT, 1976) de um enunciado, ou ainda, entre unidades menores dentro dele (BECKMAN e EDWARDS, 1990). Uma vez que, no limite, as fronteiras de palavras ou unidades maiores (sintagmas) correspondem a fronteiras sintáticas e/ou prosódicas, as medidas de duração podem refletir de forma indireta, uma vez que não há isomorfismo entre os vários níveis lingüísticos, a relação entre subsistemas hierarquicamente mais altos ou centrais e a realização fonética (NESPOR e VOGEL, 1986, SCARPA, 1997).

Ao desconsiderar a influência de fatores lingüísticos sobre a duração segmental, observa-se que, como regra, os estudos deixam de adotar critérios lingüísticos na elaboração

dos *corpora* usados para obter as medidas de duração. Apenas casualmente a possível influência de fatores lingüísticos sobre as medidas de duração e de variabilidade é apontada como algo a ser melhor investigado (KENT e FORNER, 1980, SMITH et al., 1983).

No que concerne à possibilidade de fatores lingüísticos intervirem nas diferenças entre as medidas de duração e de variabilidade para a fala adulta e infantil, a única exceção encontrada na literatura refere-se ao estudo de SMITH (1994), o qual encontrou que as unidades intrinsecamente mais curtas de duas sentenças (a aspiração [^{h}] de [p] em *He saw peak again* e [n] em *He saw fan again*) não foram as unidades menos variáveis e que as unidades intrinsecamente mais longas ([s] e [a] de *saw*, para a primeira sentença, e [f] e [ae] de *fan*, para a segunda) também não apresentaram a maior variação. Entretanto, esse caso constitui-se num exemplo do quanto fatores lingüísticos, como as fronteiras sintáticas, podem interferir em fatores fonéticos, como a duração intrínseca e, conseqüentemente, desvirtuar as conclusões tiradas sobre a relação entre duração e variabilidade. A questão que surge em relação a esses dados refere-se ao padrão de acento das sentenças e à possível influência desse padrão sobre a variabilidade das durações intrínsecas dos segmentos analisados. É possível pensar que a aspiração de [p], em *He saw peak again*, mesmo sendo, a princípio, a duração intrinsecamente mais curta entre os segmentos analisados na sentença, ao pertencer a uma sílaba acentuada e estar numa fronteira dentro da frase-veículo, torna-se sujeita a receber acento e, portanto, a apresentar maior variação. A fronteira pode surgir se o monossílabo *peak* for enfatizado em relação ao que o precedeu, pois, quando produzidas em uma frase-veículo, existe a possibilidade de as palavras-alvo serem enfatizadas, uma vez que elas constituem o que há de diferente entre as repetições. Assim, a posição de acento

lexical e a fronteira sintática podem ter feito com que a duração da aspiração de [p] ficasse mais variável, uma vez que, de acordo com BARBOSA e BAILLY (1994), numa reinterpretação de FRAISSE (1974 apud BARBOSA e BAILLY, 1994), as sílabas acentuadas são em si mesmas os locais onde há maior possibilidade de variação, ao contrário das sílabas não-acentuadas, porque essas constituem uma espécie de referência para a realização do ritmo. Da mesma forma, pode-se supor que a constância de "*He saw*", entre as várias repetições, poderia transformá-la numa espécie de bloco, tornando-a pouco sujeita a variações, e disso decorre a menor variabilidade para as durações intrinsecamente longas de [s] e [a].

Em *He saw fan again*, a questão que se nos coloca é até que ponto foi possível medir o [n] de *fan* separadamente do [a] de *again*, uma vez que tanto a ressilabificação (CLEMENTS e KEYSER, 1983) (*fa.na.gain*) como a introdução de uma pausa entre *fan* (monossílabo) e *again* (porque é elemento da frase-veículo que se repete entre as sentenças) seriam possíveis.

Um outro aspecto que chama atenção é que, ao lidar com medidas de duração, os estudos comparam os sujeitos para segmentos acústicos ou seqüências de segmentos acústicos que eles fazem corresponder a fonemas, sílabas, palavras ou sintagmas, delimitados dentro das sentenças. Contudo, não há por trás disso um modelo de análise lingüística. Já foi dito que não há correspondência estrita entre as unidades lingüísticas e os segmentos acústicos do sinal de fala (FANT, 1973). Os estudos acústicos estabelecem previamente critérios para a realização das medidas de duração de segmentos acústicos, com base na forma da onda e/ou no espectrograma de banda larga. Entre os estudos resenhados,

encontramos apenas uma sugestão de que a atuação de adultos e crianças, no que tange às durações de seqüências de segmentos acústicos de diferentes tamanhos, pode ser explicada por um modelo do ritmo. Trata-se da proposta, de SMITH et al. (1983), de que dois relógios, um responsável pelas durações segmentais e outro pela taxa de elocução ou pelo *timing* de unidades maiores, podem afetar o controle do *timing* da fala (ALLEN, 1973 apud SMITH et al., 1983). Nada, porém, é dito sobre as distinções de funcionamento dos dois relógios na fala adulta e na fala infantil.

Outro problema dos estudos resenhados diz respeito à adequação da tarefa experimental proposta para ser realizada por crianças. Uma análise de suas metodologias faz-nos cogitar se determinadas exigências podem efetivamente ser cumpridas por crianças, tais como, mudanças da taxa de elocução, fala com *bite block*, leitura com acento enfático em determinadas palavras, e até mesmo o volume da tarefa experimental que, em CHERMAK e SCHNEIDERMAN (1986) e em SMITH (1994), chegou a 120 sentenças por sujeito, demonstrando a "cegueira" dos autores provocada pela busca de indicadores do desenvolvimento da fala.

Um exemplo das dificuldades envolvidas na modificação da taxa de elocução ilustra bem o quanto a tarefa experimental pode transformar-se numa situação completamente sem sentido, tanto para crianças, como para adultos. Foi o que ocorreu no experimento de SMITH et al. (1983), com as absurdas instruções fornecidas aos sujeitos, na tentativa de eliciar as condições de taxa rápida e lenta. Na taxa de elocução rápida, os sujeitos eram instruídos a produzir a sentença-estímulo numa taxa mais rápida que a normal, com a ressalva de que eles não deixassem nada de fora, provavelmente, no sentido de não deixar de

pronunciar nada, o que de modo intuitivo parece-nos o mais provável de ocorrer, principalmente em se tratando de crianças pequenas. Na taxa de elocução lenta, a instrução dada aos sujeitos foi ainda mais despropositada, uma vez que eles não deveriam pronunciar cada palavra da sentença separadamente ou introduzir pausas durante sua produção, mas poderiam pensar em como falariam se estivessem tristes. Ora, a variação das condições experimentais de fala, no que se refere a crianças pequenas, é algo extremamente questionável e perigoso (MURRAY E ARNOTT, 1993), o que deve levar a resultados como os de SMITH (1994), em que cinco das seis correlações obtidas entre duração e variabilidade foram encontradas para as crianças de 5 anos. É interessante notar ainda que a sexta correlação, obtida para um adulto, ocorreu na condição de fala com *bite block*.

No caso do experimento de CHERMAK e SCHNEIDERMAN (1986), o absurdo com relação à tarefa experimental foi o fato de empregar leitura para crianças da faixa etária de 7 anos que ainda estão numa fase inicial do processo de alfabetização, o que vale aos autores a crítica de que a fala mais lenta do grupo mais novo pode ser conseqüência de as crianças que o compõem não serem leitores fluentes. Como se isso não bastasse, as sentenças foram lidas sob cinco condições diferentes de acento enfático, em duas taxas de elocução (conversacional e rápida). Não surpreende que apenas a sentença lida sem nenhum acento enfático e na taxa de elocução normal tenha se prestado à análise.

Um aspecto da descrição dos experimentos muito significativo é a falta de informação a respeito do modelo, oferecido aos sujeitos, para repetição. Uma questão por nós levantada desde os primeiros contatos com essa literatura, foi a seguinte: até que ponto o modelo para repetição pode influenciar nas diferenças de produção observadas entre os grupos etários,

pois, uma vez que não há nenhum controle ou análise do modelo, não fica claro se, quando oferecido às crianças pequenas, ele pode, por algum motivo, ser pronunciado numa taxa de elocução mais lenta e, caso isso ocorra, se há ou não interferência na produção das crianças. É possível imaginar as dificuldades em se obter repetições, pelas crianças pequenas, de um ou mais enunciados, dificuldades provenientes de fatores como o tempo de atenção e a motivação das crianças para a realização da tarefa experimental. Porém, no geral, a única informação trazida pelos experimentos analisados é que o modelo para repetição foi fornecido por um dos pesquisadores, a todos os grupos de sujeitos.

Na literatura também não são encontrados esclarecimentos a respeito do predomínio de estudos transversais, em relação aos estudos longitudinais. Talvez os primeiros estudos acústicos da fala infantil (KENT, 1976) tenham simplesmente repetido o que parecia ser um padrão, até então, entre os estudos acústicos da fala adulta. Os estudos subseqüentes, especialmente aqueles dedicados às medidas de duração, podem ter adotado o método transversal devido ao fato de KENT e FORNER (1980) terem tomado como ponto de partida para a investigação da maior variabilidade temporal da fala infantil a hipótese da maturação neuromotora, que pressupunha uma diminuição dos valores de duração e variabilidade em função do aumento da idade.

Na verdade, o predomínio de estudos transversais parece advir da necessidade inicial, em fonética acústica (LEHISTE, 1967), de estabelecer padrões acústicos entre grupos que diferiam de acordo com sexo e idade, numa linha de pesquisa mais descritiva. Esses padrões somente podem ser obtidos por meio do tratamento estatístico ao qual as medidas acústicas são submetidas, o que implica na necessidade de várias repetições do *corpus* pelos sujeitos e

de vários sujeitos dentro de cada grupo. Com isso, o sentido da análise acústica decorre do emprego dos testes estatísticos, os quais tornam possível falar de efeitos ou padrões significativos que emergem das medidas acústicas.

A vantagem de estudos transversais, no estudo da fala infantil, reside no fato de que, apesar de todas as variáveis que podem intervir numa tarefa como a de repetição de sentenças, ainda assim restará um material minimamente comum, para o qual adultos e crianças possam ser comparados (CORRÊA, 1996). Uma análise dos poucos estudos acústicos de dados colhidos longitudinalmente (KUBASKA e KEATING, 1981, MACK e LIEBERMAN, 1985, GOODELL e STUDDERT-KENNEDY, 1993) mostrou a dificuldade de lidar com o problema da falta de controle do contexto fonético e prosódico no qual os itens lexicais apareceram. No entanto, acreditamos que um dos interesses no estudo acústico de dados provenientes de uma coleta longitudinal é justamente acompanhar, por meio das diferentes medidas (duração, f0 e valores de formantes), o percurso das várias formas de um determinado item lexical, até que o mesmo atinja a forma mais próxima da produção adulta (GAMA-ROSSI, 1994).

1.3 Resultados de um estudo piloto e uma hipótese alternativa às hipóteses da maturação neuromotora e do artefato estatístico

Seguindo KENT e FORNER (1980), realizamos um estudo piloto (GAMA-ROSSI e ALBANO, 1996), no qual comparamos um grupo de quatro adultos a um grupo de quatro crianças, com idade média de 4 anos e 5 meses, para as durações dos segmentos acústicos, consonantais e vocálicos, de cinco sentenças simples (Eu gosto de banana, O sapato é novo,

O fusca do João é azul, O elefante é grande, Tem bolo na geladeira), repetidas pelos sujeitos a partir de modelo oferecido pela pesquisadora. Uma análise quantitativa mostrou que, para a maioria dos segmentos acústicos, não houve diferenças significativas entre adultos e crianças. As poucas diferenças que emergiram residiam em segmentos acústicos localizados em posição não-acentuada, seguindo imediatamente um acento lexical ou localizados entre dois acentos lexicais, como foi o caso de alguns clíticos.

Foi realizada ainda uma análise prosódica qualitativa dos dados, na qual construímos contornos de duração por meio de gráficos, que mostravam, para cada grupo, numa sucessão temporal, as durações médias das sílabas de cada sentença, com seus respectivos desvios-padrão. Os gráficos forneciam uma boa estimativa empírica da distribuição das proeminências relativas dentro de cada sentença (ALBANO, 1996).

De acordo com a análise intuitiva de CÂMARA (1969), os contornos acentuais do PB mostram uma queda brusca de proeminência imediatamente após o acento lexical. Assim, em seus exemplos originais, as sílabas não-acentuadas, que se seguem imediatamente à sílaba acentuada, recebem um grau de acento 0 (zero), tanto na palavra fonológica como no grupo de força. Relembremos os exemplos:

Graus de Acento	1	1	3	0
Palavra Fonológica	bra	si	lei	ro

onde: 3 = sílaba acentuada
1= sílabas não-acentuadas que precedem ao acento lexical
0 = sílabas não-acentuadas que se seguem ao acento lexical

Graus de Acento	2	0	1	1	3	0
Grupo de Força	gran	de	bra	si	lei	ro

onde: 3 = maior proeminência
2 = sílabas acentuadas das palavras precedentes
1= sílabas não-acentuadas que precedem ao acento lexical
0 = sílabas não-acentuadas que se seguem ao acento lexical

Conforme demonstrado por ALBANO (1996), a análise de CÂMARA (op. cit.) pode ser expressa por meio da notação da grade métrica (LIBERMAN e PRINCE, 1977), quando teríamos:

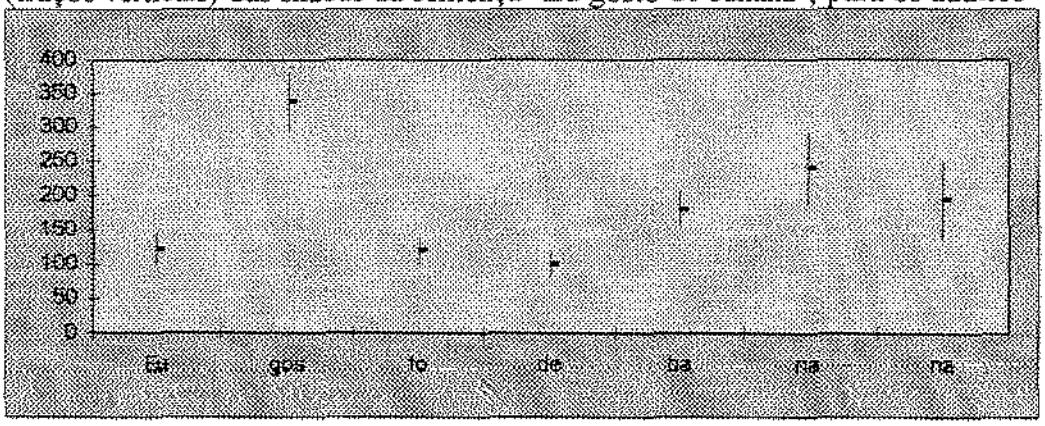
									X	
X			X		X				X	
X			X		X		X	X	X	
X	X	X	X	X	X	X	X	X	X	X
2	0	0	2	0	2	0	1	1	3	0
Câ	ma	ra	foi	um	gran	de	bra	si	lei	ro

A queda brusca da pós-tônica prevista por CÂMARA (op. cit.) ocorreu para o grupo dos adultos, ao passo que, para as crianças, a queda na duração das mesmas sílabas configurava-se numa linha menos inclinada, indicando maiores valores de duração. Vejamos um exemplo disso.

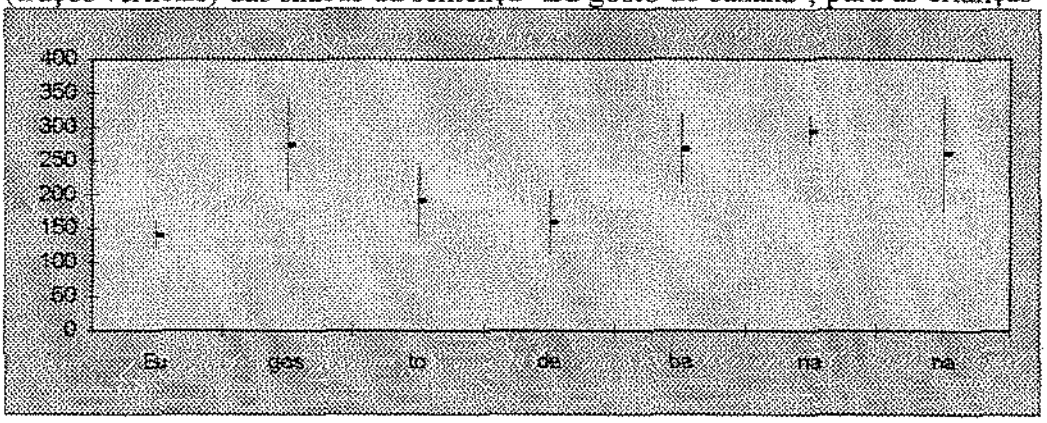
Previsões da análise de Câmara (1969) para a sentença "Eu gosto de banana", formalizadas por meio da grade métrica:

					X	
	X				X	
X	X			X	X	
X	X	X	X	X	X	X
Eu	gos	to	de	ba	na	na

Contorno de duração formado pelas durações médias (pontos) e desvios-padrão (traços verticais) das sílabas da sentença "Eu gosto de banana", para os adultos[8]



Contorno de duração formado pelas durações médias (pontos) e desvios-padrão (traços verticais) das sílabas da sentença "Eu gosto de banana", para as crianças



[8] Note que os valores de duração, apresentados no eixo vertical dos gráficos, correspondem a milésimo de segundo (ms).

Um outro aspecto importante dos contornos de duração referia-se ao fato de os adultos apresentarem desvios-padrão maiores nas sílabas tônicas e menores nas sílabas átonas, o que está de acordo com a previsão de que as sílabas não-acentuadas devem variar menos porque constituem os pontos de referência para a produção do ritmo (BARBOSA e BAILLY, 1994). As crianças, porém, mostraram desvios-padrão maiores que os dos adultos nas sílabas não-acentuadas.

Apesar da limitação dos dados do estudo piloto, começamos a suspeitar de que um fator lingüístico, a diferença na realização entre sílabas tônicas e átonas, poderia estar interagindo com a realização neuromotora dos segmentos acústicos da fala e determinando as diferenças entre adultos e crianças, observadas a partir das medidas de duração e variabilidade. Essa hipótese, se comprovada, pode fornecer o elemento que faltara aos estudos precedentes para romper com a discussão circular imposta pela hegemonia das hipóteses da maturação neuromotora e do artefato estatístico. Foi montado um novo experimento com vistas a essa terceira hipótese, o qual tentou corrigir algumas das falhas metodológicas apontadas na crítica dos experimentos da literatura precedente. Passemos, então, à discussão dos critérios usados para a montagem do *corpus* e para a coleta e análise dos dados do novo experimento.

1.4 Sujeitos e situação de gravação do novo experimento

O grupo de crianças que participou das gravações pertencia ao primeiro estágio de uma escola municipal da cidade de São Paulo. As gravações foram realizadas no primeiro

semestre de 1996. A idade das crianças variava entre aquelas que já haviam completado 4 anos e aquelas que completariam até julho daquele ano.

A pesquisadora observou o grupo de crianças nos dois meses que antecederam ao início das gravações, em duas ou três visitas semanais, com duração de quatro horas, o que correspondia ao tempo de permanência das crianças na escola. As observações foram feitas em todas as situações realizadas rotineiramente no período, desde a chegada à escola, até a espera pelos pais ou perua no horário de saída. A rotina escolar era formada basicamente de: conversa com os amigos, contagem e/ou recontagem de histórias, lanche, brincadeiras de pátio e atividades pedagógicas.

Assim, enquanto permanecia na escola, a pesquisadora acompanhava a classe em todas as atividades, e, com o tempo, passou a funcionar como assistente da professora. Isso permitiu que as crianças se acostumassem com sua presença e até formassem vínculos a ela. A observação das crianças pela pesquisadora foi feita de modo informal, sem nenhum roteiro prévio. As crianças também não foram submetidas a nenhum tipo de teste padronizado, para avaliar suas falas. Com base na suposição de que o desenvolvimento da fala ocorre em meio ao desenvolvimento perceptual (auditivo e visual), motor (global e fino), cognitivo e emocional (entre outros, GONÇALVES, 1989, GAMA, 1989, ALBANO, 1990, VIHMAN, 1991, KENT, 1996), a pesquisadora ouvia as crianças em diferentes situações, conversava com elas o mais possível, observava suas atitudes quanto ao desenvolvimento motor global e fino. A pesquisadora atentava ainda para como se dava a utilização de conceitos pelas crianças e também para aspectos de suas condutas que poderiam servir de referência para uma caracterização mais geral de seus perfis emocionais. O principal objetivo das

observações era delimitar um grupo de crianças relativamente homogêneo com relação ao conjunto de suas aquisições e com um desenvolvimento geral e de fala compatíveis com a idade de 4 anos. O fato de as crianças pertencerem à mesma escola serviu como um parâmetro para delimitar as características sociais, econômicas e culturais do grupo.

A pesquisadora participou da primeira reunião de pais, no início do semestre, na qual explicou em linhas gerais os objetivos do trabalho e os procedimentos que seriam usados nas gravações. Como forma de retribuir a participação das crianças e da professora no trabalho, a pesquisadora realizou posteriormente uma palestra aos pais sobre o desenvolvimento da motricidade oral e da fala, ressaltando aspectos preventivos. Além disso, todas as crianças da classe que apresentavam processos de simplificação dos sons da fala (YAVAS, 1988, HERNANDORENA, 1990, LAMPRECHT, 1992, TEIXEIRA, 1996), os quais não seriam mais esperados para a idade em questão, foram encaminhadas, por meio de carta, a pelo menos três locais de atendimento fonoaudiológico.

O grupo de adultos foi composto pelo pessoal técnico da escola (coordenadora, professoras, aluna de pedagogia estagiária), formado exclusivamente por mulheres. No total, foram gravadas sete crianças, as quais realizaram o mínimo de três repetições do *corpus* do experimento, e duas crianças que realizaram apenas duas repetições. Com relação ao grupo de adultos, foram gravados nove sujeitos.

As gravações foram realizadas na sala do almoxarifado da escola. Previamente, foram feitos testes, somente com a fala da pesquisadora, nas salas que estariam disponíveis para as gravações, com vistas à acústica e ao nível de ruído das mesmas. Optou-se pela sala do almoxarifado, a qual por ter suas paredes cobertas por estantes repletas de materiais

didáticos, não apresentava eco. Além disso, a sala possuía vidraças, próximas ao teto, em apenas uma das paredes, e as vidraças davam dentro de um parque, o qual fazia limite com a parte de trás da escola. Pode-se dizer, então, que as gravações foram feitas numa sala sem tratamento acústico, mas que era a mais silenciosa, disponível na escola. Foram colocadas quatro cadeiras e uma mesa baixa no centro da sala para a realização das gravações.

As gravações ocorreram ao longo de duas semanas, em uma sessão para cada sujeito. O tempo de duração das sessões foi de cerca de 15 minutos para as crianças e em torno de 10 minutos para os adultos, excluindo-se o tempo de conversa informal que precedia ao início da gravação propriamente dita, quando os técnicos de um estúdio profissional, encarregados das gravações, ajustavam o volume para a voz do sujeito e sua proximidade do microfone.

Muitas dificuldades surgiram durante a gravação das crianças. Entre elas, estava o fato de que algumas delas tornavam-se encabuladas quando chegavam à sala, trazidas pela pesquisadora, e deparavam-se com os dois técnicos e o material de gravação. A situação foi contornada com alguns minutos de conversa, quando os técnicos brincavam com as crianças em relação ao uso do microfone e a pesquisadora explorava com elas os materiais e brinquedos da sala. É importante deixar claro que as crianças foram livres para participar ou não das gravações.

Durante as gravações, eventualmente uma ou mais sentenças eram pronunciadas novamente pela pesquisadora e pelo sujeito, sempre que algum ruído interferia na gravação das mesmas. Isso ocorria quando as crianças batiam seus pés nos pés da cadeira onde estavam sentadas ou da mesa em frente a elas, sobre a qual estava o microfone em um

suporte, ou quando ruídos ambientais (avião, campainha, gritos de crianças) ocorriam durante a fala da pesquisadora ou do sujeito ou ainda quando crianças e adultos iniciavam suas repetições antes que a pesquisadora tivesse terminado de dar o modelo.

1.5 Material de gravação

Para a realização das gravações foram usados: dois microfones SHURE SM 58, um para a pesquisadora, outro para o sujeito, uma mesa de som MACKIE CR 1604, de 16 canais, um gravador *dat* SONY TCD.D7 e fitas *dat* JVC, R-120XD. Foram usados também dois fones de ouvido pelos técnicos do estúdio contratado, para a monitoração das gravações. Depois de prontas, essas foram editadas em estúdio, para eliminar ruídos entre as falas da pesquisadora e dos sujeitos e para normalizar as vozes, a partir da referência de zero dB digital. As edições do material original foram então gravadas em fitas *dat* JVC, R-120XD e em fitas cromo TDK, SD, IEC II/*type* II *high position*.

Na tarefa de repetição das sentenças, foram usadas figuras coloridas (APÊNDICE) correspondentes às sentenças do *corpus*, as quais foram coladas em cartões de 21x16 cm e plastificadas. A idéia de utilizar figuras surgiu da observação prévia do grupo de crianças, quando se notou que, de um modo geral, elas mostravam-se bastante motivadas no contato com materiais gráficos. As sentenças do *corpus* acabaram, na verdade, transformando-se em descrições das figuras correspondentes, o que foi possível devido ao processo de *scanning* e edição das figuras de um livro infantil sobre circo (LUSTIG, 1980).

Na realização da tarefa de repetição, as sentenças do *corpus* passaram a ser evocadas a partir das figuras a elas associadas. A pesquisadora aprendeu a correspondência entre as sentenças do *corpus* e as figuras, e evocava-as à medida que cada figura lhe era apresentada, sendo seguida pela repetição do sujeito do experimento, adulto ou criança. Optou-se por não incluir na ilustração de cada figura a forma escrita da sentença correspondente, com o intuito de não introduzir uma diferença na realização da tarefa pelas crianças, ainda não letradas, e pelos adultos, leitores. A questão da ordem aleatória de apresentação das sentenças do *corpus* foi resolvida com o embaralhar das figuras antes de cada repetição. Vale notar que as sentenças do experimento eram pronunciadas pela pesquisadora num estilo definido como fala clara (LINDBLOM, 1990).

Quanto ao controle do modelo oferecido pela pesquisadora aos sujeitos, o qual constituía motivo de preocupação desde o contato com os experimentos dos estudos resenhados, ele foi gravado e tornou-se disponível para análise.

1.6 Critérios adotados na realização das medidas de duração

As medidas de duração foram realizadas por meio do *software CSRE 4.5* (*the Computerized Speech Research Enviroment*) (AVAAZ, 1995), instalado em um microcomputador *Pentium*, provido de uma placa de som *Sound Blaster 16*. Os dados foram digitalizados a partir de um gravador cassete, JVC TD-V1050, conectado ao computador, utilizando-se as gravações editadas e copiadas nas fitas TDK.

As medidas de duração foram feitas com base em critérios estabelecidos a partir da forma da onda e dos *formant tracks* gerados pelo *CSRE 4.5*, para o sinal de fala digitalizado. Trata-se de *formant tracks* porque o *CSRE 4.5* gera espectrogramas via análise *LPC*. Assim, o *formant track* de cada sentença era utilizado para estabelecer primeiramente os limites aproximados de cada um dos segmentos da sentença, os quais, em seguida, eram definitivamente etiquetados, por meio de uma transcrição fonética larga, na forma da onda. Os parâmetros empregados para a obtenção dos *formant tracks* das sentenças e da janela espectral foram os seguintes: técnica de estimativa espectral: *AC* (*Autoregressive Spectral Estimation Technique*); taxa de amostragem: 16 kHz; número de pontos da janela: 256 (duração da janela = 16 ms); número de bandas: 256; percentagem de sobreposição: 50-60%; tipo de janela: *hanning*; ordem do filtro *LPC* (especifica o número de pólos necessários para a análise): 18 e pré-ênfase: 98% (AVAAZ, 1995). Esses parâmetros foram estabelecidos a partir de testes feitos com dados amostrados em diferentes taxas e variando-se o tamanho da janela, para verificar qual seria a melhor relação entre taxa de amostragem e tamanho da janela. Essa verificação constou das medidas dos formantes das vogais [i], [O], [o] e [u], em diferentes posições de acento. O objetivo era avaliar em que relação taxa de amostragem/duração da janela o algoritmo apresentava um melhor desempenho em relação às medidas esperadas para os formantes. Os valores dos formantes foram ainda pareados por meio de uma análise visual de harmônicos na análise *FFT* do mesmo dado.

A etiquetagem dos segmentos acústicos de cada sentença na forma da onda produzia um arquivo texto em *MS-DOS*, no qual encontravam-se, ao lado da transcrição referente a cada segmento, os valores correspondentes ao número de pontos naquele trecho de onda

delimitado, de acordo com a taxa de amostragem previamente estabelecida. O arquivo era então transportado para uma planilha de dados (*excel 7.0*), onde os valores em milésimo de segundo (ms), referentes à duração de cada segmento, eram obtidos a partir da divisão da quantidade de pontos pela taxa de amostragem.

Com relação aos critérios utilizados para a realização das medidas de duração, o primeiro deles foi justamente delimitar o máximo de segmentos acústicos dentro de cada sentença, com o objetivo de assegurar a comparação entre sujeitos para o maior número possível de segmentos acústicos. Assim, foram consideradas todas as descontinuidades espectrais que pudessem indicar ou servir como vestígio da presença de um dado segmento acústico, uma vez que, devido a peculiaridades inerentes à fala infantil, os espectrogramas, muitas vezes, podem não apresentar segmentos acústicos tão bem definidos como os da fala adulta, principalmente aquela correspondente a um estilo de fala clara, na qual há pouca sobreposição entre os segmentos acústicos (LINDBLOM, 1990, KENT e READ, 1992).

Para a delimitação dos segmentos acústicos de cada sentença, o principal critério adotado, a partir da utilização do *CSRE 4.5*, foi o de mudanças nos pulsos da forma da onda, somado ao critério de descontinuidades espectrais. Os extremos direitos dos segmentos acústicos de uma sentença eram primeiramente estabelecidos na análise espectral, com a anotação de seus valores no eixo do tempo, para que, de posse desses, os segmentos acústicos fossem demarcados na forma da onda. No geral, os valores da análise espectral serviam para dar uma idéia aproximada dos valores que correspondiam aos locais de corte na forma da onda. Nessa, os critérios de corte ou separação entre segmentos acústicos levavam em conta o primeiro pulso visível e o último pulso regular de cada um deles, sendo

que o corte era realizado nos vales ou na depressão máxima ou centro do vale, dos pulsos, e a irregularidade da forma de um pulso era definida por meio da comparação desse aos precedentes e seguintes. Assim, os valores obtidos na análise espectral, além de fornecer uma primeira localização das extremidades de cada segmento acústico, serviam também para identificar transições, hesitações, repetições de sons ou sílabas, pausas, trechos de aspirações, fricções ou nasalizações, que não correspondiam a um segmento específico da sentença, mas sim a algum fenômeno que ocorrera na passagem de um segmento acústico a outro, ou ainda a mudanças na qualidade de voz. Obviamente, tanto na análise espectral quanto na forma da onda, havia a verificação auditiva do segmento delimitado.

No que se refere aos segmentos acústicos que constituem padrões de formantes, esses foram delimitados por um corte no vale que antecedia ao primeiro pulso regular visível e, outro, no vale que antecedia o primeiro pulso irregular, próximo ao extremo direito do segmento, demarcado na análise espectral. Dessa forma, as transições dos segmentos acústicos, correspondentes às soantes em geral, foram incluídas na duração daqueles que os sucediam. Segmentos acústicos correspondentes a oclusões ou fricções eram então demarcados entre o último pulso regular do padrão de formantes que lhes precedia e o vale do primeiro pulso regular do padrão de formantes seguinte. Vale notar que o murmúrio de vogais nasais foi incluído na duração das consoantes seguintes, como forma de não provocar uma redução significativa das últimas.

1.7 Critérios adotados na elaboração do *corpus*

O novo *corpus* foi, como fora o do estudo piloto, formado por sentenças, isso porque, a nosso ver, a sentença possui a vantagem de ser em si mesma uma unidade de sentido, conferindo, por isso, maior naturalidade à tarefa de repetição, principalmente por parte de crianças. Além disso, enquanto um dos níveis mais altos da hierarquia de análise lingüística, a sentença constitui uma unidade sob a qual subjaz uma variedade de unidades de diferentes tamanhos. A partir dela, torna-se possível a comparação para unidades do tamanho de uma palavra ou de um sintagma, sempre que, na análise dos dados, julgar-se necessário ir além de unidades do tamanho do segmento ou da sílaba. Assim, o *corpus* é formado pelas seguintes sentenças:

A macacada faz barulho.

A macaca toca chocalho.

O macaco segura a banana.

O macaquinho anda de bicicleta.

O elefante planta bananeira.

O elefantinho pedala a motoca.

O elefante dá banho no palhaço.

A palhaçada dá muita risada.

O palhaço mergulha na bacia.

A cueca do palhaço tem muita bolinha.

O palhaço cor de rosa planta bananeira.

O palhaço Chapolin segura a calça.

A palhacinha não quer mamadeira.

A cachorrada quer osso.

O cachorrinho chora de medo.

O cachorro late para o palhaço.

O domador bate o chicote.

Como pode ser constatado a partir da observação das sentenças, elas foram construídas em torno de um léxico comum, tendo como tema "o circo". Porém, elas diferem entre si, com o objetivo de tornar o *corpus* o menos monótono possível para a tarefa de repetição pelas crianças.

O fato de o *corpus* tratar de um tema determinado permitiu que fossem empregadas várias repetições de uma mesma palavra. Um exemplo disso pode ser visto na ocorrência da palavra "palhaço". Repetições de um mesmo item lexical, entre as sentenças do *corpus*, são interessantes porque constituem uma forma de multiplicar a ocorrência dos mesmos segmentos acústicos, sem necessariamente ter de aumentar o número de repetições, que pode cansar os sujeitos.

A quantidade de sílabas das palavras que compõem o *corpus* varia de uma a cinco. O padrão de acentuação das palavras foi basicamente paroxítono, sendo que há apenas duas palavras oxítonas, "Chapolin" e "domador".

A hipótese levantada no estudo piloto, segundo a qual a diferença entre adultos e crianças residiria na realização de segmentos acústicos (correspondentes a vogais,

consoantes ou sílabas) em posições não-acentuadas, seguindo imediatamente um acento lexical (pós-tônica) ou localizados entre dois acentos lexicais (clíticos), quando os segmentos acústicos eram drasticamente reduzidos pelos adultos e não pelas crianças, fez com que, desde o início da elaboração desse *corpus*, atentássemos para a quantidade de segmentos acústicos nas posições não-acentuadas. Houve uma preocupação com a quantidade de segmentos acústicos em posição pré-tônica, os quais poderiam fornecer indícios sobre a própria realização do acento sobre a tônica. Para isso, contribuíram formas derivadas (macaca → macacada, palhaço → palhaçada, cachorro → cachorrada), diminutivos (macaco → macaquinho, palhaço → palhacinha, cachorro → cachorrinho, elefante → elefantinho) e também o uso farto de trissílabos (barulho, chocalho, segura, banana, pedala, motoca, risada, mergulha, bolinha, chicote) e quadrissílabos (bicicleta, mamadeira, bananeira). Entre os elementos que funcionaram como clíticos, estão as preposições e contrações de preposições e artigos: "de", "no" e "na".

As sentenças tornaram-se sintaticamente mais complexas a partir de ramificações dentro dos sintagmas nominal e verbal, os quais contêm outros elementos (sintagmas preposicionais, advérbios, adjetivos) além do substantivo e do verbo e do substantivo seguinte, respectivamente.

No que concerne aos critérios fonéticos subjacentes à elaboração do *corpus*, considerou-se que esse deveria ser amplamente representativo dos diferentes pontos e modos de articulação do PB, pois, uma vez que as crianças de 4 anos, em geral, já dominam praticamente todas as distinções do sistema fonológico da língua (YAVAS, 1988, HERNANDORENA, 1990, LAMPRECHT, 1992, TEIXEIRA, 1996), não seria possível

fazer afirmações mais genéricas sobre o *timing* de fala dessas crianças, a partir de uma classe restrita de sons.

O controle das relações fonotáticas dentro das sentenças foi também julgado como um aspecto importante, tanto para evitar aliterações que poderiam levar adultos e crianças a cometer simplificações que normalmente não ocorrem em suas falas, como também para assegurar o mais possível a repetição de um padrão de estrutura silábica CV ao longo de cada sentença, o que favoreceria a realização das medidas de duração, uma vez que processos de sândi externo (BISOL, 1996) podem dificultar a realização das medidas, principalmente nos casos em que ocorre ditongação, quando são gerados padrões de formantes, dentro dos quais há uma lenta transição de um segmento acústico a outro, o que compromete a delimitação entre seus limites na análise acústica. Dessa forma, locais de ocorrência de sândi externo foram controlados, mas não totalmente eliminados, como mostram as sentenças "O macaquinho anda de bicicleta" e "O domador bate o chicote", nas quais há a formação dos ditongos [waN] e [tshIw], respectivamente, pois não se sabia previamente até que ponto informações sobre as diferenças entre as formas como adultos e crianças resolveriam os possíveis pontos de reestruturação rítmica poderiam ser úteis no entendimento das diferenças rítmicas entre suas falas.

Foram ainda mantidas, dentro de algumas sentenças, posições nas quais poderia ocorrer um aparente choque entre acentos, como, por exemplo, em "A cachorrada quer osso". Contudo, somente posteriormente ficou claro que um entendimento mais amplo daquilo que ocorre nesses casos necessitaria da análise de outros parâmetros acústicos, na comparação entre os sujeitos, como, por exemplo, a análise de mudanças de *pitch* (VOGEL,

BUNNELL, HOSKINS, 1995). Do mesmo modo, nos casos de sândi externo, nos quais ocorre degeminação de vogais (BISOL, 1996), como em "O macaco segura a banana" ou "O elefantinho pedala a motoca", ou nos quais ocorre elisão (BISOL, 1996), como em "O cachorro late para o (→ paru) palhaço", o grau de perda da segunda vogal somente poderia ser melhor avaliado a partir das medidas dos valores dos formantes (ALBANO, 1997). Como será discutido, a realização de outras medidas acústicas, além da duração, ultrapassaria o escopo deste trabalho, principalmente devido ao volume de dados envolvidos.

1.8 O subconjunto de dados analisados e os critérios lingüísticos adotados na análise estatística

As medidas de duração foram feitas para um subconjunto dos dados coletados, o qual corresponde à fala de três pares de sujeitos, constituídos por: criança de 4 anos e 1 mês e pesquisadora; criança de 4 anos e 9 meses e pesquisadora; e professora e pesquisadora, doravante designados respectivamente por M-A, E-A e S-A. As duas crianças são do sexo feminino, paulistanas e suas idades são as que correspondem às datas de suas respectivas gravações. A professora era aquela da classe das crianças que participaram do experimento, porque entre os adultos gravados, ela constituía o adulto privilegiado, uma vez que sua fala representava o modelo costumeiramente oferecido às crianças. Pesquisadora e professora são paulistanas, com idades de 35 e 45 anos, respectivamente, na época da gravação.

Foi necessário reduzir os dados coletados à atuação de duas crianças e de um adulto para tornar o trabalho de medidas e análise exeqüível por apenas um pesquisador. Decidimos

diminuir o número de sujeitos e manter a análise tanto dos modelos oferecidos pela pesquisadora, como de todas as sentenças do *corpus*. As crianças M e E foram escolhidas para análise porque eram respectivamente as crianças mais nova e mais velha do grupo gravado e, dessa forma, julgamos que suas atuações poderiam mostrar momentos diferentes do processo de aquisição da duração no PB. Optamos também por não comparar um menino e uma menina, com o intuito de controlar minimamente os possíveis fatores decorrentes da diferença sexual (PERRONI, 1992).

A pesquisadora aparece três vezes como sujeito, por meio de seus modelos para repetição. Como será visto no próximo capítulo, tanto o tamanho do *corpus* como a análise dos modelos oferecidos aos sujeitos foram fatores fundamentais para a análise estatística realizada e também nos resultados obtidos.

Foram medidos todos os segmentos acústicos correspondentes às consonantes e vogais do conjunto de 51 sentenças (três repetições do *corpus* de 17 sentenças, para cada sujeito), perfazendo um total de 304 sentenças (51 sentenças x 6 sujeitos - 2 sentenças), isso porque, uma delas (O cachorrinho chora de medo) foi eliminada da análise de uma das repetições da pesquisadora e da criança de 4 anos e 1 mês, devido a problemas na gravação. Para dar uma idéia do volume de dados medidos, foram utilizadas, na análise estatística[9], para o par criança de 4 anos e 1 mês e pesquisadora, 401 vogais e 430 consoantes; para o par criança de 4 anos e 9 meses e pesquisadora, 443 vogais e 496 consoantes; e, finalmente, para o par professora e pesquisadora, 447 vogais e 487 consoantes.

---

[9] Isso significa que foram medidos muitos mais segmentos acústicos, uma vez que, como será visto, foram considerados na análise estatística apenas os segmentos para os quais os sujeitos de cada par puderam ser comparados entre si.

Com base na hipótese levantada no estudo piloto, que apontava para diferenças entre adultos e crianças na realização de segmentos em posição não-acentuada, e também a partir da análise de graus de acentuação proposta por Câmara (1969)[10], a comparação entre os sujeitos de cada par foi feita tendo como referência as diferentes posições de acento (0, 1, 2, 3, 4, 5 e 6) nas quais o segmento acústico (correspondente a vogal, consoante ou sílaba) ocorria, e, no caso das palavras (paroxítonas), com base em seu número de sílabas (di, tri ou quadrissilábicas) e em sua posição na sentença, final ou não-final. A tabela abaixo fornece um resumo das posições de acento por nós demarcadas, na palavra e na sentença:

0	Início Absoluto da Sentença
1	Primeira Pré-Tônica da Palavra
2	Segunda Pré-Tônica da Palavra
3	Terceira Pré-Tônica da Palavra
4	Tônica Lexical
5	Pós-Tônica
6	Final Absoluto da Sentença

Foram definidas então sete posições de acento dentro de cada sentença, a saber: início absoluto da sentença (0), posição pré-tônica, que, por sua vez, foi subdividida nas posições 1, 2 ou 3, dependendo do número de pré-tônicas da palavra em questão, posição tônica (4),

[10] Os graus de acento propostos por CÂMARA (op. cit.) podem ser traduzidos em termos da teoria métrica, uma vez que o autor faz referência a contornos acentuais.

posição pós-tônica (5) e final absoluto da sentença (6). Os segmentos de início e de final absolutos foram analisados separadamente de pré-tônicas e pós-tônicas, respectivamente, por estarem sujeitos ao alongamento de fronteira maior da sentença (FOUGERON e KEATING, 1997, KLATT, 1976). Vale notar que, apesar de ter sido preservada uma posição referente à terceira pré-tônica, esta somente foi preenchida no caso da palavra "elefantinho", o que tornou inviável a análise estatística dos segmentos nessa posição, uma vez que havia, nela, apenas três segmentos acústicos e a quantidade mínima necessária para considerar os resultados estatísticos como confiáveis foi estabelecida como sendo de cinco ocorrências.

Os sujeitos em cada par foram ainda comparados para o conjunto de segmentos acústicos das posições não-acentuadas, quando, no que se refere a vogais ou consoantes, agrupamos os segmentos acústicos pertencentes às três pré-tônicas e à pós-tônica (agrupamento 1235), sendo que, nas sílabas, houve uma subdivisão na análise do agrupamento constituído pelas posições não-acentuadas, com um subconjunto formado pelas sílabas que ocorrem nas posições 1235 e, o outro, por aquelas que ocorrem nas posições 01235, ou seja excluindo e incluindo, respectivamente, a posição de início absoluto. Os sujeitos foram também comparados para todos os segmentos acústicos (somatória das posições 0123456) em cada categoria analisada: vogais, consoantes e sílabas. Para uma melhor visualização dos agrupamentos de posições, observe a tabela abaixo:

1235	Pré-Tônicas e Pós-Tônicas (para vogais, consoantes e sílabas, excluindo aquelas de início absoluto)
01235	Pré-Tônicas (somente para sílabas, incluindo aquelas de início absoluto)
123456	Somatória de Todas as Ocorrências (para vogais e consoantes)
0123456	Somatória de Todas as Ocorrências (somente para sílabas)

Um exemplo de como os segmentos dentro de cada sentença do *corpus* foram demarcados a partir de suas posições de acento pode ser visto abaixo, tanto para os segmentos acústicos correspondentes a vogais e consoantes, como para os segmentos do tamanho das sílabas, para a sentença "A palhaçada dá muita risada":

A	p	a	lh	a	ç	a	d	a	d	á	m	ui	t	a	r	i	s	a	d	a
0	1	1	2	2	4	4	5	5	4	4	4	4	5	5	1	1	4	4	6	6

A	pa	lha	ça	da	dá	mui	ta	ri	sa	da
0	1	2	4	5	4	4	5	1	4	6

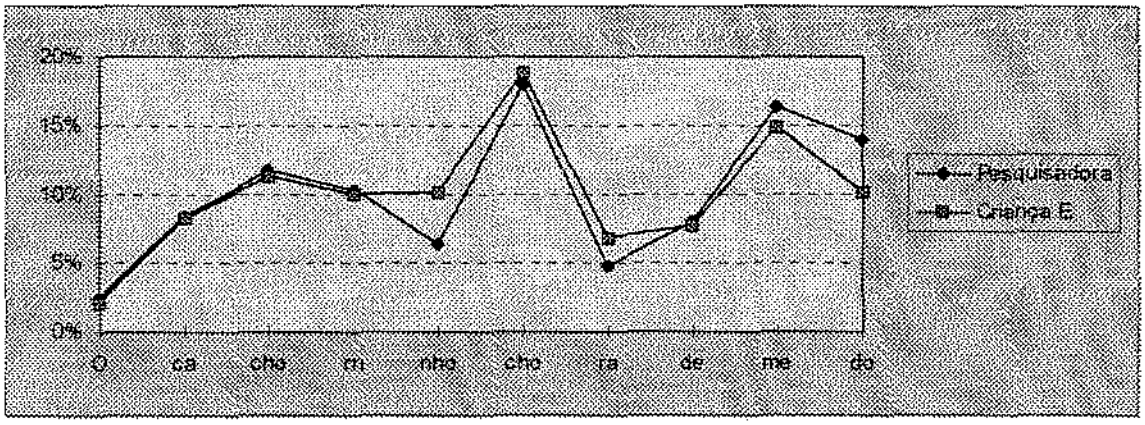
Ao observar essa demarcação para as diferentes posições de acento na sentença[11], fica evidente a concepção mattosiana por trás da análise. Entre as divergências em relação à proposta de CÂMARA (1969) está o fato de que, aqui, as pré-tônicas receberam valores diferentes e foram analisadas separadamente, isso porque, uma análise qualitativa, por meio de contornos duracionais, mostrou que as pré-tônicas de uma mesma palavra variavam entre apresentar: durações iguais, durações diferentes, ou ainda uma gradação de durações crescentes, culminando na duração maior da sílaba tônica. Para exemplificar o que está sendo dito, servimo-nos de gráficos elaborados para o par formado pela pesquisadora e criança E, os quais foram construídos com os valores percentuais[12] da duração média, calculada para as três repetições de cada sujeito, das sílabas das sentenças do *corpus*.

No gráfico exposto abaixo, podemos observar que pesquisadora e criança E realizam um aumento gradual das durações das pré-tônicas em "O cachorrinho" (O cachorrinho chora de medo), quando a criança reproduziu muito de perto o modelo dado pela pesquisadora, a não ser pela maior duração das sílabas pós-tônicas de "cachorrinho" e "chora".

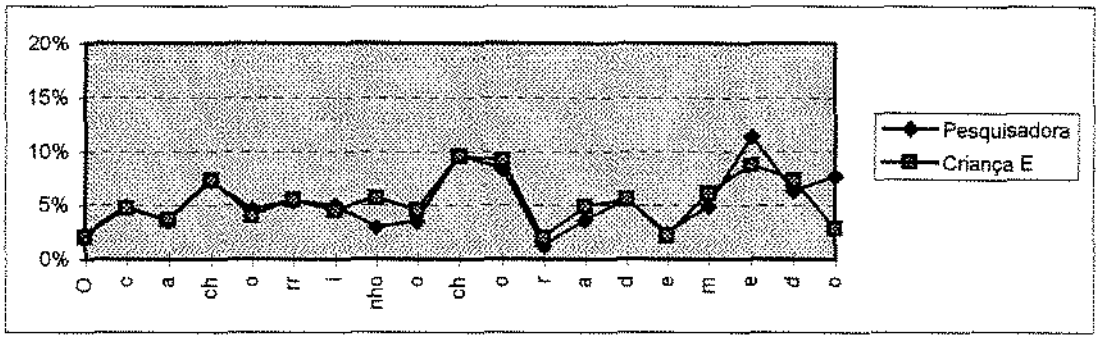
---

[11] Para a demarcação das posições de acento em cada sentença do *corpus*, veja ANEXO A, para vogais e consoantes, e ANEXO B, para sílabas.

[12] A transformação dos valores absolutos das durações em valores percentuais é feita por meio da divisão da duração de cada segmento acústico (ou conjunto de segmentos, na sílaba) pela duração da soma dos segmentos da sentença, com expressão em percentagem. Esse método tem sido bastante eficiente em reduzir a influência da taxa de elocução sobre as durações absolutas, permitindo a comparação entre sujeitos.

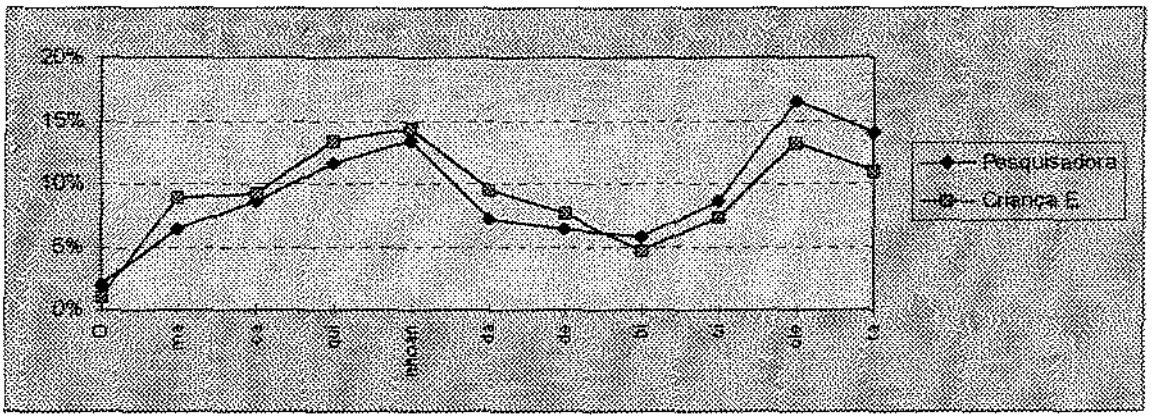


O fato de a maior proeminência da seqüência (ou do sintagma "O cachorrinho") recair sobre "cho" reflete diferenças de duração intrínseca entre os segmentos acústicos que compõem respectivamente as sílabas "cho" e "rri", como mostrado no gráfico abaixo:



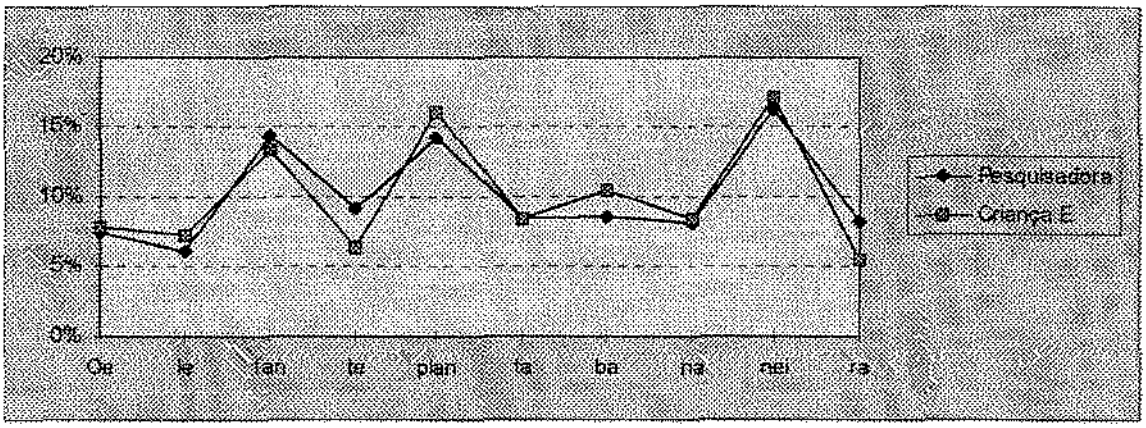
O mesmo tipo de aumento gradual das durações das pré-tônicas de uma palavra ocorre para pesquisadora e criança E em "bicicleta" (O macaquinho anda de bicicleta), como pode ser visto a seguir. Note que, em "O macaquinho", o aumento gradual das pré-tônicas é observado na fala da pesquisadora, enquanto para a criança E, há uma silabação entre "ma" e "ca", praticamente de mesma duração. O pico de duração da seqüência parece recair sobre

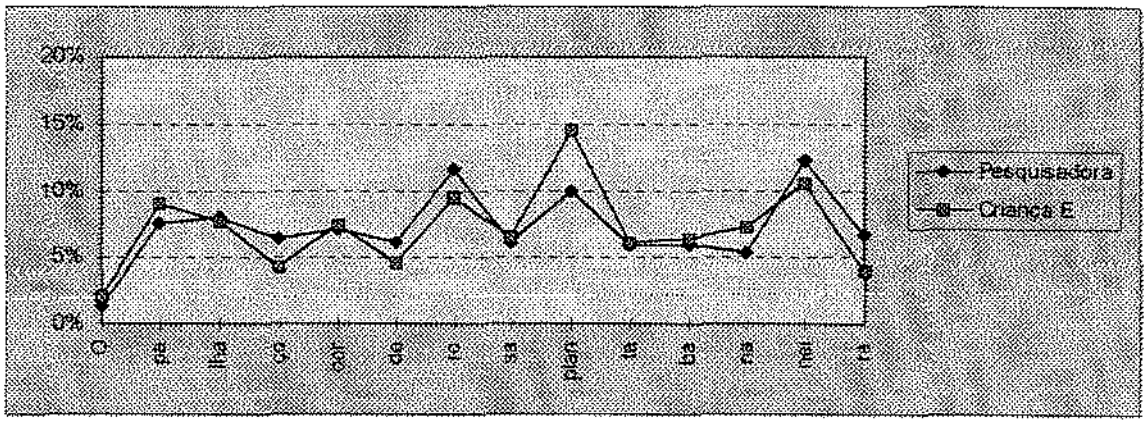
"nhoan" (junção de *macaquinhoanda*), devido à dificuldade de medir separadamente os segmentos acústicos (todos soantes) da pós-tônica (nho) e da tônica seguinte (an), uma vez que se formou aí um grande padrão de formantes.



As pré-tônicas da palavra "bananeira", ocupando a mesma posição sintática em duas sentenças diferentes (gráficos na seqüência), foram realizadas pela pesquisadora e pela criança E a partir da utilização das três estratégias discutidas (durações iguais, durações diferentes, maior na primeira pré-tônica e menor na segunda, e durações crescentes em direção à tônica[13]). Em "O elefante planta bananeira" (primeiro gráfico a seguir), observamos que a criança E adota a estratégia duração maior vs. duração menor, na passagem de "ba" para "na", ao passo que a pesquisadora realiza as duas sílabas praticamente com a mesma duração. Já em "O palhaço cor de rosa planta bananeira" (segundo gráfico a seguir), a pesquisadora reduz a sílaba "na" e a criança realiza um aumento crescente das durações de "ba" e "na" em direção à tônica ("nei").

[13] BARBOSA (1994) mostrou o mesmo crescendo duracional na realização do acento em francês.





Além de analisar as pré-tônicas separadamente, também divergimos da proposta de CÂMARA (1969) no que concerne à posição da maior proeminência da sentença, a qual, no grupo de força, recai sobre o último acento lexical à direita. Na análise aqui proposta, o último acento lexical à direita foi simplesmente incluído junto com os demais acentos lexicais, na posição 4, devido à nossa observação prévia, também a partir da análise qualitativa de contornos duracionais, de que nem sempre os sujeitos concordavam sobre a

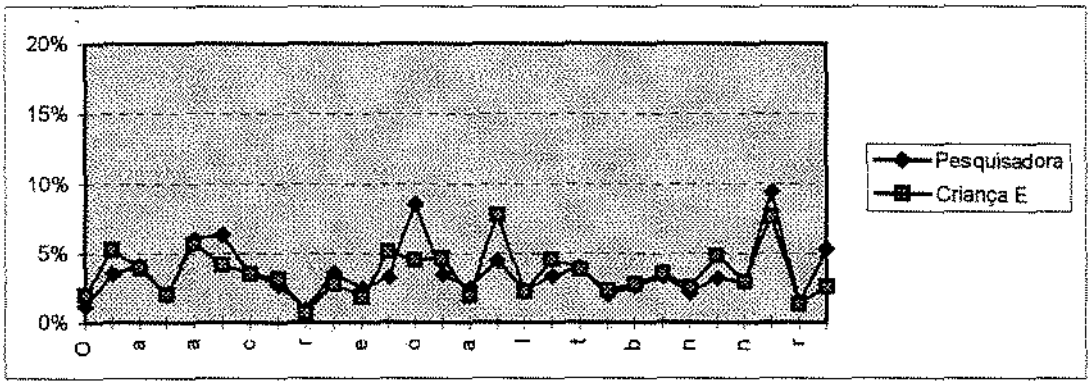
maior proeminência da sentença, pelo menos a partir do parâmetro de duração, o que é previsível via modelos atuais (NESPOR e VOGEL, 1986).

Os gráficos expostos acima já mostram algumas das contradições em relação à previsão de CÂMARA (op. cit.). No primeiro deles, que apresenta as durações médias das sílabas, em valores percentuais, da sentença "O cachorrinho chora de medo", a maior proeminência da sentença, para pesquisadora e criança E, parece recair sobre "cho" (chora). Contudo, a análise em segmentos mostra que a vogal "e" de "medo" foi mais alongada pela pesquisadora que a vogal "o" de "chora", o que sugere que talvez, para ela, a maior proeminência da sentença tenha recaído sobre o último acento lexical à direita.

Em "O palhaço cor de rosa planta bananeira" (último gráfico colocado acima), para a pesquisadora, a sílaba mais alongada na sentença é "nei" (bananeira), embora a duração dessa sílaba esteja muito próxima da duração de "ro" (rosa). Apenas com a análise dos valores percentuais das durações médias das sílabas, não é possível dizer se a pesquisadora dividiu a sentença em dois blocos ("O palhaço cor de rosa" e "planta bananeira", cada um deles com maior alongamento do último acento lexical à direita) ou se a sílaba mais distendida na sentença foi "nei". Para a criança E, a duração de "plan" (planta) é bastante distinta daquela de "ro" (rosa) e "nei"(bananeira). Porém, também não fica claro até que ponto o maior número de segmentos acústicos na sílaba "plan" pode ter causado seu maior alongamento em relação às sílabas "ro" (rosa) e "nei" (bananeira). Vejamos, então, o que ocorreu nas durações dos segmentos acústicos considerados isoladamente[14]:

---

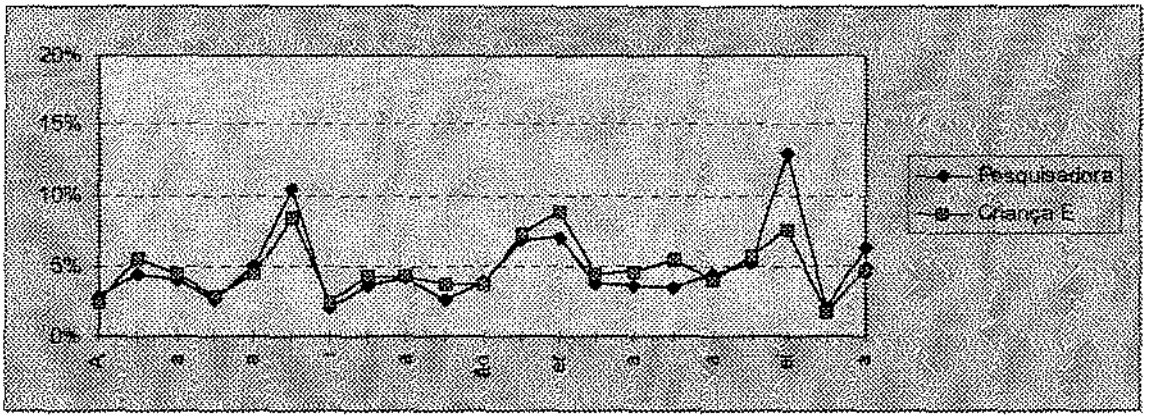
[14] Note que, no gráfico referente aos segmentos acústicos isolados da sentença "O palhaço cor de rosa planta bananeira", o segmento acústico correspondente ao "o" de "palhaço" não pôde ser medido separadamente do segmento acústico correspondente a "ç".

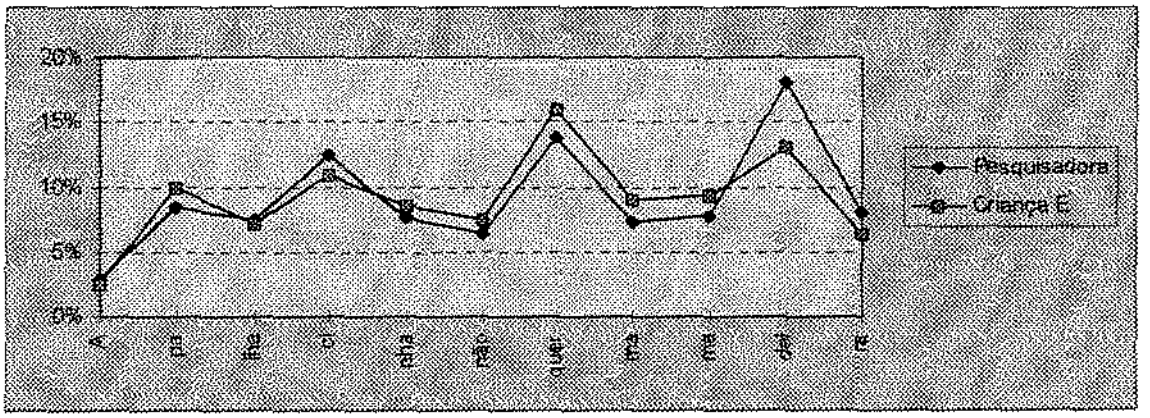


A criança E mostra um alongamento da consoante "p", em "planta", que pode ter ocorrido pela introdução de uma pausa (ou alongamento pré-fronteira), a qual, por sua vez, seria incluída na medida da duração da oclusiva, aumentando seu silêncio ou tempo de oclusão. Pode ser ainda que a maior distensão do "p" deva-se a uma hesitação durante sua pronúncia. De qualquer forma, para a criança E, o ditongo "ei" (bananeira), da última sílaba acentuada mais à direita, apresenta o maior alongamento, abaixo da duração de "p" (planta), o que nos faz supor que a sílaba "nei" recebeu a maior proeminência da sentença. A análise dos segmentos isolados mostra ainda que, para a pesquisadora, o alongamento do ditongo "ei" (bananeira) é maior que o alongamento de "o", da sílaba "ro" (rosa). Assim, nesse caso, parece que pesquisadora e criança E concordaram sobre colocar a maior proeminência da sentença sobre o último acento lexical à direita.

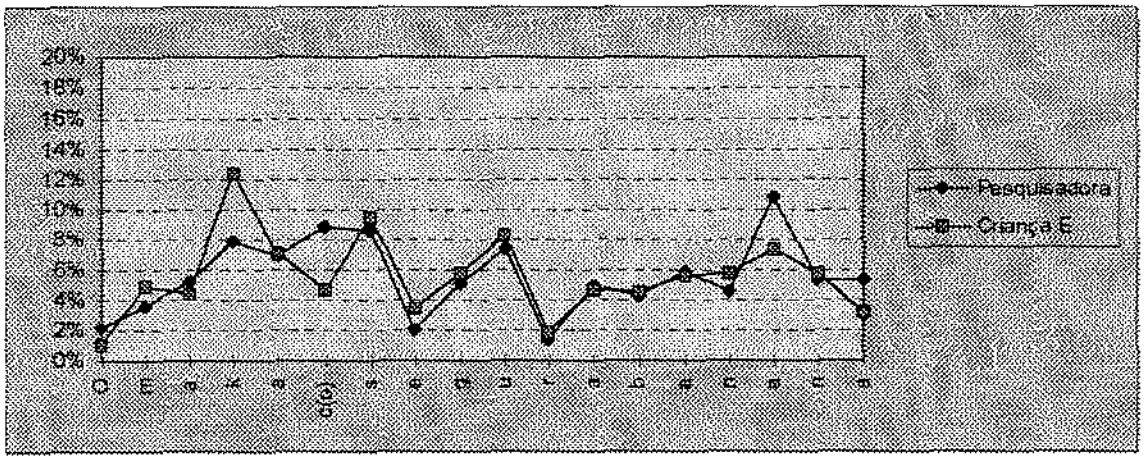
Em "A palhacinha não quer mamadeira", a observação dos gráficos, respectivamente com os valores percentuais dos segmentos isolados e das sílabas (colocados abaixo),

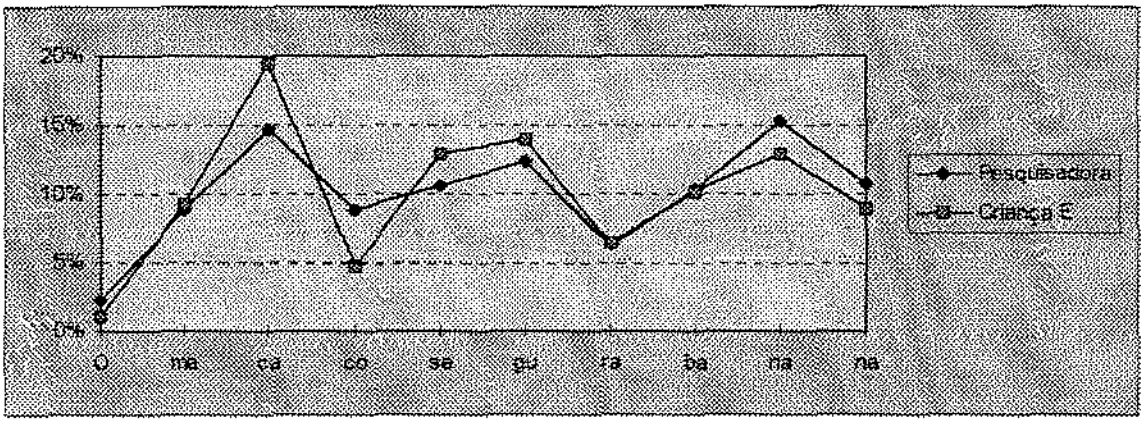
confirma que pesquisadora e criança E fizeram opções diferentes quanto à maior proeminência da sentença, que recai sobre o último acento lexical mais a direita, para a primeira, e sobre "quer", para a segunda.





Em "O macaco segura a banana", a análise dos segmentos isolados mostra que o alongamento sofrido pela vogal "u" (segura), na fala da criança E, é maior que o alongamento da vogal da sílaba que constitui o último acento lexical à direita ("na" de "banana"), ao contrário do que ocorre na fala da pesquisadora.



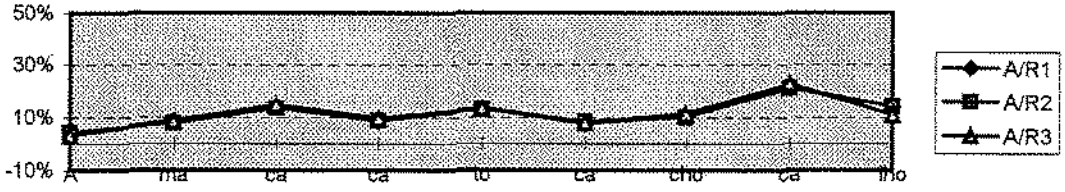


Fica claro que, a partir da análise de contornos duracionais, não é possível afirmar que a maior proeminência da sentença recai sistematicamente sobre o último acento lexical à direita, conforme fora predito por CÂMARA (op. cit.) e, por isso, essa posição de acento lexical não foi analisada separadamente das demais. Além disso, com base somente em medidas de duração, não é tampouco possível concluir de modo definitivo como se realiza a maior proeminência da sentença, pois mudanças no contorno de *pitch*, combinadas com a

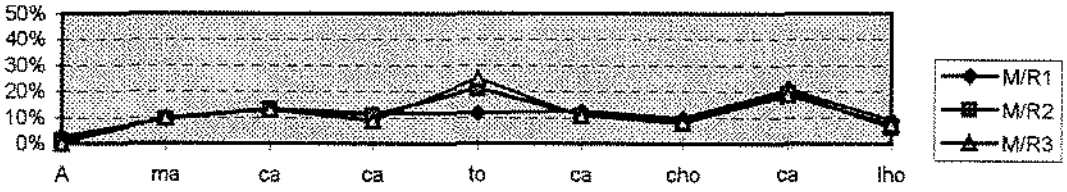
distensão das durações dos segmentos das sílabas acentuadas, podem intervir para alterá-la (VOGEL, BUNNELL e HOSKINS, 1995).

Outra observação que reitera a impossibilidade de analisar a maior proeminência da sentença separadamente, a partir da análise qualitativa de contornos duracionais, é que as crianças (M e E) nem sempre apresentavam um pico distinto de duração ao nível da sentença, mas, ao invés disso, vários acentos sobre cada uma das palavras de conteúdo, o que nos levou a supor que suas falas estariam variando entre um ritmo vocabular e um ritmo frasal. Os gráficos abaixo mostram os três modelos oferecidos pela pesquisadora e as três repetições, de cada sujeito, para a sentença "A macaca toca chocalho". Note que, na segunda repetição da criança M, as sílabas "to" (toca) e "ca" (chocalho) apresentam durações muito próximas, e, para a criança E, o mesmo ocorre, na segunda repetição, para as sílabas "ca" de "macaca" e "chocalho". Porém, nada parecido é observado nos modelos da pesquisadora aos três sujeitos, nem nas repetições da professora, quando a maior proeminência da sentença recai sempre sobre o último acento lexical à direita.

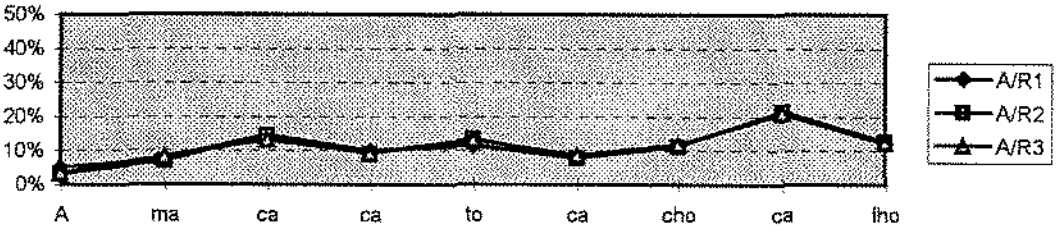
Modelos oferecidos pela pesquisadora à criança M



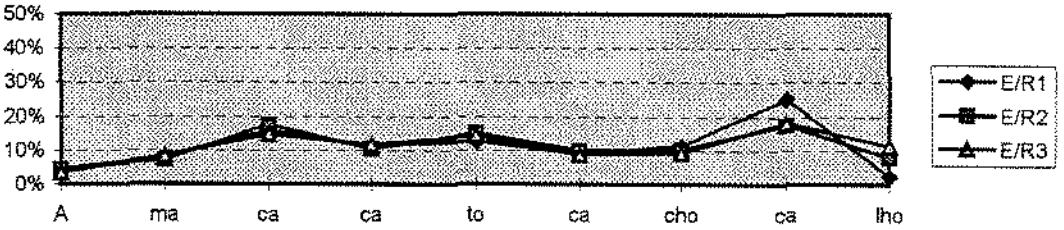
Repetições da criança M



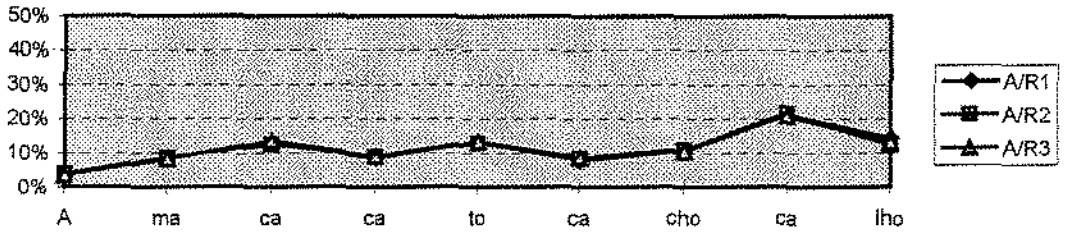
Modelos oferecidos pela pesquisadora à criança E



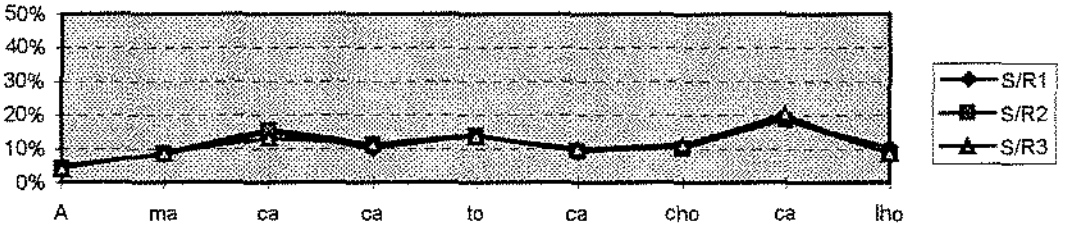
Repetições da criança E



Modelos oferecidos pela pesquisadora à professora (S)



Repetições da professora (S)



Voltando à análise estatística dos dados apresentados neste trabalho, primeiramente, os sujeitos de cada par foram comparados entre si, para o mesmo segmento acústico, por meio de um teste t, para verificar se poderia ou não ser mantida a hipótese de que a média das diferenças de duração entre os sujeitos de cada par fosse estatisticamente igual a zero. Em segundo lugar, foram feitas comparações entre os dois adultos e as duas crianças, entre as duas crianças, entre estas e a professora, entre os modelos oferecidos pela pesquisadora a cada um dos sujeitos e entre os dois adultos, por meio do teste de Duncan (MONTGOMERY, 1996). Nele, os sujeitos foram comparados entre si com base em todas as suas repetições de cada sentença. Em cada caso, os sujeitos foram comparados para os segmentos acústicos vocálicos e consonantais e do tamanho das sílabas de cada sentença. O

nível de significância previamente fixado no teste foi de 5%, ou seja, para todos os pares possíveis, calcula-se quais são significativos e quais não são, tomando precauções para que o erro não passe de 5%, ou seja, cria-se um fator de correção para o nível de significância, o qual corrige ou não permite que ao se fazer as múltiplas comparações o nível de significância seja aumentado.

No teste *t*, somente puderam ser computadas as durações dos segmentos medidos para ambos os sujeitos, uma vez que a comparação entre eles era feita para a mesma produção, ou seja, o mesmo segmento acústico, da mesma repetição. Portanto, quando não foi possível delimitar a duração de um segmento na fala de um dos sujeitos do par ou quando havia uma disparidade no valor de sua duração, causada, por exemplo, pela introdução de pausa ou hesitação na fala de um dos sujeitos, o segmento acústico era automaticamente eliminado da análise estatística.

Ainda no teste *t*, os sujeitos de cada par foram comparados entre si para as durações de segmentos acústicos correspondentes a vogais, consoantes, sílabas e palavras. Apesar de a sílaba ser tida como o domínio do acento (EDWARDS e BECKMAN, 1988), a análise qualitativa nos contornos duracionais dos segmentos acústicos que compõem a sentença mostrou-se bastante útil, nos dados analisados acima, para supor diferenças entre a contribuição de vogais e consoantes na formação dos contrastes de duração, no PB. Por isso, em cada posição de acento (0, 1, 2, 3, 4, 5, 6), ou agrupamento de posições (1235 e 123456, no caso de segmentos acústicos correspondentes a vogais ou consoantes, e 1235, 01235 e 0123456, no caso de segmentos acústicos correspondentes a sílabas), e para cada tipo de segmento acústico (vocálico, consonantal ou silábico), foi calculada a diferença entre

as durações dos sujeitos do par e, a seguir, a média e o desvio padrão das diferenças de duração obtidas para o par. Note que a diferença para as durações do sujeitos do par é obviamente feita dentro de cada repetição (primeira, segunda ou terceira, sendo que não é calculada, por exemplo, uma média para a duração do segmento entre as repetições), porque se trata da comparação entre os sujeitos do par para a mesma produção. A diferença entre os sujeitos de cada par é sempre calculada como: a duração do segmento na fala do sujeito (M, E ou S) menos a duração do mesmo segmento na fala da pesquisadora (A), o que significa que a diferença é expressa em um valor negativo quando a pesquisadora (A) leva mais tempo que o outro membro do par para produzir o segmento acústico, e vice-versa.

Para dar um exemplo do que foi dito, quando o par M-A (criança de 4 anos e 1 mês e pesquisadora) foi comparado para as vogais produzidas na posição tônica (posição 4), a média resultante de todas as diferenças entre o par para as vogais nessa posição foi de 10 ms, o que indica que, a criança M apresentou uma duração ligeiramente maior que a da pesquisadora, nas vogais acentuadas. Note ainda que essa média corresponde à média das diferenças entre o par para todos os tipos de vogais (baixas, altas, orais, nasalizadas, etc) e de encontros vocálicos que ocorreram na posição 4. Assim, a hipótese por trás da análise estatística, aqui proposta, é a de que os segmentos acústicos (sejam eles vogais, consoantes ou sílabas) irão sofrer maior ou menor alongamento em função de suas posições de acento dentro da sentença, independentemente de suas durações intrínsecas.

Essa hipótese leva em conta que a combinação entre posição e acento, conforme será visto na discussão dos modelos dinâmicos apresentada no capítulo 3, principalmente aquele

da Fonologia Articulatória, pode favorecer tanto gestos articulatórios de maior amplitude, como mais reduzidos.

2. NOVOS DADOS E REFLEXÕES SOBRE A INTERAÇÃO ENTRE DURAÇÃO E VARIABILIDADE NO ESTUDO DA AQUISIÇÃO DA DURAÇÃO NO PORTUGUÊS BRASILEIRO

Para a apresentação dos resultados obtidos na análise de pares, por meio do teste t, foram elaborados gráficos que mostram, para cada par (M-A, E-A, S-A), para cada tipo de segmento acústico (correspondente a vogal, consoante e sílaba) e em cada posição (0, 1, 2, 4, 5, 6, e, para as palavras, em posição não-final e final de sentença) ou agrupamento de posições de acento (1235,123456, para segmentos acústicos vocálicos ou consonantais, e 1235, 01235 e 0123456, para segmentos acústicos do tamanho de sílabas), a média das diferenças de duração (em ms), expressa por meio de barras, e o desvio-padrão (em ms) ou variabilidade absoluta dessas diferenças, o qual aparece na curva sobreposta às barras.

Vale lembrar que:

(1) Por conveniência, o que denominamos abaixo de vogais, consoantes e sílabas são segmentos acústicos no sentido de FANT (1973) (vide nota 2, Capítulo 1, páginas 7 e 8) e dos critérios aqui utilizados na realização das medidas de duração (Capítulo 1, páginas 44 a 47);

(2) A posição 3, referente à terceira pré-tônica, não aparece isoladamente, porque nela ocorreram apenas os segmentos acústicos correspondentes àqueles da sílaba "fan" das três repetições de "elefantinho" em "O elefantinho pedala a motoca", e somente foram analisados dados com um mínimo de cinco ocorrências;

(3) A diferença de duração de um segmento acústico, para cada par, é sempre calculada como: a duração do segmento na fala do sujeito (M, E ou S) menos a duração do segmento na fala da pesquisadora (A), o que significa que, se a média das diferenças é positiva, o sujeito (M, E ou S) apresentou durações maiores que a pesquisadora naquela posição (ou agrupamento de posições), e vice-versa;

(4) As vogais de início absoluto de sentença, que correspondem ao artigo que precede o substantivo no sintagma nominal, foram analisadas separadamente na posição pré-tônica 0 das sílabas, devido ao fato de estarem mais sujeitas às variações de fronteira inicial, decorrentes da demarcação pelo sujeito de um limite de fronteira principal, nesse caso, o início da sentença. Da mesma forma, os segmentos acústicos correspondentes à sílaba final foram analisados separadamente na posição 6, por estarem sujeitos ao fenômeno de alongamento da sílaba final da sentença. Foram incluídos ainda, na posição 0, os padrões de formantes resultantes do sândi externo das vogais correspondentes ao artigo "O" e à primeira pré-tônica de "elefante" e de "elefantinho", o que deu origem a um ditongo ([we]);

(5) O agrupamento de segmentos acústicos correspondente às sílabas somente foi considerado quando foi possível medir as durações de todos os segmentos que compunham a sílaba, para ambos os sujeitos do par;

(6) Nas sentenças nas quais havia possibilidade de ressilabificação, essa foi levada em conta na demarcação das posições de acento, tanto dos segmentos acústicos correspondentes a vogais e consoantes (ANEXO A) como do conjunto de segmentos acústicos que constituíam a sílaba proveniente da ressilabificação (ANEXO B);

(7) Da posição 4, correspondente aos acentos lexicais, constam também as maiores proeminências das sentenças, isso porque, apesar de CÂMARA (1969) sugerir que a maior proeminência dos grupos de força recai sobre o último acento lexical à direita, a análise de contornos duracionais, apresentada no capítulo anterior, mostrou que não há conformidade entre os sujeitos para a predição de CÂMARA (op. cit.) e, além disso, consideramos que outros parâmetros acústicos (como, por exemplo, f0) devem ser investigados para que se possa efetivamente fazer afirmações sobre a maior proeminência da sentença;

(8) Em apenas uma sentença (O palhaço mergulha na bacia), o último acento lexical à direita coincidiu com a última sílaba da sentença. Nesse caso, os segmentos acústicos (isolados ou agrupados), correspondentes à sílaba "cia" (bacia), foram incluídos na posição 6 e não na posição 4. Vale lembrar que o núcleo da sílaba foi medido como um único padrão de formantes, uma vez que, como já foi dito, quando ocorreram encontros vocálicos, seus segmentos acústicos não foram considerados separadamente, com relação às medidas de duração;

(9) Todos os valores numéricos, desde as medidas de duração até os valores das médias e desvios-padrão das diferenças de duração, foram arredondados somente para sua apresentação final, nas tabelas presentes no trabalho. O arredondamento foi feito no *Excel 7.0*;

(10) Finalmente, um aspecto muito importante, e que deve ficar claro desde já, é que a análise estatística dos modelos oferecidos pela pesquisadora aos três sujeitos (criança M, criança E e professora) mostrou que não havia diferenças significativas entre os enunciados da pesquisadora nas três situações de apresentação do modelo. A análise foi feita por meio do teste de Duncan, no qual os enunciados da pesquisadora foram comparados entre si tanto para a duração dos segmentos acústicos que correspondiam a vogais ou consoantes, como para aqueles do tamanho de sílabas.

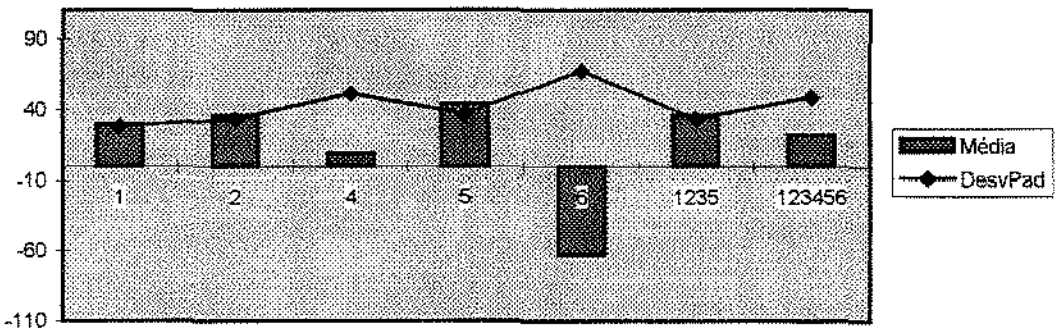
Os gráficos referentes a cada tipo de segmento acústico serão expostos na seguinte seqüência: vogais, consoantes e sílabas, e, dentro de cada tipo de segmento, preferencialmente para os pares M-A, E-A e S-A, respectivamente da criança menor (M) para a professora (S). Além disso, os comentários a respeito dos resultados obtidos serão feitos da seguinte forma: diferenças entre os sujeitos do par e diferenças entre pares.

2.1 Comparações entre os sujeitos de cada par para o teste t

2.1.1 Duração e variabilidade de segmentos acústicos vocálicos

No que concerne às diferenças quanto à duração dentro de cada par, observe-se o Gráfico 1, referente ao par M-A:

Gráfico 1 - Médias e desvios-padrão das diferenças de duração para os segmentos acústicos vocálicos do par M-A



O Gráfico 1 mostra que as durações das vogais (gráfico de barras) da criança M (4 anos e 1 mês) são maiores (valores positivos) que as da pesquisadora, em todas as posições ou agrupamento de posições de acento, exceto na posição 6, a qual corresponde à vogal da sílaba final da sentença, e que, na fala da pesquisadora, sofre um alongamento considerável (valor negativo, com média = -63 ms, ANEXO C, Tabela 1). Esse alongamento parece

constituir-se numa característica do estilo de fala da pesquisadora, já que é observado na comparação entre ela e os outros dois sujeitos (Gráficos 2 e 3).

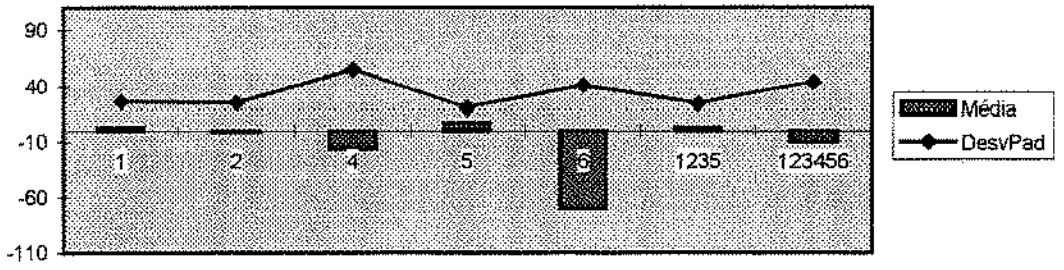
As diferenças entre a criança M e a pesquisadora são significativas nas várias posições de acento consideradas separadamente (vide os valores de p, para as posições 1, 2, 4, 5 e 6, ANEXO C, Tabela 1) e também nos agrupamentos que incluem somente as posições não-acentuadas (p=0.0001, para o agrupamento 1235, ANEXO C, Tabela 1) ou todas as posições (p=0.0001, para o agrupamento 123456, ANEXO C, Tabela 1). É interessante notar que, a menor média das diferenças de duração entre a criança M e a pesquisadora refere-se justamente à posição 4 (média = 10 ms, ANEXO C, Tabela 1), a qual corresponde à posição de acento lexical e que inclui a maior proeminência da sentença, o que nos leva à hipótese de que a implementação do parâmetro de duração é primeiramente dominada, pela criança M, para a posição de acento. O fato de que a diferença entre criança M e pesquisadora também diminui quando ambas são comparadas para todas as vogais (agrupamento de posições 123456, média = 22 ms, ANEXO C, Tabela 1) deve-se provavelmente ao grande alongamento das vogais da posição 6 (de final de sentença), realizado pela pesquisadora.

Observe-se que quase não há diferença entre as médias das diferenças de duração das pré-tônicas 1 e 2 para o par M-A (respectivamente, 30 e 36 ms , ANEXO C, Tabela 1) da mesma forma que não é observada uma mudança considerável na variabilidade absoluta ou desvio-padrão entre essas pré-tônicas (respectivamente 28 e 33 ms, para posições 1 e 2, ANEXO C, Tabela 1).

A posição para a qual a criança M apresenta os maiores valores de duração na comparação com a pesquisadora é justamente a posição 5 (média = 44 ms, ANEXO C, Tabela 1), referente às vogais que ocorrem em pós-tônicas e em clíticos. Vale notar que a variabilidade absoluta ou desvio-padrão das vogais pertencentes às posições não-acentuadas ou ao agrupamento dessas é muito semelhante (ou seja, 28 ms, 33 ms, 37 ms e 33 ms, respectivamente para as posições 1, 2, 5, e para o agrupamento de posições 1235).

Observemos, agora, o par E-A para as mesmas medidas.

Gráfico 2 - Médias e desvios-padrão das diferenças de duração para os segmentos acústicos vocálicos do par E-A

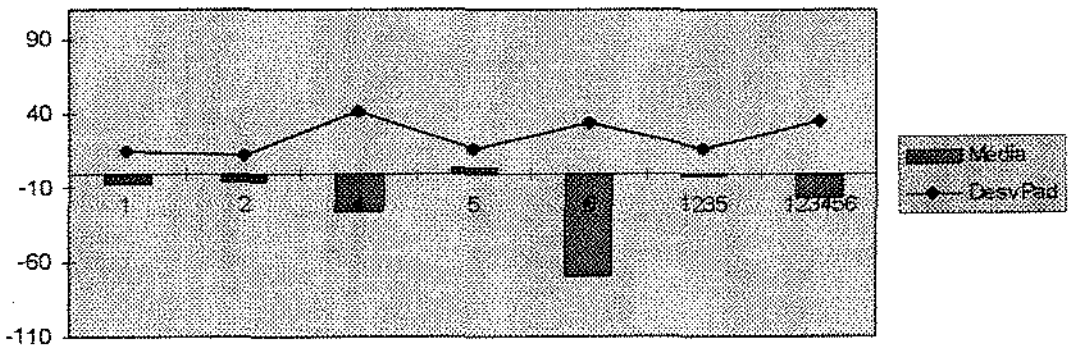


No que se refere ao par E-A, Gráfico 2, há uma diminuição das médias das diferenças de duração dos segmentos acústicos vocálicos entre a criança E (4 anos e 9 meses) e a pesquisadora, sendo que não há diferença significativa entre elas nas vogais das posições pré-tônicas 1 e 2 (respectivamente, p=0.30281 e p=0.8711, ANEXO C, Tabela 2). Nas demais posições, a diferença entre elas é sempre significativa (ANEXO C, Tabela 2). Além disso, a pesquisadora apresenta durações maiores (valores negativos) que a criança E não

apenas na posição 6, mas também na posição 4, referente ao acento lexical e que inclui a maior proeminência da sentença, e ainda no agrupamento de todas as vogais (123456) (respectivamente, -70 ms, na posição 6; -16 ms, na posição 4; e -10 ms, no agrupamento 123456, ANEXO C, Tabela 2).

A falta de diferença significativa entre pesquisadora e criança E para as posições pré-tônicas 1 e 2 e a diferença significativa entre elas para a posição 5 poderiam levar à hipótese de que, para a criança E, a estabilidade das vogais não-acentuadas (BARBOSA e BAILLY, 1994) pode ser primeiramente observada para aquelas em posição pré-tônica em relação às vogais pertencentes à posição pós-tônica. Entretanto, como veremos abaixo, foram encontradas diferenças significativas entre os adultos (par S-A) tanto para as vogais pertencentes às posições 1 e 2 como para aquelas da posição 5.

Gráfico 3 - Médias e desvios-padrão das diferenças de duração para os segmentos acústicos vocálicos do par S-A

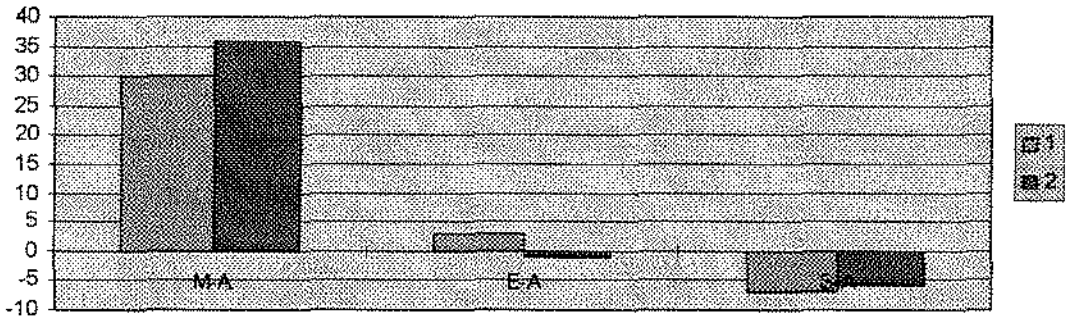


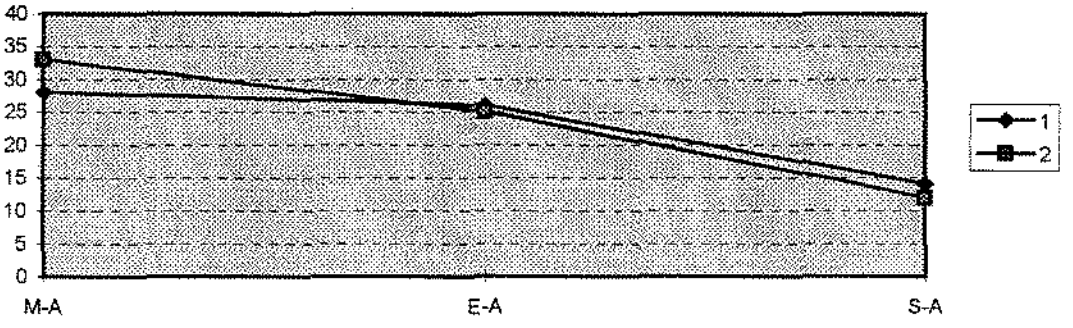
A observação do Gráfico 3, referente ao par S-A (professora e pesquisadora), mostra que a pesquisadora apresenta vogais mais longas (valores negativos) que a professora em todas as posições, exceto na posição referente às vogais pós-tônicas e pertencentes a clíticos (posição 5), na qual se observa uma média ligeiramente maior, 4 ms, a favor da professora (valor positivo) (ANEXO C, Tabela 3), o que sugere que a pesquisadora reduz um pouco mais suas vogais nessa posição. Note que, apesar de a média das diferenças entre ambas ser pequena (4 ms), a diferença entre elas é significativa (p=0.0156, ANEXO C, Tabela 3). Pesquisadora e professora diferem significativamente para os segmentos acústicos vocálicos em todas as posições de acento e também nos agrupamentos de posições (1235 e 123456).

Poder-se-ia questionar por que, então, os sujeitos dos três pares diferem significativamente entre si na posição 5, referente a vogais pós-tônicas e pertencentes a clíticos. Vale lembrar que, nessa posição, encontram-se vogais que pertencem a sílabas que formam fronteiras sintáticas menores dentro das sentenças, ou seja, sílabas que correspondem às fronteiras entre sintagma nominal e sintagma verbal. Assim, essas vogais podem, eventualmente, estar sujeitas ao fenômeno de alongamento pré-pausal (BECKMAN e EDWARDS, 1990) e, além disso, o fenômeno, em si mesmo, pode constituir-se numa característica estilística da fala de um determinado sujeito mas não de outro, como ocorre, no caso da demarcação de fronteiras maiores, com o alongamento da vogal pertencente à sílaba final da sentença, na fala da pesquisadora, o qual não é observado, pelo menos na mesma intensidade, na fala dos outros três sujeitos.

Para o par S-A (da mesma forma que para os pares M-A e E-A), não há diferença entre as vogais nas posições pré-tônicas 1 e 2, o que pode ser visto tanto em relação à média das

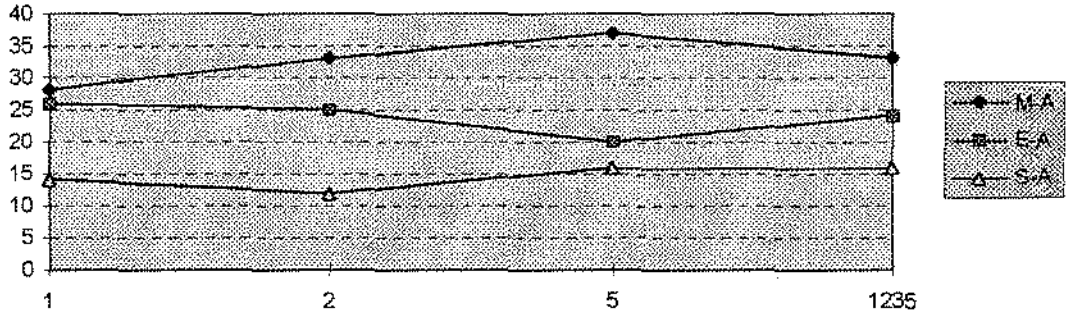
diferenças de duração (respectivamente, -7 e - 6 ms, ANEXO C, Tabela 3) como à variabilidade (respectivamente, 14 e 12 ms, ANEXO C, Tabela 3), apontando para uma certa estabilidade das vogais pré-tônicas. Para uma melhor visualização da estabilidade das vogais pré-tônicas 1 e 2, entre os pares, observe-se os gráficos abaixo, para os valores de duração (barras) e variabilidade (linhas):





Uma comparação visual entre os Gráficos 1 (par M-A) e 3 (par S-A) e os Gráficos 2 (par E-A) e 3 (par S-A) sugere que, no que concerne às médias das diferenças de duração dos segmentos acústicos vocálicos em cada posição, o par E-A (criança de 4 anos e 9 meses

e pesquisadora) assemelha-se mais ao par S-A (professora-pesquisadora) que o par M-A (criança de 4 anos e 1 mês-pesquisadora). Para os três pares, a variabilidade das vogais pertencentes às posições ou agrupamento de posições não-acentuadas é muito semelhante. Para as posições 1, 2 e 5 e para o agrupamento de posições 1235, para o par E-A, temos respectivamente 26, 25, 20 e 24 ms (ANEXO C, Tabela 2) e para o par S-A, respectivamente, 14, 12, 16 e 16 ms (ANEXO C, Tabela 3), havendo, portanto, uma diminuição da variabilidade das posições não-acentuadas do par M-A (quando tínhamos respectivamente 28, 33, 37 e 33 ms, ANEXO C, Tabela 1) para o par S-A, como pode ser visto por meio do gráfico de linhas abaixo:



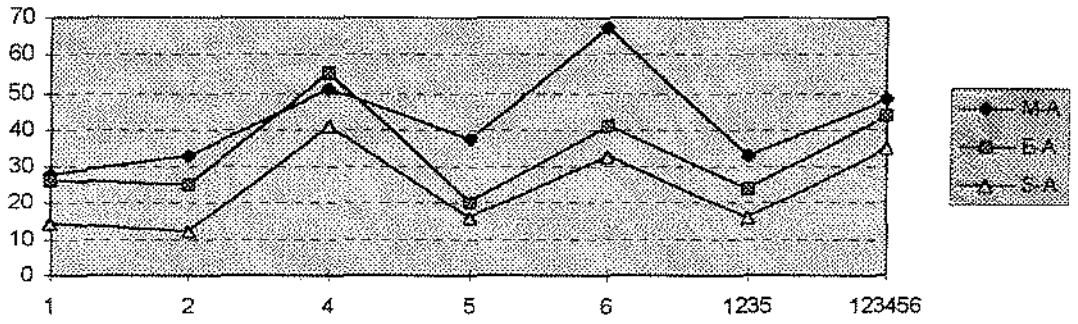
Na comparação do Gráfico 1 (par M-A) com os Gráficos 2 (par E-A) e 3 (par S-A), observa-se ainda que, na posição 4, referente a acento lexical e que inclui também a maior proeminência da sentença, apenas a criança M apresenta uma média das diferenças de duração maior que a pesquisadora (valor positivo). Como essa criança apresenta um maior alongamento das vogais em todas as posições não-acentuadas, a média das diferenças de

duração das vogais na posição 4, a seu favor, parece refletir uma espécie de alongamento compensatório, ou seja, um alongamento até certo ponto necessário para demarcar as vogais que participam da formação dos picos de duração, ou proeminências das palavras e das sentenças, em relação às demais vogais. Vale notar que a média das diferenças de duração das vogais, na posição 4, aumenta em módulo, na fala da pesquisadora, da comparação com a criança M para a comparação com o adulto S (respectivamente, 10 ms, -16 ms e -26 ms, para os pares M-A, E-A e S-A). Quanto à variabilidade absoluta ou desvio-padrão, na posição 4, ela é ligeiramente maior para o par E-A (55 ms) em comparação com o par M-A (51 ms), mas menor para o par S-A (41 ms).

Na posição 6, observa-se também uma diminuição da variabilidade absoluta ou desvio-padrão do par M-A para o par S-A (respectivamente, 67 ms, 41 ms e 33 ms, para os pares M-A, E-A e S-A). Entretanto, a essa diminuição da variabilidade não corresponde uma diminuição das médias das diferenças de duração das vogais, as quais são muito próximas entre os pares (-63 ms, -70 ms e -69 ms, respectivamente, para os pares M-A, E-A e S-A). A diminuição da variabilidade absoluta ou desvio-padrão entre os pares, do par formado pela criança menor (sujeito M) para o par formado pela professora (sujeito S), pode ser sugestiva de que se trata de um recurso estilístico cuja adoção, ou não, ainda está por ser estabilizada na fala das crianças, isso porque, apesar de a professora não fazer uso desse recurso da mesma forma que a pesquisadora, a variabilidade entre elas é a menor entre os pares. Vale notar que apenas para o par M-A a variabilidade absoluta é maior na posição 6 que na posição 4 (respectivamente, 67 ms para a primeira e 51 ms para a segunda).

Quando os pares foram comparados para todas as vogais (agrupamento 123456), somente a criança M mostrou uma média das diferenças da duração de segmentos acústicos vocálicos maior que a pesquisadora (22 ms, ANEXO C, Tabela 1), sendo que a pesquisadora apresenta vogais mais longas que a professora (-17 ms, ANEXO C, Tabela 3) e que a criança E (-10 ms, ANEXO C, Tabela 2). Esse fato pode ser devido ao maior alongamento, na fala da pesquisadora, das vogais tanto em posição tônica (posição 4) como em posição final de sentença (posição 6), o que parece caracterizar seu estilo de fala clara (LINDBLOM, 1990, KENT e READ, 1992). Porém, é interessante considerar também a diferença de variabilidade absoluta ou desvio-padrão entre os pares no agrupamento de todas as vogais (123456), diferença essa que diminui da criança M para a professora (respectivamente, 48 ms, 44 ms e 35 ms), o que sugere uma queda da variabilidade absoluta para as durações vocálicas, em função da idade, conforme previsão da hipótese da maturação neuromotora, agora, livre das distorções que a falta de controle de variáveis lingüísticas causara a experimentos anteriores. Isso pode ser confirmado pela observação das curvas de desvios-padrão ou variabilidade absoluta, sobrepostas aos gráficos de barra, em termos de suas configurações gerais, as quais não somente assemelham-se bastante para os três pares, mas apresentam um abaixamento como um todo, do par M-A para o par S-A (para todas as posições, com exceção da posição 4, no par E-A), como pode ser apreciado no gráfico abaixo:

Curvas de variabilidade absoluta para as vogais dos três pares nas várias posições e agrupamentos de posições de acento

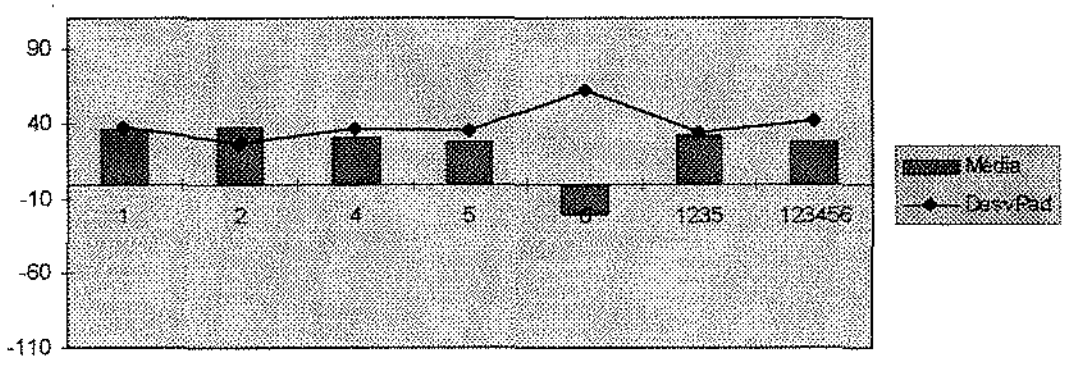


Os pontos de maior variabilidade dentro da curva de cada par são justamente aqueles correspondentes às posições de acento lexical, e que inclui a maior proeminência da sentença, e de alongamento final (posições 4 e 6, respectivamente), o que está de acordo com BARBOSA e BAILLY (1994). Vale lembrar que, para os três pares, a diferença entre os sujeitos nas posições 4 e 6 é sempre significativa (ANEXO C, Tabela 1, Tabela 2 e Tabela 3).

2.1.2 Duração e variabilidade de segmentos acústicos consonantais

Como será visto na seqüência, a análise dos dados de duração e variabilidade das consoantes, tanto entre os sujeitos de cada par como entre pares, não pôde deixar de levar em conta o que já fora visto para as vogais.

Gráfico 4 - Médias e desvios-padrão das diferenças de duração para os segmentos acústicos consonantais do par M-A



Assim, a observação do par M-A para as médias das diferenças de duração das consoantes, por meio do Gráfico 4, mostra, da mesma forma que para as vogais, que a criança M apresenta consoantes mais longas em todas as posições (ANEXO D, Tabela 4), exceto na posição 6, referente à consoante da sílaba final da sentença, a qual também é alongada pela pesquisadora, apesar de a diferença entre o alongamento da vogal e da consoante da sílaba final ser bastante grande (-63 ms para a vogal e -21 ms para a consoante), o que aponta na direção de que a vogal é o elemento que pode ser mais distendido durante o alongamento. Pesquisadora e criança M (4 anos e 1 mês) são significativamente diferentes para as consoantes em todas as posições e agrupamentos de posições de acento (ANEXO D, Tabela 4).

No que concerne às posições 1 e 2, referentes às pré-tônicas, da mesma forma que para as vogais, observa-se que a média das diferenças de duração das consoantes nas duas posições é praticamente a mesma (respectivamente, 36 e 37 ms, ANEXO D, Tabela 4), mas

há uma diminuição da variabilidade absoluta ou desvio-padrão, da posição 1 para a posição 2 (respectivamente, 37 e 26 ms, ANEXO D, Tabela 4). Poder-se-ia cogitar que, na fala da criança M, começa a haver uma estabilização das consoantes pertencentes à segunda pré-tônica (posição 2), ou ainda que as consoantes pertencentes à primeira pré-tônica variam mais por estarem muito sujeitas a fatores que interferem em sua produção, em início de enunciado, como forma de demarcar o limite de uma fronteira principal (FOUGERON e KEATING, 1997). Note que, na posição 1, são encontradas apenas consoantes oclusivas surdas, cujos *VOT*s (*Voice Onset Time*) podem ser afetados pela maior ou menor força utilizada no ataque vocal do sujeito, no início da sentença. Isso pode ser visto quando ocorre mais de um estouro ou, até mesmo, quando há uma certa quantidade de aspiração que se segue ao estouro.

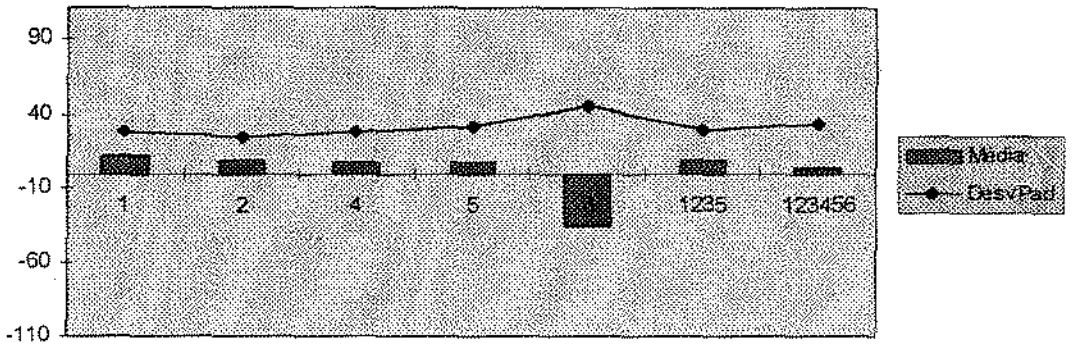
Contrariamente ao que foi observado na análise das vogais para o par M-A (Gráfico 1, ANEXO C, Tabela 1), há uma diminuição da média das diferenças de duração das consoantes na posição 5 (44 ms para as vogais e 27 ms para as consoantes), ou seja, das consoantes pertencentes a pós-tônicas e clíticos, sendo que parece não haver mudança da variabilidade absoluta ou desvio-padrão (para as vogais, 37 ms, e 34 ms para as consoantes). Isso faz supor que, na fala da criança M, as consoantes na posição 5 (pós-tônicas e clíticos) começam a ser reduzidas antes das vogais, apesar de nem por isso as primeiras variarem significativamente menos que as últimas. Porém, ao mesmo tempo, é importante ressaltar que talvez o controle das consoantes em posição pós-tônica pareça ocorrer antes daquele das vogais na mesma posição devido ao fato, já comentado anteriormente, de que, em nossos dados, as vogais na posição pós-tônica (que ocupam a posição 5 juntamente com

vogais pertencentes a clíticos) podem estar veiculando informação sobre fronteiras menores dentro da sentença, como aquelas entre os sintagmas nominal e verbal.

Observa-se, ainda, que as variabilidades absolutas ou desvios-padrão de vogais e consoantes, pertencentes às posições ou agrupamentos de posições não-acentuadas, são muito semelhantes na fala da criança M (28 ms, 33 ms, 37 ms e 33 ms, para as vogais das posições 1, 2, 5 e do agrupamento de posições 1235, e, respectivamente, 37 ms, 26 ms, 34 ms e 33 ms, para as consoantes).

Na posição 4, a de acento lexical, incluindo a maior proeminência da sentença, observa-se, no par M-A (Gráfico 4), que a média das diferenças de duração das consoantes é maior que a das vogais (respectivamente, 31 ms e 10 ms), enquanto para a variabilidade absoluta ou desvio-padrão ocorre o contrário, a variabilidade é maior para as vogais (respectivamente, 51 ms e 36 ms, para vogais e consoantes). Considerados em conjunto, esses resultados parecem sugerir que, na posição de acento, que aqui inclui o acento lexical e a maior proeminência da sentença, a criança M domina primeiramente a produção da vogal, o que se reflete numa menor média das diferenças de duração na comparação com a pesquisadora, ao mesmo tempo que, para a realização do acento, a vogal sofre maior variação (BARBOSA e BAILLY, 1994). Ainda para a criança M, vale ressaltar que, com exceção da posição 2, quase não há diferença entre a variabilidade absoluta das consoantes pertencentes à posição de acento (posição 4, 36 ms) e às posições não-acentuadas (posições 1, 2 e 5, respectivamente, 37, 26 e 34 ms) ou agrupamento delas (1235, 33 ms) (ANEXO D, Tabela 4).

Gráfico 5 - Médias e desvios-padrão das diferenças de duração para os segmentos acústicos consonantais do par E-A

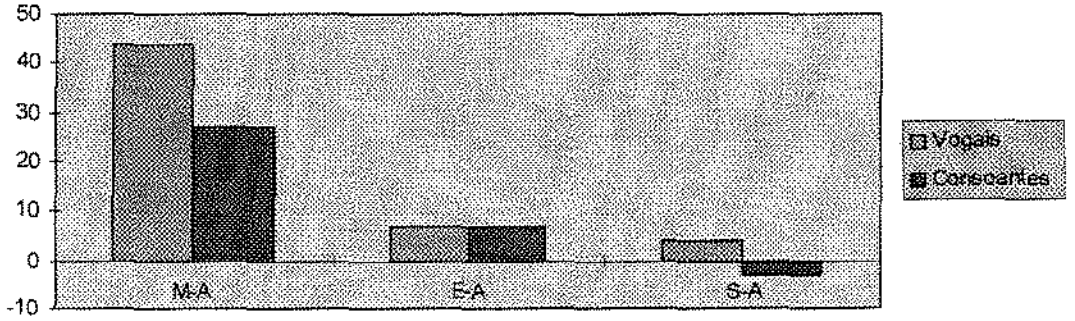


O Gráfico 5, referente ao par E-A, mostra uma situação um pouco diferente daquela observada para as vogais, uma vez que, para as consoantes, a criança E (4 anos e 9 meses) apresenta durações maiores (valores positivos) que a pesquisadora em todas as posições de acento, exceto na posição 6 (consoante da sílaba final da sentença), sendo que em todas as posições ou agrupamentos de posições foi encontrada uma diferença significativa entre criança E e pesquisadora (ANEXO D, Tabela 5). É interessante notar que, entre as posições não-acentuadas, ao contrário do que ocorreu na análise das vogais para esse par, a média das diferenças de duração das consoantes é maior nas posições 1 e 2 (respectivamente, 12 ms e 9 ms, ANEXO D, Tabela 5, contra 3 ms e -1 ms, para as vogais nas mesmas posições, ANEXO C, Tabela 2) que na posição 5 (7 ms, tanto para consoantes como para vogais, ANEXO D, Tabela 5, e ANEXO C, Tabela 2), ainda que a variabilidade seja quase a mesma para consoantes e vogais nas posições 1 e 2 (27 e 24 ms, para consoantes, e 26 e 25 ms,

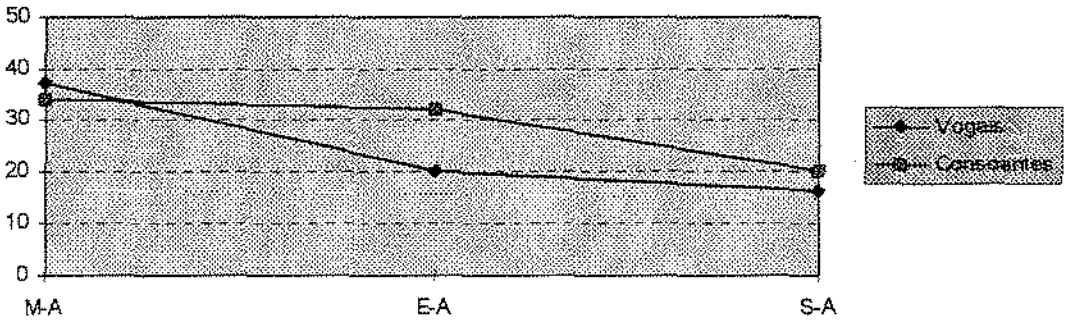
para vogais). No entanto, na posição 5, apesar de a média das diferenças de duração ser a mesma para consoantes e vogais (7 ms), a variabilidade é maior para as primeiras (32 ms para as consoantes contra 20 ms para as vogais). Portanto, no que concerne à criança E, parece que a redução esperada no PB para a posição pós-tônica (CÂMARA, 1969), que aqui inclui pós-tônicas e clíticos, é primeiramente consolidada para a vogal, como demonstra sua menor variabilidade. Observa-se ainda que, para o agrupamento de posições não-acentuadas, 1235 (pré-tônicas, pós-tônicas e clíticos), o par E-A apresenta média e desvio-padrão das diferenças de duração menores para as vogais que para as consoantes (respectivamente, 3 ms e 24 ms, para as vogais, e 9 ms e 29 ms, para as consoantes), o que, no que concerne à aquisição dos pontos de referência ou elementos átonos para a construção das alternâncias rítmicas (BARBOSA e BAILLY, 1994), parece apontar para uma estabilização primeiramente das vogais, ao menos na fala da criança E.

Assim, em relação à posição 5, para a criança M (4 anos e 1 mês), observa-se primeiramente a redução da consoante (27 ms contra 44 ms da vogais), apesar de a variabilidade de consoantes e vogais, nessa posição, ser muito próxima (respectivamente, 34 e 37 ms). Por outro lado, para a criança E (4 anos e 9 meses), na posição 5, obteve-se a mesma duração para vogais e consoantes, apesar de a variabilidade das consoantes ser maior (respectivamente, 32 ms para as consoantes e 20 ms para as vogais). A criança E parece estar, no geral, mais próxima do padrão adulto de produção vocálica, principalmente no que se refere à redução e menor variabilidade das vogais não-acentuadas, o mesmo não ocorrendo em relação às suas consoantes.

Médias das diferenças de duração de vogais e consoantes na posição 5



Desvios-padrão das diferenças de duração de vogais e consoantes na posição 5

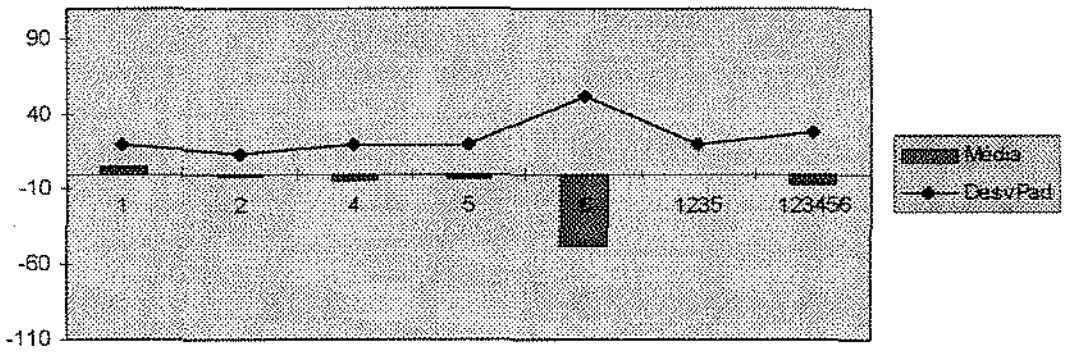


Esses achados parecem ir de encontro à hipótese de SMITH (1992) sobre a independência entre os parâmetros de duração e variabilidade, a qual nos leva ainda a supor que, em termos da produção do ritmo da fala, uma aquisição está efetivamente consolidada quando a variabilidade do segmento acústico em questão é compatível com seu grau de acentuação (em outras palavras, com a posição por ele ocupada em na cadeia de alternâncias que caracteriza o padrão rítmico da língua). Nesse sentido, o domínio ou controle da

variabilidade reflete o domínio que uma categoria lingüística (o grau de acentuação do segmento produzido) exerce sobre a realização motora.

Quando se observa o par E-A para todas as vogais e todas as consoantes (123456), encontra-se que a média das diferenças de duração é maior na fala da pesquisadora para as vogais que para as consoantes (-10 ms para vogais contra 4 ms, respectivamente). Há também uma maior variabilidade absoluta ou desvio-padrão para vogais (44 ms) em comparação às consoantes (33 ms). A maior média das vogais na fala da pesquisadora parece ser dada, como já foi visto, por seu maior alongamento das vogais nas posições 4 (acento lexical inclusive maior proeminência da sentença) e 6 (alongamento da vogal da sílaba final da sentença) e pela observação de que, nessa última posição, a vogal sofre um alongamento muito maior que a consoante. Contudo, o fato de que, no conjunto, as vogais variam mais que as consoantes parece refletir a maior flexibilidade das primeiras em relação às últimas.

Gráfico 6 - Médias e desvios-padrão das diferenças de duração para os segmentos acústicos consonantais do par S-A

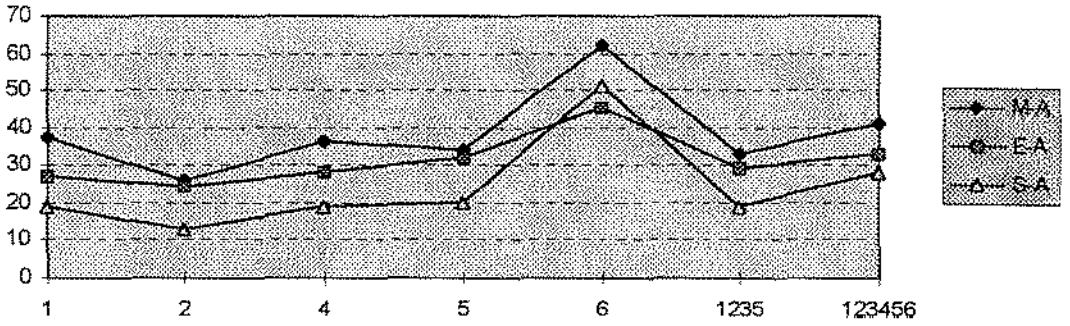


No que se refere ao par S-A (Gráfico 6), o que chama atenção é a estabilidade das consoantes que emerge da comparação entre as falas da pesquisadora e da professora. Com exceção da posição 6, na qual se observa o alongamento da consoante da sílaba final da sentença na fala da pesquisadora, as médias das diferenças de duração entre elas, para todas as posições e agrupamentos de posições de acento, apresentam valores muito baixos, sendo que quando ambas são comparadas para o agrupamento de posições não-acentuadas (1235), a média é igual a zero e não há diferença significativa entre elas (p=0.9172, ANEXO D, Tabela 6). Também não são encontradas diferenças significativas entre pesquisadora e professora para as consoantes nas posições 2 (segunda pré-tônica, com p=0.2133, ANEXO D, Tabela 6) e 5 (pós-tônicas e clíticos, com p=0.0985, ANEXO D, Tabela 6). Note que na posição 4, referente a acento lexical e que inclui a maior proeminência da sentença, há uma considerável diminuição da média das diferenças de duração dos segmentos acústicos consonantais em relação à duração de segmentos acústicos vocálicos (respectivamente, -5 ms e -26 ms), a qual é acompanhada de uma diminuição da variabilidade absoluta ou desvio-padrão (19 ms para as consoantes e 41 ms para as vogais), o que parece sugerir que na realização do acento lexical, e que aqui inclui a maior proeminência da sentença, não se observa o alongamento da consoante, sendo a vogal a principal responsável pela implementação do acento de um modo geral.

A observação das curvas de desvio-padrão dos Gráficos 4, 5 e 6, referentes à análise das consoantes para os pares M-A, E-A e S-A, evidencia novamente um padrão muito semelhante entre as três curvas, com um abaixamento das mesmas, do par M-A para o par

S-A, sugerindo uma diminuição da variabilidade absoluta ou desvio-padrão, em todas as posições, da comparação entre pesquisadora e criança menor para a comparação entre pesquisadora e professora. O único pico encontra-se sobre a posição 6 (consoante da sílaba final, a qual é alongada na fala da pesquisadora), uma vez que a consoante na posição de acento lexical (que inclui a maior proeminência da sentença, posição 4) não apresenta a mesma variação que a vogal.

Curvas de variabilidade absoluta para as consoantes dos três pares nas várias posições ou agrupamentos de posições de acento



2.1.3 Duração e variabilidade de segmentos acústicos do tamanho de sílabas

Na comparação entre os pares para os segmentos acústicos do tamanho de sílabas, foram incluídas duas novas posições, além daquelas já utilizadas na análise de vogais e consoantes. Na verdade, a primeira posição, que corresponde às vogais de início de sentença, designada por posição 0, fez com que houvesse uma subdivisão na análise do agrupamento das posições não-acentuadas, o qual passou a incluir ou não a posição 0. Isso porque, não se sabia, a princípio, até que ponto essa posição poderia interferir na análise do agrupamento de posições não-acentuadas, uma vez que os segmentos acústicos em início absoluto de sentença, nesse caso, vogais que correspondiam a artigos, constituem em si mesmos fontes de variação, devido a diferenças no ataque vocal entre os sujeitos ou à demarcação ou não, por eles, do início de uma fronteira sintática principal, como a sentença.

Na posição 0, criança M e pesquisadora somente puderam ser comparadas para dezenove segmentos acústicos, porque muitas vezes a criança M iniciava sua repetição antes de a pesquisadora ter concluído totalmente sua fala, o que tornava difícil a delimitação entre a última vogal da fala da pesquisadora e a primeira vogal da fala da criança M. Um aspecto interessante é que, apesar de as gravações terem sido monitoradas auditivamente, por meio de fones de ouvido, pelos técnicos do estúdio contratado, em muitos casos, a constatação da junção entre o final da fala da pesquisadora e o início da repetição do próximo sujeito apenas foi percebida quando da digitalização do sinal de fala.

Para o par M-A, a média das diferenças de duração das sílabas na posição 0 é -3 ms, enquanto nas vogais e sílabas na posição pré-tônica 1, as médias das diferenças de duração

são respectivamente 30 ms e 67 ms. O grande alongamento observado na fala da pesquisadora (média negativa), entre as posições silábicas 0 e 1, em relação à fala da criança M, parece sugerir que a pesquisadora também demarca a fronteira de início de sentença (FOUGERON e KEATING, 1997).

Gráfico 7 - Médias e desvios-padrão das diferenças de duração para o agrupamento de segmentos acústicos correspondentes a sílabas do par M-A

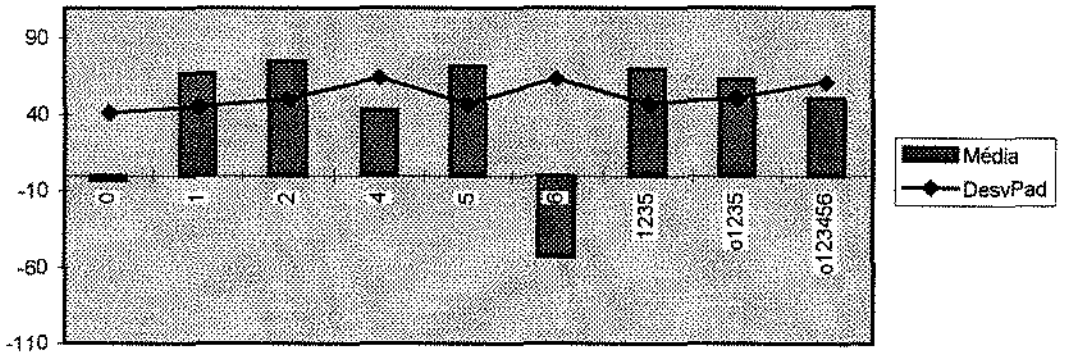


Gráfico 8 - Médias e desvios-padrão das diferenças de duração para o agrupamento de segmentos acústicos correspondentes a sílabas do par E -A

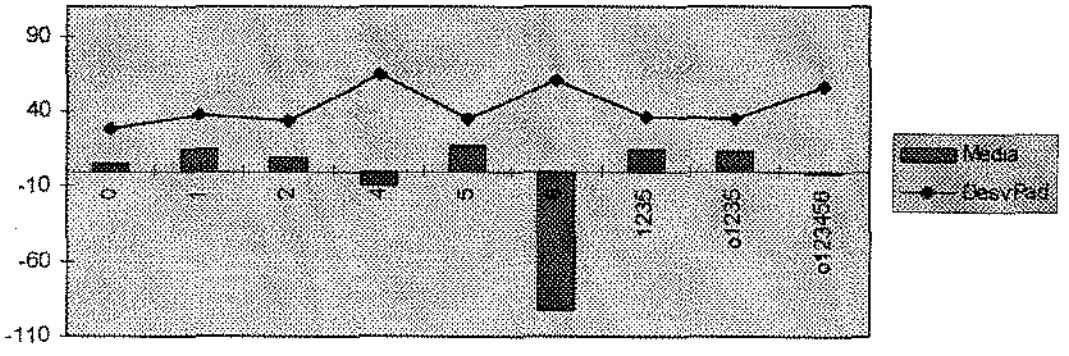
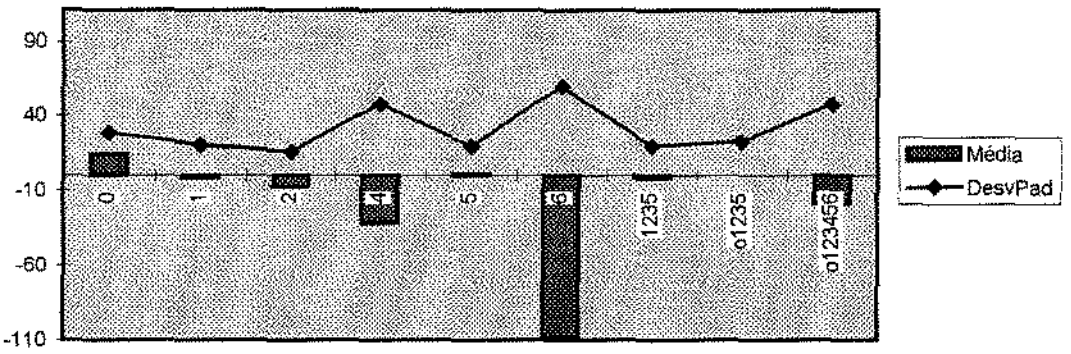


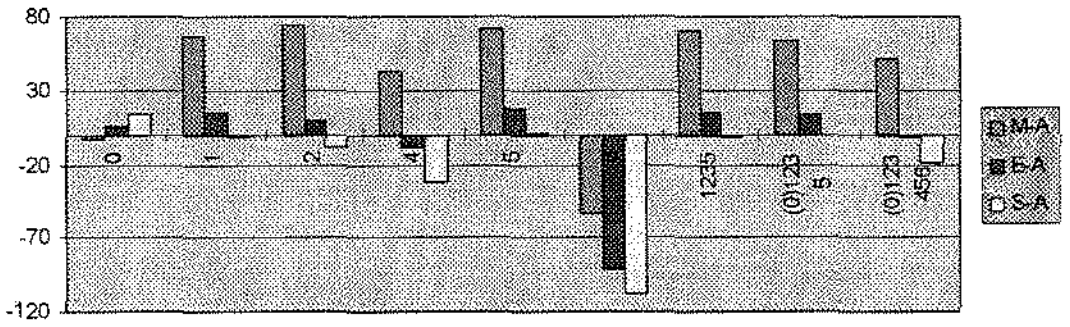
Gráfico 9- Médias e desvios-padrão das diferenças de duração para o agrupamento de segmentos acústicos correspondentes a sílabas do par S -A



No que concerne às diferenças significativas entre a criança M e a pesquisadora (Gráfico 7 e ANEXO E, Tabela 7), a única posição de acento na qual não foi encontrada diferença significativa entre elas foi a posição 0. Entre o par E-A (Gráfico 8 e ANEXO E, Tabela 8), não foram encontradas diferenças significativas para os segmentos acústicos correspondentes a sílabas pertencentes às posições: 0 (sílabas de início absoluto de sentença, com p=0.2369, ANEXO E, Tabela 8); 2 (sílabas pré-tônicas, quando p=0.1024, ANEXO E, Tabela 8); 4 (sílabas tônicas, referentes a acento lexical inclusive maior proeminência da sentença, quando p=0.0834, ANEXO E, Tabela 8); e também para o agrupamento de todas as sílabas (0123456, quando p=0.425, ANEXO E, Tabela 8). Já para o par S-A (Gráfico 9 e ANEXO E, Tabela 9), não foram encontradas diferenças significativas para a posição 1,

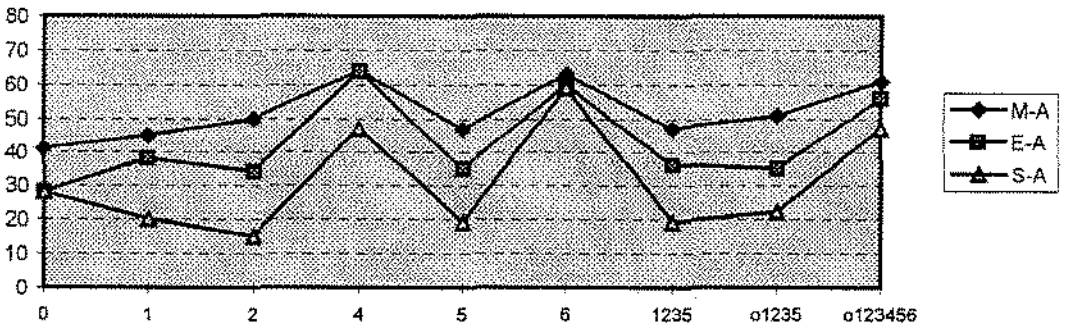
referente à primeira pré-tônica (com p=0.2554, ANEXO E, Tabela 9), para a posição 5[1] (pós-tônicas e clíticos, quando p=0.6485, ANEXO E, Tabela 9) e para os agrupamentos de posições não-acentuadas, incluindo (01235) ou não (1235) a posição de início absoluto da sentença (com p=0.7478 e p=0.0638, respectivamente, ANEXO E, Tabela 9). Portanto, quanto aos segmentos acústicos correspondentes a sílabas, os achados para o par E-A apontam na direção de que a criança E (4 anos e 9 meses) ainda não se assemelha ao adulto para os elementos átonos e mais estáveis do ritmo de fala, mas, ao contrário, para os pontos de maior variação, enquanto os achados para o par S-A confirmam a estabilidade da fala adulta com relação à produção dos elementos átonos. O gráficos abaixo ilustram o que acabou de ser dito, por meio da comparação entre os três pares para as médias (gráfico de barras) e os desvios-padrão (gráfico de linhas) das diferenças de duração das sílabas em cada posição de acento.

Comparação entre as médias das diferenças de duração para as sílabas, produzidas pelos pares M-A, E-A e S-A, em cada posição de acento



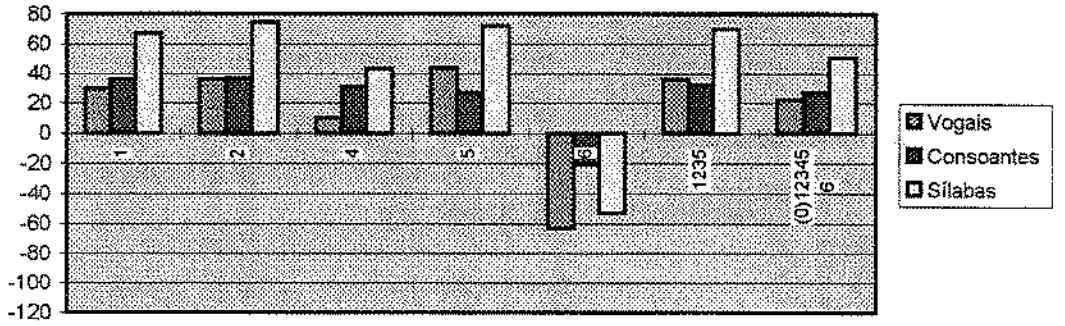
[1] Plínio A. Barbosa (comunicação pessoal) sugeriu que a falta de diferença significativa entre a pesquisadora e a professora para a média das diferenças de duração das sílabas nas posições 1 e 5 pode querer dizer que a professora *sincronizou* sua taxa de elocução à do modelo apresentado pela pesquisadora, o que parece não ter ocorrido com as crianças.

Curvas dos valores de desvios-padrão para as sílabas dos pares M-A, E-A e S-A, nas várias posições de acento

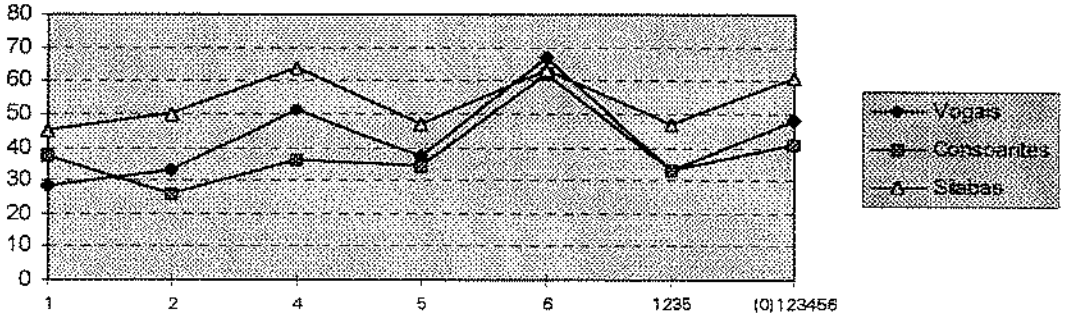


A comparação entre os valores obtidos para as sílabas, nas diferentes posições de acento ou agrupamentos de posições, e os valores obtidos separadamente para vogais e consoantes mostra que o aumento do número de segmentos acústicos na sílaba é diretamente proporcional ao aumento das médias e dos desvios-padrão das diferenças de duração nos pares formados pelas crianças e pesquisadora, mas não para o par formado por ela e pela professora. A partir dessa constatação, poder-se-ia cogitar que há uma maior coarticulação (FOWLER e SALTZMAN, 1993) (ou efeito compensatório, nas palavras de BARBOSA, comunicação pessoal) entre os segmentos acústicos que compõem a sílaba na fala adulta em comparação com a fala infantil.

Comparação entre as médias das diferenças de duração para vogais, consoantes e sílabas, produzidas pelo par M-A, em cada posição de acento



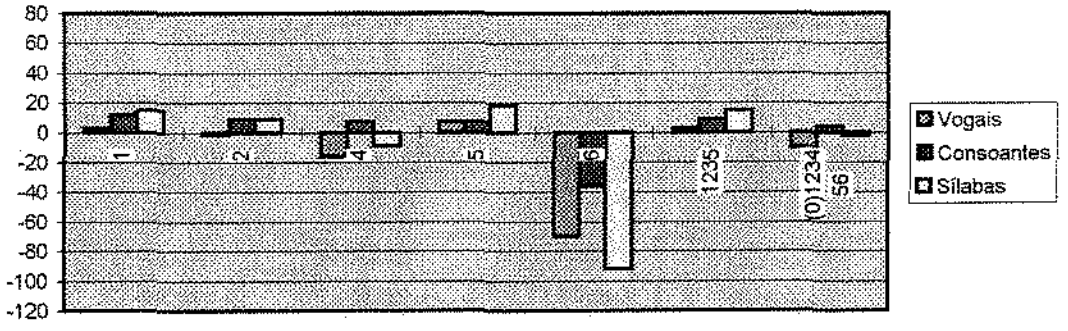
Comparação entre os desvios-padrão das diferenças de duração para vogais, consoantes e sílabas, produzidas pelo par M-A, em cada posição de acento



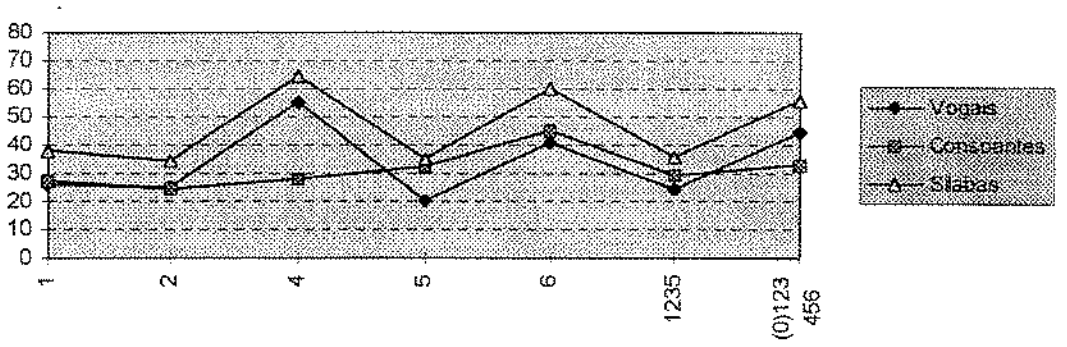
Nos gráficos acima, referentes ao par M-A, observa-se uma certa equivalência entre variabilidade absoluta ou valores de desvios-padrão somente para vogais, consoantes e sílabas na posição 6, sílaba final da sentença (respectivamente, 67 ms, 62 ms e 63 ms), com médias das diferenças de duração relativamente mais próximas entre vogais e sílabas

(respectivamente, -63 ms e -53 ms) mas não entre consoantes e sílabas (respectivamente, -21 ms e -53 ms), o que vai de encontro à hipótese já levantada de que a vogal desempenha o principal papel no alongamento da sílaba final da sentença, observado na fala da pesquisadora. Para as sílabas nas posições não-acentuadas, 1, 2 e 5, e também para o agrupamento de posições não-acentuadas, 1235 e 01235, os valores de desvio-padrão ou variabilidade absoluta saltam de algo ao redor de 30 ms, para consoantes e vogais, para 50 ms, para sílabas. No que concerne às médias das diferenças de duração nas mesmas posições, ainda para o par M-A, essas saltam de 30 a 40 ms, para vogais e consoantes, para valores em torno de 70 ms, nas sílabas. Na posição 4, referente a acento lexical e que inclui a maior proeminência da sentença, também observa-se um aumento da média das diferenças de duração e da variabilidade absoluta, no par M-A, das vogais e consoantes para as sílabas (respectivamente, médias de 10 ms, 31 ms e 43 ms, acompanhadas dos seguintes valores de desvios-padrão: 51 ms, 36 ms e 64 ms).

Comparação entre as médias das diferenças de duração para vogais, consoantes e sílabas, produzidas pelo par E-A, em cada posição de acento



Comparação entre os desvios-padrão das diferenças de duração para vogais, consoantes e sílabas, produzidas pelo par E-A, em cada posição de acento



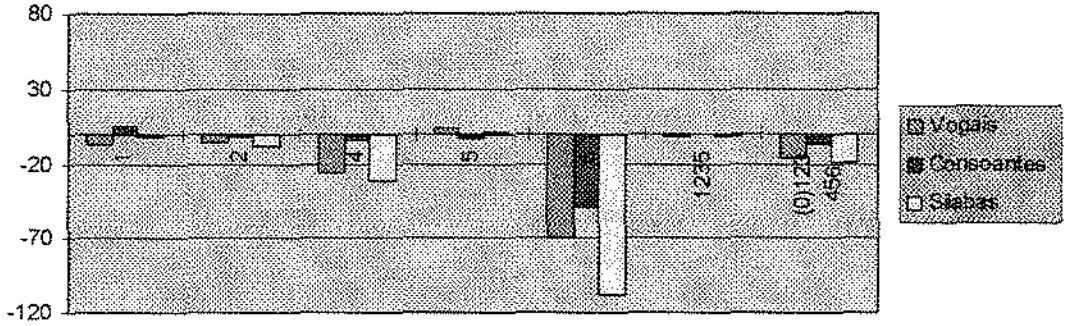
Para o par E-A, na posição 6, final de sentença, há um aumento na variabilidade de vogais e consoantes (respectivamente, 41 ms e 45 ms) para sílabas (60 ms), sendo que a média das diferenças de duração diminui das vogais para as consoantes (respectivamente, -70 e -36 ms) e aumenta das vogais para as sílabas (respectivamente, -70 ms e -92 ms). Na posição 4, de acento lexical, que inclui a maior proeminência da sentença, a variabilidade absoluta aumenta das vogais e consoantes para as sílabas (respectivamente, 55 ms, 28 ms e 64 ms), com as variabilidades absolutas de vogais e sílabas mais próximas entre si e bem maiores em relação às consoantes. Além disso, o alongamento nessa posição é ligeiramente maior, na fala da pesquisadora, nas vogais, com uma média das diferenças de duração igual a -16 ms, que nas sílabas (média das diferenças de duração igual a -9). Porém, na fala da criança E, o alongamento na posição 4 é maior nas consoantes que nas sílabas (respectivamente, 7 ms para -9 ms). Entre as posições não-acentuadas, não se encontra diferença nas médias das diferenças de duração na comparação de consoantes e sílabas das

posições 1 e 2 (respectivamente, 12 ms e 9 ms e 15 ms e 9 ms), mas sim na comparação entre vogais e sílabas para as mesmas posições (respectivamente, 3 ms e -1 ms e 15 e 9 ms), quando há um aumento considerável das médias para as sílabas, na fala da criança E. Quanto à variabilidade absoluta ou desvio-padrão, nota-se um aumento, com a variabilidade de vogais e consoantes das posições pré-tônicas 1 e 2 saltando respectivamente de 26 e 25 ms e 27 e 24 ms, para 38 ms e 34 ms, para sílabas nas mesmas posições. Esses resultados parecem reafirmar a suspeita, já levantada anteriormente, de que a criança E encontra-se mais próxima do padrão adulto de fala no que concerne às vogais.

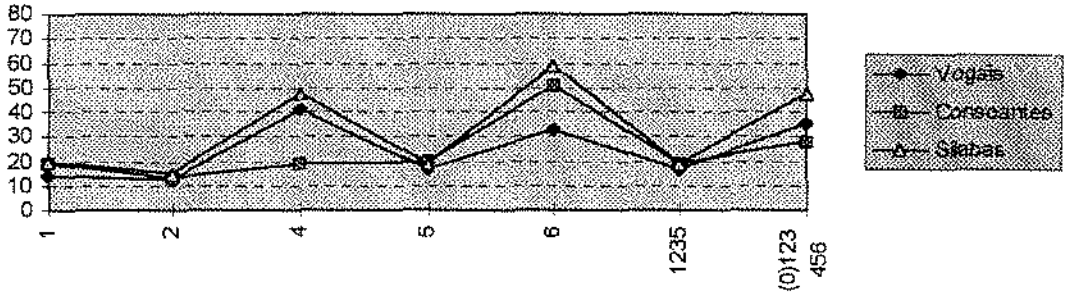
Na posição 5, de pós-tônicas e clíticos, há um aumento da média das diferenças de duração de vogais e consoantes para sílabas (para as últimas, 18 ms, enquanto para vogais e consoantes a média era de 7 ms) e um aumento da variabilidade absoluta ou desvio-padrão das vogais para as sílabas (respectivamente, 20 ms e 35 ms), mas não das consoantes para as sílabas (respectivamente, 32 ms e 35 ms), o que novamente vai de encontro à suposição de que a criança E, também na posição 5, domina primeiramente a produção das vogais. Em relação ao agrupamento de posições não-acentuadas (1235), o qual nas sílabas pode ou não incluir as vogais de início absoluto (1235 ou 01235), há um aumento tanto das médias das diferenças de duração (3 ms para vogais, 9 ms para consoantes e 15 ms para as sílabas do agrupamento 1235, e 13 ms para as sílabas do agrupamento 01235) como da variabilidade absoluta das vogais e consoantes para as sílabas (24 ms e 29 ms, para vogais e consoantes, e 36 ms e 35 ms para os agrupamentos 1235 e 01235, respectivamente). Para todas as sílabas (agrupamento de posições 0123456), observa-se uma diminuição da média das diferenças de duração (- 2ms), em relação a vogais e consoantes (respectivamente, -10 ms e 4 ms), mas

um aumento da variabilidade (56 ms para as sílabas, 44 ms para vogais e 33 ms para consoantes). Por tudo que foi exposto, parece possível concluir que a criança E aproxima-se mais da pesquisadora na produção das vogais.

Comparação entre as médias das diferenças de duração para vogais, consoantes e sílabas, produzidas pelo par S-A, em cada posição de acento



Comparação entre os desvios-padrão das diferenças de duração para vogais, consoantes e sílabas, produzidas pelo par S-A, em cada posição de acento



Como foi dito anteriormente, para o par S-A, não se observam disparidades, nas posições ou agrupamentos de posições não-acentuadas, entre as médias das diferenças de duração e os valores de desvios-padrão, das vogais e consoantes para as sílabas, em função

do aumento do número de segmentos acústicos nas últimas. O par apresenta médias e desvios-padrão das diferenças de duração muito próximos para o agrupamento de posições não-acentuadas, 1235, de vogais e consoantes (respectivamente, -2 ms e 16 ms, para vogais, e 0 e 19 ms, para consoantes), e também para os agrupamentos de sílabas, 1235 ou 01235 (respectivamente, -2 ms e 19 ms, para 1235, e 0 e 22 ms, para 01235), o que reforça a suposição de que os adultos realizam uma produção mais estável no que concerne aos elementos átonos que servem como referência à produção das alternâncias rítmicas. A mesma proximidade entre as médias e os desvios-padrão das diferenças de duração pode ser observada quando se compara separadamente as posições 1, 2 e 5 para vogais, consoantes e sílabas e encontra-se para as médias das diferenças de duração, respectivamente: -7 ms, 5 ms e -2 ms, para a posição 1; -6 ms, -2 ms e -8 ms, para a posição 2; e 4 ms, -3 ms e 1 ms, para a posição 5. No que se refere às variabilidades absolutas ou valores de desvios-padrão, tem-se, para vogais, consoantes e sílabas, respectivamente, na posição 1: 14 ms, 19 ms e 20 ms; na posição 2: 12 ms,13 ms e15 ms; e na posição 5: 16 ms, 20 ms e 19 ms.

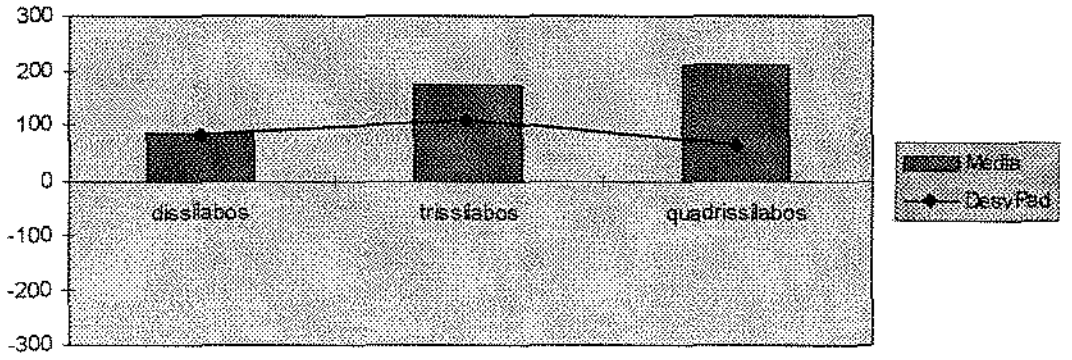
Entretanto, surgem diferenças, para o par S-A, entre a realização de vogais, consoantes e sílabas justamente quando se compara a produção desses diferentes segmentos acústicos para os picos de duração, correspondentes à posição 4 (acento lexical, incluindo maior proeminência da sentença) e à posição 6 (sílaba final da sentença, sujeita a alongamento). Na posição 4, observa-se uma maior proximidade entre os valores das médias das diferenças de duração e dos desvios-padrão de vogais e sílabas (respectivamente, -26 ms e 41 ms, para as primeiras, e -32 ms e 47 ms, para as segundas), com uma diminuição desses valores em relação às consoantes (média de -5 ms e desvio-padrão de 19 ms). Isso vai de encontro a

uma hipótese já levantada anteriormente de que, na fala adulta, a vogal parece desempenhar o papel principal na realização do acento, o que se confirma pela proximidade entre os valores de média e variabilidade absoluta de vogais e sílabas, ao passo que as consoantes são elementos mais pontuais. Vale lembrar ainda que, no par S-A, as consoantes na posição 4 apresentam média e desvio-padrão das diferenças de duração semelhantes às posições não-acentuadas 1, 2, e 5. Já, na posição 6, observa-se que, entre vogais e consoantes, há uma diminuição das médias das diferenças de duração, sempre a favor da pesquisadora, que é quem realiza o alongamento (ou o maior alongamento) da sílaba final da sentença. Assim, a média passa de -69 ms para -49 ms, apesar de ocorrer o contrário em relação ao desvio-padrão, ou seja, ele aumenta das vogais (33 ms) para as consoantes (51 ms). Talvez isso possa ser explicado pelo fato de haver diferenças entre pesquisadora e professora quanto ao alongamento da consoante da sílaba final. Por outro lado, há um aumento tanto da média como do desvio-padrão das diferenças de duração para as sílabas na posição 6 (respectivamente, -108 ms e 59 ms). Os adultos parecem concordar com a previsão (BARBOSA e BAILLY, 1994) de que os locais de acento são os locais de maior variação em termos da produção do ritmo de fala. Quando o par S-A foi comparado para os valores das médias e dos desvios-padrão das diferenças de duração de todas as vogais, todas as consoantes e todas as sílabas, os menores valores foram obtidos para as consoantes (média de -7 ms e desvio-padrão de 28 ms), havendo uma maior proximidade entre vogais e sílabas (respectivamente, -17 ms e 35 ms para vogais, e -19 ms e 47 ms para sílabas).

2.1.4 Duração e variabilidade de segmentos acústicos do tamanho de palavras

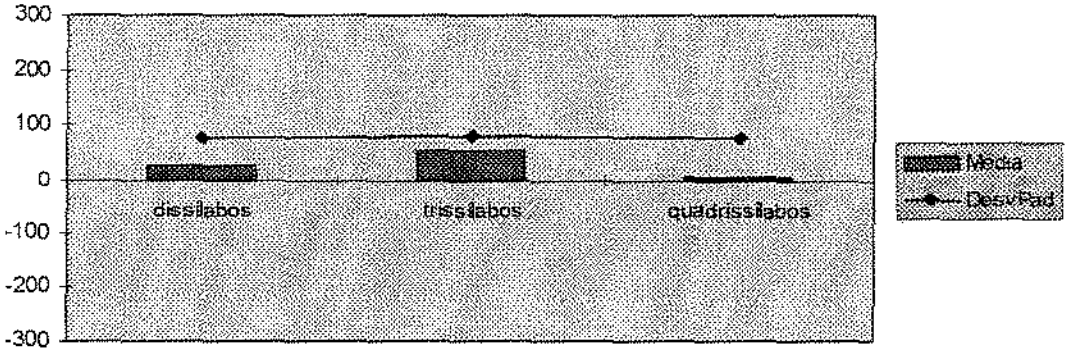
Os sujeitos em cada par foram ainda comparados para os agrupamentos de segmentos acústicos correspondentes a palavras. Da mesma forma que para as sílabas, isso ocorreu somente quando fora possível medir todos os segmentos acústicos de cada palavra para ambos os sujeitos do par. A comparação foi feita para palavras paroxítonas, cujas extensões variavam de duas a quatro sílabas. Além disso, foi feita uma divisão entre as palavras que ocorriam em posição não-final e aquelas que pertenciam à posição final de enunciado. Vale notar que a categoria gramatical à qual as palavras pertenciam não foi controlada. Assim, em posição não-final, entre os dissílabos paroxítonos (ANEXO F), encontram-se substantivos (banho, rosa e cueca), verbos (late, chora, planta, bate, toca e anda), preposição (para) e advérbio (muita), mas com uma maior quantidade de verbos. Já entre os trissílabos paroxítonos, na mesma posição, há substantivos (cachorro, palhaço, macaca e macaco) e verbos (mergulha, segura e pedala) e, entre os quadrissílabos, somente substantivos (cachorrinho, palhacinha, elefante, cachorrada e macacada). No que concerne às palavras em posição final (ANEXO F), ocorrem apenas substantivos, em cada uma das extensões vocabulares analisadas.

Gráfico 10 - Médias e desvios-padrão das diferenças de duração do agrupamento de segmentos acústicos correspondentes a palavras em posição não-final para o par M-A



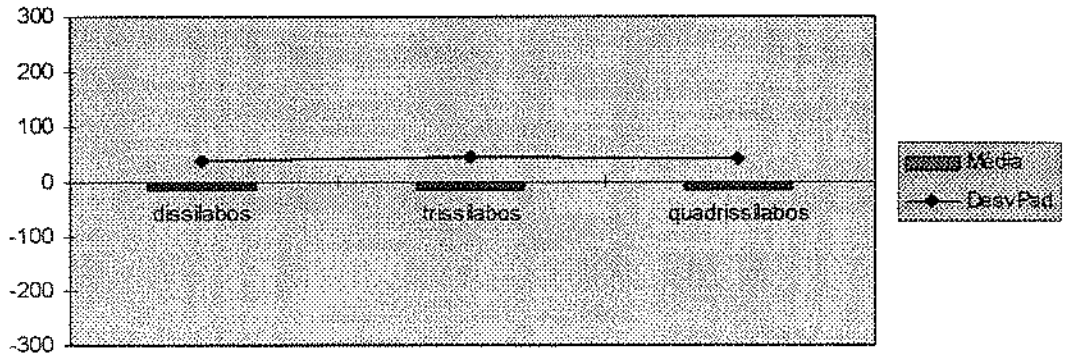
Assim, o Gráfico 10 (ANEXO G, Tabela 10), referente a dissílabos, trissílabos e quadrissílabos em posição não-final, para o par M-A, mostra que a média das diferenças de duração entre os sujeitos do par aumenta na fala da criança M (valores positivos) em função do aumento do número de sílabas da palavra (87 ms para dissílabos, 174 ms para trissílabos e 213 ms para quadrissílabos, ou seja, não há efeito compensatório, conforme BARBOSA, comunicação pessoal), enquanto a variabilidade absoluta é maior para trissílabos, seguida por dissílabos e quadrissílabos (respectivamente 110 ms, 84 ms e 64 ms). Os sujeitos do par diferem significativamente entre si para as três extensões vocabulares analisadas.

Gráfico 11 - Médias e desvios-padrão das diferenças de duração do agrupamento de segmentos acústicos correspondentes a palavras em posição não-final para o par E-A



Para o par E-A, Gráfico 11 (ANEXO G, Tabela 11), a média das diferenças de duração entre os sujeitos do par é menor para quadrissílabos, sendo seguida por dissílabos e trissílabos (respectivamente, 7 ms, 27 ms e 53 ms), não havendo diferença na variabilidade absoluta ou valores de desvios-padrão em função da extensão vocabular (75 ms, 78 ms e 75ms, respectivamente, para dissílabos, trissílabos e quadrissílabos). Criança E e pesquisadora diferem significativamente entre si apenas para trissílabos ($p=0.007$).

Gráfico 12 - Médias e desvios-padrão das diferenças de duração do agrupamento de segmentos acústicos correspondentes a palavras em posição não-final para o par S-A



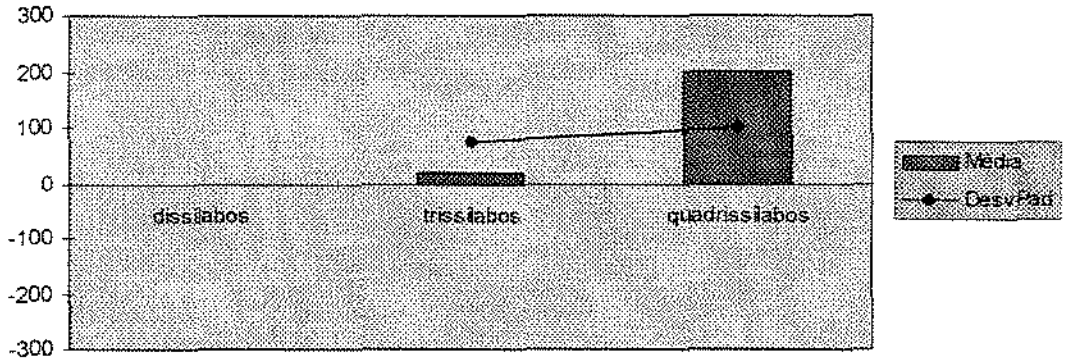
Finalmente, na análise do par S-A (Gráfico 12, ANEXO G, Tabela 12), observa-se que a média das diferenças de duração entre os sujeitos do par é sempre negativa, o que sugere durações maiores na fala da pesquisadora. Entretanto, as variações na extensão vocabular não resultam em diferenças nem entre as médias das diferenças de duração (respectivamente, -17 ms, -15 ms e -13 ms, para dissílabos, trissílabos e quadrissílabos) nem entre as variabilidades absolutas ou valores de desvios-padrão (respectivamente, 36 ms, 46 ms e 41 ms, para dissílabos, trissílabos e quadrissílabos), apesar de haver uma diferença um pouco maior entre a variabilidade de dissílabos, de um lado, e trissílabos e quadrissílabos, de outro. Pesquisadora e professora diferem significativamente entre si apenas para dissílabos ($p=0.0282$).

Uma consideração a ser feita sobre os dados expostos acima refere-se à observação de que, nos agrupamentos de segmentos acústicos correspondentes a sílabas, para os adultos, o aumento do seu número não implicou um aumento da média das diferenças de duração ou

da variabilidade absoluta em relação aos segmentos acústicos considerados isoladamente (vogais ou consoantes), principalmente aqueles em posições átonas. Da mesma forma, na análise das palavras, as diferenças de extensão vocabular não implicaram diferenças nas médias ou nos desvios-padrão das diferenças de duração. Poder-se-ia supor que, na fala adulta, há compensação a nível articulatório, ou seja, deve haver uma maior compressão de segmentos acústicos em determinados locais, o que, por sua vez, gera uma maior quantidade de coarticulação (FOWLER e SALTZMAN, 1993). A questão da maior ou menor quantidade de coarticulação envolvida na produção da fala provavelmente liga-se à maior ou menor estabilidade do controle neural que subjaz à realização da fala a nível motor, e reflete também uma questão relacionada ao próprio *timing*, uma vez que o fenômeno da coarticulação prevê tanto a modificação de gestos articulatórios em função dos gestos que a eles se seguem (coarticulação antecipatória), quanto o inverso, ou seja, a modificação de gestos articulatórios pela influência dos gestos que lhes precedem (coarticulação perseveratória).

Ainda quanto às palavras em posição não-final, para os pares M-A e E-A, há uma discrepância entre os dados da análise de quadrissílabos, de um lado, e dissílabos e trissílabos, de outro. Para o par M-A, os quadrissílabos apresentam a menor variabilidade absoluta ou desvio-padrão. Já para o par E-A, os quadrissílabos apresentam a menor média das diferenças de duração.

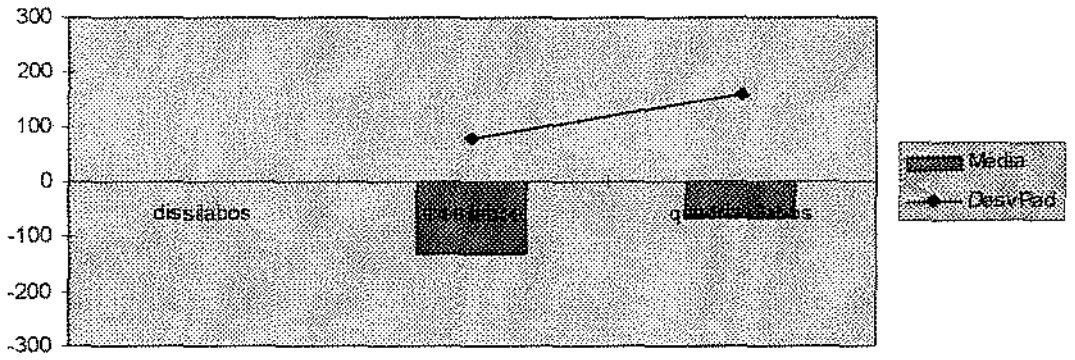
Gráfico 13 - Médias e desvios-padrão das diferenças de duração do agrupamento de segmentos acústicos correspondentes a palavras em posição final para o par M-A



No que concerne às palavras em posição final da sentença, não foi possível comparar os pares M-A e E-A para dissílabos, uma vez que para ambos havia menos de cinco ocorrências. Com relação a trissílabos e quadrissílabos, na posição final, a comparação entre criança M e pesquisadora (Gráfico 13, ANEXO H, Tabela 13) mostra um aumento considerável na média das diferenças de duração dos primeiros para os segundos. Vale notar que para ambos a média é sempre positiva, ou seja, para as palavras em final de sentença a criança M alonga mais que a pesquisadora, o que não ocorria para segmentos acústicos considerados isoladamente (vogais e consoantes) ou agrupados em unidades menores que a palavra (ou seja, em sílabas). Há também um aumento da variabilidade, nesse par, de trissílabos para quadrissílabos (respectivamente, 74 ms e 98 ms), apesar de a discrepância entre as médias das diferenças de duração ser muito maior (20 ms para trissílabos e 200 ms

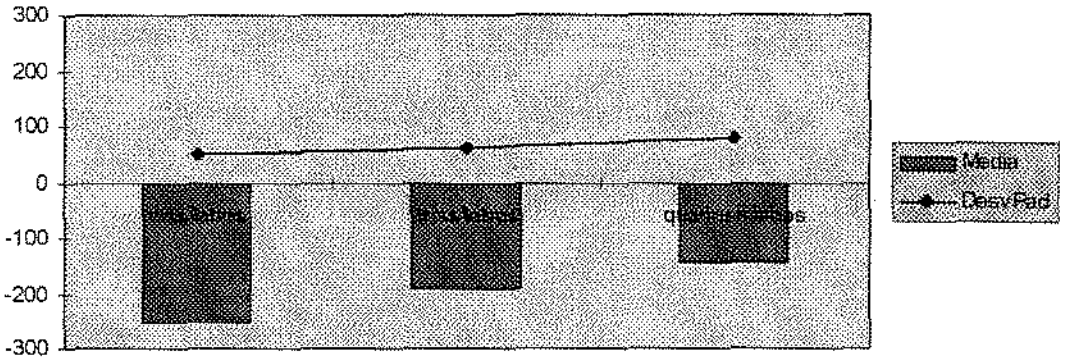
para quadrissílabos). Não há diferença significativa entre criança M e pesquisadora para trissílabos em posição final de sentença (p=0.5382), mas sim entre quadrissílabos (p=0.0103).

Gráfico 14 - Médias e desvios-padrão das diferenças de duração do agrupamento de segmentos acústicos correspondentes a palavras em posição final para o par E-A



Para o par E-A (Gráfico 14, ANEXO H, Tabela 14), a média das diferenças de duração é maior para trissílabos e ambos, trissílabos e quadrissílabos, são mais longos na fala da pesquisadora (respectivamente, -131 ms e -66 ms). Entretanto, ao contrário do que ocorria para o par nas palavras em posição não-final, há um aumento considerável da variabilidade absoluta ou desvio-padrão, dos trissílabos para os quadrissílabos (respectivamente, 77 ms e 158 ms), havendo diferença significativa entre criança E e pesquisadora apenas nos trissílabos (p=0.0001).

Gráfico 15 - Médias e desvios-padrão das diferenças de duração do agrupamento de segmentos acústicos correspondentes a palavras em posição final para o par S-A



Por fim, para o par S-A (Gráfico 15, ANEXO H, Tabela 15), as médias das diferenças de duração são sempre maiores na fala da pesquisadora, diminuindo dos dissílabos para os quadrisssílabos (-252 ms para dissílabos, -191 ms para trissílabos e - 145 para quadrissílabos). No que concerne à variabilidade absoluta ou valores de desvios-padrão, ocorre o inverso, ou seja, há um aumento dos dissílabos para os quadrissílabos (50 ms para dissílabos, 61 ms para trissílabos e 77 ms para quadrissílabos). Professora e pesquisadora são significativamente diferentes para as três extensões vocabulares analisadas (p=0.0001 nos três casos).

2.2 Comparações entre os sujeitos para o teste estatístico de Duncan

As comparações feitas por meio do teste de Duncan (MONTGOMERY, 1996) tiveram como objetivo testar as hipóteses levantadas a partir do teste t realizado entre os sujeitos de

cada par. No teste de Duncan, os sujeitos ou pares de sujeitos foram comparados para as durações dos segmentos acústicos isolados (vogais e consoantes) e para o agrupamento de segmentos acústicos correspondentes às sílabas de cada sentença. Os sujeitos individualmente ou os pares de sujeitos são agrupados sob uma mesma letra se não há diferença significativa entre eles e, sob letras diferentes, se diferenças emergem. O nível de significância previamente fixado foi de 5%.

Quando as crianças (M e E) e a professora (sujeito S) foram comparadas entre si para as durações dos segmentos acústicos vocálicos e consonantais dentro de cada sentença (QUADRO 1), não foram encontradas diferenças significativas entre criança E e professora, mas sim entre ambas e criança M, para quinze das dezessete sentenças. Em uma sentença (O domador bate o chicote), não houve diferença significativa entre as crianças, mas entre elas e a professora. Em outra sentença (O macaco segura a banana), criança M, criança E e professora diferiram significativamente entre si. É interessante notar que, apesar de, na maioria das sentenças, a criança E ter sido agrupada com a professora (mesma letra), quando comparadas para os segmentos acústicos de todas as sentenças, criança M, criança E e professora diferiram significativamente entre si.

Quadro 1: Diferenças de duração entre criança M, criança E e professora para os segmentos acústicos vocálicos e consonantais de cada sentença

Criança M	Criança E	Professora	Sentenças
A	B	B	O cachorro late para o palhaço
A	B	B	O cachorrinho chora de medo
A	B	B	A palhacinha não quer mamadeira
A	B	B	O elefante planta bananeira
A	A	B	O domador bate o chicote
A	B	B	O elefante dá banho no palhaço

A	B	B	O palhaço mergulha na bacia
A	B	B	A macaca toca chocalho
A	B	B	O palhaço cor de rosa planta bananeira
A	B	B	A palhaçada dá muita risada
A	B	B	O macaquinho anda de bicicleta
A	B	B	A cachorrada quer osso
A	B	B	A macacada faz barulho
A	B	B	O palhaço Chapolin segura a calça
A	B	B	O elefantinho pedala a motoca
A	B	B	A cueca do palhaço tem muita bolinha
A	B	C	O macaco segura a banana
A	B	C	Todas

Diferenças de duração entre criança M, criança E e professora para os agrupamentos de segmentos acústicos correspondentes às sílabas de cada sentença

Criança M	Criança E	Professora	Sentenças
A	B	B	O cachorro late para o palhaço
A	B	B	O cachorrinho chora de medo
A	B	B	A palhacinha não quer mamadeira
A	AB	B	O elefante planta bananeira
A	AB	B	O domador bate o chicote
A	B	B	O elefante dá banho no palhaço
A	B	B	O palhaço mergulha na bacia
A	B	B	A macaca toca chocalho
A	B	B	O palhaço cor de rosa planta bananeira
A	B	B	A palhaçada dá muita risada
A	B	B	O macaquinho anda de bicicleta
A	B	B	A cachorrada quer osso
A	AB	B	A macacada faz barulho
A	B	B	O palhaço Chapolin segura a calça
A	B	B	O elefantinho pedala a motoca
A	B	B	A cueca do palhaço tem muita bolinha
A	B	C	O macaco segura a banana
A	B	C	Todas

Quando comparadas para os segmentos acústicos correspondentes a sílabas (QUADRO 1), não foram encontradas diferenças significativas, entre criança E e professora, em treze sentenças, para as quais, entretanto, ambas diferiram significativamente da criança M. A criança E não diferiu de modo significativo tanto da criança M como da professora, ou

seja, agrupou-se simultaneamente a ambas, em três das dezessete sentenças (O elefante planta bananeira, O domador bate o chicote e A macacada faz barulho). Novamente, na sentença "O macaco segura a banana", os três sujeitos diferiram significativamente entre si. Quando foram comparadas para os segmentos acústicos correspondentes às sílabas de todas as sentenças, criança M, criança E e professora também diferiram significativamente.

Assim, talvez uma conclusão que pode advir do emprego do teste estatístico de Duncan é que, na comparação para os segmentos acústicos isolados (vogais e consoantes) ou agrupados (sílabas) de cada sentença, a criança E aproxima-se mais da professora que a criança M. Isso vai de encontro ao que foi visto na análise entre os sujeitos de cada par, a qual mostrou que, no geral, as durações da criança E aproximam-se mais daquelas produzidas pela pesquisadora que as da criança M, e que a aproximação é também maior entre os pares E-A e S-A que entre os pares M-A e S-A ou, até mesmo, M-A e E-A.

No que concerne aos casos que constituem minoria, nos quais ou a criança E iguala-se à criança M ou a criança E iguala-se simultaneamente à criança M e à professora, ou ainda, quando criança M, criança E e professora diferem entre si (O macaco segura a banana), não parece ser possível fazer suposições a respeito dos fatores que teriam influenciado esses agrupamentos. Poder-se-ia levantar até que ponto fatores relacionados ao significado ou à complexidade sintática das sentenças estariam intervindo nas produções infantis. No que se refere aos significados, poder-se-ia cogitar que, na sentença "O domador bate o chicote", diferenças nas produções das crianças estariam relacionadas à sua maior ou menor familiaridade com a palavra "domador", apesar de a sentença estar associada a uma figura, em cuja cena aparece o "domador". Quanto à influência de um fator sintático, tal como a

complexidade sintática da sentença, as sentenças envolvidas (O domador bate o chicote, O macaco segura a banana, O elefante planta bananeira e A macacada faz barulho) não apresentam ramificações no sintagma nominal ou no sintagma verbal. Quanto à métrica, elas variam de nove a dez sílabas, quando as sentenças do *corpus* variam de oito a quatorze sílabas, ou seja, elas não estão entre as sentenças mais longas, que são justamente aquelas que contêm ramificações em seus sintagmas. De qualquer forma, quando comparadas para o total dos segmentos acústicos consonantais e vocálicos das sentenças e para o total dos agrupamentos de segmentos acústicos (sílabas), criança M, criança E e professora diferiram significativamente entre si. Assim, uma conclusão possível da comparação dos três sujeitos entre si é a de que, apesar de a criança E estar mais próxima do padrão adulto de produção, ao menos na comparação feita dentro dos pares e, em especial, para os segmentos acústicos vocálicos, ela ainda não se equipara àquele padrão. Um outro exemplo que aponta para essa conclusão é que quando as duas crianças, de um lado, e os dois adultos, de outro, foram também comparados tanto para os segmentos acústicos consonantais e vocálicos como para os segmentos acústicos correspondentes às sílabas de todas as sentenças (QUADRO 2), os pares diferiram significativamente.

Quadro 2: Diferenças de duração entre Crianças e Adultos para os segmentos acústicos vocálicos e consonantais de cada sentença

Crianças	Adultos	Sentenças
A	B	O cachorro late para o palhaço
A	B	O cachorrinho chora de medo
A	B	A palhacinha não quer mamadeira
A	B	O elefante planta bananeira
A	B	O domador bate o chicote
A	A	O elefante dá banho no palhaço

A	A	O palhaço mergulha na bacia
A	B	A macaca toca chocalho
A	B	O palhaço cor de rosa planta bananeira
A	B	A palhaçada dá muita risada
A	A	O macaquinho anda de bicicleta
A	A	A cachorrada quer osso
A	A	A macacada faz barulho
A	B	O palhaço Chapolin segura a calça
A	B	O elefantinho pedala a motoca
A	B	A cueca do palhaço tem muita bolinha
A	B	O macaco segura a banana
A	B	Todas

Diferenças de duração entre Crianças e Adultos para os agrupamentos de segmentos acústicos correspondentes às sílabas de cada sentença

Crianças	Adultos	Sentenças
A	B	O cachorro late para o palhaço
A	B	O cachorrinho chora de medo
A	A	A palhacinha não quer mamadeira
A	B	O elefante planta bananeira
A	A	O domador bate o chicote
A	A	O elefante dá banho no palhaço
A	A	O palhaço mergulha na bacia
A	A	A macaca toca chocalho
A	B	O palhaço cor de rosa planta bananeira
A	A	A palhaçada dá muita risada
A	A	O macaquinho anda de bicicleta
A	A	A cachorrada quer osso
A	A	A macacada faz barulho
A	A	O palhaço Chapolin segura a calça
A	A	O elefantinho pedala a motoca
A	A	A cueca do palhaço tem muita bolinha
A	B	O macaco segura a banana
A	B	Todas

Ainda na comparação entre crianças (M e E) e adultos (pesquisadora e professora), para os segmentos acústicos de cada sentença, observa-se que crianças e adultos diferem de modo significativo entre si na maior parte das sentenças do *corpus*, sendo que as cinco sentenças nas quais não há diferenças significativas entre eles são: "O elefante dá banho no

palhaço", "O palhaço mergulha na bacia", "O macaquinho anda de bicicleta", "A cachorrada quer osso" e "A macacada faz barulho". Ao contrário, quando comparados para os segmentos acústicos correspondentes a sílabas (QUADRO 2), crianças e adultos igualam-se na maior parte das sentenças e diferem em apenas cinco delas (O cachorro late para o palhaço, O cachorrinho chora de medo, O elefante planta bananeira, O palhaço cor de rosa planta bananeira e O macaco segura a banana). Poder-se-ia questionar até que ponto, nesse caso, a diminuição das diferenças entre os dois grupos para os segmentos acústicos correspondentes às sílabas, dentro de cada sentença, não ocorreu pelo fato de que, por tratar-se de um agrupamento de segmentos acústicos, as diferenças entre os grupos tenderiam a anular-se, apesar de, na análise de pares por meio do teste t, ter sido mostrado que o maior número de segmentos acústicos na sílaba parece fazer diferença nos pares que envolvem as crianças, mas não no par referente aos adultos. De qualquer forma, quando comparados para o conjunto dos segmentos acústicos do tamanho de sílabas, de todas as sentenças, os dois grupos voltaram a diferenciar-se.

Na comparação das crianças entre si (QUADRO 3), elas diferiram para a maioria das sentenças, tanto quando comparadas para os segmentos acústicos isoladamente (vogais e consoantes) como quando comparadas para os agrupamentos de segmentos acústicos correspondentes às sílabas, o que foi confirmado, em cada caso, quando as crianças foram comparadas para todas as sentenças.

Quadro 3: Diferenças de duração entre Criança M e Criança E para os segmentos acústicos vocálicos e consonantais de cada sentença

Criança M	Criança E	Sentenças
A	B	O cachorro late para o palhaço
A	B	O cachorrinho chora de medo
A	B	A palhacinha não quer mamadeira
A	B	O elefante planta bananeira
A	A	O domador bate o chicote
A	B	O elefante dá banho no palhaço
A	B	O palhaço mergulha na bacia
A	B	A macaca toca chocalho
A	B	O palhaço cor de rosa planta bananeira
A	B	A palhaçada dá muita risada
A	B	O macaquinho anda de bicicleta
A	B	A cachorrada quer osso
A	B	A macacada faz barulho
A	B	O palhaço Chapolin segura a calça
A	B	O elefantinho pedala a motoca
A	B	A cueca do palhaço tem muita bolinha
A	B	O macaco segura a banana
A	B	Todas

Diferenças de duração entre Criança M e Criança E para os agrupamentos de segmentos acústicos correspondentes às sílabas de cada sentença

Criança M	Criança E	Sentenças
A	B	O cachorro late para o palhaço
A	B	O cachorrinho chora de medo
A	B	A palhacinha não quer mamadeira
A	A	O elefante planta bananeira
A	A	O domador bate o chicote
A	B	O elefante dá banho no palhaço
A	B	O palhaço mergulha na bacia
A	B	A macaca toca chocalho
A	B	O palhaço cor de rosa planta bananeira
A	B	A palhaçada dá muita risada
A	B	O macaquinho anda de bicicleta
A	B	A cachorrada quer osso
A	A	A macacada faz barulho
A	B	O palhaço Chapolin segura a calça
A	B	O elefantinho pedala a motoca
A	B	A cueca do palhaço tem muita bolinha
A	B	O macaco segura a banana

A	B	Todas

Por fim, a comparação entre pesquisadora e professora (QUADRO 4) mostrou que elas não diferem entre si para a maioria das sentenças, tanto nos segmentos acústicos vocálicos e consonantais de cada sentença como nos seus agrupamentos em sílabas, mas que para o total das sentenças, nos dois casos, elas diferem entre si. Esse fato leva-nos a cogitar que, no conjunto das sentenças, podem estar em jogo diferenças estilísticas entre ambas.

Quadro 4: Diferenças de duração entre Pesquisadora e Professora para os segmentos acústicos vocálicos e consonantais de cada sentença

Pesquisadora	Professora	Sentenças
A	A	O cachorro late para o palhaço
A	A	O cachorrinho chora de medo
A	A	A palhacinha não quer mamadeira
A	A	O elefante planta bananeira
A	A	O domador bate o chicote
A	A	O elefante dá banho no palhaço
A	A	O palhaço mergulha na bacia
A	A	A macaca toca chocalho
A	A	O palhaço cor de rosa planta bananeira
A	B	A palhaçada dá muita risada
A	A	O macaquinho anda de bicicleta
A	A	A cachorrada quer osso
A	A	A macacada faz barulho
A	A	O palhaço Chapolin segura a calça
A	A	O elefantinho pedala a motoca
A	A	A cueca do palhaço tem muita bolinha
A	B	O macaco segura a banana
A	B	Todas

Diferenças de duração entre Pesquisadora e Professora para os agrupamentos de segmentos acústicos correspondentes às sílabas de cada sentença

Pesquisadora	Professora	Sentenças
A	A	O cachorro late para o palhaço
A	A	O cachorrinho chora de medo
A	A	A palhacinha não quer mamadeira

A	A	O elefante planta bananeira
A	A	O domador bate o chicote
A	B	O elefante dá banho no palhaço
A	A	O palhaço mergulha na bacia
A	A	A macaca toca chocalho
A	B	O palhaço cor de rosa planta bananeira
A	A	A palhaçada dá muita risada
A	A	O macaquinho anda de bicicleta
A	A	A cachorrada quer osso
A	A	A macacada faz barulho
A	A	O palhaço Chapolin segura a calça
A	A	O elefantinho pedala a motoca
A	A	A cueca do palhaço tem muita bolinha
A	B	O macaco segura a banana
A	B	Todas

2.3 Os novos fatos sobre a interação entre duração e variabilidade a partir do estudo da aquisição da duração no português brasileiro

Cabe fazermos uma recapitulação dos principais resultados apresentados. Quanto à relação entre os segmentos acústicos ou conjunto de segmentos acústicos analisados (vogais, consoantes e sílabas) e as posições de tonicidade por eles ocupadas (pré-tônica, tônica e pós-tônica), na formação das alternâncias rítmicas do contorno acentual do PB (CÂMARA, 1969, ALBANO, 1996), a observação geral dos pares permite-nos dizer que:

1) A nova coleta de dados confirmou a hipótese levantada durante o estudo piloto (GAMA-ROSSI e ALBANO, 1996), uma vez que as duas crianças estudadas, de 4 anos e 1 mês e de 4 anos e 9 meses, diferem da pesquisadora principalmente quanto à produção dos segmentos acústicos nas posições não-acentuadas e, em especial, no que se refere à redução para os segmentos acústicos em posição pós-tônica, que, segundo CÂMARA (1969), é a posição que apresenta a redução mais drástica entre as não-acentuadas;

2) Quanto às pré-tônicas, na análise de cada par, foi sempre apontada uma certa estabilidade dos segmentos acústicos nessa posição, não havendo diferenças notáveis nos valores de suas médias e desvios-padrão das diferenças de duração, o que parece fortalecer a intuição mattosiana (CÂMARA, 1969), segundo a qual, todas as pré-tônicas, independentemente de seu número, recebem um mesmo valor de acento. Para o par S-A, no entanto, observa-se uma ligeira queda na variabilidade da pré-tônica 2 nos contornos de desvios-padrão obtidos para os diferentes segmentos acústicos. A queda é menor nas vogais e mais evidente nas consoantes e sílabas. A observação de que a pré-tônica 2, para o par S-A, constitui-se no ponto de menor variação dos contornos dos desvios-padrão para vogais, consoantes e sílabas faz-nos cogitar se a maior estabilidade dessa pré-tônica, ao menos no par de adultos estudado, não se deve ao fato de ela constituir o ponto que precede imediatamente ao acento lexical (BARBOSA e BAILLY, 1994). Em outras palavras, será que a menor variabilidade da pré-tônica 2 tem um papel na construção das alternâncias rítmicas do contorno duracional do PB? Esta pergunta somente poderá ser respondida a partir de futuros estudos, como o de AQUINO (em preparação);

3) Quanto à variação da pré-tônica 2 nos pares formados pelas crianças e pesquisadora, a observação dos contornos de desvios-padrão mostra que, para o par M-A, há uma ligeira subida na variação da pré-tônica 2 em vogais e sílabas, quando se observa também um aumento de duração, ou seja, há um contorno ascendente da pré-tônica 1 (ou da pré-tônica 0, nas sílabas) em direção à posição de acento (posição 4). Para o par E-A, pode-se dizer que as pré-tônicas 1 e 2 variam igualmente nas vogais e sílabas. Portanto, a comparação visual entre os três pares, por meio de seus gráficos, mostra uma pequena diferença no

modo como eles estão variando no trecho do contorno dos desvios-padrão que vai das pré-tônicas para a tônica, com uma linha ascendente da pré-tônica 1 para a posição de acento, no par M-A, uma linha reta entre as pré-tônicas 1 e 2, para o par E-A, e uma linha descendente, para o par S-A. Os três pares concordam no que diz respeito à menor variação da pré-tônica 2 nas consoantes. Se a diferença entre eles no trecho do contorno que vai da pré-tônica 1 ou 0 para a posição de acento lexical, respectivamente em vogais e sílabas, reflete ou não uma diferença de desenvolvimento na realização do contorno duracional do PB é uma questão que somente outros estudos na área conseguirão esclarecer. Resta-nos ainda especular o que faz com que a criança M apresente um aumento gradual e crescente das pré-tônicas, tanto nas durações como nos desvios-padrão de vogais e sílabas, em direção ao acento. Por tudo que foi mostrado, M parece estar num momento anterior da aquisição da duração na comparação com E. Com isso, poder-se-ia pensar que a atuação de M sugere que as crianças têm de construir, num primeiro momento, os picos de duração, o que será mais bem discutido abaixo;

4) Para vogais e sílabas, os pontos de maior variação da linha constituída pelos desvios-padrão, de cada par, nos gráficos apresentados, são aqueles das posições de acento lexical (posição 4), que inclui a maior proeminência da sentença, e dos segmentos acústicos pertencentes à última sílaba da sentença (posição 6), sujeita ao fenômeno de alongamento da sílaba final. Esses achados corroboram a predição de que os locais de acento constituem os pontos de maior variabilidade dos contornos acentuais (BARBOSA e BAILLY, 1994);

5) A linha de desvios-padrão para consoantes, em contraste com vogais e sílabas, mostra que a consoante não é distendida na posição de acento lexical (posição 4), mas apenas na sílaba final da sentença (posição 6), o que ocorre especialmente na fala da pesquisadora;

6) No PB, a vogal parece ser a principal responsável pela implementação do acento lexical (que, relembrando, em nossa análise, inclui a maior proeminência da sentença) e também pelo alongamento da sílaba final, uma vez que, aqui, a distensão observada na fala da pesquisadora é maior para vogais que para consoantes. Um aspecto interessante da análise dos segmentos acústicos vocálicos do par formado pelos adultos (S-A) é que, apesar de a média das diferenças de duração para esses segmentos acústicos, nas várias posições de acento consideradas, ser relativamente baixa, assim como seus desvios-padrão, especialmente nas posições não-acentuadas e agrupamento destas (posição 1235), professora e pesquisadora diferem significativamente para os segmentos acústicos vocálicos em todas as posições de acento. Poder-se-ia atribuir esse fato a uma diferença dos estilos de fala de ambas. Parece possível supor que a pesquisadora veicula mais seus acentos por meio das vogais. Ela apresenta vogais mais longas que a professora em todas as posições de acento, exceto na posição 5, referente à pós-tônica. Vimos que ela também apresenta vogais mais longas que a criança E na posição de acento lexical (posição 4). Note-se, ainda, que, na média das diferenças de duração de todas as vogais (ou agrupamento de vogais em todas as posições de acento, 123456), essas são mais longas para a pesquisadora que para a criança E e para a professora;

7) Vale frisar que as menores médias e os menores desvios-padrão das diferenças de duração, em todas as posições de acento analisadas para o par formado pelos adultos, foram

obtidos justamente para as consoantes, o que corrobora seu caráter mais pontual, conseqüência de um movimento balístico, já bastante discutido na literatura (RODRIGUES, 1989, BROWMAN e GOLDSTEIN, 1990);

8) Quanto às diferenças de desenvolvimento entre as crianças estudadas, a análise dos dados permite-nos fazer as seguintes considerações:

8.1) A criança M (4 anos e 1 mês) aproxima-se mais do padrão adulto de produção de fala (representado pela pesquisadora) na posição de acento lexical (ou posição 4, que inclui a maior proeminência da sentença), tanto para vogais quanto para sílabas, quando diminui a média das diferenças de duração entre M e A. Na verdade, apenas a criança M possui vogais e sílabas mais longas que a pesquisadora na posição 4, uma vez que uma característica do estilo de fala clara da última, o qual não se pode afirmar se idiossincrásico ou conseqüência de seu papel de modelo no experimento, é a produção de vogais e sílabas mais longas na posição de acento, como se vê nas comparações com a criança E e com a professora. Como a criança M apresenta durações maiores em todas as posições não-acentuadas, fomos levados a supor que M realiza um alongamento compensatório dos acentos lexicais, para demarcar picos de duração em sua fala. Assim, uma conclusão possível seria a de que a criança M (4 anos e 1 mês) já domina a implementação do parâmetro de duração para a realização do acento, embora ainda não o faça para os pontos de referência ou elementos átonos (BARBOSA e BAILLY, 1994) que participam da construção das alternâncias rítmicas do PB (CÂMARA, 1969), como demonstram suas durações bem maiores que as da pesquisadora para os segmentos acústicos em posição pré e pós-tônica. Pode ser ainda que os maiores valores de duração e desvio-padrão desses segmentos acústicos estejam de algum

modo relacionados com o fato de que, ao construir os picos de duração (ao longo de seqüências de sílabas), M alongue o que ocorre depois deles como forma de demarcar blocos (possivelmente unidades do tamanho de palavras ou sintagmas) dentro do enunciado. Novamente, isso apenas poderá ser melhor avaliado por outros estudos na área;

8.2) Na verdade, no que tange à posição de acento, a criança M apresenta um pico na linha dos desvios-padrão de vogais e sílabas, mas não de consoantes, apesar de, nessa posição, as últimas serem bem mais longas que as da pesquisadora. Ou seja, pode ser que, por já ter adquirido a implementação da posição de acento, a criança M *saiba* que o elemento a ser mais distendido ou variado é justamente a vogal. É claro que isso somente poderá ser confirmado por um estudo que considere as condutas de crianças em idades anteriores à de M (4 anos e 1 mês), ou por um estudo longitudinal, quando poderá ser mais bem avaliado se, em algum momento do desenvolvimento, as crianças alongam igualmente vogais e consoantes na realização de acentos, ou se as consoantes, devido ao caráter mais pontual e balístico dos gestos que as constituem, nunca são alongadas como as vogais. Com isso, apesar de a criança M apresentar durações maiores e valores de desvios-padrão mais altos na comparação com a pesquisadora, que a criança E e a professora, podemos dizer que, resguardadas algumas diferenças, como as já citadas, os contornos de desvios-padrão do par M-A, considerados globalmente, não são dispares em relação aos contornos obtidos para os outros dois pares na análise de cada tipo de segmento acústico. Assim, nossos dados mostram que, embora haja um rebaixamento dos valores de duração e desvio-padrão do par M-A (criança menor e pesquisadora) para o par S-A (adultos), parece que, ao menos para as duas crianças aqui estudadas, as distinções mais globais de variação entre as diferentes

posições de acento já foram adquiridas, faltando apenas ajustar, no sentido de BROWMAN e GOLDSTEIN (1989), os valores dessa variação para a faixa adulta. Esta observação difere daquela feita pela literatura precedente, resenhada no capítulo 1, a qual, por comparar grupos de crianças a um grupo de adultos para os valores de duração e desvio-padrão de segmentos acústicos isolados, sem relação com parâmetros lingüísticos, como a posição de acento ocupada pelos segmentos acústicos analisados, restringia-se simplesmente a afirmar que as crianças abaixo de 6 anos apresentavam durações maiores e eram mais variáveis que os grupos etários mais velhos, embora já se percebesse que algumas crianças do grupo etário mais novo apresentavam valores de duração e desvio-padrão dentro da faixa adulta de variação;

8.3) A criança E (4 anos e 9 meses) está mais próxima da atuação do adulto, principalmente no que se refere às médias das diferenças de duração dos segmentos acústicos vocálicos, sendo que suas vogais em posições não-acentuadas são mais reduzidas que suas consoantes e sílabas nas mesmas posições, como pôde ser visto na comparação com a pesquisadora. Vale lembrar que E não difere da pesquisadora para as vogais nas posições pré-tônicas 1 e 2 e nas sílabas pertencentes à posição pré-tônica 2, à posição 4 (acento lexical e que inclui a maior proeminência da sentença) e também para o conjunto de todas as sílabas. Ou seja, para a criança E, as vogais parecem ser os segmentos acústicos que estão caminhando mais rápido em direção à redução ou estabilidade encontrada na fala adulta para as posições não-acentuadas. Por outro lado, as sílabas aproximam-se mais do adulto quando consideradas em conjunto (0123456) e na posição de acento (4), mas não nas posições não-acentuadas.

Vale lembrar ainda que, na comparação entre M, E e a professora a partir do teste de Duncan, a criança E agrupou-se, na maior parte das vezes, com a professora;

9) Por fim, um fato muito importante, e que deverá ser discutido no próximo capítulo, à luz de modelos dinâmicos, refere-se ao achado de que o aumento do número de segmentos acústicos na sílaba causou, nos pares formados pelas crianças, um aumento tanto das médias como dos desvios-padrão das diferenças de duração, o que nos levou a supor que, na fala adulta, há uma maior coordenação com sobreposição de gestos (BROWMAN e GOLDSTEIN, 1989), levando, por vezes, a uma maior coarticulação (FOWLER e SALTZMAN, 1993), ou, ainda, a um maior efeito compensatório (BARBOSA, comunicação pessoal) entre os segmentos acústicos que constituem a sílaba. Vale lembrar que um aumento proporcional ao número de sílabas é ainda ser observado para a criança M, na análise das palavras em posição não-final, quando ocorre o aumento das médias das diferenças de duração em função do aumento do número de sílabas da palavra.

3. MODELOS DINÂMICOS NO ESTUDO DA AQUISIÇÃO DA LINGUAGEM

3.1 O problema das unidades de análise no estudo da fala infantil

KENT (1996) coloca que é necessário cautela ao empregar métodos desenvolvidos para o estudo da fala adulta na compreensão da fala infantil. Um exemplo bastante conhecido disso, segundo ele, são as transcrições fonéticas das vocalizações infantis. Conclusões contraditórias têm sido tiradas a partir dessas transcrições, indo desde a afirmação de que os infantes podem produzir todos os sons necessários a uma língua natural, até a afirmação de que os bebês apresentam um repertório fonético limitado, que aumenta gradualmente em função da idade e da experiência lingüística.

MENN (1983) argumenta que é difícil falar em fonemas na análise da produção de proto-palavras e palavras porque contrastes fonêmicos e fonéticos não se desenvolvem *pari passu*. A criança pode ser capaz de distinguir um contraste fonêmico sem que seja capaz de lidar com a fonética relevante à produção do mesmo, ou pode apresentar o controle fonético de um determinado contraste antes de ser capaz de usá-lo para veicular uma distinção

fonêmica. Um exemplo do primeiro caso, fornecido pela autora, é o das crianças falantes do inglês que contrastam o vozeamento em posição final de palavra por meio do apagamento das oclusivas vozeadas e da produção de oclusivas glotais no lugar das oclusivas não-vozeadas. O segundo caso é ilustrado também por crianças aprendizes do inglês que produzem oclusivas vozeadas em posição inicial, representadas pelos fones [b, d, g], quando ainda não adquiriram os fonemas /b, d, g/, uma vez que não possuem o contraste entre estes e os fonemas /p, t, k/. Contudo, MENN (1983) ressalta que o problema mais intrigante em relação à fala infantil é a dificuldade de lidar com o que ela chama de variação arbitrária[1] entre exemplos e repetições de um mesmo alvo fonético ou item lexical.

STUDDERT-KENNEDY (1987) afirma que a primeira unidade segmental da fala infantil não é nem o traço, nem o fonema, mas a palavra. Como evidência, ele cita que (1) as formas fonéticas dominadas em uma palavra não são necessariamente dominadas em outra; (2) assimilações de ponto e modo de articulação são encontradas entre consoantes de uma mesma palavra, mesmo quando a criança pode executar a consoante assimilada corretamente em outras palavras; e (3) as formas fonéticas de palavras individuais variam amplamente de um momento para outro (o que MENN (op. cit.) designou por "variação arbitrária").

O autor diz que uma abordagem gestual (BROWMAN e GOLDSTEIN, 1986 apud STUDDERT-KENNEDY, 1987) é preferível a uma descrição de traços da fala infantil porque fornece um conteúdo físico observável, ao passo que traços e fonemas são descritores abstratos, definidos por lingüistas, e não podem ser especificados em termos

[1] A expressão "variação arbitrária"é empregada por MENN (op. cit.) em oposição à variação encontrada na fala adulta, que, na maior parte das vezes, pode ser explicada pelo contexto fonético, pelo padrão de acentuação, por fatores morfossintáticos, sociais, etc.

articulatórios ou acústicos. Ao empregar gestos articulatórios, segundo STUDDERT-KENNEDY, lida-se com padrões de movimento no tempo e no espaço, que podem ser tratados de acordo com princípios gerais do controle motor.

A palavra é vista então como um padrão coordenado de gestos articulatórios. A especificação de seus gestos componentes é fornecida pela fala adulta, mas a organização temporal dos gestos tem de ser descoberta pela própria criança. O autor coloca que o processo de desenvolvimento fonológico é guiado pelo conflito entre facilidade articulatória e crescimento lexical. Ele explica que, quando a criança possui poucas palavras, ela tem apenas uma ou duas rotinas articulatórias, e novas palavras são adicionadas a seu repertório inicial em função da similaridade de seus padrões gestuais às rotinas articulatórias já dominadas, sendo que palavras que envolvem rotinas diferentes são evitadas. Somente quando as primeiras rotinas estão consolidadas é que novas rotinas começam a surgir, pressionadas pela necessidade, por parte da criança, de expandir seu vocabulário.

Assim, para STUDDERT-KENNEDY (op. cit.), a criança parte das rotinas articulatórias que participam do padrão coordenado de gestos das primeiras palavras, até chegar às rotinas articulatórias de cada um dos segmentos fonéticos de sua língua. O conjunto de rotinas diversifica-se, para incluir cada vez mais novas classes do modelo adulto. O autor afirma que a emergência do segmento fonético, enquanto uma unidade perceptomotora[2], faz com que o léxico adulto torne-se disponível para a criança, a partir de suas possibilidades fonéticas.

[2] STUDDERT-KENNEDY (op. cit.) considera que a primeira evidência de um vínculo perceptomotor, no infante, é dada pela imitação de um som adulto.

BROWMAN e GOLDSTEIN (1989) afirmam que, quando a criança começa a produzir palavras, unidades complexas que têm de ser distinguidas umas das outras, usando os gestos disponíveis como alicerce, o inventário fonológico passa a ser constituído. Para os autores, a comparação das primeiras produções infantis, compostas de um pequeno conjunto de gestos, com a estrutura gestual das formas adultas deixa claro que há pelo menos dois desenvolvimentos necessários para ir de uma à outra, a saber: diferenciação e ajustamento de gestos individuais e coordenação deles na palavra.

No que concerne à diferenciação de gestos individuais, os gestos balbuciados, segundo os autores, envolvem um único movimento, provavelmente grosseiro. Nos gestos de constrição oral, as línguas empregam conjuntos de articuladores que contrastam quanto ao grau (por exemplo, gestos de fechamento, gestos fricativos e gestos mais amplos ou aproximantes) e ao local (por exemplo, dental, alveolar ou pós-alveolar) da constrição formada. Assim, o gesto único da constrição do balbucio diferencia-se, posteriormente, numa variedade de gestos potencialmente contrastivos, ajustados com diferentes valores para grau e local de constrição. A partição dos contínuos físicos, constituídos pelo grau e local de constrição, em categorias discretas provavelmente é auxiliada, segundo os autores, por relações articulatórias e auditivas quânticas, ou não-lineares, do tipo daquelas propostas por STEVENS (1972, 1989 apud BROWMAN e GOLDSTEIN, 1989, ALBANO, 1999).

No que se refere ao ajustamento de gestos individuais, BROWMAN e GOLDSTEIN (1989) comentam que esse pode ocorrer mesmo quando a diferenciação entre gestos não é necessária. Por exemplo, se uma língua possui um único gesto de fechamento da ponta da

língua, seu local de constrição pode ser ajustado de acordo com um valor específico à língua, como, por exemplo, alveolar para o inglês e dental para o francês.

Quanto à coordenação entre gestos, os autores afirmam que os vários gestos ou ações primitivas do trato vocal, que compõem uma palavra, têm de ser organizados adequadamente. Inicialmente, observa-se que, embora a criança muitas vezes já conheça os gestos relevantes de uma palavra, ela os produz sem que ainda seja capaz de organizá-los como na fala adulta (GAMA-ROSSI, 1994).

3.2 Uma unidade com tempo intrínseco na explicação de processos fônicos da fala corrente: a Fonologia Articulatória

O gesto articulatório (BROWMAN e GOLDSTEIN, 1985, 1986, 1989, 1990, 1992)[3] é entendido como ação primitiva dos articuladores do trato vocal, ação essa que corresponde à formação e soltura de uma constrição, dentro de um dos subsistemas articulatórios do trato, a saber: oral, laríngeo e vélico. Enquanto caracterizações de eventos físicos, os gestos ocorrem no tempo e no espaço.

Porque os gestos têm tempo intrínseco, a coordenação entre eles é dada pela sincronização de seus estados dinâmicos[4], e não por um controle extrínseco de tempo, e é

---

[3] Para um tratamento em maior profundidade da Fonologia Articulatória (BROWMAN e GOLDSTEIN, 1985, 1986, 1989, 1990, 1992), em português, veja os trabalhos de D'ANGELIS (1998) e ALBANO (1999). Para nós, o importante é apontar a vantagem de uma unidade com tempo intrínseco, como o gesto articulatório, no estudo da fala infantil. Porém, na medida em que for necessário, para facilitar o entendimento da resenha, traremos, em notas de rodapé, alguns dos conceitos que subjazem ao modelo.

[4] A Fonologia Articulatória partiu da análise do traçado de movimentos articulatórios contínuos (FUJIMURA et al., 1973 apud BROWMAN e GOLDSTEIN, 1986) e do desenvolvimento de um arcabouço teórico para a descrição analítica e matemática de movimentos coordenados (SALTZMAN e KELSO, 1983 apud BROWMAN e GOLDSTEIN, 1986), o chamado modelo de dinâmica de tarefa.

especificada por meio da descrição de suas fases relativas, nas quais, o início da ativação de um gesto é motivado por um determinado valor da ativação ou fase de outro.

BROWMAN e GOLDSTEIN (1992) mostram como contrastes entre itens lexicais, explicados pela presença de um segmento ou de um traço diferente, podem ser entendidos em função da presença ou ausência de um único gesto. Assim, a diferença entre os itens *add*, *bad* e *had* reside no fato de que os dois últimos apresentam respectivamente um gesto de oclusão labial e um gesto de abertura glotal, da mesma forma que o contraste entre *bad* e *pad* é explicado pela presença de um gesto de abertura glotal em *pad*, enquanto a diferença entre *pad* e *pan* está na presença de um gesto de abertura vélica no último.

Segundo BROWMAN e GOLDSTEIN (1985, 1990), a estrutura articulatória de um enunciado é especificada por meio (1) da seleção de seus gestos adequados e (2) dos valores dos parâmetros dinâmicos a serem empregados no modelo de dinâmica de tarefa, usados como coeficientes de equações massa-mola amortecidas. Um exemplo físico de um sistema massa-mola, de acordo com os autores, é aquele de um objeto móvel (uma massa) preso por uma mola a uma superfície rígida. Quando a massa é puxada e a mola esticada para além de sua posição (ou comprimento) de equilíbrio, a massa começa a oscilar. A equação que caracteriza o movimento é:
$mx''+bx' + k(x-x_0) = 0$, onde m = massa do objeto, b = amortecimento do sistema, k = rigidez da mola, x_0 = comprimento da mola em repouso (posição de equilíbrio), x = deslocamento instantâneo do objeto, x' = velocidade instantânea do objeto e x'' = aceleração instantânea do objeto. A equação descreve os padrões de movimento resultantes da massa como trajetórias assintóticas.
BROWMAN e GOLDSTEIN (1990) colocam que o parâmetro referente à posição de equilíbrio, na equação, especifica o alvo a ser atingido pelo articulador, ao passo que o parâmetro rigidez especifica o tempo necessário para atingir o alvo. O parâmetro que especifica o amortecimento do gesto é sempre ajustado para obter um amortecimento crítico, ou seja, o gesto deve atingir seu alvo, mas não tocá-lo ou ultrapassá-lo. Segundo eles, a trajetória gerada por um gesto criticamente amortecido aproxima-se cada vez mais e lentamente da posição de equilíbrio, ao invés de oscilar ao redor dela, uma vez que o alvo é uma assíntota. Assim, uma quantidade infinita de tempo é necessária para atingir o alvo num sistema criticamente amortecido e, por isso, os autores afirmam que, por especificação deles, o alcance do alvo ocorre em 240°com relação ao ciclo subjacente abstrato de um gesto, que é de 360°.
BROWMAN e GOLDSTEIN (op. cit.) explicam que a dinâmica de tarefa serve para coordenar os articuladores dentro de um determinado gesto e também para coordenar os efeitos dos vários gestos concorrentes sobre um único articulador. Entretanto, a dinâmica de tarefa não coordena os gestos em si mesmos, o que é feito na pauta gestual, por meio da sincronização dos estados dinâmicos dos gestos, uns em relação aos outros. As equações na dinâmica de tarefa regulam a coordenação dos articuladores de tal modo que a formação de uma constrição é alcançada conforme os movimentos dos articuladores desdobram-se no tempo, enquanto as relações de fase entre os gestos na pauta gestual de um enunciado regulam a coordenação dos gestos entre si. Uma vez que o modelo assume uma massa unitária e um amortecimento

Outro modo de sinalizar contrastes lexicais na abordagem gestual é dado pelos diferentes conjuntos de articuladores[5] que participam na realização de itens contrastantes. *Bad* envolve uma oclusão labial, alcançada pelo movimento de lábio superior, lábio inferior e mandíbula, ao passo que *dad*, uma oclusão de ponta da língua, que inclui os movimentos de ponta da língua, dorso da língua e mandíbula.

Itens lexicais podem ainda ser contrastados em função dos valores que os parâmetros dinâmicos podem assumir. Diferentes valores do alvo de constrição da ponta da língua

---

crítico, a rigidez e a posição de equilíbrio são os parâmetros que podem variar na equação para veicular informação lingüística, como identidade fonética e acento.

[5] De acordo com BROWMAN e GOLDSTEIN (1990), os articuladores que participam da realização de uma determinada tarefa, por exemplo, o gesto de fechamento labial, que envolve a mandíbula, o lábio inferior e o lábio superior, têm de estar atrelados de uma forma específica à realização da tarefa. Portanto, são as características do movimento das variáveis da tarefa, também designadas por variáveis do trato, que são controladas na dinâmica de tarefa. No exemplo citado, o fechamento labial refere-se a um único objetivo para a variável de abertura labial, e não a um conjunto de objetivos individuais de mandíbula, lábio inferior e lábio superior. BROWMAN e GOLDSTEIN (1989, 1990) propõem oito variáveis do trato e em cada uma delas um determinado articulador ou conjunto de articuladores, conforme relacionados abaixo. As siglas usadas para cada uma das variáveis do trato referem-se às suas designações em inglês (BROWMAN e GOLDSTEIN, 1990), por nós mantidas, a saber: LP (*lip protrusion*), LA (*lip aperture*), TTCL (*tongue tip constriction location*), TTCD (*tongue tip constriction degree*), TBCL (*tongue body constriction location*), TBCD (*tongue body constriction degree*) VEL (*velic aperture*) e GLO (*glottal aperture*).

Variáveis do Trato	Articuladores Envolvidos
LP (protrusão labial)	lábio superior, lábio inferior e mandíbula
LA (abertura labial)	lábio superior, lábio inferior e mandíbula
TTCL (local de constrição da ponta da língua)	ponta da língua, dorso da língua e mandíbula
TTCD (grau de constrição da ponta da língua)	ponta da língua, dorso da língua e mandíbula
TBCL (local de constrição do dorso da língua)	dorso da língua e mandíbula
TBCD (grau de constrição do dorso da língua)	dorso da língua e mandíbula
VEL (abertura vélica)	véu palatino
GLO (abertura glotal)	glote

Cada variável do trato associada a um gesto é modelada por meio de uma ou duas equações dinâmicas separadas (BROWMAN e GOLDSTEIN, 1990), uma delas associada a grau de constrição e a outra a local de constrição. Grau e local de constrição referem-se às dimensões vertical e horizontal do plano sagital do trato vocal, entendido como um tubo. O grau de constrição é vertical em relação à dimensão horizontal estacionária das paredes superior ou posterior do trato vocal. O local de constrição refere-se à dimensão horizontal do trato, ou seja, ao local sobre a parede superior ou posterior do trato onde ocorre a constrição gestual e, portanto, segundo os autores, é distinto do conjunto de articuladores que se movem para fazer a constrição.

podem levar a uma oclusão completa, como em *dad*, ou à produção de uma turbulência, como em *span*.

A variação alofônica ou contextual, na Fonologia Articulatória (doravante, FAR), é entendida como resultado ou da sobreposição de unidades gestuais invariantes ou da diminuição da magnitude dos gestos individuais no tempo e no espaço. Como um exemplo, BROWMAN e GOLDSTEIN (1992) discutem o caso das oclusivas surdas do inglês, que tradicionalmente eram descritas como tendo alofones aspirados em início de sílaba e não-aspirados nos demais contextos, a partir de uma regra que lhes atribuía respectivamente os traços [glote+aberta] e [glote-aberta]. Entretanto, segundo os autores, em muitos contextos nos quais o *output* parece ser uma oclusiva não-aspirada, um gesto de abertura e fechamento glotal está presente no *input*, e, portanto, a presença ou ausência de aspiração no *output* não é uma função discreta de se a glote está ou não aberta, mas sim uma função do *timing* ou do padrão de sobreposição entre o gesto glotal e o gesto oral associado ou ainda uma função gradiente da magnitude do gesto glotal[6].

A análise tradicional propunha também que as oclusivas surdas do inglês eram não-aspiradas em meio de palavra, antes de vogais não-acentuadas, porque aí elas seriam ambissilábicas e representadas com o traço [glote-aberta]. Para BROWMAN e GOLDSTEIN (1992), o problema dessa representação é que oclusivas surdas na mesma posição, diante de vogais acentuadas, eram consideradas aspiradas e representadas pelo traço [glote+aberta], porque, nesse caso, estavam em início de sílaba. Pelo mesmo motivo,

[6] Um efeito semelhante foi mostrado por LEVY (1993) na análise acústica da fala de crianças que, aos ouvidos adultos, pareciam desvozear sistematicamente as oclusivas sonoras do português brasileiro.

independentemente se diante de vogal acentuada ou não, as oclusivas surdas em início de palavra eram consideradas sempre aspiradas e representadas com o traço [glote+aberta].

Para mostrar a inadequação da análise tradicional, BROWMAN e GOLDSTEIN citam um estudo de COOPER (1991, apud BROWMAN e GOLDSTEIN, 1992) que utilizou transiluminação para medir a abertura glotal em quatro contextos: inicial vs. medial, antes de vogais acentuadas e não-acentuadas. COOPER encontrou que havia 1) um gesto de abertura glotal nos quatro contextos, contrariando a predição de que nas oclusivas não-aspiradas a glote está [-aberta], e 2) efeitos do grau de acentuação e da posição na palavra sobre a magnitude do gesto de abertura glotal, no tempo e no espaço. Em outras palavras, a posição inicial e a tonicidade favoreciam gestos mais amplos, enquanto a posição medial e a atonicidade favoreciam gestos glotais de magnitude reduzida. COOPER concluiu então que o grau de acentuação e a posição na palavra, e não a estrutura da sílaba e a categoria de aspiração, são as variáveis relevantes na regulação do comportamento laríngeo das consoantes surdas do inglês. BROWMAN e GOLDSTEIN (1992) afirmam que, do ponto de vista gestual, não há nada de especial ou categoricamente diferente com as oclusivas surdas em posição medial e diante de vogal não-acentuada, uma vez que este é simplesmente o contexto no qual ocorre a maior redução do gesto glotal, devido ao efeito combinado de grau de acentuação e posição.

STERIADE (1990) mostra o quanto uma unidade de análise com tempo intrínseco pode ser vantajosa para a fonologia. Ela discute a lei de Dorsey, a qual se assemelha a uma regra de inserção vocálica que transforma sílabas CrV(C) em seqüências CvrV(C), nas quais v é uma cópia de V. Segundo a autora, o problema que a lei de Dorsey coloca para a análise

autossegmental é o de assegurar que a qualidade da vogal inserida combine com a da vogal tautossilábica do grupo consonantal, na representação subjacente, uma vez que o processo de inserção vocálica consiste de duas etapas: a inserção de uma casa (*slot*) para V (ou de uma unidade de *timing* equivalente) e a associação da casa com o segmento adequado. Dessa forma, a inserção de uma casa V, em uma seqüência como /i.tra/, cria uma nova sílaba, /i.tV.ra/, e o problema é que, a partir daí, não é possível determinar de onde veio a nova sílaba /tV/, ou seja, não se pode decidir se V em /tV/ deveria estar associada aos traços do /i/ precedente ou do /a/ seguinte.

Segundo STERIADE (op. cit.), dois aspectos da FAR podem explicar que os grupos consonantais separados por epêntese são *onsets* subjacentes e que a qualidade da vogal inserida combina com aquela do núcleo tautossilábico subjacente, com *onset* complexo, a saber: 1) os gestos têm tempo intrínseco e, dentro de uma sílaba, 2) as articulações consonantais são superpostas ao gesto vocálico[7] (BROWMAN e GOLDSTEIN, 1990 apud

---

[7] Segundo BROWMAN e GOLDSTEIN (1990), dados de raio X (ÖHMAN, 1966 e FOWLER, 1983, apud BROWMAN e GOLDSTEIN, 1990) têm sustentado o ponto de vista de que as articulações consonantais sobrepõem-se às articulações vocálicas, contínuas e minimamente sobrepostas entre si. Para dar um exemplo, reproduzimos uma figura de BROWMAN e GOLDSTEIN (op. cit.) que mostra as relações entre as camadas C e V para o enunciado *piece plots* ([pis# 'plats]).

(BROWMAN e GOLDSTEIN (op. cit.) comentam que a última consoante, *s*, não está associada ao gesto vocálico, uma vez que este não co-ocorre com o gesto consonantal final.)
BROWMAN e GOLDSTEIN (op. cit.) afirmam que o plano bidimensional das camadas V e C mostra a seqüencialização dos gestos ocorrendo ao longo da dimensão horizontal, enquanto o tipo de função é mostrado ao longo da dimensão vertical. A seqüencialização horizontal deve capturar a mesma seqüência de gestos que aquela disposta na camada oral. Uma restrição é que as camadas V e C devem retratar as relações seqüenciais canônicas entre os gestos, quando projetados numa única camada. As linhas de associação entre as duas camadas indicam que os gestos consonantais associados ocorrem juntamente com o gesto vocálico. A adjacência de gestos na camada V indica que as articulações vocálicas são efetivamente contínuas, com possibilidade mínima de sobreposição.

STERIADE, 1990). Assim, a FAR permite, segundo a autora, que a lei de Dorsey seja vista como uma mudança no *timing* relativo de três gestos. Ela exemplifica com o *input* /pra/, que começa com dois gestos consonantais sobrepostos um ao outro. Um atraso no *onset* da líquida pode criar uma seqüência na qual os dois gestos deixam de se sobrepor e um atraso ainda maior faz surgir uma seqüência na qual o gesto vocálico começa a aparecer entre os gestos consonantais. STERIADE propõe a seguinte representação:

Camadas	Gestos no *Input*	Gestos no *Output*
dorso da língua	[----------a----------]	[----------a----------]
ponta da língua	[----r----]	→ [----r----]
lábios	[-p-]	[-p-]
	(pra)	(para)

A autora coloca que o deslocamento da segunda consoante do grupo poderia ainda levá-la à extremidade oposta da sílaba, quando a consoante deslocada tornar-se-ia periférica e o resultado do deslocamento seria interpretado como um monossílabo. A representação proposta é a seguinte:

De acordo com os autores, há uma diferença funcional entre vogais e consoantes, com as vogais agindo como um pano de fundo para as figuras das consoantes. A contigüidade de gestos na camada V é totalmente independente da camada C, ou seja, as vogais são contíguas ou parcialmente sobrepostas independentemente do número de consoantes intervenientes. Já a contigüidade da camada C é sensível aos gestos vocálicos intervenientes, uma vez que os gestos consonantais sobrepõem-se menos, ou não se sobrepõem, quando um gesto vocálico intervém. Um efeito dessa diferença funcional entre as camadas é a mudança na rigidez conforme mais consoantes são inseridas entre vogais, quando diminui a rigidez dos gestos vocálicos e aumenta a dos gestos consonantais.

Camadas	Gestos no *Input*		Gestos no *Output*
dorso da língua	[----------a----------]		[----------a----------]
ponta da língua	[----r----]	→	[----r----]
lábios	[-p-]		[-p-]
	(pra)		(par)

Para STERIADE (1990), as análises acima[8] sugerem que o movimento intra-silábico dos gestos consonantais possui dois parâmetros: a direção (para a esquerda ou para a direita) e a posição alvo (periférica ou não), uma vez que os movimentos das líquidas para a esquerda, na periferia (CVr(C) → CrV(C)) ou não (CVr(C) → CVrV(C)), também são encontrados. Ela salienta que, sob o ponto de vista da FAR, a lei de Dorsey e a metátese líquida/vogal constituem de fato um único e mesmo fenômeno, a saber: um atraso significativo no *onset* do segundo gesto consonantal de um *onset* complexo. Ainda segundo a autora, (1) a posição alvo do gesto consonantal movido deve ser sempre um ponto na duração da vogal à qual ele está superposto, o que explica porque seqüências como /Mi.tra/ produzem /Mi.ta.ra/ e não */Mi.ti.ra/, e (2) há um aspecto categórico no movimento intra-silábico, na medida em que diferentes dialetos selecionam consistentemente o movimento para uma posição periférica ou não-periférica, mas nunca uma mistura de ambos.

[8] No português brasileiro, é comum presenciarmos, enquanto falantes e ouvintes, lapsos de língua envolvendo grupos consonantais do tipo CRV ou monossílabos do tipo CVR, onde R corresponde ao *tap* ou vibrante simples. Além disso, SILVA (1996) mostrou por meio de um estudo acústico que, nos grupos consonantais CRV, nos quais R é igual ao *tap* ou ao /l/, ocorre a "inserção" de uma vogal entre C e R, de qualidade mais central que V, embora de duração muito menor. É interessante notar que tanto a descoberta de SILVA como os lapsos no português brasileiro fazem sentido à luz da análise oferecida por STERIADE (1990) para a lei de Dorsey e para a metátese líquida/vogal, a partir da FAR.

Como o tamanho do deslocamento da líquida para uma posição não-periférica pode permanecer indeterminado, STERIADE supõe ainda que a diferença entre a aplicação da lei de Dorsey em /patiri/ e em /pateri/, ambas derivadas de /patri/, poderia ser atribuída a uma diferença entre um deslocamento suficientemente amplo na primeira, para deixar para trás uma qualidade vocálica identificável, ao passo que, na segunda, haveria um deslocamento pequeno demais para isso, o qual resultaria em /pateri/ e, desse modo, o tamanho real do ajustamento do *timing* que produz os efeitos encontrados sob a designação de lei de Dorsey poderia variar livremente, permanecendo constante apenas a posição alvo do movimento (periférico ou não). A autora considera então que o movimento intra-silábico assemelha-se a uma regra fonológica e que uma condição de boa formação da sílaba resultante é que tanto a lei de Dorsey quanto a metátese intra-silábica à direita aplicam-se somente a *onsets* complexos, uma vez que mudanças como /ra/ → /ara/ ou /ra/ → /ar/ não são atestadas.

3.3 Necessidade do tempo intrínseco na explicação de processos fônicos da fala corrente

Cabe ressaltar que, dada a superioridade da FAR para dar conta dos processos alofônicos, uma comparação entre ela e a Fonologia Autossegmental, como propomos abaixo, a partir de KINGSTON e COHEN (1992), tem como único intuito responder a algumas contestações de que o modelo seria excessivamente poderoso ou faria previsões incorretas. Contudo, essas contestações podem ser replicadas se forem entendidas num sentido um pouco mais abstrato do que o faz a FAR atualmente.

KINGSTON e COHEN (1992) colocam que, para certos fenômenos, não fica claro a necessidade de se postular uma unidade fonológica com tempo intrínseco, já que os mesmos são igualmente tratáveis a partir de uma unidade sem tempo. Para exemplificar seu ponto, eles discutem as explicações dadas pela Teoria Autossegmental e pela FAR para o apagamento da consoante nasal, no inglês, seguida por uma oclusiva surda (*limp*, *lint*, *link*), quando a vogal precedente permanece nasalizada. KINGSTON e COHEN mostram que a diferença entre as explicações reside em considerar o fenômeno como categórico ou gradiente, e, portanto, as diferenças entre as teorias encontram-se nas predições empíricas possíveis a partir de cada uma delas.

COHN (1990 apud KINGSTON e COHEN, 1992) mostrou que, quando a consoante nasal é apagada, o véu está totalmente abaixado já no início da vogal, o que a fez considerar a nasalização da vogal como fonológica. Com base na Teoria Autossegmental, a autora propôs que o apagamento da consoante nasal ocorre devido ao desligamento e remoção da casa que a consoante ocupava na camada de *timing*. Entretanto, esse apagamento deixa livre o traço [+ nasal], o qual se liga à vogal precedente.

Segundo KINGSTON e COHEN (op. cit.), na FAR, a consoante nasal não seria apagada[9], mas seu fechamento oral seria abreviado. Com isso, a abertura glotal, que subjaz

[9] Enquanto unidades lingüísticas dinâmicas, discretas e abstratas, os gestos são invariantes através de diferentes conteúdos. Gestos não podem ser inseridos ou apagados da pauta gestual de um dado item lexical. Mudanças na pauta gestual de um item lexical resultam de (1) redução na magnitude dos gestos individuais, tanto no tempo como no espaço, e (2) aumento da sobreposição entre os gestos. Esses dois processos gerais podem dar conta, segundo BROWMAN e GOLDSTEIN (1989), dos processos fônicos observados na fala corrente e descritos como: apagamento, inserção, assimilação e enfraquecimento de segmentos. De acordo com BROWMAN e GOLDSTEIN (1989), o que decorre do fato de considerar os gestos como unidades básicas é que a representação lexical resultante é inerentemente subespecificada, isto é, ela não contém especificações para traços irrelevantes. Quando um dado articulador não está envolvido na especificação de

ao intervalo de desvozeamento da oclusiva seguinte, começaria mais cedo, e, por sua vez, obrigaria o abaixamento velar a ocorrer muito antes, tanto para preservar a ordem das articulações como para evitar a incompatibilidade entre glote aberta e nasalização. O resultado seria então uma sobreposição cada vez maior entre o abaixamento do véu e a vogal precedente.

As explicações da Fonologia Autossegmental e da FAR, de acordo com KINGSTON e COHEN (op. cit.), resultam em predições completamente diferentes, o que dificulta escolher, de antemão, entre uma ou outra teoria. Para eles, à explicação de COHN subjaz uma diferença categórica entre os casos nos quais a nasal está presente e o véu abaixa gradualmente durante a vogal precedente e aqueles em que a nasal está ausente e o abaixamento do véu já ocorreu quando do início da vogal. Já à explicação com base na FAR subjaz um contínuo de graus de sobreposição entre o abaixamento do véu e a articulação da vogal. KINGSTON e COHEN (op. cit.) colocam que, para COHN, o abaixamento do véu seria mais rápido quando a nasalidade constitui uma propriedade fonológica da vogal, que quando a nasalidade decorre de uma consoante nasal não realizada, enquanto, para a FAR, não haveria mudanças no curso do movimento de abaixamento do véu, uma vez que tudo decorre da maior ou menor sobreposição entre o abaixamento do véu e a vogal precedente. Parece-nos que a diferença entre as duas previsões poderia ser eventualmente testada com experimentos fonéticos.

---

um gesto, ele é atraído a uma posição neutra, específica àquele articulador (SALTZMAN et al, 1988 apud BROWMAN e GOLDSTEIN, 1989).

3.4 A unidade com tempo intrínseco e a abstração das representações subjacentes

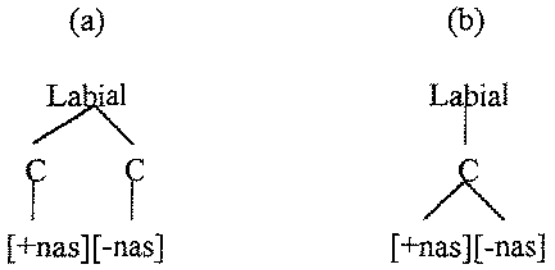
KINGSTON e COHEN (1992) discutem um exemplo no qual a não-ordenação no tempo, dos traços de um segmento, pode constituir uma restrição importante. Na Geometria de Traços (McCARTHY, 1988 apud KINGSTON e COHEN, 1992), numa oclusiva labiovelar aspirada, como [kp^h], a abertura da glote representada por [glote+aberta] não é ordenada no tempo com relação às especificações de ponto (nós labial e dorsal) ou modo ([-soante, +consonantal, -contínuo]), nem há duas especificações de ponto representadas pelos nós labial e dorsal ordenados no tempo um em relação ao outro. Segundo McCARTHY (comunicação pessoal apud KINGSTON e COHEN, 1992), a ordem temporal não precisa ser especificada, uma vez que as oclusivas raramente contrastam para a abertura da glote com relação ao fechamento da oclusiva e nenhuma língua contrasta labiovelares nas quais o fechamento labial precede o velar com labiovelares nas quais o fechamento velar precede o labial.

Na FAR, porque os gestos têm tempo intrínseco, segundo KINGSTON e COHEN (op. cit.), as relações temporais entre eles são explicitadas na representação de um enunciado. Com isso, no caso das oclusivas labiovelares aspiradas, a FAR vê-se obrigada a excluir contrastes no *timing* relativo dos gestos, por meio do acréscimo de restrições desnecessárias à Geometria de Traços. KINGSTON e COHEN (1992) mostram que, na FAR, a abertura glotal pode alinhar-se ao fechamento oral de três formas, a saber: 1) a abertura pode estar totalmente contida no fechamento da oclusiva, resultando numa oclusiva surda não-aspirada; 2) a glote pode permanecer aberta além da soltura da oclusiva, resultando numa oclusiva

surda pós-aspirada; e, finalmente, 3) a glote pode abrir antes que o fechamento ocorra, resultando numa oclusiva surda pré-aspirada. O problema, colocam KINGSTON e COHEN, é que embora esses três tipos de oclusivas ocorram nas línguas do mundo, as oclusivas surdas pré-aspiradas são muito raras. Vale dizer, no entanto, que a FAR prevê que algumas combinações de gestos articulatoriamente possíveis podem resultar em segmentos acústicos difíceis de perceber.

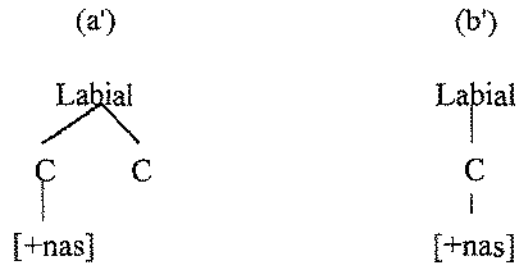
Um outro exemplo em que se alega que uma unidade com tempo intrínseco pode ser menos parcimoniosa que uma unidade sem tempo é apresentado por STERIADE (1990). Ela mostrou como o fato de que a Fonologia Autossegmental não permite distinguir entre sobreposição e simultaneidade, sob certas circunstâncias, pode constituir uma vantagem sobre a FAR, num nível mais abstrato de representação.

Para dar um exemplo, STERIADE coloca que, na Fonologia Autossegmental, a sobreposição articulatória entre nasalidade e fechamento labial pode ser representada tanto em seqüências longas, como em seqüências curtas, das seguintes formas, respectivamente:



Entretanto, segundo a autora, se o que estiver envolvido não for o contraste entre dois tipos de autossegmentos ([+nasal] e [-nasal]), mas entre um único tipo de autossegmento

([+nasal]) e sua ausência, a noção de sobreposição entre nasalidade e fechamento labial não é perdida para a seqüência longa, como visto em (a'), mas sim para a seqüência curta, como visto em (b'), uma vez que, na última, o fato de haver apenas um valor de nasalidade faz com que não se possa mais distinguir entre sobreposição e simultaneidade:



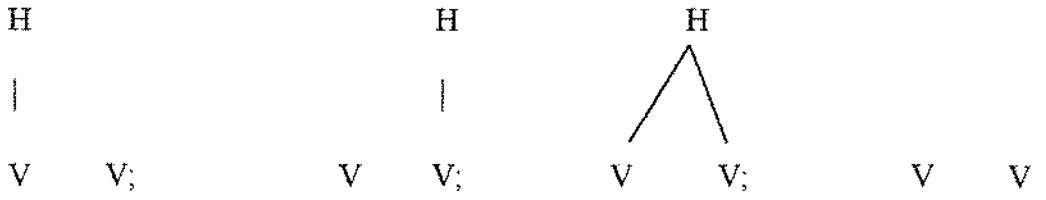
Na FAR, onde os gestos têm tempo intrínseco, a distinção entre sobreposição parcial e simultaneidade ocorre independentemente da duração da seqüência e de se cada uma das camadas envolvidas contém um ou mais gestos. STERIADE (1990) demonstra isso por meio da seguinte representação:

Sobreposição na seqüência curta	Sobreposição na seqüência longa
Camada 1 [----------]	[----------]
Camada 2 [--------]	[----------]

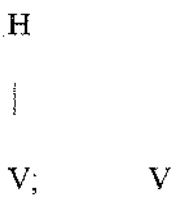
STERIADE discute então a distribuição de contornos tonais no grego antigo e no lituano, línguas nas quais as vogais breves são especificadas com um tom alto (H) ou baixo (L), enquanto as vogais longas e os ditongos podem ser especificados com tons alto (H),

baixo (L), descendente (HL) ou ascendente (LH), e afirma que se pode assumir que, a nível subjacente, há apenas um autossegmento, um tom alto (H), e que o tom baixo (L), no estágio adequado da representação, é simplesmente a ausência do alto (H). Com isso, segundo STERIADE, não apenas as formas das representações subjacentes e intermediárias são simplificadas, mas torna-se possível explicar a assimetria entre os contornos tonais sobre vogais longas e breves, da seguinte forma:

a) vogais longas:



b) vogais breves:



Segundo a autora, essa análise prediz que toda língua na qual o tom alto (H) é o único tom subjacente exibirá a mesma assimetria que o grego, ou seja, vogais longas podem mostrar um contraste de três ou quatro tons, ao passo que vogais breves apresentam apenas um contraste binário.

Entretanto, na FAR, onde os gestos estão em fase diretamente um em relação ao outro, sem uma camada de *timing* intermediária (o que pode ser entendido como sem um controle extrínseco de tempo), a autora alega que a neutralização dos contrastes entre os tons ascendente (LH), descendente (HL) e alto (H) sobre as vogais breves tem de ser estipulada: o gesto equivalente a um tom alto (H) deve ter uma duração mínima, que não seja significativamente mais curta que aquela do gesto de uma vogal breve. STERIADE aponta que a estipulação deve ainda ser específica à língua, já que há línguas que não apresentam esse tipo de restrição, nas quais vogais longas e breves exibem igualmente contornos tonais.

A interpretação de STERIADE (1990) impede que a sua análise seja traduzida nos termos da FAR, porque assume que a longa vai ser representada como gesto longo. Entretanto, essa posição não é pacífica. Há trabalhos, no quadro gestual, que interpretam as geminadas como dois gestos. Do ponto de vista físico, isso faz mais sentido, pois um gesto arbitrariamente longo tenderia a demorar para atingir o alvo, ficando distorcido. Dessa forma, faz muito mais sentido pensar na sustentação do alvo pelo advento do segundo gesto.

3.5 A falta de uma unidade de controle auditivo e de um controle extrínseco de tempo na Fonologia Articulatória

Os exemplos discutidos por KINGSTON e COHEN (1992) e STERIADE (1990), embora superficialmente pareçam demolidores, conduzem-nos à discussão de duas críticas internas muito importantes para o aperfeiçoamento da FAR.

O caso apresentado por KINGSTON e COHEN (op. cit.) traz à tona a necessidade de um elo auditivo para excluir possibilidades, menos freqüentes nas línguas, de variação alofônica. KINGSTON e COHEN (op. cit.) sugerem que deve haver razões perceptuais restringindo a ocorrência de oclusivas pré-aspiradas.

KOHLER (1992) afirma que aquilo que não é distintivo para o ouvinte pode ser mais facilmente reduzido pelo falante, de acordo com o princípio do mínimo esforço articulatório (*principle of economy of effort*). Segundo ele, diferentes situações comunicativas impõem diferentes demandas para os ouvintes, obrigando os falantes a ajustarem suas produções, de modo a garantir que a interação lingüística seja bem sucedida. Portanto, os vários estilos de fala apresentam graus também variados de redução em função das necessidades do ouvinte. O autor propõe que a FAR seja então ampliada com uma unidade de controle auditivo, porque, em sua opinião, os falantes não apenas controlam os gestos com relação aos aspectos fisiológicos e articulatórios, envolvidos na dinâmica da produção de sons, mas também levam em conta as necessidades dos ouvintes ao adaptar suas produções (LINDBLOM, 1990 apud KHOLER, 1992). Por fim, KHOLER (op. cit.) coloca que a FAR

não pode negar as conseqüências acústicas dos gestos[10] e que ela deve tornar-se um modelo orientado ao *input* e ao *output*[11].

O segundo caso, discutido por STERIADE (1990), leva-nos à crítica sobre a necessidade de um controle extrínseco de tempo na FAR, além do tempo intrínseco dos gestos.

A única referência feita pela FAR ao que pareceria um controle extrínseco de tempo[12] diz respeito à camada rítmica (BROWMAN e GOLDSTEIN, 1990), sobreposta à pauta gestual[13], e que contém a informação sobre acento. Assim, cada constelação de gestos da camada gestual, do tamanho de uma sílaba, está associada a um nó de acento da camada rítmica. Por sua vez, cada nó de acento pode afetar o grau de constrição e a rigidez dos gestos a ele associados. BROWMAN e GOLDSTEIN (op. cit.) fornecem a seguinte representação para o exemplo *piece plots* ([pis# 'plats]) (já discutido na explicação das associações entre as camadas C e V):

---

[10] ALBANO (1999), no ensaio **O gesto e suas bordas: para uma fonologia acústico-articulatória do português brasileiro**, mostra como as conseqüências acústicas dos gestos articulatórios podem restringir sua quantidade de sobreposição ou de redução no tempo e no espaço.

[11] BROWMAN e GOLDSTEIN (1992) afirmam que a análise gestual é uma análise do *input*, enquanto as análises mais tradicionais são análises do *output*, uma vez que *input* e *output* concernem respectivamente à descrição da organização gestual articulatória e à forma global resultante do trato vocal (acústica). Para eles, as análises segmentais tradicionais são descrições de efeitos combinados de gestos sobrepostos em uma constelação gestual e, portanto, são descrições da acústica ou *output*. As descrições de traços, por sua vez, referem-se a atributos dos segmentos e também são, freqüentemente, descrições do *output*.

[12] Uma proposta, em caráter inicial, para um controle extrínseco de tempo na FAR, foi apresentada por E. SALTZMAN na *Fifth Conference on Laboratory Phonology*, realizada em Evanston, na Northwestern University, em julho de 1996. O texto da conferência deverá sair publicado em mais um livro da série *Papers in Laboratory Phonology*, correspondente ao evento, e terá como organizadores J. PIERREHUMBERT e M. BROE.

[13] Apesar de o termo pauta gestual já ter ocorrido anteriormente no texto, vale lembrar que ele foi primeiramente proposto por ALBANO (1997) e posteriormente adotado por D'ANGELIS (1998) para referirem-se ao que na FAR tem sido designado como *gestural score*.

Camada	Gestos	
Rítmica	$X_{secundário}$	$X_{primário}$
Dorso da Língua	i	a
Ponta da língua	s	l t s
Lábios	p	p

Diferenças de acento, na FAR, são veiculadas por mudanças na rigidez dos gestos, as quais determinam o tempo necessário para um articulador atingir seu alvo. BROWMAN e GOLDSTEIN (1985) encontraram que o grau de acento tem um efeito consistente sobre a rigidez, ou seja, mudanças na rigidez dos articuladores provocam mudanças nas suas freqüências de oscilação[14]. Nas sílabas acentuadas, a freqüência de oscilação da massa dos articuladores diminui, devido à diminuição na rigidez, e a conseqüência é que a duração do *output* acústico aumenta. Ao contrário, nas sílabas não-acentuadas ou sílabas com vogais reduzidas, a freqüência de oscilação da massa dos articuladores aumenta, devido a um aumento na rigidez, e a conseqüência é que a duração do *output* acústico diminui. O parâmetro rigidez é então o responsável por tornar os gestos articulatórios mais lentos ou mais rápidos.

O que é possível depreender a partir da FAR é que há uma interação entre o parâmetro rigidez (k), presente nas equações ($mx''+ bx' + k(x-x_0) = 0$) associadas a grau e local de constrição, e a camada rítmica sobreposta à pauta gestual. Porém, como ocorre essa interação não está claro.

BROWMAN e GOLDSTEIN (1989) colocam que a rigidez (k) de um gesto é um parâmetro dinâmico, inferido a partir dos movimentos articulatórios, e que varia em função do grau de constrição, da acentuação e da taxa de elocução[15]. Eles supõem que a rigidez possa ser ajustada independentemente, de tal forma a servir para distinguir entre dois gestos. O exemplo dado refere-se aos gestos envolvidos na produção de /j/ e /w/, os quais teriam, respectivamente, o mesmo grau e local de constrição das vogais /i/ e /u/, mas com um valor de rigidez maior que o das vogais, similarmente à descrição articulatória proposta por CATFORD (1977 apud BROWMAN e GOLDSTEIN, 1989) para *glides*. Os autores colocam ainda que é possível que o valor da rigidez que governa a taxa de movimento de uma constrição esteja relacionado à rigidez biomecânica dos tecidos envolvidos, mas isto não é desenvolvido para que saibamos como.

3.6 Um modelo dinâmico do ritmo

PORT, CUMMINS e GASSER (1995) propõem um modelo dinâmico do ritmo no qual há controladores de tempo, os osciladores adaptativos (*Adaptative Oscillator*, McAULEY, 1994, 1995 apud PORT et al., 1995), dispostos nos diferentes níveis da percepção e da produção de fala. Segundo os autores, o cérebro somente poderia perceber padrões duracionais quase-periódicos ou parcialmente regulares como regulares se o sistema lingüístico empregasse um oscilador adaptativo, ou seja, um sistema que gera uma função de ativação senoidal em uma determinada freqüência. Os *inputs* ao sistema são pulsos. Cada

---

[14] A fórmula que relaciona freqüência e massa é: $\varpi_0 = \sqrt{k/m} = 2\pi f_0$ (onde m=1, como proposto pela FAR).

vez que um pulso forte ocorre, o oscilador reajusta sua fase a zero e reinicia imediatamente sua função cosseno. Se os pulsos externos são periódicos, ou muito próximos de periódicos, o oscilador adaptativo responde ajustando seu período intrínseco ao período do *input*. Parâmetros do sistema determinam a tolerância do oscilador para a variação no período do *input* e o número de ciclos necessários para que a adaptação seja quase completa. Uma pequena quantidade de variação não evita a indução[16] do oscilador a uma média do período do *input*.

PORT et al. (1995) afirmam que esse sistema pode ser implementado em termos computacionais por meio de um pequeno conjunto de equações, podendo aumentar ou diminuir sua velocidade rapidamente, em poucos ciclos, para aproximar-se da freqüência dos *inputs*. Se estes cessam, após alguns ciclos, o sistema começa a decair para sua taxa intrínseca original.

Se o oscilador adaptativo é capaz de ser induzido a oscilar de uma determinada forma, de acordo com uma série de pulsos externos, então os ouvintes de uma língua devem relatar uma alta regularidade temporal, porque osciladores adaptativos, na percepção de fala, forneceriam a base para a regularidade do *timing* da língua. A "gramática"[17] usada pelos falantes de uma língua incorpora, segundo PORT et al. (op. cit.), um sistema oscilatório que gera, durante a produção da fala, a estrutura rítmica e, quando se escuta a fala, um ritmo

---

[15] A tradução "taxa de elocução" foi proposta por BARBOSA, em SCARPA (no prelo), para *speech rate*.
[16] O termo *entrainment*, no original (PORT et al., op. cit.), vem sendo traduzido, no LAFAPE, por indução (Plínio Barbosa, comunicação pessoal), que fornece a idéia de que "algo externo induz uma dada coisa a variar de determinada forma".
[17] Colocamos aspas em *gramática* porque, da forma usada por PORT et al. (1995), diz respeito a um sistema físico e não mental, como geralmente se supõe na lingüística gerativa.

perceptual similar. É esse sistema oscilatório que explica o fato de que os ouvintes impõem regularidade ao sinal de fala porque conseguem prever o quê e quando vai acontecer.

Devido à dificuldade de encontrar intervalos de tempo regularmente espaçados no inglês[18], PORT et al. (1995) investigaram o *timing* moraico do japonês. As análises tradicionais da mora, segundo eles, concordam que geralmente uma mora corresponde a uma sílaba CV, como na palavra *doko*, a qual possui duas sílabas e duas moras. Porém, a primeira consoante em um grupo e a segunda parte de uma vogal longa também são consideradas moras. Desse modo, nas palavras *Honda* e *chotto*, há três moras, com o *n* em *Honda* e o primeiro *t* em *chotto* tendo a duração de uma mora cada um, apesar de as palavras terem somente duas sílabas. Na palavra *Tookyoo*, por sua vez, há quatro moras e duas sílabas. A partir dessas afirmações sobre a mora, os autores colocam que uma investigação experimental esperaria que, se todas as moras têm a mesma duração, as fronteiras acústicas que as separam deveriam ser igualmente espaçadas, ou seja, o intervalo do *onset* do *n* ao *onset* do *d*, em *Honda*, deveria ter a mesma duração acústica que *Ho* e *da*. Porém, essa predição não se confirma.

Partindo de uma teoria mais sofisticada da mora (PORT, AL-ANI e MAEDA, 1980, PORT, DALBY e O'DELL, 1987 e HAN, 1994 apud PORT et. al., 1995), PORT et al. (op. cit.) examinaram seqüências mais longas de fala, contendo várias moras, para saber até que

[18] PORT et al. (op. cit.) dizem que o fato de que a predição de uma isocronia perfeita não sobreviveu a medidas físicas fez com que a questão passasse a ser o quanto de isocronia é necessário para sustentar a classificação proposta por PIKE (1945 apud PORT et al., 1995) de línguas *stress-timed*, nas quais o intervalo de tempo entre sílabas proeminentes seria esticado ou comprimido para manter os *onsets* das sílabas acentuadas igualmente espaçados, e línguas *syllable-timed*, nas quais cada sílaba seria pronunciada com um mesmo espaçamento em relação às vizinhas. Seguindo PIKE (op. cit.), ABERCROMBIE (1967 apud PORT et al., 1995) propôs que os ouvintes do primeiro tipo de língua tinham expectativas quanto à regularidade da

ponto moras individuais poderiam desviar da duração média da mora de um enunciado, sempre que a mora incluísse consoantes e vogais cujas durações intrínsecas fossem longas ou curtas, ou quando a mora fosse constituída de um único segmento, C ou V. Um experimento mostrou que (1) todas as palavras com o mesmo número de moras, enunciadas numa dada taxa de elocução, tinham aproximadamente a mesma duração e (2) a adição de uma mora a uma palavra alongava-a numa quantidade aproximadamente constante, assumindo uma taxa de elocução também constante. Segundo os autores, cada mora, independentemente de seu conteúdo segmental, parece acrescentar a mesma quantidade de duração à palavra e, se, por um lado, esses resultados confirmam a regularidade do *timing* moraico, sugerido pelas análises tradicionais intuitivas, por outro, mostram que essa regularidade não pode ser observada se as moras são medidas isoladamente, mas somente se o forem no contexto das moras circundantes. Eles concluem, portanto, que: (1) tanto os segmentos dentro das moras como as moras adjacentes ajustam suas durações para compensar o acréscimo de uma nova mora e (2) a regularidade do *timing* moraico surge somente quando o tempo é medido de modo a nivelar as medidas de duração numa janela de tempo que inclui, pelo menos, várias moras.

As medidas de produção mostram que os falantes japoneses tentam compensar moras intrinsecamente curtas ou longas por meio de alongamento ou encurtamento, tanto antecipatório como perseveratório, das moras adjacentes. Para PORT et al. (1995), a curto prazo, o efeito desses ajustamentos, na taxa de elocução, é manter a duração média da mora constante. Se há um mecanismo oscilatório adaptativo que é excitado pela fala japonesa,

---

sucessão de acentos, ao passo que os ouvintes do segundo tipo de língua tinham expectativas quanto à

produzida num *timing* moraico, espera-se, segundo os autores, que os ouvintes japoneses possam usar o ângulo de fase dentro da mora como uma medida de tempo para fazer julgamentos perceptuais. A medida do ângulo de fase seria obtida a partir de certas propriedades físicas das moras que precedem a sílaba que está sendo escutada. Eles propõem então que é a gramática do japonês que define o sistema oscilatório baseado na mora e, desse modo, a estrutura temporal distintiva da língua, que restringe tanto a produção como o sistema perceptual do falante, é atribuída à gramática da própria língua[19]. Assim, o oscilador adaptativo seria um modelo plausível do relógio empregado pelos ouvintes japoneses para medir o *timing* de fala.

PORT et al. (1995) afirmam que, nas línguas com características de *stress-timed*, deve haver osciladores acomodados em mais de uma escala de tempo, uma vez que a relação hierárquica de sílabas e pés implica a existência de um oscilador no nível da sílaba acoplado a um oscilador no nível do pé, o qual mede o intervalo entre acentos. Segundo os autores, é isso que normalmente é concebido como métrica, ou seja, um sistema hierárquico de ciclos, e que, na lingüística, tem sido definido em relação a uma escala ordinal de tempo, vista como seqüências ordenadas de unidades fortes e fracas, dispostas em níveis hierárquicos. Os autores presumiram que, se o inglês fosse rítmico no nível das sílabas acentuadas, então períodos fonológicos entre acentos deveriam interagir com qualquer outro período que se impusesse à realização, como, por exemplo, o período da frase[20] como um todo.

---

regularidade da sucessão de sílabas.

[19] Plínio Barbosa (comunicação pessoal) sugeriu que, aquilo que PORT et al. (op. cit.) chamam de "gramática do japonês" (*grammar of Japanese*) é, na verdade, o ritmo do japonês, e o que eles chamam de "gramática da própria língua" (*the grammar of the language itself*) é o próprio sistema lingüístico.

[20] No original, *phrase*, que julgamos poder ser entendido como sentença ou frase.

Para comprovar essa suposição, foi realizado um experimento no qual havia um conflito entre a taxa de repetição de uma frase (*Take a pack of cards*), imposta por um estímulo semelhante a um metrônomo, e a periodicidade hipotética de acentos no nível frasal do inglês. Os falantes mostraram grande dificuldade em colocar os acentos num ângulo de fase arbitrário, ao repetir a frase muitas vezes. Havia três fases altamente atrativas localizadas nos harmônicos mais baixos do período definido pela repetição da frase inteira ou próxima a eles. Segundo os autores, isso forneceu evidência a favor de um processo dinâmico de indução, ao invés da mera concatenação de durações individuais. Eles concluíram que os resultados sugerem fortemente uma estruturação hierárquica do *timing* das frases repetidas no experimento, e que, como no *timing* musical, os sujeitos mostraram preferência por uma unidade de medida, a frase como um todo, e uma unidade de batida, o *onset* da sílaba acentuada.

3.7 Modelos dinâmicos e os fatos sobre as relações entre desenvolvimento lingüístico e neuromotor na aquisição da duração no português brasileiro

À primeira vista, nossos resultados, expostos no capítulo anterior, coadunam-se com as previsões mais gerais da literatura resenhada no capítulo 1, de que as crianças abaixo de 6 anos apresentam durações maiores e mais variáveis que os adultos e de que há diferenças entre as atuações de crianças pertencentes à mesma faixa etária. Porém, **o que nossos dados acrescentam à literatura é a observação de que as duas crianças estudadas, com menos de 6 anos, já produzem o contorno duracional do PB, apesar de implementarem-no dentro de suas possibilidades neuromotoras**. É isto que faz com que crianças mais novas

exagerem os picos de duração, possivelmente como M o faz, porque ainda não reduzem suficientemente os elementos não-acentuados, que, como vimos, constituem os pontos de referência para a construção das alternâncias rítmicas (BARBOSA e BAILLY, 1994). Por sua vez, crianças mais velhas já reduzem alguns elementos, porém não todos. É o que vemos no que concerne à atuação de E, quando as vogais são reduzidas antes das consoantes. Parece haver, aí, de um lado, um fator relacionado à prática ou exercício da vogal, que, como já foi mostrado (GONÇALVES, 1989, GAMA, 1989), precede a prática ou exercício das obstruções do trato vocal, e, de outro, um fator relacionado ao próprio amadurecimento neuromotor, uma vez que os gestos de realização das vogais são mais lentos e envolvem constrições mais amplas que aqueles de realização das consoantes, mais rápidos e pontuais (RODRIGUES, 1989, BROWMAN e GOLDSTEIN, 1990), e que por isso exigem movimentos mais rápidos e precisos.

Os critérios utilizados na elaboração do *corpus* e da análise estatística, que tentaram levar em conta as diferenças lingüísticas subjacentes à realização dos contrastes de acento implementados via duração no PB, conduziram-nos a observações que rebatem as hipóteses da maturação neuromotora e do artefato estatístico, mostrando-as como muito simplistas e, portanto, insuficientes para explicar a complexa interação entre desenvolvimento lingüístico e neuromotor na aquisição da duração. A hipótese da maturação neuromotora prevê que as durações na fala de crianças abaixo de 6 anos serão sempre maiores e mais variáveis, independentemente de serem intrinsecamente maiores ou não. Essa hipótese não explica que, apesar de as crianças M e E apresentarem valores de duração maiores em muitas das posições de acento

analisadas, seus contornos de variabilidade, para os segmentos acústicos correspondentes a vogais, consoantes e sílabas, não são inteiramente distintos daqueles do par S-A. A hipótese do artefato estatístico, por sua vez, prediz que durações intrinsecamente maiores ou menores serão também maiores ou menores e mais ou menos variáveis na fala infantil, o que não explica como a criança pode dominar a implementação da duração de um tipo de segmento antes de outro, como ocorre com E para as vogais, em oposição às consoantes.

O que faz com que as crianças produzam macroscopicamente contornos de variabilidade (valores de desvio-padrão) semelhantes aos dos adultos, embora elas ainda não consigam reduzir e estabilizar, como eles, os elementos não-acentuados desses contornos? É possível pensar que o controle neuromotor dos movimentos mais globais da produção da fala seja inicialmente mais fácil ou anteceda ao controle dos movimentos mais refinados, como se observa habitualmente para o desenvolvimento da motricidade de membros superiores e inferiores.

Uma análise de caráter mais qualitativo, na qual construímos contornos com as durações médias (pontos, nos gráficos) e desvios-padrão (traços verticais, nos gráficos) das sílabas de uma sentença do *corpus*, fornece um bom exemplo daquilo que está sendo discutido. Note que as médias e desvios-padrão foram calculados para as três repetições de cada sujeito e que as sílabas são dispostas em sua sucessão temporal.

Gráfico 1 - Médias e desvios-padrão dos três modelos da pesquisadora oferecidos à criança M para a sentença "A palhaçada dá muita risada"

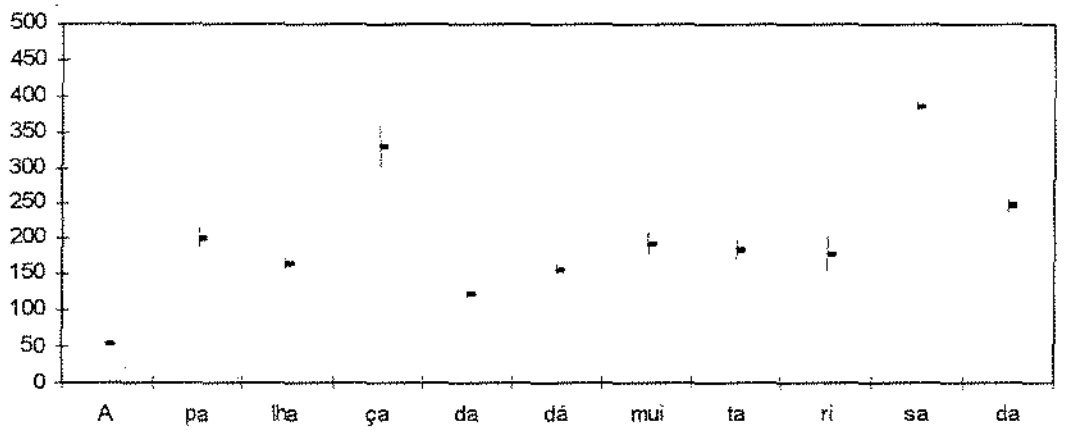
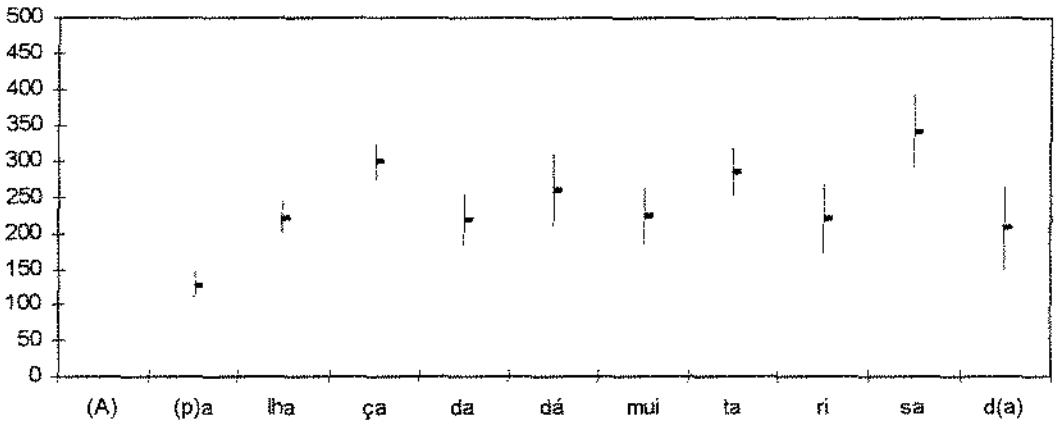


Gráfico 2 - Médias e desvios-padrão das três repetições da criança M para a sentença "A palhaçada dá muita risada"



A observação dos gráficos 1 e 2[21], respectivamente dos modelos oferecidos pela pesquisadora à criança M e das repetições desta para a sentença "A palhaçada dá muita risada", mostra-nos que, apesar de suas durações e desvios-padrão mais longos, M apresenta um contorno das médias e desvios-padrão das durações das três repetições da sentença

também com dois picos, sobre as sílabas tônicas de "palhaçada" e "risada", respectivamente, sendo que a maior proeminência da sentença, tanto para M como para A, ao menos em relação à duração, recai sobre o último acento lexical mais à direita, ou seja, sobre a tônica de "risada". A maior diferença entre ambas está justamente no trecho do contorno entre os dois picos de duração. A partir do gráfico 1, poder-se-ia supor que o contorno da pesquisadora apresenta basicamente três blocos: "A palhaçada", "dá muita" e "risada", e em cada um deles a maior proeminência incide sobre o acento lexical mais à direita. No contorno de M, por sua vez, além dos picos de duração das tônicas de "palhaçada" e "risada", parece haver um pico de duração sobre o verbo "dá" e outro sobre a sílaba pós-tônica "ta", de "muita". No trecho que compreende desde a pós-tônica de "palhaçada" até pré-tônica de "risada", não se observa um bloco, mas sim acentos alternados.

[21] Os segmentos acústicos que nos gráficos das crianças (Gráficos 2 e 4) aparecem entre parênteses não puderam ser medidos em uma ou mais repetições da sentença.

Gráfico 3 - Médias e desvios-padrão dos três modelos da pesquisadora oferecidos à criança E para a sentença "A palhaçada dá muita risada"

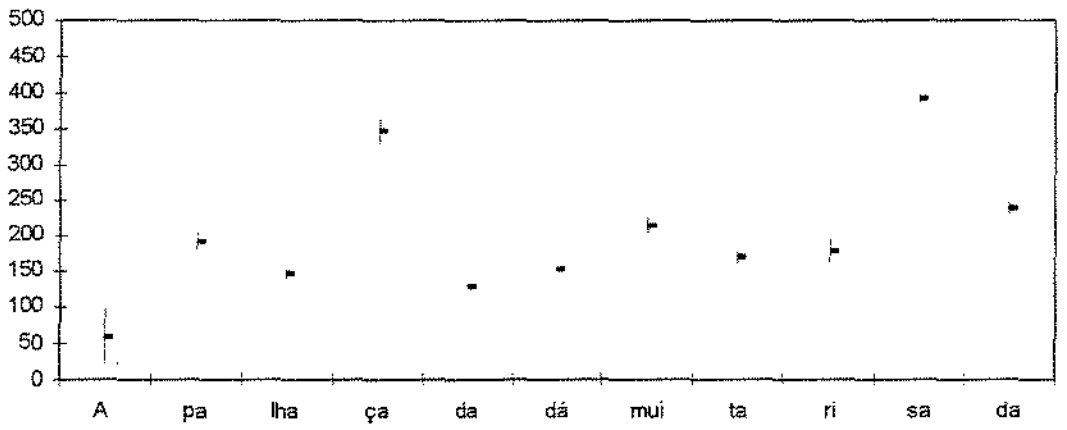
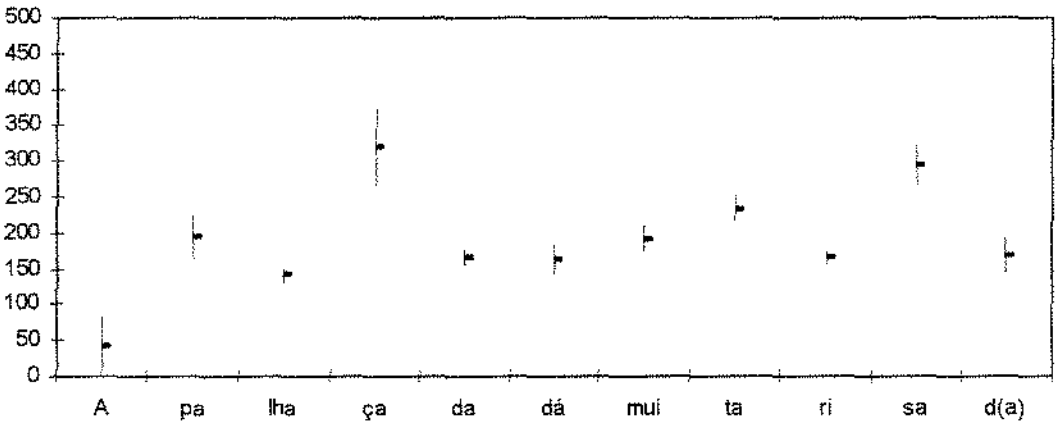


Gráfico 4 - Médias e desvios-padrão das três repetições da criança E para a sentença "A palhaçada dá muita risada"



Um contorno parecido com aquele colocado no gráfico 1 é observado no gráfico 3, onde estão as médias e desvios-padrão das durações dos três modelos oferecidos pela pesquisadora à criança E. Nesse contorno, é possível distinguir os mesmos três blocos acima descritos no contorno dos modelos oferecidos a M (Gráfico 1), talvez com uma queda um pouco maior da pós-tônica "ta" de "muita". Para a criança E, também obtém-se um contorno

macroscopicamente semelhante àqueles da pesquisadora (Gráficos 1 e 3) e da criança M (Gráfico 2). Contudo, note que, no trecho entre a pós-tônica de "palhaçada" e a pré-tônica de "risada", há uma linha ascendente desde o verbo "dá" até a pós-tônica de "muita", a qual constitui o ponto mais alto da linha. A maior proeminência da sentença parece recair sobre o acento lexical mais à esquerda, ou seja, sobre a tônica de "palhaçada".

Gráfico 5 - Médias e desvios-padrão dos três modelos da pesquisadora oferecidos à professora para a sentença "A palhaçada dá muita risada"

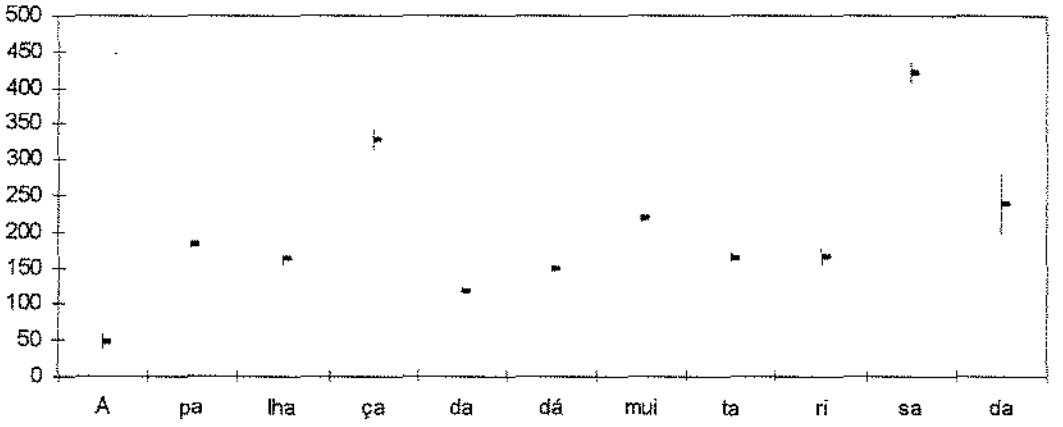
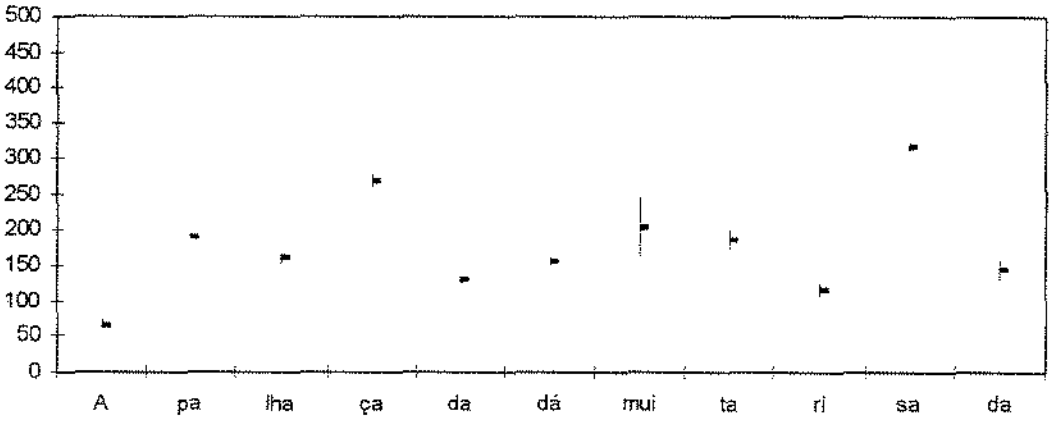


Gráfico 6 - Médias e desvios-padrão das três repetições da professora (sujeito S) para a sentença "A palhaçada dá muita risada"



Por fim, os gráficos dos contornos de duração referentes aos modelos oferecidos pela pesquisadora à professora e das repetições desta (respectivamente, Gráficos 5 e 6) concordam quanto aos três blocos observados nos contornos dos modelos oferecidos pela pesquisadora a cada criança. O maior desvio-padrão observado na fala da professora sobre a tônica de "muita" deve-se a uma hesitação durante a produção do "m" em uma de suas repetições.

A análise realizada acima mostra as crianças produzindo o contorno de duração da sentença analisada, na sua forma mais global, a partir de seus recursos, já bastante discutidos. A diferença entre os contornos das crianças e aqueles dos adultos reside na produção do elementos não-acentuados, mais pontuais, que constituem as batidas de referência, entre os dois picos de duração. Pode-se supor que a produção desses elementos de referência está relacionada a uma motricidade mais finamente ajustada (BROWMAN e GOLDSTEIN, 1989) e estável. A instabilidade do controle neuromotor nas crianças pode ser mais bem apreciada quando se compara a produção delas à produção adulta para os segmentos acústicos vocálicos e consonantais, de uma dada sentença. Para tanto, elaboramos gráficos (GAMA-ROSSI e ALBANO, 1998) com os valores percentuais dos segmentos acústicos, em cada uma das três repetições da sentença "A palhaçada dá muita risada". Note que a dispersão na produção dos segmentos acústicos diminui do gráfico 7, referente à atuação da criança M, para o gráfico 9, que mostra a atuação da professora, em suas três repetições. Vale lembrar que os segmentos que ocorrem entre parênteses não puderam ser medidos em uma ou mais repetições e que houve uma hesitação na fala da professora, no "m" de "muita", em uma das repetições da sentença.

Gráfico 7 - Valores percentuais dos segmentos de cada uma das três repetições da criança M para a sentença "A palhaçada dá muita risada"

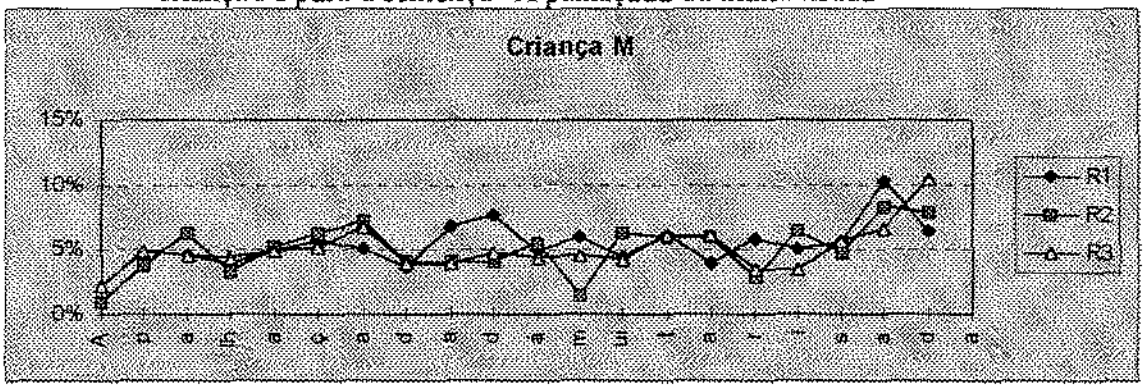


Gráfico 8 - Valores percentuais dos segmentos de cada uma das três repetições da criança E[22] para a sentença "A palhaçada dá muita risada"

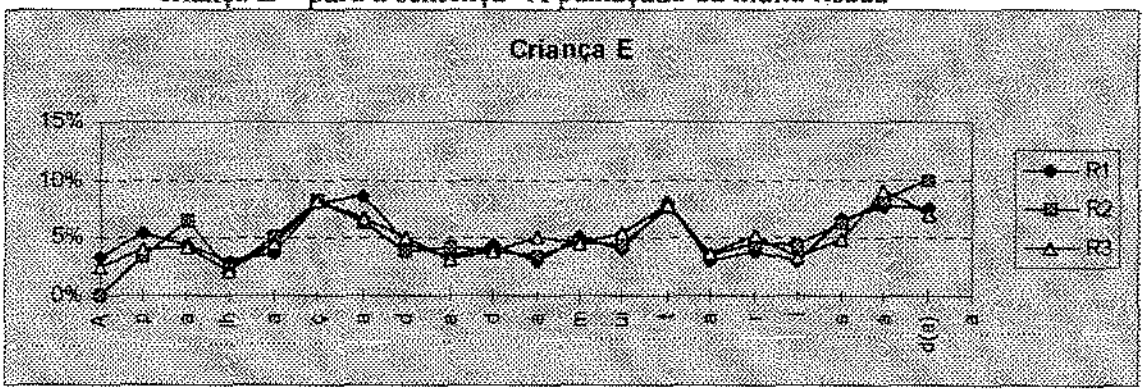
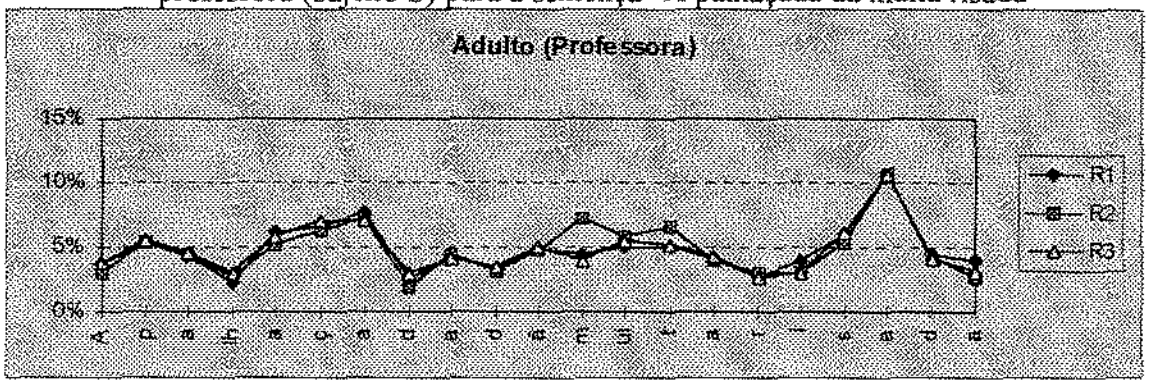


Gráfico 9 - Valores percentuais dos segmentos de cada uma das três repetições da professora (sujeito S) para a sentença "A palhaçada dá muita risada"



[22] O local do evento correspondente ao "a" do final da sentença, e que não foi medido em nenhuma das repetições, foi mantido no gráfico para o alinhamento deste com os demais.

De que modo a FAR (BROWMAN e GOLDSTEIN, 1985, 1986, 1989, 1990, 1992) e o modelo de PORT, CUMMINS e GASSER (1995) contribuem para a compreensão dos dados aqui apresentados?

Para a FAR, parece fácil explicar a instabilidade observada na produção dos segmentos acústicos vocálicos e consonantais, observada na fala infantil (Gráficos 7 e 8), recorrendo à afirmação de que, durante um período considerável de tempo, após o balbucio, a criança tem de dar continuidade à tarefa de diferenciar, ajustar e coordenar gestos (BROWMAN e GOLDSTEIN, 1989) na produção da fala. Talvez a FAR possa aclarar ainda a relação entre o desenvolvimento neuromotor e os maiores valores de duração e desvio-padrão das crianças pequenas na produção dos segmentos em posições não-acentuadas. De um lado, BROWMAN e GOLDSTEIN (1989, 1990) cogitam que o valor da rigidez que governa a taxa de movimento em uma constrição pode estar relacionado com a rigidez biomecânica dos tecidos envolvidos, embora isso não tenha sido desenvolvido ou explicado pelo modelo. De outro, eles mostram (BROWMAN e GOLDSTEIN, 1985) que mudanças na freqüência de oscilação da massa do articulador são diretamente proporcionais a mudanças na rigidez e inversamente proporcionais a mudanças na duração do *output* acústico. Assim, poderíamos supor que, devido a uma *neuromotricidade* em construção, as crianças abaixo de 6 anos não conseguem aumentar suficientemente a freqüência de oscilação de seus articuladores de modo a produzir segmentos reduzidos e estáveis nas posições não-acentuadas, como ocorre para a fala adulta. Nesse sentido, a FAR também é capaz de explicar porque os contornos de desvios-padrão obtidos para os pares M-A e E-A são mais altos que aqueles do par S-A, em todos os segmentos analisados (vogais, consoantes e sílabas). Também nosso achado sobre a

fala infantil de que ao aumento do número de segmentos na sílaba corresponde um aumento das médias e dos desvios-padrão das diferenças de duração pode ser explicado por uma menor coordenação entre gestos (BROWMAN e GOLDSTEIN, 1989), o que BARBOSA (comunicação pessoal) designou por "menor efeito compensatório".

Contudo, torna-se mais difícil para a FAR dar conta da observação de que, apesar de apresentar valores mais altos, os contornos de desvios-padrão ou variabilidade absoluta dos pares formados pelas crianças assemelham-se aos do par formado pelos adultos, porque, para isso, seria necessário fazer referência a uma noção de tempo que extrapola o tempo gestual.

Um modelo como o de PORT et al. (1995) explica com maior facilidade o fato de que, embora as crianças ainda não produzam os elementos não-acentuados como os adultos, nem por isso seus contornos de variabilidade são totalmente distintos daqueles. É preciso supor apenas que as crianças possuem osciladores no nível dos pontos de maior variação das frases, ou seja, osciladores ajustados para os principais pontos de variação macrorrítmica do contorno duracional, como mostrado pelos Gráficos 2 e 4, correspondentes às atuações de M e E para a sentença "A palhaçada dá muita risada", mas que não possuem osciladores finamente ajustados para as variações microrrítmicas. Um exemplo disso é visto tanto nas sílabas não-acentuadas dispostas entre os picos de duração constituídos por "ça" (de "palhaçada") e "sa" (de "risada"), dos Gráficos 2 e 4, como na produção dos segmentos da sentença ("A palhaçada dá muita risada"), nas suas três repetições, Gráficos 7 e 8. Assim, parece que, para as crianças aqui examinadas, está em andamento, de um lado, o aumento da coordenação entre os gestos de um enunciado, e, de outro, a aquisição de um oscilador ao

nível da sílaba. Se for possível fazer referência à terminologia de PORT et al. (1995), as crianças possuem uma unidade de medida, a frase como um todo, à qual elas podem ou não induzir seus osciladores ao nível da palavra e do sintagma, ao passo que elas ainda não adquiriram uma unidade de batida, ou seja, a sílaba, da qual possivelmente dependerá também a maior redução e estabilidade dos elementos nas posições não-acentuadas.

Concluindo, a análise quantitativa dos dados da fala adulta, apresentada neste trabalho, permite-nos afirmar que um modelo na forma daquele proposto por PORT et al. (op. cit.) é necessário para dar conta do ritmo do PB[23], uma vez que foram encontradas evidências daquilo que corresponde aos osciladores adaptativos ao nível da sílaba e da palavra. Avaliações da necessidade ou não de postular osciladores acima desses níveis somente poderão ser feitas com o emprego de outras medidas acústicas, além da duração. No entanto, análises qualitativas das durações de vogais, consoantes e sílabas, expressas por meio de gráficos que as mostram numa sucessão temporal, na forma de contornos duracionais (GAMA-ROSSI, 1995, 1996a, 1996b, GAMA-ROSSI e ALBANO, 1996), sugerem a existência de osciladores ao nível do sintagma e também da sentença.

Da mesma forma, a análise quantitativa da duração de segmentos acústicos que correspondem a consoantes e vogais, para a fala adulta e infantil, corrobora a necessidade de empregar uma unidade como o gesto articulatório no estudo da fala infantil. Análises qualitativas desses dados (GAMA-ROSSI e ALBANO, 1998) mostraram porque uma

[23] No LAFAPE, BARBOSA está elaborando um modelo dinâmico do ritmo com base na avaliação de dados acústicos do PB, além de outras línguas.

unidade como o gesto articulatório pode dar conta daquilo que foi designado por MENN (1983) como "variação arbitrária" da fala infantil.

Portanto, parece-nos que, enquanto não é possível dispor de um modelo dinâmico do ritmo do PB, um modelo para dar conta dos dados oferecidos por este trabalho seria uma espécie de acoplamento entre a FAR (BROWMAN e GOLDSTEIN, 1985, 1986, 1989, 1990, 1992) e o modelo de PORT et al. (1995), com várias modificações no sentido de tornar ambos os modelos mais flexíveis e abstratos, conforme vimos propondo em nossas discussões no LAFAPE, cuja primeira apresentação de maior envergadura é ALBANO (1999).

ANEXO A

Demarcação das posições de acento para vogais e consoantes de cada sentença

O cachorro late para o palhaço

O	c	a	ch	o	rr	o	l	a	t	e	p	a	r	ao	p	a	lh	a	ç	o
0	1	1	4	4	5	5	4	4	5	5	4	4	5	5	1	1	4	4	6	6

O cachorrinho chora de medo

O	c	a	ch	o	rr	i	nh	o	ch	o	r	a	d	e	m	e	d	o
0	1	1	2	2	4	4	5	5	4	4	5	5	5	5	4	4	6	6

A palhacinha não quer mamadeira

A	p	a	lh	a	c	i	nh	a	n	ão	qu	e	r	m	a	m	a	d	ei	r	a
0	1	1	2	2	4	4	5	5	4	4	4	4	4	1	1	2	2	4	4	6	6

O elefante planta bananeira

Oe	l	e	f	an	t	e	p	l	an	t	a	b	a	n	a	n	ei	r	a
0	2	2	4	4	5	5	4	4	4	5	5	1	1	2	2	4	4	6	6

O domador bate o chicote

O	d	o	m	a	d	o	r	b	a	t	eo	ch	i	c	o	t	e
0	1	1	2	2	4	4	4	4	4	5	5	1	1	4	4	6	6

O elefante dá banho no palhaço

Oe	l	e	f	an	t	e	d	a	b	a	nh	o	n	o	p	a	lh	a	ç	o
0	2	2	4	4	5	5	4	4	4	4	5	5	5	5	1	1	4	4	6	6

O palhaço mergulha na bacia

O	p	a	lh	a	ç	o	m	e	r	g	u	lh	a	n	a	b	a	s	ia
0	1	1	4	4	5	5	1	1	1	4	4	5	5	5	5	1	1	6	6

A macaca toca chocalho

A	m	a	c	a	c	a	t	o	c	a	ch	o	c	a	lh	o
0	1	1	4	4	5	5	4	4	5	5	1	1	4	4	6	6

O palhaço cor de rosa planta bananeira

O	p	a	lh	a	ç	o	c	o	r	d	e	r	o	s	a
0	1	1	4	4	5	5	4	4	4	5	5	4	4	5	5

p	l	an	t	a	b	a	n	a	n	ei	r	a
4	4	4	5	5	1	1	2	2	4	4	6	6

A palhaçada dá muita risada

A	p	a	lh	a	ç	a	d	a	d	a	m	ui	t	a	r	i	s	a	d	a
0	1	1	2	2	4	4	5	5	4	4	4	4	5	5	1	1	4	4	6	6

O macaquinho anda de bicicleta

O	m	a	c	a	qu	i	nh	oan	d	a	d	e	b	i	c	i	c	l	e	t	a
0	1	1	2	2	4	4	4	4	5	5	5	5	1	1	2	2	4	4	4	6	6

A cachorrada quer osso

A	c	a	ch	o	rr	a	d	a	qu	e	r	o	ss	o
0	1	1	2	2	4	4	5	5	4	4	4	4	6	6

A macacada faz barulho

A	m	a	c	a	c	a	d	a	f	a	z	b	a	r	u	lh	o
0	1	1	2	2	4	4	5	5	4	4	4	1	1	4	4	6	6

O palhaço Chapolin segura a calça

O	p	a	lh	a	ç	o	Ch	a	p	o	l	in	s	e	g	u	r	aa	c	al	ç	a
0	1	1	4	4	5	5	1	1	2	2	4	4	1	1	4	4	5	5	4	4	6	6

O elefantinho pedala a motoca

Oe	l	e	f	an	t	i	nh	o	p	e	d	a	l	aa	m	o	t	o	c	a
0	2	2	3	3	4	4	5	5	1	1	4	4	5	5	1	1	4	4	6	6

A cueca do palhaço tem muita bolinha

A	c	ue	c	a	d	o	p	a	lh	a	ç	o	t	em	m	ui	t	a	b	o	l	i	nh	a
0	4	4	5	5	5	5	1	1	4	4	5	5	4	4	4	4	5	5	1	1	4	4	6	6

O macaco segura a banana

O	m	a	c	a	c	o	s	e	g	u	r	aa	b	a	n	a	n	a
0	1	1	4	4	5	5	1	1	4	4	5	5	1	1	4	4	6	6

ANEXO B

Demarcação das posições de acento para sílabas de cada sentença

O cachorro late para o palhaço

O	ca	cho	rro	la	te	pa	rao	pa	lha	ço
0	1	4	5	4	5	4	5	1	4	6

O cachorrinho chora de medo

O	ca	cho	rri	nho	cho	ra	de	me	do
0	1	2	4	5	4	5	5	4	6

A palhacinha não quer mamadeira

A	pa	lha	ci	nha	não	quer	ma	ma	dei	ra
0	1	2	4	5	4	4	1	2	4	6

O elefante planta bananeira

Oe	le	fan	te	plan	ta	ba	na	nei	ra
0	2	4	5	4	5	1	2	4	6

O domador bate o chicote

O	do	ma	dor	ba	teo	chi	co	te
0	1	2	4	4	5	1	4	6

O elefante dá banho no palhaço

Oe	le	fan	te	da	ba	nho	no	pa	lha	ço
0	2	4	5	4	4	5	5	1	4	6

O palhaço mergulha na bacia

O	pa	lha	ço	mer	gu	lha	na	ba	cia
0	1	4	5	1	4	5	5	1	6

A macaca toca chocalho

A	ma	ca	ca	to	ca	cho	ca	lho
0	1	4	5	4	5	1	4	6

O palhaço cor de rosa planta bananeira

O	pa	lha	ço	cor	de	ro	sa
0	1	4	5	4	5	4	5

plan	ta	ba	na	nei	ra
4	5	1	2	4	6

A palhaçada dá muita risada

A	pa	lha	ça	da	da	mui	ta	rí	sa	da
0	1	2	4	5	4	4	5	1	4	6

O macaquinho anda de bicicleta

O	ma	ca	qui	nhoan	da	de	bi	ci	cle	ta
0	1	2	4	4	5	5	1	2	4	6

A cachorrada quer osso

A	ca	cho	rra	da	que	ro	sso
0	1	2	4	5	4	4	6

A macacada faz barulho

A	ma	ca	ca	da	faz	ba	ru	lho
0	1	2	4	5	4	1	4	6

O palhaço Chapolin segura a calça

O	pa	lha	ço	Cha	po	lin	se	gu	raa	cal	ça
0	1	4	5	1	2	4	1	4	5	4	6

O elefantinho pedala a motoca

Oe	le	fan	ti	nho	pe	da	laa	mo	to	ca
0	2	3	4	5	1	4	5	1	4	6

A cueca do palhaço tem muita bolinha

A	cue	ca	do	pa	lha	ço	tem	mui	ta	bo	lin	nha
0	4	5	5	1	4	5	4	4	5	1	4	6

O macaco segura a banana

O	ma	ca	co	se	gu	raa	ba	na	na
0	1	4	5	1	4	5	1	4	6

ANEXO C

TABELA 1- Resultados do teste t para os segmentos acústicos vocálicos do par M-A

Posição	N	Média	Desvio Padrão	t	p
1	94	30	28	10,32	0,0001
2	44	36	33	7,2181	0,0001
4	158	10	51	2,558	0,0115
5	85	44	37	10,784	0,0001
6	17	-63	67	-3,8484	0,0014
1235	226	36	33	16,229	0,0001
123456	401	22	48	9,0601	0,0001

TABELA 2 - Resultados do teste t para os segmentos acústicos vocálicos do par E-A

Posição	N	Média	Desvio Padrão	t	p
1	97	3	26	1,036	0,3028
2	44	-1	25	-0,1633	0,8711
4	173	-16	55	-3,775	0,0002
5	92	7	20	3,1665	0,0021
6	34	-70	41	-9,8485	0,0001
1235	236	3	24	2,114	0,0356
123456	443	-10	44	-4,6093	0,0001

TABELA 3 - Resultados do teste t para os segmentos acústicos vocálicos do par S-A

Posição	N	Média	Desvio Padrão	t	p
1	99	-7	14	-5,1808	0,0001
2	45	-6	12	-3,3222	0,0018
4	175	-26	41	-8,6054	0,0001
5	113	4	16	2,4556	0,0156
6	42	-69	33	-13,761	0,0001
1235	260	-2	16	-2,4753	0,014
123456	477	-17	35	-10,703	0,0001

ANEXO D

TABELA 4 - Resultados do teste t para os segmentos acústicos consonantais do par M-A

Posição	N	Média	Desvio Padrão	t	p
1	74	36	37	8,5142	0,0001
2	41	37	26	9,2871	0,0001
4	172	31	36	11,367	0,0001
5	99	27	34	7,9221	0,0001
6	41	-21	62	-2,1155	0,0407
1235	217	32	33	14,138	0,0001
123456	430	27	41	13,513	0,0001

TABELA 5 - Resultados do teste t para os segmentos acústicos consonantais do par E-A

Posição	N	Média	Desvio Padrão	t	p
1	96	12	27	4,3272	0,0001
2	45	9	24	2,5246	0,0153
4	191	7	28	3,2717	0,0013
5	112	7	32	2,2294	0,0278
6	49	-36	45	-5,6782	0,0001
1235	256	9	29	5,1329	0,0001
123456	496	4	33	2,4992	0,0128

TABELA 6 - Resultados do teste t para os segmentos acústicos consonantais do par S-A

Posição	N	Média	Desvio Padrão	t	p
1	97	5	19	2,7084	0,008
2	45	-2	13	-1,2629	0,2133
4	186	-5	19	-3,619	0,0004
5	107	-3	20	-1,6667	0,0985
6	49	-49	51	-6,7201	0,0001
1235	252	0	19	0,1041	0,9172
123456	487	-7	28	-5,3581	0,0001

ANEXO E

TABELA 7 - Resultados do teste t para os segmentos acústicos do tamanho de sílabas para o par M-A

Posição	N	Média	Desvio Padrão	t	p
0	19	-3	41	-0,3299	0,7453
1	69	67	45	12,284	0,0001
2	41	75	50	9,5918	0,0001
4	139	43	64	7,9451	0,0001
5	81	72	47	13,725	0,0001
6	16	-53	63	-3,3524	0,0044
1235	194	70	47	20,897	0,0001
01235	213	63	51	18,323	0,0001
0123456	368	51	61	15,869	0,0001

TABELA 8 - Resultados do teste t para os segmentos acústicos do tamanho de sílabas para o par E-A

Posição	N	Média	Desvio Padrão	t	p
0	45	5	28	1,1991	0,2369
1	93	15	38	3,7941	0,0003
2	44	9	34	1,6689	0,1024
4	167	-9	64	-1,7415	0,0834
5	91	18	35	4,9269	0,0001
6	34	-92	60	-8,9656	0,0001
1235	231	15	36	6,1545	0,0001
01235	276	13	35	6,1868	0,0001
0123456	477	-2	56	-0,7985	0,425

TABELA 9 - Resultados do teste t para os segmentos acústicos do tamanho de sílabas para o par S-A

Posição	N	Média	Desvio Padrão	t	p
0	51	14	28	3,4103	0,0013
1	98	-2	20	-1,144	0,2554
2	45	-8	15	-3,6663	0,0007
4	172	-32	47	-8,9096	0,0001
5	112	1	19	0,4571	0,6485
6	42	-108	59	-11,905	0,0001
1235	258	-2	19	-1,8617	0,0638
01235	309	0	22	0,3219	0,7478
0123456	523	-19	47	-9,2735	0,0001

ANEXO F

Relação das palavras paroxítonas, di, tri e quadrissilábicas, em posição não-final

Dissílabos	Trissílabos	Quadrissílabos
late	cachorro	cachorrinho
para (o)	palhaço	palhacinha
chora	mergulha	elefante
planta	macaca	cachorrada
bate (o)	segura	macacada
banho	pedala	
toca	macaco	
rosa		
muita		
(o) anda		
cueca		

Relação das palavras paroxítonas, di, tri e quadrissilábicas, em posição final

Dissílabos	Trissílabos	Quadrissílabos
medo	palhaço	mamadeira
osso	chicote	bananeira
calça	palhaço	bicicleta
	chocalho	
	risada	
	barulho	
	motoca	
	bolinha	
	banana	

ANEXO G

TABELA 10 - Resultados do teste t para os segmentos acústicos do tamanho de palavras em posição não-final para o par M-A

Extensão Vocabular	N	Média	Desvio Padrão	t	p
di	15	87	84	4,0294	0,0012
tri	12	174	110	5,4836	0,0002
quadri	10	213	64	10,576	0,0001

TABELA 11 - Resultados do teste t para os segmentos acústicos do tamanho de palavras em posição não-final para o par E-A

Extensão Vocabular	N	Média	Desvio Padrão	t	p
di	25	27	75	1,8035	0,0839
tri	20	53	78	3,0218	0,007
quadri	17	7	75	0,3669	0,7185

TABELA 12 - Resultados do teste t para os segmentos acústicos do tamanho de palavras em posição não-final para o par S-A

Extensão Vocabular	N	Média	Desvio Padrão	t	p
di	26	-17	36	-2,3307	0,0282
tri	31	-15	46	-1,8095	0,0804
quadri	20	-13	41	-1,4453	0,1647

ANEXO H

TABELA 13 - Resultados do teste t para os segmentos acústicos do tamanho de palavras em posição final para o par M-A

Final	N	Média	DesvPad	t	p
tri	6	20	74	0,6605	0,5382
quadri	5	200	98	4,5611	0,0103

TABELA 14 - Resultados do teste t para os segmentos acústicos do tamanho de palavras em posição final para o par E-A

Final	N	Média	DesvPad	t	p
tri	16	-131	77	-6,8238	0,0001
quadri	10	-66	158	-1,3109	0,2224

TABELA 15 - Resultados do teste t para os segmentos acústicos do tamanho de palavras em posição final para o par S-A

Final	N	Média	DesvPad	t	p
di	8	-252	50	-14,21	0,0001
tri	18	-191	61	-13,278	0,0001
quadri	12	-145	77	-6,5013	0,0001

REFERÊNCIAS BIBLIOGRÁFICAS

ALBANO, E. C. **Da fala à linguagem Tocando de Ouvido**. São Paulo, Martins Fontes, 1990.

ALBANO, E. C. "Demarcative" features specification in phonology and phonetics: the case of Portuguese allophony/allomorphy. Poster apresentado na **Fifth Conference on Laboratory Phonology**. Evanston, Northwestern University, 1996.

ALBANO, E. C. O português brasileiro e as controvérsias da Fonética atual: pelo aperfeiçoamento da Fonologia Articulatória. **Revista da ALFAL**, submetido, 1997.

ALBANO, E.; BARBOSA, P.; GAMA-ROSSI, A.; MADUREIRA, S.; SILVA, A. A interface fonética-fonologia e a interação prosódia-segmentos. **Estudos Lingüísticos XXVII Anais de Seminários do GEL** (Trabalhos apresentados no XLV Seminário/1997 - UNICAMP, Campinas, SP), São José do Rio Preto, UNESP-IBILCE, p.135-43, 1998.

ALBANO, E. C. Dependências entre unidades fônicas na fonética e no léxico: o papel do contorno vocálico-acentual. **Relatório individual apresentado ao CNPq**, processo no 52.4110/96-4 LL, referente ao período de março de 1997 a julho de 1998.

ALBANO, E. C. **O gesto e suas bordas: para uma fonologia acústico-articulatória do português brasileiro**. Manuscrito inédito, 1999.

ALLEN, G. D. Segmental timing control in speech production. **Journal of Phonetics**, v.1, p.219-37, 1973.

AQUINO, P. A. **O papel das vogais reduzidas pós-tônicas na construção de um sistema de síntese concatenativa para o português brasileiro**. Campinas, 1997. Dissertação (Mestrado) - Laboratório de Fonética Acústica e Psicolingüística Experimental, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.

AVAAZ Innovations Inc. Sound Scientific Solutions. **CSRE User's Guide**. Ontario, 1995.

BARBOSA, P. A. **Caractérisation et génération automatique de la structuration rythmique du français**. Grenoble, 1994. Tese (Doutorado) - Institut de la Communication Parlée/ Institut National Polytechnique de Grenoble.

BARBOSA, P.; BAILLY, G. Characterisation of rhythmic patterns for text-to-speech synthesis. **Speech Communication**, v.15, p.127-37, 1994.

BARBOSA, P. A. At least two macrorythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation. **Cadernos de Estudos Lingüísticos**, v.31, p.33-53, 1996.

BECKMAN, M. E.; EDWARDS, J. Lengthenings and shortenings and the nature of prosodic constituency. In: KINGSTON, J.; BECKMAN, M. E. (eds.) **Papers in laboratory phonology I. Between the grammar and physics of speech**. Cambridge, Cambridge University Press, 1990.

BISOL, L. Sândi externo: o processo e a variação. Em: KATO, M. (org.) **Gramática do Português Falado Volume V: Convergências**. Campinas, Editora da UNICAMP/São Paulo, FAPESP, 1996.

BROWMAN, C.; GOLDSTEIN, L. Dynamic Modeling of Phonetic Structure. In: FROMKIN, V. A. (ed.) **Phonetic Linguistics Essays in honor of Peter Ladefoged**. Orlando, Academic Press, 1985.

BROWMAN, C.; GOLDSTEIN, L. Towards an articulatory phonology. **Phonology Yearbook**, v.3, p.219-52, 1986.

BROWMAN, C.; GOLDSTEIN, L. Articulatory gestures as phonological units. **Phonology**, v.6, p.201-51, 1989.

BROWMAN, C.; GOLDSTEIN, L. Tiers in articulatory phonology, with some implications for casual speech. In: KINGSTON, J.; BECKMAN, M. E. (eds.) **Papers in laboratory phonology I. Between the grammar and physics of speech**. Cambridge, Cambridge University Press, 1990.

BROWMAN, C.; GOLDSTEIN, L. Articulatory Phonology: An overview. **Phonetica**, v.49, p.155-80, 1992.

CÂMARA, J. M. **Problemas de lingüística descritiva**. 2.ed. Petrópolis, Vozes, 1969.

CHERMAK, G. D.; SCHNEIDERMAN, C. R. Speech timing variability of children and adults.[Letter to the editor]. **Journal of Phonetics**, v.13, p.477-80, 1986.

CLEMENTS, G. N.; KEYSER, S. J. **CV Phonology A generative theory of the syllable**. Cambridge, The MIT Press, 1983.

CORRÊA, L. M. S. Dificuldades e potencialidades do uso do método experimental no estudo da aquisição da linguagem. Em: CASTRO, M. F. P.(org.) **O Método e o Dado no Estudo da Linguagem**. Campinas, Editora da Universidade Estadual de Campinas, 1996.

D'ANGELIS, W. R. **Traços de modo e modos de traçar geometrias: línguas Macro-Jê e teoria fonológica**. Campinas, 1998. Tese (Doutorado) - Laboratório de Fonética Acústica e Psicolingüística Experimental, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.

EDWARDS, J.; BECKMAN, M. E. Articulatory timing and the prosodic interpretation of syllable duration. **Phonetica**, v.45, p.156-74, 1988.

FANT, G. **Speech Sounds and Features**. Cambridge, The MIT Press, 1973.

FLEGE, J. E. Anticipatory and carry-over nasal coarticulation in the speech of children and adults. **Journal of Speech and Hearing Research**, v. 31, p.525-36, 1988.

FOUGERON, C.; KEATING, P. A. Articulatory strengthening at edges of prosodic domains. **Journal of the Acoustical Society of America**, v.101, p.3728-40, 1997.

FOWLER, C.; SALTZMAN, E. Coordination and coarticulation in speech production. **Language and Speech**, v.36, p.171-95, 1993.

GAMA, A. J. A. **Fala e ação no cuidado materno ao bebê**. Campinas, 1989. Dissertação (Mestrado) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.

GAMA-ROSSI, A. J. A. A contribuição de uma abordagem motora para a aquisição de linguagem. **ANPOLL 10 anos, Boletim Informativo 22**. Brasília, 1994.

GAMA-ROSSI, A. J. A. Aspectos da coordenação de gestos articulatórios no desenvolvimento do português brasileiro. Palestra apresentada no **Simpósio "Coordenação da motricidade fonatória: estudos fonético-experimentais e clínicos"**, realizado durante o **II Congresso Brasileiro de Neuropsicologia**, no Centro de Convenções da Universidade Estadual de Campinas, de 18 a 20 de maio de 1995.

GAMA-ROSSI, A. J. A. Durações de segmentos e sílabas na fala de adultos e crianças em PB: Um estudo piloto. Comunicação apresentada no **XLIV Seminário do Grupo de Estudos Lingüísticos do Estado de São Paulo** (GEL), realizado na Universidade de Taubaté, de 23 a 25 de maio de 1996(a).

GAMA-ROSSI, A. J. A. Sobre as diferenças das durações de segmentos e sílabas na fala de adultos e crianças em PB. Comunicação apresentada no **II Seminário Teses em Andamento**, realizado no Instituto de Estudos da Linguagem, da Universidade Estadual de Campinas, de 30 de setembro a 2 de outubro de 1996(b).

GAMA-ROSSI, A. J. A.; ALBANO, E. C. Possible phonological influences on segment duration differences between adults and children: a pilot study. Poster apresentado na **Fifth Conference on Laboratory Phonology**. Evanston, Northwestern University, 1996.

GAMA-ROSSI, A. J. A.; ALBANO, E. C. Aspectos neuromotores e lingüísticos das diferenças de durações entre a fala adulta e infantil no PB. **Letras de Hoje**, v.33, p.115-121, junho de 1998.

GONÇALVES, M. J. **A construção da fala por uma criança**. Campinas, 1989. Dissertação (Mestrado) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.

GOODELL, E. W.; STUDDERT-KENNEDY, M. Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: a longitudinal study. **Journal of Speech and Hearing Research**, v.36, p.707-27, 1993.

HERNANDORENA, C. L. M. **Aquisição da fonologia do Português: estabelecimento de padrões com base em traços distintivos**. Porto Alegre, 1990. Tese (Doutorado). Pontifícia Universidade Católica do Rio Grande do Sul.

KENT, R. D. Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. **Journal of Speech and Hearing Research**, v.19, p.421-47, 1976.

KENT, R. D.; FORNER, L. L. Speech segment durations in sentence recitations by children and adults. **Journal of Phonetics**, v.8, p.157-68, 1980.

KENT, R. D.; READ, C. **The acoustic analysis of speech**. San Diego, Singular Publishing Group, Inc., 1992.

KENT, R. D. Desenvolvimento fonológico como biologia e comportamento. Em: CHAPMAN, R. S. (org.) **Processos e Distúrbios na Aquisição da Linguagem**. Porto Alegre, Artes Médicas, 1996.

KINGSTON, J.; COHEN, A. H. Extending Articulatory Phonology. **Phonetica**, v.49, p.194-204, 1992.

KLATT, D. H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. **Journal of the Acoustical Society of America**, v.59, p.1208-21, 1976.

KOHLER, K. J. Gestural Reorganization in Connected Speech: A Functional Viewpoint on 'Articulatory Phonology'. **Phonetica**, v.49, p.205-11.

KUBASKA, C. A.; KEATING, P. A. Word duration in early child speech. **Journal of Speech and Hearing Research**, v.24, p.615-21, 1981.

LAMPRECHT, R. R. Influência de fatores fonéticos e fonológicos na aquisição das obstruintes sonoras do Português. Em: Anais do **II Encontro Nacional sobre Aquisição da Linguagem**, Porto Alegre, CEAAL/PUCRS, p.165-83, 1992.

LEHISTE, I. **Readings in acoustic phonetics**. Cambridge, The MIT Press, 1967.

LEVY, I. P. **Uma nova face da nau dos insensatos: a dificuldade de vozear obstruintes em crianças de idade escolar**. Campinas, 1993. Tese (Doutorado) - Laboratório de Fonética Acústica e Psicolingüística Experimental, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.

LIBERMAN, M.; PRINCE, A. On stress and linguistic rhythm. **Linguistic Inquiry**, v.8, p.249-336, 1977.

LINDBLOM, B. The status of phonetic gestures. **PERILUS XI**, p.21-39, 1990.

LUSTIG, L.; PENICK, I. **The pop-up book of the circus**. Great Britain, Hutchinson Junior Books Ltd., 1980.

MACK, M.; LIEBERMAN, P. Acoustic analysis of words produced by a child from 46 to 149 weeks. **Journal of Child Language**, v.12, p.527-50, 1985.

MASSINI, G. **A duração no estudo do acento e ritmo do português**. Campinas, 1991. Dissertação (Mestrado) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.

MENN, L. Development of articulatory, phonetic and phonological capabilities. In: BUTTERWORTH, B.(ed.) **Language Production, Volume 2: Development, Writing and Other Processes**. London, Academic Press, 1983.

MONTGOMERY, D. C. **Design and analysis of experiments**. 4th. ed., John Wiley and Sons, 1996.

MORAES, J. A. **Recherches sur l'intonation modale du portugais brésilien parlé à Rio de Janeiro. Analyse acoustique, perceptive et fonctionnelle**. Paris, 1984. Tese (Doutorado) - Université de la Sorbonne Nouvelle Paris III.

MURRAY, I. R.; ARNOTT, J. L. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. **Journal of the Acoustical Society of America**, v.93, p.1097-108, 1993.

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

NITTROUER, S.; STUDDERT-KENNEDY, M.; McGOWAN, R. S. The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. **Journal of Speech and Hearing Research**, v.32, p.120-32, 1989.

PERRONI, M. C. Diferenças individuais em aquisição da linguagem: um estudo sobre gêmeos. Em: Anais do **II Encontro Nacional sobre Aquisição da Linguagem,** Porto Alegre, CEAAL/PUCRS, p.203-11, 1992.

PORT, R.; CUMMINS, F.; GASSER, M. A dynamic approach to rhythm in language toward a temporal phonology. In: DAINOR, A.; HEMPHILL, R.; LUKA, B.; NEED, B. AND PARGMAN, S. (eds.) **Papers from the 31th Meeting of the Chicago Linguistic Society**. Chicago, Chicago Linguistic Society,1995.

REPP, B. Some observations on the development of anticipatory coarticulation. **Journal of the Acoustical Society of America**, v.79, p.1616-9, 1986.

RODRIGUES, N. **Neurolingüística dos Distúrbios da Fala**. São Paulo, Editora Cortez, 1989.

SCARPA, E. M. Interfaces entre componentes e representação na aquisição da prosódia. Conferência apresentada no **IV Encontro Nacional de Aquisição da Linguagem**, PUCRS, Porto Alegre, 1997.

SCARPA, E. M. (org.) **Estudos de prosódia**. Campinas, Editora da UNICAMP [no prelo].

SERENO, J. A.; BAUM, S. R.; MAREAN, G. C.; LIEBERMAN, P. Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. **Journal of the Acoustical Society of America**, v.81, p.512-9, 1987.

SERENO, J. A. ; LIEBERMAN, P. Developmental aspects of lingual coarticulation. **Journal of Phonetics**, v.15, p.247-57, 1987.

SILVA, A. H. P. **Para a descrição fonético-acústica das líquidas no português brasileiro: dados de um informante paulistano**. Campinas, 1996. Dissertação (Mestrado) - Laboratório de Fonética Acústica e Psicolingüística Experimental, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.

SMITH, B. L.; SUGARMAN, M. D.; LONG, S. H. Experimental manipulation of speaking rate for studying temporal variability in children's speech. **Journal of the Acoustical Society of America**, v.74, p.744-9, 1983.

SMITH, B. Relationships between duration and temporal variability in children's speech. **Journal of the Acoustical Society of America**, v.91, p.2165-74, 1992.

SMITH, B. Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adult's speech. **Journal of Phonetics**, v.22, p.155-75, 1994.

SMITH, B. L.; KENNEY, M. K. An assessment of several acoustic parameters in children's speech production development: longitudinal data. **Journal of Phonetics**, v. 26, p.95-108, 1998.

SOUZA, E. M. G. **Para a caracterização fonético-acústica da nasalidade no português do Brasil**. Campinas, 1994. Dissertação (Mestrado) - Laboratório de Fonética Acústica e Psicolingüística Experimental, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.

STATHOPOULOS, E. T. Variability revisited: an acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech. **Journal of Phonetics**, v.23, p.67-80, 1995.

STERIADE, D. Gestures and autosegments: comments on Browman and Goldstein's paper. In: KINGSTON, J.; BECKMAN, M. E. (eds.) **Papers in laboratory phonology I Between the grammar and physics of speech**. Cambridge, Cambridge University Press, 1990.

STUDDERT-KENNEDY, M. The phoneme as a perceptuomotor structure. In: ALLPORT, A.; MACKAY, D. G.; PRINZ, W.; SCHEERER, E. (eds.) **Language, Perception and Production. Relationships between listening, speaking, reading and writing**. Orlando, Academic Press, 1987.

TEIXEIRA, E. R. A aquisição das classes de sons e a aplicação dos processos de simplificação fonológica. Trabalho apresentado na mesa-redonda "A aquisição da Fonologia", GT de Psicolingüística, no **XI Encontro Nacional da ANPOLL**. João Pessoa, 1996.

TURNBAUGH, K. R.; HOFFMAN, P. R.; DANILOFF, R. G.; ABSHER, R. Stop-vowel coarticulation in 3-year-old, 5-year-old and adult speakers. **Journal of the Acoustical Society of America**, v.77, p.1256-7, 1985.

VIHMAN, M. M. Ontogeny of Phonetic Gestures: Speech Production. In: MATTINGLY, I. G.; STUDDERT-KENNEDY, M. (eds.) **Modularity and the Motor Theory of Speech Perception. Proceedings of a Conference to Honor Alvin M. Liberman**. Hillsdale, Lawrence Erlbaum Associates, Publishers, 1991.

VOGEL, I.; BUNNELL, T.; HOSKINS, S. The phonology and phonetics of the Rhythm Rule. In: CONNELL, B.; ARVANITI, A. (eds.) **Phonology and phonetic evidence. Papers in Laboratory Phonology IV**. Cambridge, Cambridge. University Press, 1995.

YAVAS, M. S. Padrões na Aquisição da Fonologia no Português. **Letras de Hoje**, v.23, p.7-30, 1988.

APÊNDICE

Na seqüência, são mostradas as figuras correspondentes a cada uma das sentenças do *corpus*, na seguinte ordem:

A macacada faz barulho.

A macaca toca chocalho.

O macaco segura a banana.

O macaquinho anda de bicicleta.

O elefante planta bananeira.

O elefantinho pedala a motoca.

O elefante dá banho no palhaço.

A palhaçada dá muita risada.

O palhaço mergulha na bacia.

A cueca do palhaço tem muita bolinha.

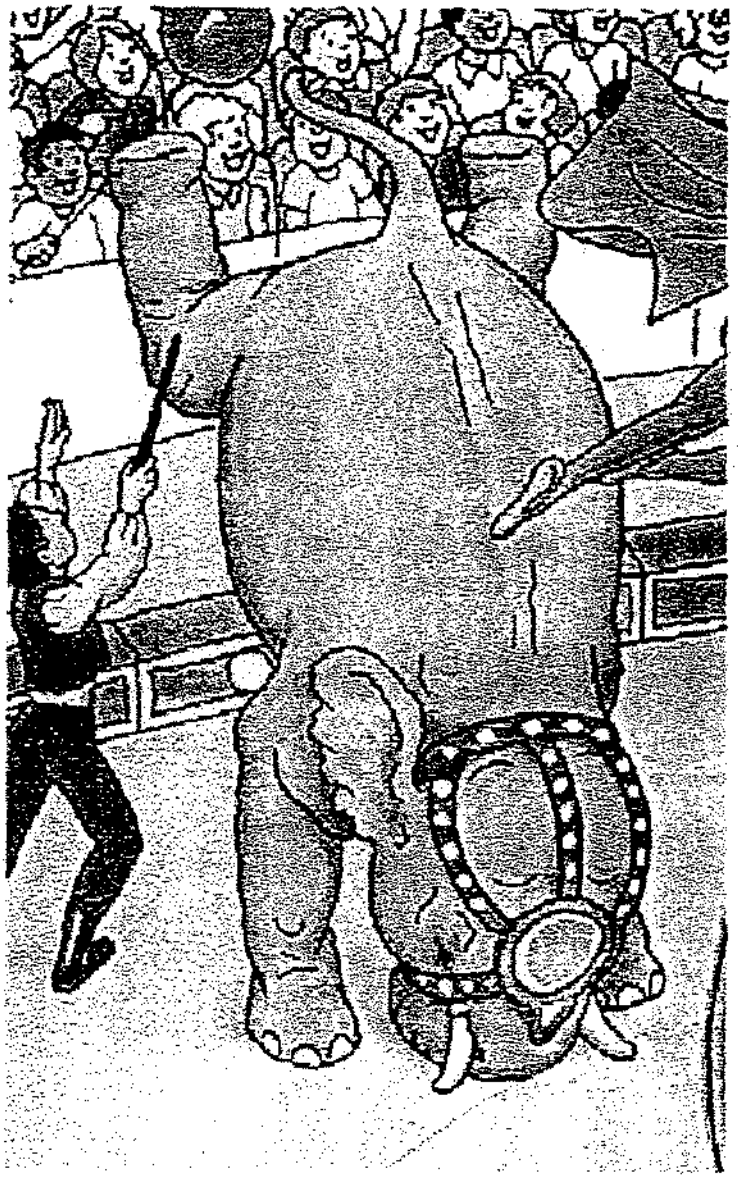
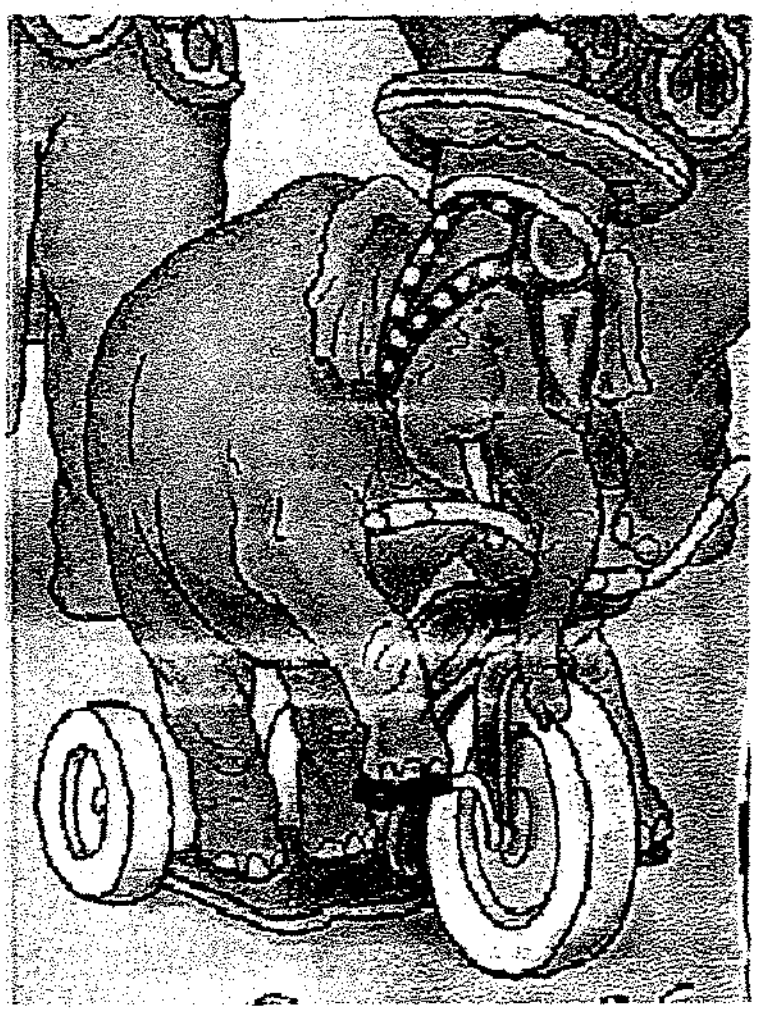
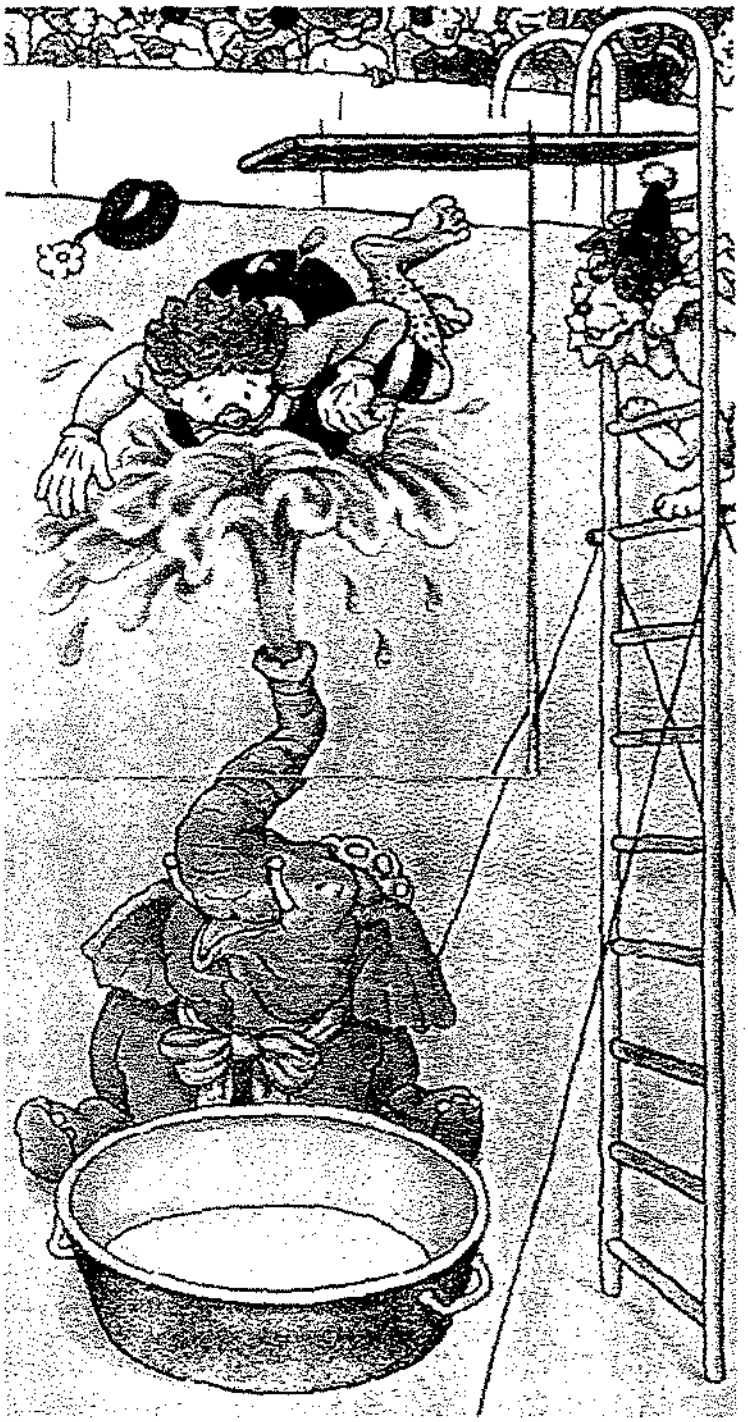
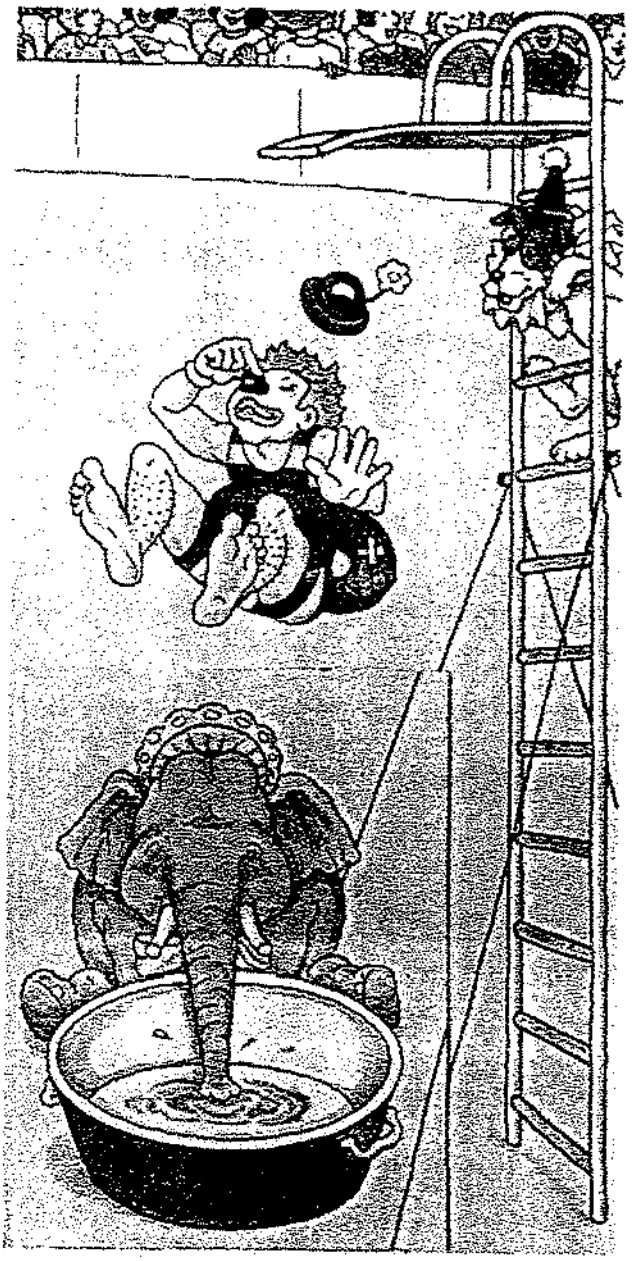
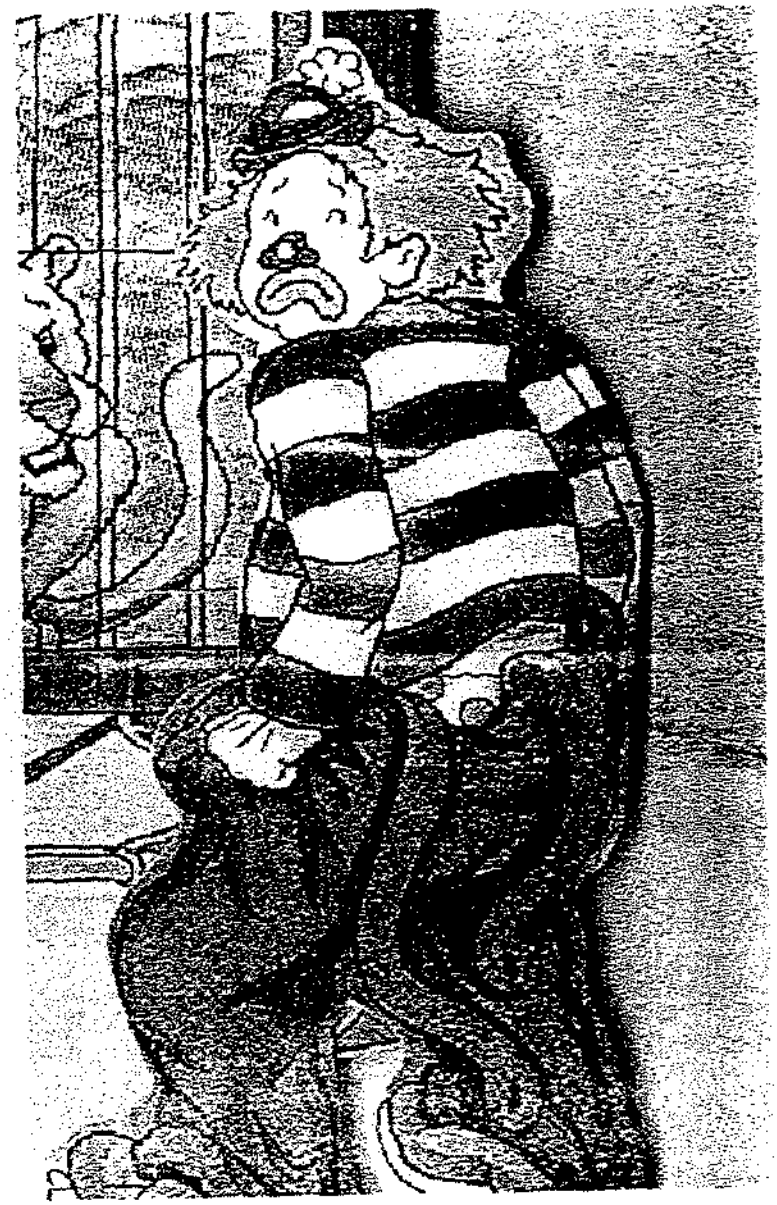
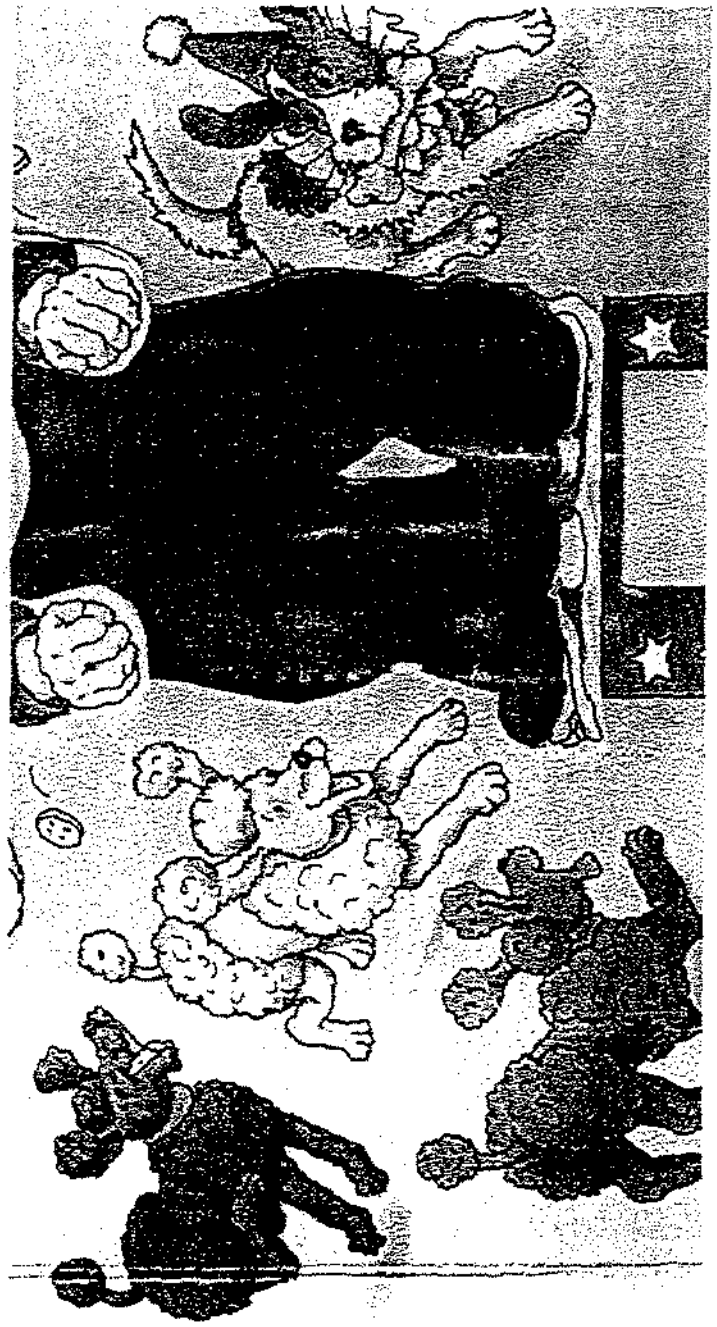
O palhaço cor de rosa planta bananeira.

O palhaço Chapolin segura a calça.

A palhacinha não quer mamadeira.

A cachorrada quer osso.

O cachorrinho chora de medo.

O cachorro late para o palhaço.

O domador bate o chicote.