

## A Romaria

*Para Dom Casaldáliga e para  
todos os que caminham*

Fogo e diamante no céu azul  
de São Félix do Araguaia,  
no coração do Brasil,  
no meu peito, nos olhos dos caminhantes,  
em seu bispo Dom Pedro,  
no encontro do povo todo  
e de todas as galáxias,  
para recordar a dor dos mártírios  
e a força da vontade do homem,  
do Padre João Bosco Penido Burnier.  
Sóis e estrelas para iluminar  
o caminho da caminhada.  
Neste caminho de gente  
incendiada de amor,  
a pomba da paz, cruzeiro de luz,  
junta o perfume da selva  
e o desejo do povo, harmonizando  
até o mais longínquo espaço do cosmo.

*Teresinka Pereira*

# Relações entre percepção e produção na aquisição da duração da vogal no português brasileiro\*

Aglael Gama-Rossi\*\*

## 1 Diferenças entre fala adulta e infantil para a duração da vogal

Recentemente apresentamos um estudo (Gama-Rossi, 1999) sobre a aquisição da duração no português brasileiro, do qual participaram duas crianças de idade média de 4 anos e 5 meses e um adulto, todos do sexo feminino e nascidos na cidade de São Paulo. As medidas de duração foram feitas para os segmentos acústicos correspondentes a vogais, consoantes e sílabas, dentro de um conjunto de 17 sentenças, enunciadas pela pesquisadora e repetidas por cada sujeito, no mínimo três vezes, sendo que a ordem do conjunto de sentenças diferia a cada repetição. Os dados foram colhidos em sessões individuais de gravação, realizadas na escola das crianças.

A fala da pesquisadora durante a apresentação do conjunto de sentenças a cada sujeito foi também gravada e analisada. Constatou-se, por meio do teste de Duncan, que não havia diferenças significativas entre os modelos apresentados aos sujeitos para repetição. Em outras palavras, as diferenças entre a realização da tarefa pelo adulto e pelas crianças não estavam condicionadas por diferenças dos modelos fornecidos a cada um deles. Isso possibilitou uma análise comparativa dos sujeitos dentro de pares forma-

\* Agradeço o diálogo sempre inspirador com Eleonora Cavalcante Albano, Plínio Almeida Barbosa e Sandra Madureira; a Paulo Rehder, pela análise estatística; e a Benedito Mauro Rossi, pela diagramação dos gráficos.

\*\* PUCSP - Pontifícia Universidade Católica de São Paulo.

dos por: professora das crianças e pesquisadora (par S-A), criança de 4 anos e 1 mês e pesquisadora (par M-A), e criança de 4 anos e 9 meses e pesquisadora (par E-A). Dentro de cada par, a comparação entre os sujeitos foi feita controlando-se: (1) o tipo de segmento acústico (correspondente a vogal, consoante e sílaba) e (2) a posição de acento por ele ocupada na palavra (pré-tônica, tônica e pós-tônica) ou na sentença (início e final).

A comparação entre os sujeitos de cada par deu-se para a média das diferenças de duração entre suas falas, sendo as diferenças de duração calculadas da seguinte forma: a duração do segmento acústico na repetição do sujeito (M, E ou S) menos a duração do mesmo segmento acústico no modelo oferecido pela pesquisadora. A hipótese nula era de que, se o sujeito e a pesquisadora levassem o mesmo tempo para realizar o segmento acústico, a média das diferenças de duração seria zero. Esse procedimento foi feito para cada tipo de segmento acústico, em cada posição de acento. Se a média fosse positiva, as durações tinham sido maiores na fala do sujeito; se a média fosse negativa, na fala da pesquisadora. Um teste t mostrava se havia ou não diferenças significativas entre os sujeitos.

Os gráficos abaixo mostram os resultados, para cada par, da média (expressa em barras) das diferenças de duração de segmentos acústicos correspondentes a vogais, consoantes e sílabas, nas posições designadas por "pré 1" (primeira pré-tônica de palavras tri ou quadrissilábicas), "pré 2" (segunda pré-tônica de palavras quadrissilábicas), "tônica" (tônicas de todas as palavras, incluindo a maior proeminência da sentença) e "pós" (pós-tônicas de palavras em posição não-final de sentença), sendo que os segmentos acústicos pertencentes à pós-tônica de final da sentença foram analisados separadamente sob "final", uma vez que estão sujeitos ao alongamento de final de sentença. Na posição designada por "átomas", são mostradas as médias das diferenças de duração do conjunto de pré e pós-tônicas. Finalmente, na posição "todas", estão todas as ocorrências de cada tipo de segmentos acústicos em todas as posições de acento.

Gráfico 1  
Par S-A (professora e pesquisadora)

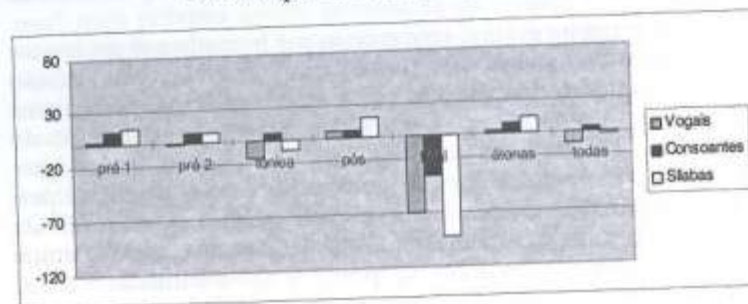


Gráfico 2  
Par M-A (criança de 4 anos e 1 mês e pesquisadora)

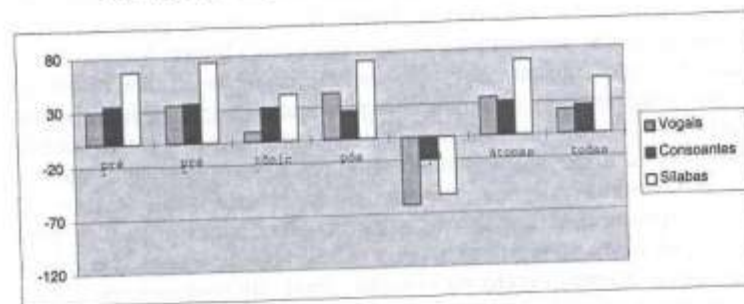
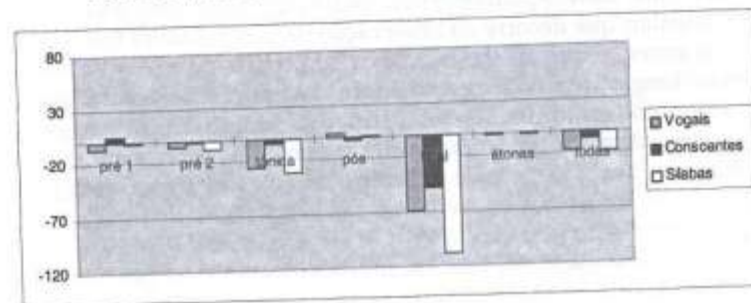


Gráfico 3  
Par E-A (criança de 4 anos e 9 meses e pesquisadora)



Nesta comunicação, interessa-nos especificamente a implementação da duração dos segmentos acústicos correspondentes às vogais.

No primeiro gráfico, referente ao par formado por professora e pesquisadora, ambas igualam-se na implementação da duração de consoantes pertencentes às posições de atonicidade, conforme demonstra a média (igual a zero) das diferenças de duração do conjunto de consoantes das posições pré e pós-tônica (sob "átomas"), não havendo diferença significativa entre professora e pesquisadora ( $n=252$ , média=0, desvio-padrão=19ms e  $p=0,9172$ ). Ou seja, esse resultado demonstra que ambas levaram o mesmo tempo para produzir as consoantes nas posições não-acentuadas. Para as vogais, em cada uma das posições de acento consideradas, são sempre encontradas diferenças significativas entre professora e pesquisadora ("pré 1",  $p=0,0001$ ; "pré 2",  $p=0,0018$ ; "tônica"=0,0001; "pós"=0,0156; "final"=0,001; "átomas"=0,014; "todas"=0,0001). Juntos, esses dois dados apontam respectivamente (1) para a estabilidade neuromotora da fala adulta, observada na maior precisão, entre repetições, das consoantes não-acentuadas, e (2) para a vogal como um elemento que está a serviço da realização de diferenças estilísticas, uma vez que, as vogais no geral mais longas na fala da pesquisadora levam-nos a supor que ela tende a apresentar uma fala mais "clara", com acentos mais exagerados, do que a professora.

No segundo gráfico, observa-se que a criança M (4 anos e 1 mês) apresenta durações mais longas que a pesquisadora para todos os tipos de segmentos acústicos analisados, em todas as posições de acento, exceto na posição "final" de sentença, na qual a pesquisadora realiza de modo consistente o alongamento da sílaba final (barras invertidas). Porém, a menor média das diferenças de duração na comparação entre a criança M e a pesquisadora é obtida justamente para as vogais em posição "tônica" ( $n=158$ , média=10ms, desvio-padrão=51ms,  $p=0,0115$ ). Um outro dado complementar, que decorre da observação da pesquisadora nos outros dois pares, é que apenas a criança M produz durações vocálicas mais longas do que a pesquisadora na posição de acento lexical (que, como já foi dito, inclui a maior proeminência da sentença). Ou seja, uma conclusão possível é a de que, para compensar suas durações mais longas nas posições de atonicidade, a criança M "exagera" seus picos de acento e o faz por meio de um maior alongamento da duração vocálica.

A observação do terceiro gráfico, referente ao par formado pela criança E (4 anos e 9 meses) e pela pesquisadora, sugere imediatamente que a criança E, embora apenas 8 meses mais velha do

que a criança M, encontra-se num momento distinto da aquisição da duração. O Gráfico 3 apresenta uma configuração de barras no geral mais próxima àquela encontrada no Gráfico 1, referente ao par de adultos. Porém, a maior aproximação entre a criança E e a pesquisadora não está, como ocorre para os adultos, nas consoantes átonas, mas nas vogais, sendo que não são encontradas diferenças significativas entre a criança E e a pesquisadora para as vogais da primeira e segunda pré-tônicas (respectivamente,  $n=97$ , média=3ms, desvio-padrão=26ms,  $p=0,3028$  e  $n=44$ , média=-1ms, desvio-padrão=25ms,  $p=0,8711$ ). Ou seja, a criança E apresenta, nas posições de atonicidade, segmentos acústicos bem mais reduzidos do que a criança M; porém, a redução da vogal antecede a redução de consoantes e sílabas.

Portanto, em se tratando da implementação da duração da vogal, as seguintes conclusões podem ser delineadas a partir dos dados apresentados: (1) no que concerne aos adultos, a vogal serve para apresentar distinções estilísticas na realização do contorno duracional do português brasileiro; (2) no que concerne às crianças, apesar de elas estarem em momentos distintos da aquisição do parâmetro de duração e de ainda apresentarem restrições neuromotoras que impedem-nas de realizar as reduções dos elementos átonos, do modo previsto pela língua (Câmara, 1969; Gama-Rossi, 1999), para ambas, a aquisição da duração da vogal antecede a de consoantes e sílabas.

A questão trazida a esta comunicação refere-se a de que forma o fato de as crianças ajustarem primeiramente a implementação do parâmetro de duração para a vogal serve, dentro de modelos dinâmicos de produção e percepção de fala, para pensar a relação entre percepção e produção na aquisição da língua.

## 2 Um modelo dinâmico de produção de fala e a aquisição da duração da vogal

Ao explicar as diferenças entre os pares de sujeitos analisados a partir da implementação do parâmetro de duração (Gama-Rossi, 1999), adotamos um modelo dinâmico, a Fonologia Articulatória (Browman e Goldstein, 1990), que tem como unidade de análise, o gesto articulatório, uma unidade definida simultaneamente de forma abstrata e dinâmica.

Em linhas muito gerais, na sua face mais abstrata, o gesto articulatório organiza os itens lexicais em pautas gestuais, que contém gestos tais como: abertura e fechamento do véu palatino ( $(-)\mu$ ),

construção faríngea ([a]), fechamento labial ([β]), abertura e fechamento da glote ([γ]), etc., organizados na sua ordem canônica de ocorrência. Na sua face dinâmica, os gestos abstratos da pauta gestual estão associados às variáveis do trato, ou seja, a conjuntos de articuladores (fisicamente) atrelados entre si para a realização de uma dada tarefa de fala ou constrição no trato vocal (por exemplo, o gesto de fechamento bilabial está associado às variáveis do trato de abertura labial e protrusão labial, sendo que em ambas são coordenados os movimentos conjuntos dos articuladores de lábio superior, lábio inferior e mandíbula). O movimento de cada variável do trato é implementado por meio de uma equação que descreve o movimento de oscilação de uma massa presa a uma mola, e que pode gerar, pela substituição dos valores de seus parâmetros, infinitas trajetórias para cada variável do trato.

A coordenação entre os gestos de um enunciado é especificada na pauta gestual por meio de relações de fase entre os estados dinâmicos dos gestos. As relações de fase estabelecem em que momento (início, alvo ou fim) de um gesto em andamento inicia a atividade de um outro gesto. Com isso, os gestos articulatorios podem sobrepor-se em diferentes graus aos gestos articulatorios que os precedem ou sucedem, em função de variações na taxa de elocução ou nos padrões prosódicos da fala. Processos fônicos tais como o apagamento e a inserção de segmentos são explicados pela maior ou menor sobreposição entre gestos articulatorios ou por mudanças em sua magnitude.

Também para a coordenação de gestos dentro de um enunciado e entre enunciados, o modelo propõe duas camadas funcionais: a camada da vogal e a camada da consoante. A separação entre gestos vocálicos e consonantais em camadas distintas baseia-se em evidências acústicas e articulatorias (Fowler e Saltzman, 1993) a favor da idéia de que as vogais são produzidas de modo mais ou menos contínuo na fala, formando, no sentido de Öhman (1966), extensos "ditongos" (no original *diphthongally*) e, portanto, as vogais são apenas parcialmente coarticuladas entre si, independentemente do número de consoantes intervenientes. Ao contrário, as consoantes têm sido entendidas como superpostas ao fluxo vocálico em andamento. Evidências estatísticas significativas de coarticulação antecipatória e perseveratória entre vogais, em enunciados do tipo VCV, foram encontradas (Recasens, 1984b, apud Fowler e Saltzman, 1993) em dados eletropalatográficos, mesmo quando influências coarticulatórias antecipatórias ou perseveratórias das vogais sobre a consoante eram bloqueadas devido à consoante ser

uma palatal (/k/), e assim colocar fortes restrições sobre o corpo da língua (um articulador primário na produção das vogais).

Voltando aos dados discutidos na seção anterior, é possível hipotetizar que as crianças, na tarefa de adquirir sua língua materna, comecem por coarticular os gestos vocálicos entre si, o que explicaria que, mesmo para crianças que se encontram em momentos distintos da aquisição do parâmetro de duração, caso de M e E, o manejo da duração vocálica precede o da duração consonantal.

De acordo com Browman e Goldstein (1989), o gesto único de constrição das primeiras vocalizações infantis tem de ser diferenciado numa variedade de gestos potencialmente distintos, ajustados com diferentes valores para o grau de constrição (por exemplo, gestos de fechamento completo ou parcial, ou ainda gestos mais amplos) e para o local de constrição (por exemplo, dental, alveolar, pós-alveolar, etc.). Os autores colocam que a partição dos contínuos físicos, constituídos pelo grau e local de constrição, em categorias discretas é provavelmente auxiliada por relações articulatorias e auditivas quânticas, ou não-lineares. O problema é o que os autores não explicam como isso ocorre.

Albano (no prelo), numa reformulação da Fonologia Articulatória, propõe "definir um gesto como uma manobra motora que produz efeitos acústicos que identificam tipos distintos de constrição e as suas respectivas regiões no trato vocal. Assim, por exemplo, o que confere unidade ao gesto de fechamento dental não é o articulador ativo da ponta da língua, mas a amplitude baixa ou zero da oclusão acoplada ao caráter inequivocamente agudo das ressonâncias produzidas na região coronal". Mais adiante, a autora afirma que "a unidade acústica do gesto é construída *off line*, primeiro durante o processo de aquisição de linguagem e, depois, na mudança linguística socialmente motivada".

Evidências não faltam a favor da construção *off line* da contraparte acústica do gesto articulatorio na aquisição de linguagem. Werker e Polka (1993) colocam que é possível localizar três fases distintas, no desenvolvimento da percepção de fala, durante o primeiro ano de vida. Do nascimento até os 6 meses, a criança é capaz de discriminar muitos ou quase todos os contrastes fonéticos das línguas do mundo. Essa capacidade pode perdurar até cerca dos 8 meses, mas torna-se geralmente indisponível entre 10 e 12 meses, quando a língua nativa passa a influenciar decisivamente a percepção de fala. Não por acaso, nessa época também, os bebês começam a balbuciar sílabas reduplicadas e a imitar sons da fala adulta, o que, segundo Studdert-Kennedy (1987), seria a primeira evidência de um vínculo perceptomotor na criança. Além disso, as

crianças reproduzem contornos entonacionais da fala adulta antes de produzirem suas contrapartes segmentais (Gonçalves, 1989; Gama, 1989).

Assim, o exercício a que nos propomos na seção seguinte é o de tentar relacionar um fato articulatório (a precedência da produção da duração da vogal na fala de duas crianças brasileiras) a uma compreensão dinâmica da percepção de fala e a uma possível evidência a favor de um invariante motor (uma vez que a busca de invariantes na percepção de fala tem sido restrita a alguns exemplos, não havendo, portanto, uma relação unívoca entre os sons da fala e suas representações no sinal acústico).

### 3 A aquisição da duração da vogal a partir da redução da percepção à produção

Como foi dito anteriormente, esta comunicação pretende discutir, à luz de modelos dinâmicos de produção e percepção de fala, de que forma a precedência da implementação da duração da vogal na fala infantil pode levar-nos a pensar a relação entre percepção e produção na aquisição da língua. Para isso, nesta última seção, valemo-nos da Teoria da Percepção Direta (ou Realismo Direto) (Fowler e Rosenblum, 1991) e de um achado relativo a um aspecto invariante nas relações de fase entre gestos articulatórios (Harris, Tuller e Kelso, 1986).

O modelo da Percepção Direta parte da Psicologia Ecológica (Gibson, 1966), para a qual todo tipo de percepção envolve a recuperação de eventos distais a partir da informação sobre eles na estimulação proximal. Ou seja, na medida em que a estrutura encontrada na onda sonora (percepção auditiva) ou na radiação eletromagnética (percepção visual) é específica às propriedades do evento distal que a causou, a onda sonora ou a radiação eletromagnética serve de informação sobre a fonte distal (como exemplo, na onda sonora, a fonte distal pode ser uma campainha ou um trato vocal). O papel da estimulação proximal é estimular os sistemas perceptuais (audição ou visão) e informá-los sobre o evento distal. Portanto, na Psicologia Ecológica, ouvir ou ver, é ouvir ou ver eventos distais com base na informação sobre eles contida na estimulação proximal dos órgãos sensoriais.

No que concerne à percepção da fala, a fonte distal é o trato vocal. Segundo Fowler e Rosenblum (op. cit.), se o trato vocal produz gestos articulatórios, então o evento distal é o conjunto de gestos que compõe o enunciado do falante. O estímulo proximal é o sinal acústico estruturado pelo movimento do trato vocal e, na

medida em que a estrutura no sinal tende a ser específica ao evento que a causou, o sinal acústico serve de informação ao ouvinte sobre os eventos ocorridos no trato vocal do falante. Em outras palavras, a informação fornecida pela estimulação proximal dos "ouvidos" do ouvinte refere-se aos gestos articulatórios produzidos no trato vocal do falante.

Com base nessa teoria, podemos inferir que o que a criança percebe são os gestos articulatórios produzidos no trato vocal do adulto e que as vogais, por serem gestos que servem de fundo à fala, têm um papel primordial na sua percepção.

Vimos na seção anterior que a trajetória de um gesto articulatório no trato vocal pode ser dividida pelo menos em três pontos: início, alvo e fim; e que os gestos coordenam-se entre si por meio de relações de fase, ou seja, num dado momento da ativação de um gesto em andamento, um outro inicia ou extingue-se. Harris, Tuller e Kelso (1986) afirmam que há, nas estruturas articulatórias, bases aferentes para a informação do ângulo de fase (ou seja, se o movimento completo de um gesto pode ser visto abstratamente como um ciclo subjacente de 360°, numa seqüência VC, o alvo do gesto consonantal, atingido em 240°, ocorre a 330° do gesto vocálico). A informação do ângulo de fase é dada pela sensibilidade à posição e velocidade dos fusos musculares (*muscle spindles*) e das articulações. Segundo Harris et al. (op. cit.), pode ser que certos ângulos de fase forneçam informação para orquestrar o fluxo temporal de atividade entre articuladores e/ou configurações do trato vocal. Tais ângulos de fase serviriam como fontes de informação natural, dinamicamente especificadas para coordenar a fala.

Concluindo, uma tentativa de unir os dados às teorias aqui expostas seria supor que, durante o processo de aquisição de uma língua, a criança constrói *representações motoras*, a partir das informações (auditivas e/ou visuais) que recebe sobre a forma como os adultos (ou falantes mais experientes da língua) coordenam seus gestos ou conjuntos de gestos em sons significativos e das informações proprioceptivas sobre as possibilidades motoras de seu trato vocal, decorrentes da experiência própria em seu manejo. As vogais, por serem gestos com inícios e finais mais lentos, minimamente coarticulados entre si, servindo de "fundo" à coarticulação consonantal, são excelentes candidatos para serem tomados pelas crianças como os "tijolos" de tais *representações motoras*.

## Referências bibliográficas

- ALBANO, E. C. *O gesto e suas bordas: para uma fonologia acústico-articulatória do português brasileiro*. Campinas: Mercado Aberto, no prelo.
- BROWMAN, C.; GOLDSTEIN, L. Articulatory gestures as phonological units. *Phonology*, v. 6, p. 201-51, 1989.
- . 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.
- CÂMARA, J. M. *Problemas de lingüística descritiva*. 2. ed. Petrópolis: Vozes, 1969.
- FOWLER, C.; ROSENBLUM, L. D. The perception of phonetic gestures. 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, 1991.
- ; SALTZMAN, E. Coordination and coarticulation in speech production. *Language and Speech*, v. 36, p. 171-95, 1993.
- GAMA-ROSSI, A. J. A. *Relações entre desenvolvimento lingüístico e neuromotor: a aquisição da duração no português brasileiro*. Tese de doutorado, LAFAPE-IEL-UNICAMP, Campinas, 1999.
- GAMA, A. J. A. *Fala e ação no cuidado materno ao bebê*. Dissertação de Mestrado, IEL-UNICAMP, Campinas, 1989.
- GIBSON, J. J. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.
- GONÇALVES, M. J. *A construção da fala por uma criança*. Dissertação de Mestrado, IEL-UNICAMP, Campinas, 1989.
- HARRIS, K. S.; TULLER, B.; KELSO, J. A. S. Temporal invariance in the production of speech. In: PERKELL, J. S.; KLATT, D. H. (eds.) *Invariance and variability in speech processes*. Hillsdale: Lawrence Erlbaum, 1986.
- ÖHMAN, S. Coarticulation in VCV utterances: spectrographic measures. *Journal of the Acoustical Society of America*, 41, p. 151-168, 1966.
- 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.
- WERKER, J. F.; POLKA, L. Developmental changes in speech perception: new challenges and new directions. *Journal of Phonetics*, v. 21, n. 1-2, p. 83-101, 1993.