

Barbosa, P. A. (2004) Elementos para uma tipologia do ritmo (lingüístico) da fala à luz de um modelo de osciladores acoplados. *In Cognito-Cadernos Românicos em Ciências Cognitivas. 2 (1), 31-58.*

---

## **ELEMENTOS PARA UMA TIPOLOGIA DO RITMO (LINGÜÍSTICO) DA FALA A LUZ DE UM MODELO DE OSCILADORES ACOPLADOS**

**Plínio A. BARBOSA**

*Lab. de Fonética e Psicolingüística e Dep. de Lingüística – Instituto de Estudos da Linguagem/Unicamp  
CP 6045, 13084-971 Campinas (SP), Brasil  
E-mail: plinio@iel.unicamp.br*

---

### **Resumo**

*O modelamento dinâmico do ritmo da fala fundamentado na teoria de osciladores acoplados possibilita abordar o estudo dos padrões duracionais dos enunciados de uma língua a partir da consideração de dois níveis temporais : um tempo extrínseco ou prosódico e um tempo intrínseco ou segmental. O tempo extrínseco, considerado de relevância nesse trabalho, é implementado pelo acoplamento de dois osciladores acoplados : um oscilador acentual que delimita grupos acentuais cuja estruturação é determinada parcialmente pela amplitude das batidas do oscilador, e um oscilador silábico que, induzido pelo acentual, gera uma seqüência de instantes de tempo em torno dos quais se situa o onset das vogais do enunciado. No quadro do modelo, uma metodologia para a investigação da tipologia rítmica das línguas é proposta. Algumas simulações do modelo apresentadas aqui permitem testar sua capacidade em prever a variabilidade dos padrões acentuais da fala natural.*

### **Abstract**

*The study of durational patterns of utterances in a language is viewed under the framework of a couple-oscillators dynamical model of speech rhythm production. This model has two main components: a two-coupled-oscillator model (a syllabic and a phrase stress oscillator) and an integrated gestural score specifying the coordination between abstract gestures. The phrase stress oscillator deals with the metrical component of the model by delimitating stress groups, the syllabic oscillator imposes a vowel flow continuum along the entire utterance, and the integrated gestural score is the domain where non-derivational lexical contrasts take place. In the model framework, two techniques allowing the systematisation of rhythm typology research under a dynamical perspective are presented. The first technique is the automatic detection of stress group boundaries from analysis of patterns of syllable-sized durations, and the second is the estimation of model parameters from inversion of the equations describing period coupling.*

### **Résumé**

*Fondée sur la théorie des oscillateurs couplés, la modélisation du rythme de la parole rend possible l'étude des patrons de durée des énoncés d'une langue à partir de la considération de deux niveaux temporels : un temps extrinsèque ou prosodique et un temps intrinsèque ou segmental. Le temps extrinsèque, qui joue un rôle prépondérant dans ce travail, s'organise à partir du couplage entre deux oscillateurs : un oscillateur accentuel qui délimite les groupes accentuels dont la structuration est déterminé partiellement par l'amplitude des frappes de l'oscillateur, et un oscillateur syllabique qui, entraîné par l'oscillateur accentuel, génère une suite d'instantanés autour desquels se place l'onset des voyelles de l'énoncé. Dans le cadre du modèle, une méthodologie d'investigation de la typologie rythmique des langues est proposée. Quelques simulations des patrons de durée avec le modèle sont présentées ici qui rendent possible de tester la capacité du modèle pour prédire la variabilité des patrons accentuels de la parole naturelle.*

## 1. Introdução

O modelo dinâmico de produção do ritmo da fala apresentado aqui está fundamentado na Teoria de Osciladores Acoplados, que é discutida de forma didática por Strogatz e Stewart (1994), que traçam um breve percurso histórico da pesquisa com osciladores acoplados, na busca de uma explicação para o fenômeno. Essa mesma teoria está inserida numa teoria mais geral, a Teoria dos Sistemas Dinâmicos. Para uma aproximação a essa teoria para dados comportamentais ligados aos movimentos humanos, sugerimos a leitura de Kelso (1995), que apresenta diversas investigações experimentais que podem ser acomodadas sob a forma de modelos de osciladores acoplados, incluindo dados de motricidade e de percepção. Saltzman e Munhall (1989) aplicam a noção de acoplamento em seu modelo funcional – *task dynamics* – para transduzir a informação abstrata das pautas gestuais de Browman e Goldstein (1986; 1989; 1992) em movimentos efetivos dos articuladores da fala.

O modelo proposto aqui gera a duração segmental para línguas que manipulam a duração como parâmetro de controle prosódico para assinalar o acento frasal (*phrase stress*) ao longo do enunciado. Tais acentos frasais delimitam os grupos acentuais (*stress groups*) de um enunciado. A presente versão do modelo considera que a língua em questão tem grupos acentuais com cabeça à direita. O modelo pode então ser aplicado da forma apresentada aqui para línguas como o espanhol, o francês, o italiano e o português brasileiro – doravante PB. A geração da duração segmental se realiza pela consideração de dois níveis de organização temporal, o tempo extrínseco e o tempo intrínseco.

Nesse primeiro estágio de implementação, o modelo lida com o que estamos chamando de tempo extrínseco ou prosódico pela ação conjunta e acoplada de dois osciladores. O caráter extrínseco de nosso modelo não diz respeito a um controle temporal que informa precisamente os instantes de tempo de ocorrência de eventos articulatorios subsilábicos durante a enunciação, tal como nos modelos criticados por Fowler (1980). O tempo extrínseco é compreendido aqui como inerentemente prosódico e, portanto, resultado de uma negociação entre um sistema lingüístico e um sistema de produção de fala, cada um dos mesmos especificando um conjunto de restrições que lhe é próprio. O modelo é composto de dois osciladores universais

(hipoteticamente presentes em todas as línguas): um oscilador silábico que trata da sucessão de unidades do tamanho da sílaba e portanto está relacionado com a noção de silabidade entendida como um fluxo contínuo de vogais interrompido pelos gestos consonantais (Öhman, 1966; Fujimura, 1995; Browman e Goldstein, 1990), e um oscilador acentual, que trata da sucessão dos acentos frasais, isto é, das proeminências quase-periódicas ao longo de um enunciado e delimitadoras dos grupos acentuais.

O modelo também integra um tempo intrínseco, este sim compreendido como nos trabalhos de Fowler (*ibidem*), de Browman e Goldstein (1986; 1989) e Albano (2001). O tempo intrínseco diz respeito à especificação de durações isomorfas a gestos segmentais de um léxico gestual. Segundo a proposta de Albano (*ibidem*), é nesse léxico que está especificado o contraste entre segmentos plenos e reduzidos, contraste este que assinala a posição do acento lexical (correspondendo à posição dos últimos segmentos plenos do item lexical, no caso do PB. Essa solução também serviria para o espanhol e o italiano. Como o francês não tem acento lexical, sílabas iguais seriam especificadas da mesma forma, não importa a posição que ocupem na palavra).

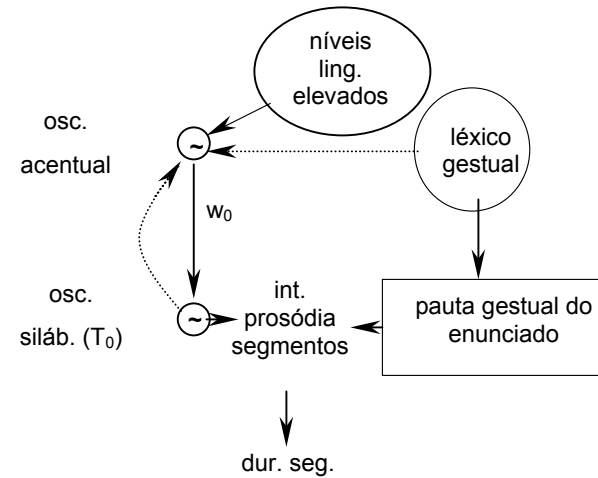
A complexidade dos padrões temporais produzidos pelo modelo na superfície é portanto o resultado de duas ordens de especificação temporal: (1) a de um modelo de dois osciladores que se acoplam funcionalmente para a implementação da estruturação rítmica do enunciado; (2) a de uma especificação de duração intrínseca, que perturba (modifica localmente) a organização temporal anterior a fim de produzir a organização temporal final. Note que, nessa visão, o controle macroscópico da duração (possibilidade de modificação substancial da duração segundo informação lingüística superior ao segmento quase isomórfico ao fonema) está nas mãos do tempo extrínseco. Essa mesma ordem temporal está relacionada à constituição e à execução de um mecanismo de acentuação (o sufixo indica um caráter dinâmico) ao longo dos grupos acentuais, culminando – para línguas com cabeça à direita no domínio do grupo acentual –, com o acento frasal.

Em seguida o modelo será descrito detalhadamente, situando-se o mesmo em relação a outros modelos que se propõem a dar conta do ritmo lingüístico.

## **2. Descrição do modelo de produção da organização temporal da fala**

A figura que segue (figura 1) apresenta o modelo dinâmico de produção do ritmo da fala em desenvolvimento. A implementação do mesmo foi fundamentada em um *corpus* de frases isoladas lidas em laboratório, de forma neutra, por um falante masculino de cerca de 35 anos, da região do Recife (Pernambuco), e que morava há dez anos em Campinas. O modelo de osciladores acoplados pode ser visto à esquerda da figura e é sobre ele que falaremos mais amiúde neste artigo. O oscilador acentual é implementado por um trem de pulsos quase-periódico cujas amplitudes e desvios da periodicidade são determinados por níveis lingüísticos mais elevados. O oscilador silábico é, por sua vez, implementado por uma cossenóide levantada de amplitude arbitrária igual a 1, cujo período na condição de não-acoplamento especifica abstratamente a taxa de elocução do enunciado (o advérbio

'abstratamente' é usado aqui com dois sentidos. Em primeiro lugar, porque o período do oscilador silábico, quando da geração das durações de unidades do tamanho da sílaba, se modifica sistematicamente sob controle do oscilador acentual e não retorna, em princípio jamais, ao período da situação de não-acoplamento. Em segundo lugar, porque as durações segmentais finais geradas para um enunciado só são obtidas após a interação com o nível segmental, a partir da especificação de duração intrínseca dos gestos do léxico). Depreende-se da figura que o oscilador acentual recebe como entrada informação de níveis lingüísticos mais elevados e do



léxico gestual. Além disso, pela sua característica oscilatória, ele impõe restrições eurítmicas à ocorrência de seus pulsos, o que lhe confere uma quase-periodicidade.

**Figura 1** — Modelo dinâmico de produção do ritmo da fala

Embora os níveis lingüísticos mais elevados digam respeito à especificação, ao longo do enunciado, de informação sintática e semântica, somente a posição de palavras funcionais na frase está sendo levada em conta na versão atual do modelo (essas posições determinam onde o acento frasal não pode incidir, para uma situação de leitura neutra), pois uma descrição sintático-semântica da frase exigiria um *parser*, de que não dispomos no momento. A posição dos acentos lexicais, a partir de sua especificação no léxico gestual, é levada em conta e desvia os pulsos do oscilador acentual, num primeiro momento considerados periódicos, para coincidirem com o *onset* da vogal lexicalmente acentuada mais próxima. A sucessão de pulsos do oscilador acentual especifica assim os locais de acento frasal, que gatilham por sua vez um mecanismo de acentuação, considerado universal. A amplitude dos pulsos do oscilador acentual varia de acordo com a força da ligação sintática entre os constituintes adjacentes, anterior e posterior ao pulso, bem como com o grau de ênfase, ligado a uma especificação semântica, conferido ao enunciado naquele ponto.

O oscilador silábico, por sua vez, tem os máximos de seus pulsos coincidentes com os *onsets* das vogais do enunciado. Após o acoplamento por sua interação com o

oscilador acentual, o oscilador silábico determina extrinsecamente para a pauta gestual a posição dos *onsets* vocálicos ao longo do enunciado. A unidade fônica que dois *onsets* vocálicos consecutivos delimitam é referida aqui por unidade V-V. Esta unidade está relacionada à importância da transição CV para a produção e percepção da fala, conforme discorrido extensamente em Barbosa (2000a) a partir da análise de dados do PB iluminados pelos trabalhos de Allen (1972), Dogil e Braun (1988) e pela literatura sobre o *perceptual-center* (Morton *et al.*, 1976; Pompino-Marschall, 1989; 1991; Scott, 1993 e Janker, 1995, entre outros). À luz desses trabalhos (mas não somente, veja o caráter precursor da obra de Classé, 1939), diversos autores atentaram para a importância da delimitação do intervalo entre dois *p-centers* consecutivos como fundamentais para a pesquisa sobre o ritmo da fala, assegurando a adoção de unidades fônicas isomórficas aos efeitos acentuais. Entre esses autores destacam-se Roach (1982), Barbosa e Bailly (1994), Cummins e Port (1998), Port *et al.* (1995), Barbosa (1996; 2000a; 2000b) e Cummins (2002). A sucessão de transições CV funciona como seqüência de pontos de ancoramento para a produção e percepção da fala, implementando um fluxo vocálico contínuo (perturbado localmente por consoantes) diretamente correlato à oscilação mandibular. A hipótese de que a enunciação se dá por um fluxo vocálico interrompido por perturbações consonantais é bastante antiga, pois remonta a observações gerais como as de Charles de Brosses em *Traité de la formation mécanique des langues*, de 1765 (apud Eco, 1995: 92), e foi avançada a partir de abordagem experimental quando das primeiras observações espectrográficas (Potter *et al.*, 1947), que culminaram com o estudo de Öhman (1966). Essa hipótese é defendida vivamente por pesquisadores como Fujimura (1995; 2002), que implementou a *base function* de seu modelo C/D e a associação local de gestos consonantais a partir dessa hipótese de trabalho. O oscilador silábico implementa assim a silabidade entendida nos termos aqui explicitados e, portanto, também uma característica universal.

Os efeitos do acoplamento entre os dois osciladores para o oscilador silábico se manifestam a partir de duas vias. A primeira a partir da influência das amplitudes dos pulsos do oscilador acentual sobre os períodos do oscilador silábico, ao longo de todo o grupo acentual que casa pulso finaliza (posição do acento frasal). A segunda a partir da força de acoplamento relativa ( $w_0$ ), que especifica o grau de influência do oscilador acentual sobre o silábico. O acoplamento de período, que traduz concretamente essa influência, será abordado detalhadamente na próxima seção.

O modelo de ritmo apresentado aqui não ignora o chamado ritmo lingüístico, mas o entende de forma mais ampla do que aquela, baseada num alinhamento de unidades discretas em vários níveis, com uma sucessão de itens lexicais, apresentada pela Fonologia Métrica (Lieberman e Prince, 1977).

O primeiro sentido do lingüístico no modelo dinâmico de produção do ritmo da fala está relacionado à maneira pela qual os pulsos do oscilador acentual se alinham com os *onsets* de vogais acentuadas lexicalmente, isto é, com material lingüístico. Essa atribuição de ritmo lingüístico ao tipo de estruturação rítmica concebida pelo modelo com que trabalhamos tem o mesmo valor que a proposta de Lieberman e Prince (*ibidem*). Porém, em nosso caso, as amplitudes do oscilador acentual podem assumir qualquer valor real positivo, diferentemente da proposta discreta da fonologia métrica.

O segundo sentido do lingüístico neste modelo refere-se à força de acoplamento relativa entre os dois osciladores que, embora possa variar de indivíduo para indivíduo, teria uma variabilidade intra- e inter-individual bem menor que aquela entre línguas distintas. Essa noção de variabilidade é estatística, isto é, os valores da força de acoplamento num mesmo indivíduo e entre indivíduos de uma mesma variedade lingüística proviriam de amostras de uma mesma população estatística, representativa dos enunciados de uma língua. Essa variabilidade é, em princípio, bem menor que aquela provinda de duas línguas (populações) distintas.

O terceiro sentido do lingüístico refere-se ao léxico gestual em que são especificadas as coordenações entre gestos abstratos da língua, bem como são contrastados gestos plenos e reduzidos (em línguas que o fazem), conforme a proposta de Albano (*ibidem*).

Um quarto sentido do lingüístico refere-se ainda à maneira como se dá a interação prosódia-segmentos. A hipótese subjacente aqui é a de que essa interação se dá de forma distinta de língua para língua. Essa interação foi implementada com dados do locutor já citado, pelo uso de uma rede neural recorrente (Barbosa, 2001) que demonstrou a capacidade do modelo em gerar automaticamente as durações acústicas segmentais para um sistema de síntese da fala (Barbosa *et al.*, 1999).

Neste modelo, os acentos lexicais frasalmente proeminentes são assinalados como tais pelas posições dos pulsos de um oscilador acentual. Cada pulso desencadeia um mecanismo de acentuação no grupo acentual que ele finaliza, através de um mecanismo de desaceleração progressiva (indução) parametrizado pelos parâmetros força de acoplamento relativa (embutida numa função de sincronismo), taxa de indução e amplitude do pulso. Esse mecanismo de desaceleração progressiva convive com um mecanismo de aceleração local durante as duas primeiras unidades V-V do grupo acentual, mecanismo que atrai a duração da unidade V-V para um retorno à situação de não-acoplamento. Esses aspectos são explicitados na próxima seção.

### 3. Implementação do modelo

Para a implementação do modelo de acoplamento entre os dois osciladores, inspiramos-nos no modelo de osciladores adaptativos para a percepção de intervalos de tempo de McAuley (1995), embora tenhamos introduzido várias modificações quanto à maneira como se dá o acoplamento e a relação com variáveis lingüísticas. O primeiro componente comum entre nosso modelo e o de McAuley, é a função de sincronismo entre os osciladores,  $s(.)$  (a função de saída – *output function* – postulada por McAuley). Em nosso modelo, ao contrário da decisão do autor citado, a forma da função foi determinada empiricamente a partir de um *corpus* de frases lidas em três taxas de elocução pelo mesmo locutor citado acima (num total de 108 frases para as três taxas). A função é dada pela fórmula (1) e foi determinada por regressão não-linear (via regressão linear dos valores transformados por funções específicas da variável em questão) das durações V-V ao longo das frases do *corpus*. A evolução do aumento relativo da duração da unidade V-V ( $\Delta VV_i/VV_i$ ) ao longo de cada frase foi computada

e transformada por meio de algumas funções não-lineares, sendo a melhor regressão aquela obtida com a função exponencial ( $R^2 = 0,999$ ).

$$s(0) = w_0 \cdot \exp(-N + 2), \text{ e } s(N-1) = \exp(-5,81 + 0,016 \cdot T_0) \quad (1a)$$

$$s(n) = (1 - w_0) \cdot s(n-1) + w_0 \cdot \exp(-N + n + 2), \text{ para } 0 < n < N-1 \quad (1b)$$

Na fórmula 1,  $N$  é o número de unidades V-V no grupo acentual, e  $n$  é o índice da unidade V-V atual (começando de 0 para a primeira unidade V-V do grupo acentual). A força de acoplamento relativa entre os osciladores,  $w_0$ , é um parâmetro real positivo, específico a uma língua (num sentido estatístico, como explicitado na seção anterior) e especificando o grau de acoplamento entre os osciladores acentual e silábico. Essa força está relacionada, como já discutimos em trabalho anterior (Barbosa, 2002a), com as noções de ritmo silábico (*syllable timing*) e acentual (*stress timing*): quanto maior o valor da força de acoplamento, maior a influência do oscilador acentual sobre o silábico, simulando um ritmo acentual. O inverso simularia uma tendência a um ritmo silábico.  $T_0$  é o período do oscilador silábico na condição de não-acoplamento.

O *reset* de fase no modelo de osciladores acoplados é completamente diferente daquele proposto por McAuley (em seu trabalho, esse *reset* determina a direção e o momento de iniciar um acoplamento de período). Em nosso modelo esse *reset* de fase ocorre a cada pico positivo do oscilador silábico, pico que está alinhado com o *onset* de cada vogal. O acoplamento de período no oscilador silábico é realizado via indução progressiva a partir do cálculo do aumento de duração abstrata da unidade V-V ( $\Delta T$ ) na fórmula (2).

$$\Delta T = \alpha \cdot T \cdot s(n) \cdot i(m) - \beta \cdot (T - T_0) \cdot i(m-1) \quad (2)$$

Os parâmetros  $\alpha$ ,  $\beta$ , e  $T_0$  são respectivamente a taxa de indução, a taxa de decaimento e o período do oscilador silábico na condição de não-acoplamento. As funções  $s(\cdot)$  e  $i(\cdot)$  são respectivamente a função de sincronismo (dada pela fórmula 1) e a amplitude do pulso do oscilador acentual que domina o grupo acentual atual (no caso do PB, é o pulso que está por vir).  $T$  é o período atual do oscilador silábico (a duração abstrata da unidade V-V). O *reset* de período é realizado via decaimento, pelo segundo termo da equação acima,  $-\beta \cdot (T - T_0) \cdot i(m-1)$ , que só está presente durante os dois primeiros ciclos do oscilador silábico dominados por um pulso de oscilador acentual por vir. A taxa de decaimento é especificada por  $\beta$ . Esse termo da equação não está presente para os ciclos do oscilador silábico (correspondendo a unidades V-V pós-tônicas) após o último pulso do oscilador acentual (coincidente com a última sílaba tônica do enunciado), pois não há um pulso do oscilador acentual ainda por vir. Essa característica do modelo explica o alongamento final ao final do enunciado, incluindo aquele dos segmentos pós-tônicos. O período atual  $T$  é atualizado a cada ciclo do oscilador silábico, iterativamente, segundo (3).

$$T(n) = \text{duração (abstrata) da unidade V-V} = T(n-1) + \Delta T \quad (3)$$

Os valores de  $w_0$ ,  $\alpha$  e  $\beta$  foram otimizados de forma a minimizar o erro entre os valores do período  $T$  do oscilador silábico induzido e a média das durações V-V para os grupos acentuais do *corpus* gravado com o locutor citado, nas três taxas de elocução. Esse procedimento gerou os seguintes valores:  $w_0 = 0,78$ , e  $(\alpha, \beta) = (0,31 ; 1,13)$  para uma taxa de elocução rápida ( $T_0$  até 128 ms),  $(\alpha, \beta) = (0,38 ; 1,13)$  para

uma taxa normal ( $T_0 = 129$  ms a  $176$  ms) e  $(\alpha, \beta) = (0,43; 1,04)$  para a taxa lenta ( $T_0$  acima de  $177$  ms). O algoritmo de otimização foi implementado de tal forma a possibilitar uma correção de valores de  $w_0$  com ordem de grandeza dez vezes menor que a ordem de grandeza da correção de valores das taxas  $\alpha$  e  $\beta$ . A convergência foi considerada atingida quando a diferença entre duração abstrata predita pelo modelo e a duração média de unidades V-V no corpus (para grupos acentuais de mesmo tamanho), em qualquer posição no grupo acentual foi menor que  $25$  ms.

Uma vez completada a fase de otimização, os parâmetros acima foram perturbados (modificados) a fim de reproduzir dados comportamentais de outros sujeitos, em diversas situações que não aquela da leitura de frases isoladas. A finalidade dessa etapa é a de não somente testar a versatilidade do modelo em reproduzir a variabilidade da fala natural como também a de introduzir uma técnica de estimação de (pelo menos) parte desses parâmetros sem a necessidade de *corpora* ad hoc da extensão daquele usado na fase de otimização. Para tanto é necessário apresentar critérios de produção para a delimitação de grupos acentuais.

#### **4. Aspectos metodológicos para a pesquisa tipológica: delimitação de grupos acentuais nos enunciados de uma língua natural**

Para determinar as fronteiras dos grupos acentuais a partir de critérios de produção de fala<sup>1</sup>, fundamentados, no caso do PB, em dados de duração acústica, é necessário detectar os picos de evolução da duração das unidades V-V, tendo em vista que esse parâmetro acústico é primordial para a implementação do acento lexical (Massini, 1991; Barbosa, 1996; Fernandes, 1976) e frasal (Barbosa, 1996) em PB (é o caso também do francês e do português europeu, e possivelmente, aquele do italiano e espanhol europeu e de suas variedades americanas). Para essa determinação, utilizamos um procedimento semi-automático que tem demonstrado alto grau de reprodutibilidade. O procedimento pode ser executado para um *corpus* de fala qualquer a partir dos quatro passos descritos em seguida.

- Delimite as unidades V-V a partir da marcação dos *onsets* de vogal para todo o *corpus*, calculando as durações das mesmas a partir das fronteiras. Essa marcação pode ser feita semi-automaticamente usando um programa para detectar *onsets* de vogal, como o *Beat Extractor* (vide anexo 3) implementado por Cummins e Port (1998) e adaptado recentemente para o *software* Praat por

---

<sup>1</sup> É preciso salientar que os picos de duração de unidades V-V ao longo do enunciado não correspondem necessariamente a proeminências acentuais percebidas (embora a coincidência seja altamente freqüente em PB), tendo em vista que a percepção de proeminências depende do concurso de outros parâmetros acústicos, notadamente a freqüência fundamental (visto que em PB a intensidade e a duração costumam co-variarem). Do ponto de vista de um modelo de produção do ritmo da fala voltado para a implementação da duração, são os picos de duração que precisam ser gerados automaticamente da maneira como são observados nos enunciados naturais. Sendo assim, a questão da percepção da fala é ortogonal aos interesses da predição dos padrões de duração aqui apresentados.

mim (essa versão está disponível livremente na forma de programa aberto, via solicitação ao autor);

- Calcule durações normalizadas para cada unidade V-V ( $z_{V-V}$ ) por uma técnica estendida a partir do z-score clássico (cf. Campbell e Isard, 1991; Barbosa, 1994), usando a fórmula (4):

$$z_{V-V} = (dur_{V-V} - \Sigma\mu_i) / \sqrt{\Sigma\sigma_i^2} \quad (4)$$

Para a qual,  $dur_{V-V}$  é a duração da unidade V-V,  $\Sigma\mu_i$  é a soma das médias das durações dos segmentos da unidade V-V tomadas de um outro *corpus* (o índice  $i$  se refere a cada segmento que compõe a unidade V-V). Essas médias foram computadas a partir de um *corpus* ad hoc de referência contendo todas as seqüências de segmentos que a fonotaxe da língua admite, incluindo seqüências de segmentos entre palavras. Para esse fim, temos usado o *corpus* gravado para o obtenção do inventário de unidades de um sistema de síntese concatenativo, como descrito em Barbosa *et al.*, 1999). A expressão  $\sqrt{\Sigma\sigma_i^2}$  é a raiz quadrada da soma das variâncias das durações dos segmentos da unidade V-V, computados a partir do mesmo *corpus* ad hoc.

- Aplique aos valores de z-score das unidades V-V calculados anteriormente uma técnica de suavização em cinco pontos a fim de minimizar as oscilações locais devidas à realização do acento lexical e ao efeito remanescente de durações intrínsecas dos segmentos. Esse procedimento faz ressaltar o mecanismo de realização do acento frasal, que marca a culminância do grupo acentual. A fórmula (5) foi usada para essa finalidade, em que  $z_{suav\ i}$  é o z-score suavizado da unidade V-V na posição  $i$  e  $z_i$ , o z-score da unidade V-V da posição  $i$ .

$$z_{suav\ i} = (5.z_i + 3.z_{i-1} + 3.z_{i+1} + 1.z_{i-2} + 1.z_{i+2}) / 13 \quad (5)$$

- Detecte os picos de máximo dos z-scores suavizados via cálculo da derivada discreta. Tais picos correspondem às fronteiras dos grupos acentuais. O acento frasal, construído ao longo do grupo acentual culmina nesses pontos.

Para realização desses quatro passos um programa em MatLab foi implementado (o programa também está disponível via solicitação ao autor).

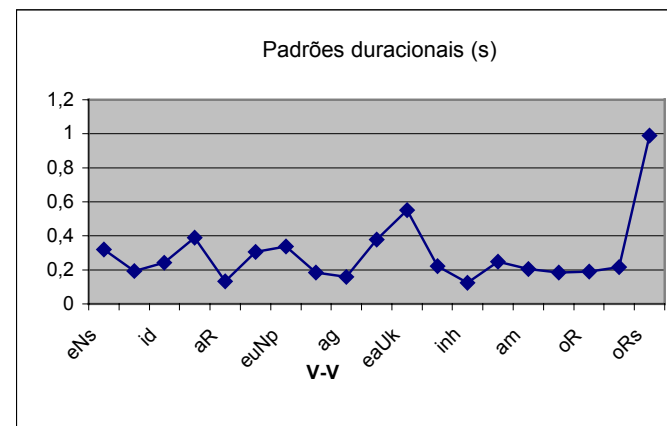
Os passos acima são em seguida ilustrados com dados de um locutor do dialeto paulista, de sexo masculino e com cerca de 30 anos no momento da gravação, que leu um trecho (dois parágrafos) da obra de Monteiro Lobato (1920), de cerca de 200 unidades V-V (ou sílabas)<sup>2</sup>. O texto se encontra no anexo 1. Ele tem a mesma ordem de freqüência de ocorrência dos fonemas do PB (coeficiente de *Spearman* > 0,95;  $p = 0$ ), cuja ordem de referência foi computada a partir da estatística descritiva de bases de dados do PB de grande porte (Albano *et al.*, 1995). Esse trecho foi lido

---

<sup>2</sup> O texto tem aproximadamente o mesmo tamanho da tradução do texto "The North Wind and the Sun" ("O Vento Norte e o Sol"). O texto e suas traduções são usados para ilustrar o fonetismo das línguas pela Associação Fonética Internacional, tanto no periódico *Journal of the International Phonetic Association* quanto no manual da associação, lançado em 1999 pela Cambridge University Press (*Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet*).

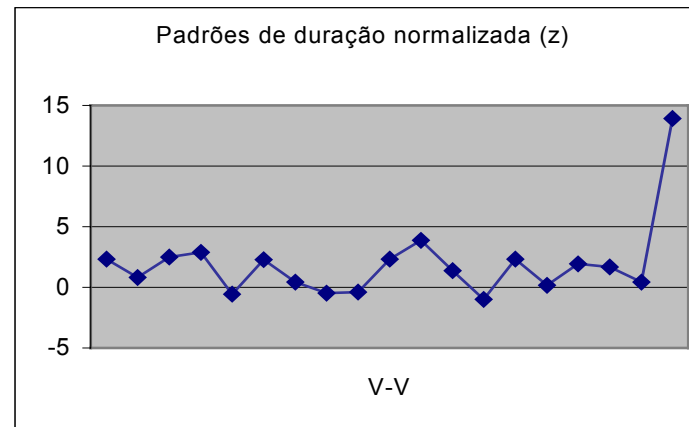
em três taxas de elocução pelo locutor paulista. As taxas foram etiquetadas *grosso modo* de “normal”, “lenta” e “rápida”. A taxa normal foi aquela mais confortável para o locutor (correspondendo a sua taxa natural de enunciação). As demais foram eliciadas a partir das seguintes instruções: “fale o mais rápido que possa sem introduzir distorções” (taxa rápida) e “fale o mais lentamente que possa conservando a estrutura lingüística – a mensagem transmitida – do trecho sendo lido (taxa lenta). Apenas os dados correspondentes à taxa normal para esse locutor são mostradas nesta ilustração (o tempo total de leitura foi de 47,76 s).

A figura seguinte (figura 2) mostra a evolução da duração bruta das unidades V-V ao longo do trecho inicial da leitura para a taxa normal. O trecho se refere à primeira frase do texto: “Em seguida apareceu um papagaio real que tinha fama de orador.” Devido ao contributo da duração intrínseca dos segmentos que compõem cada unidade V-V, os picos de duração expressos desta maneira assinalam de maneira inseparável idiosincrasias segmentais e efeitos acentuais, tanto lexicais quanto frasais. Na figura é possível constatar picos para as seqüências “eNs” (**Em seguida**), “ap” (seguida **apareceu**), “euNp” (**apareceu um papagaio**), “eaUk” (**real que**), “Af” (**tinha fama**), “oRs” (**orador. Subiu [...]**). (A notação aqui usada, apresentada por Albano e Moreira, 1996, resgata a noção de arqui fonema e a integra àquela de redutibilidade dos segmentos fônicos: as letras maiúsculas representam segmentos reduzidos e as minúsculas, segmentos plenos. Assim, as quatro consoantes de coda do português – analisadas como arqui fonemas – são notadas /S, R, N, L/, bem como as consoantes em *onset* complexo /R, L/ – em palavras como “prato” e “Platão”. Por esse critério, as vogais pós-tônicas são notadas com letras maiúsculas. Esse sistema notacional permite diferenciar do ponto de vista da duração, fones com durações médias oriundas de populações estatísticas distintas – vide anexo 2. Não há correspondência bijetiva – um a um – entre esse sistema notacional e aquele do A. F. I.)



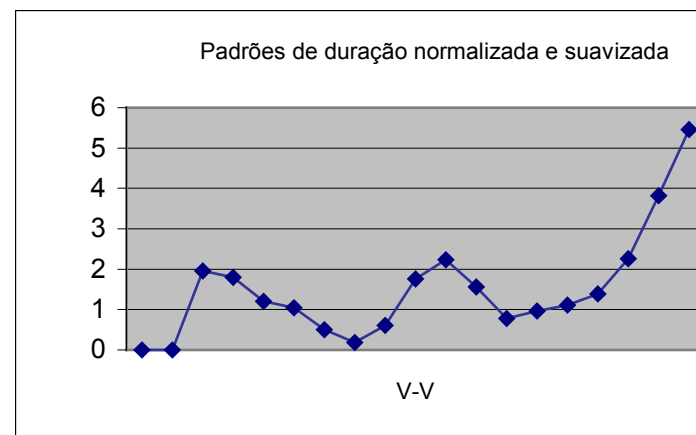
**Figura 2** – Padrões duracionais brutos das unidades V-V do trecho “Em seguida apareceu um papagaio real que tinha fama de orador.”, lido pelo locutor AP na taxa de elocução normal.

As durações V-V da figura 2 foram obtidas pela aplicação do primeiro passo descrito acima. Para minimizar o efeito da duração intrínseca, a normalização por z-score é aplicada (segundo passo acima), produzindo os padrões da figura seguinte (figura 3), para o mesmo trecho acima.



**Figura 3** – Padrões de duração normalizada (z-score) das unidades V-V do trecho “Em seguida apareceu um papagaio real que tinha fama de orador.”, lido pelo locutor AP na taxa de elocução normal.

A aplicação da técnica de normalização eliminou o efeito da duração intrínseca da unidade V-V “euNp”, composto de três segmentos, mas acrescentou outros picos em “es” (apareceu) e “Ad” (fama de). A técnica de suavização em cinco pontos, descrita no terceiro passo, será fundamental para eliminar esses efeitos duracionais locais remanescentes. O resultado é apresentado na figura que segue (figura 4).



**Figura 4** – Padrões de duração normalizada e suavizada das unidades V-V do trecho “Em seguida apareceu um papagaio real que tinha fama de orador.”

Os picos de duração normalizada e suavizada (figura 4) revelam apenas três trechos proeminentes: “id” (Em seguida), “eaUk” (real que) e “oRs” (orador. Subiu), que correspondem a fronteiras sintáticas e prosódicas fortes (respectivamente depois de introdutor, antes de oração subordinada e final de frase). Todos os picos também coincidem com V-V lexicalmente acentuados. (É necessário observar apenas que os dois primeiros V-V foram arbitrariamente zerados, em virtude da ausência de contextos anteriores para aplicação da técnica de suavização.)

A automatização do quarto passo acima permite a detecção de todas as fronteiras dos grupos acentuais para qualquer *corpus* do PB. Tendo em vista que o locutor que serviu de referência para cálculo das médias e desvios-padrão dos segmentos do PB é do dialeto pernambucano, pode ser desejável obter um *corpus* de referência para um locutor do dialeto em estudo. A gravação de um locutor masculino e um feminino também é desejável. Ressalte-se, no entanto, que o objetivo da normalização é minimizar os efeitos sistêmicos da relação entre durações intrínsecas dos segmentos do PB. Em relação a esse aspecto o locutor que usamos tem sido eficaz.

Uma vez tendo detectados os grupos acentuais, é possível calcular a duração bruta dos mesmos ( $dur_{GA}$ ) e contar o número de unidades V-V em cada um ( $n_{V-V}/GA$ ). A regressão linear entre as variáveis  $dur_{GA}$  e  $n_{V-V}/GA$  permitiu uma primeira caracterização da força de acoplamento relativa entre os dois osciladores, pelo cômputo da razão entre o ponto de interseção da reta de regressão com o eixo das ordenadas e a inclinação da mesma. Esse procedimento, proposto inicialmente por O'Dell e Nieminen (1999) foi aplicado criticamente por Barbosa (2000a) para o PB, revelando a necessidade de garantir a comparabilidade interlingüística da estruturação rítmica pelo controle, em primeiro lugar, da taxa de elocução. Essa técnica parte da hipótese de equilíbrio dos efeitos entre os osciladores e não considera o que ocorre ao longo de cada grupo acentual (a não ser como parte de um efeito geral em todo o *corpus*).

Uma técnica mais precisa para a estimação da força de acoplamento relativa, que considera a evolução ponto a ponto da duração das unidades V-V ao longo dos grupos acentuais será apresentada em seguida.

## **5. Aspectos metodológicos para a pesquisa tipológica: estimativa da força de acoplamento relativa**

A força de acoplamento relativa ( $w_0$ ) pode ser estimada ( $\hat{w}_0$ ) a partir de algumas considerações ligadas ao próprio modelo de produção do ritmo da fala apresentado anteriormente.

Partindo de uma duração normalizada, obtida a partir da duração bruta de dados comportamentais, mas que pudesse se abstrair de efeitos de acento lexical e duração intrínseca, é possível obter uma aproximação dos padrões de duração V-V gerados pelo modelo. Ora, uma duração satisfazendo essa descrição foi apresentada anteriormente pelo cálculo da duração normalizada e suavizada. Só restaria empregar uma transformação inversa, considerando um V-V típico, para se obter uma duração abstrata expressa em milissegundos. Para tanto, reconstruímos a

partir das durações do *corpus* de referência (anexo 2) a média e desvio-padrão do V-V “aC” (em que “a” é a vogal mais baixa e mais freqüente em PB e C é uma consoante oclusiva surda). Obteve-se o par  $(\mu_r, \sigma_r) = (212 \text{ ms}; 56 \text{ ms})$ , em que a média  $\mu_r$  foi obtida pela soma da média da duração de [a] com a média das médias de duração das três oclusivas surdas do PB e o desvio-padrão foi obtido pela raiz quadrada da soma das variâncias de duração dos quatro segmentos referidos. Aplicando-se a fórmula (6) para cada duração normalizada/suavizada do *corpus* ( $Z_{\text{suav.V-V}}$ ), obtêm-se as durações abstratas desejadas.

$$\text{dur}_{\text{V-Vabst.}} = Z_{\text{suav.V-V}} \cdot \sigma_r + \mu_r \quad (6)$$

Caso essa transformação deva ser trabalhada com corpora em taxas de elocução distintas, recomenda-se variar a duração média em função de cada taxa, mas tendo o cuidado de atribuir um valor de desvio-padrão que conserve a razão 212/56 entre média e desvio-padrão, para garantir a independência do aumento relativo de duração V-V ( $\Delta T/T$ ) (fórmula 8) em relação aos valores de referência adotados.

Considerando as durações abstratas obtidas por (6) como aquelas que seriam geradas pelo modelo de ritmo com parâmetros que simulassem o locutor do *corpus* em estudo, utiliza-se um procedimento de inversão fundamentado nas seguintes considerações. Se para cada grupo acentual considerarmos os valores de duração abstrata entre a terceira e penúltima posições, a fórmula (2) que caracteriza o acoplamento de período se simplificaria como em (7). A decisão de não tomar a última posição garante que a fórmula correspondente à função  $s(\cdot)$  seja apenas a (1b), expressa apenas sob a forma exponencial, o que permitirá o uso da transformação abaixo (fórmula 7).

$$\Delta T = \alpha \cdot T \cdot s(n) \cdot i(m) \quad (7)$$

Se considerarmos, numa primeira aproximação (como ponto de partida necessário enquanto não se dispuser de um analisador gramatical adequado para o PB), que os valores de amplitude dos pulsos do oscilador acentual são constantes e iguais a 1, e introduzirmos o índice  $n$  da iteração no cálculo do aumento relativo de duração abstrata ( $\Delta T/T$ ) para as posições 3 a  $N-1$  (em que  $N$  é o número de unidades V-V no grupo acentual) obtêm-se a fórmula (8) a partir de (7).

$$(\Delta T/T)(n) = \alpha s(n) \quad (8)$$

A expressão acima (fórmula 8) atesta que os valores podem ser estimados ( $\alpha \cdot \hat{s}(n)$ ) pelo cálculo das médias, para cada posição ao longo de um grupo acentual de tamanho  $N$ , dos aumentos relativos de duração abstrata. De posse dessas estimativas, é possível obter a estimativa da força de acoplamento relativa,  $\hat{w}_0$ , a partir da inversão da fórmula (1b), calculando-se os  $N-4$  valores de  $\hat{w}_0$  para cada grupo acentual de tamanho  $N > 4$  (fórmula 9). O valor estimado final para a força de acoplamento relativa a partir do *corpus* é computado pela média de todas as estimativas parciais de cada grupo acentual do *corpus* (fórmula 10).

$$(\hat{w}_0)_j = (\hat{s}_n - \hat{s}_{n-1}) / (\alpha \cdot \exp(-N+1+n) - \hat{s}_{n-1}), \text{ com } j = N - 1 \text{ a } 3, \text{ regressivamente} \quad (9)$$

$$\hat{w}_0 = \sum_j (\hat{w}_0)_j / M \quad M = \text{número de estimativas parciais} \quad (10)$$

O procedimento será exemplificado abaixo com dois locutores que leram o *corpus* do anexo 1 em três taxas de elocução, eliciadas pelo mesmo procedimento descrito acima. O valor da taxa de indução  $\alpha$  para as taxas de elocução normais de cada um deles é arbitrado em 1. Para as taxas lenta e rápida o valor de  $\alpha$  é estimado a partir da razão entre os aumentos relativos de duração abstrata para cada taxa em questão e a taxa normal ( $(\Delta T/T_{\text{taxa } i})/(\Delta T/T_{\text{taxa normal}})$ ), hipotetizando que o valor das amplitudes dos pulsos do oscilador acentual bem como a força de acoplamento sejam as mesmas entre a taxa em questão e a taxa normal, para um grupo de mesmo tamanho idêntico nas duas taxas e formado pelas mesmas unidades V-V. O valor estimado final para  $\alpha$  é calculado pela média das razões dos aumentos relativos de duração entre os *corpora*. (Note que a duração abstrata não é usada no cômputo dessa razão e sim a duração bruta de cada unidade V-V nos *corpora*.)

Antes de apresentar os resultados, serão elencados os cinco passos para a obtenção do valor estimado da força de acoplamento relativa.

- Restaure um valor de duração abstrata para a unidade V-V a partir da duração normalizada e suavizada  $z_{\text{suav.V-V}}$  aplicando-se a fórmula (6) com o par de valores (212 ms, 56 ms) para  $(\mu_r, \sigma_r)$ , como valores de referência de um V-V canônico. Outro valor de  $\mu_r$  também é possível, desde que se adote  $\sigma_r = \mu_r \cdot 56/212$ ;
- Calcule os N-3 valores de aumento relativo de duração abstrata  $(\Delta T/T)(n)$  para todo grupo acentual de tamanho  $N > 4$ ;
- Estime os valores  $\alpha.s(n)$  (para  $n = 3$  a  $N-1$ ) para todos os grupos acentuais de tamanho  $N$ , com  $N > 4$ , calculando-se a média de  $\Delta T/T$  para cada posição  $n$ . Por exemplo, para dez grupos acentuais de tamanho 6, calculam-se os dez valores de  $\Delta T/T$  nas posições 3 a 5. A média dos dez valores de  $\Delta T/T$  na posição 3 estimam  $\alpha.s(3)$  do grupo de tamanho 6, a média dos dez valores de  $\Delta T/T$  na posição 4, estimam  $\alpha.s(4)$ , e assim por diante;
- Escolha (ou estime, conforme explicado anteriormente) o valor de  $\alpha$ ;
- Calcule as N-4 estimativas parciais de  $w_0$  (fórmula 9) para todo grupo de tamanho  $N$  ( $N > 4$ ) e calcule a média para o *corpus*, considerando as estimativas parciais (fórmula 10).

Para realização desses cinco passos, dois programas em MatLab foram implementados, um para estimar a força de acoplamento relativa,  $\hat{w}_0$ , dado o valor da taxa de indução  $\alpha$ , e o outro para estimar a própria taxa de indução  $\alpha$  tendo-se um *corpus* de taxa normal como referência e um *corpus* gravado em outra taxa, para o mesmo locutor, lendo o mesmo trecho.

A técnica explicitada acima foi aplicada para dois locutores, AP e AC, que leram o *corpus* do anexo 1 em três taxas de elocução<sup>3</sup>. Depois de estimar a força de

---

<sup>3</sup> Há uma diferença significativa entre as taxas de elocução do locutor AC ( $p < 0,03$ ) e do locutor AP ( $p < 10^{-4}$ ). O teste de Scheffé mostra uma diferença significativa apenas entre as taxas lenta e rápida para AC, e entre rápida e lenta, bem como entre normal e lenta para AP.

acoplamento relativa para a taxa normal, adotando-se  $\alpha = 1$ , estimou-se um valor de  $\alpha$  para as taxas lenta e rápida para cada locutor. Procedendo dessa forma, os valores que seguem (tabela 1) foram obtidos para os dois locutores.

**Tabela 1** – Parâmetros taxa de indução e força de acoplamento estimados para dois locutores em três taxas de elocução (valores numéricos das taxas fornecidos).

	AC			AP		
	$\alpha$	$\hat{w}_0$	T.E. (V-V/s)	$\alpha$	$\hat{w}_0$	T.E. (V-V/s)
<b>normal</b>	1,00	1,12	4,6	1,00	0,70	4,3
<b>lenta</b>	1,57	0,61	4,1	1,06	0,46	3,4
<b>rápida</b>	1,18	0,92	5,1	1,23	0,58	4,9

A média das forças de acoplamento estimadas para o locutor AC, considerando as três taxas, é de 0,88 e a do locutor AP, de 0,58. Os resultados para os dois locutores indicam que o locutor AP tende mais ao ritmo silábico (menos diferenciação entre as durações das unidades V-V dos enunciados por ele produzidos) do que AC. A média das forças de acoplamento relativas entre os dois locutores pode ser usada como primeira aproximação de uma força de acoplamento média para locutores masculinos do dialeto paulista. Calculando-se a média, obtém-se o valor de 0,73. O intervalo de confiança da média a 95 % está compreendido entre 0,47 e 0,99.

Os parâmetros estimados acima podem ser usados para simular pelo modelo as durações abstratas das unidades V-V do mesmo *corpus* lido pelos locutores e, em seguida, para comparar as durações preditas com as durações observadas no *corpus*, mas modificadas a partir do procedimento já descritos de normalização, suavização e restauração de duração de unidade canônica. Isso será feito em seguida com o locutor AP.

## 6. Simulações da estruturação rítmica com o modelo

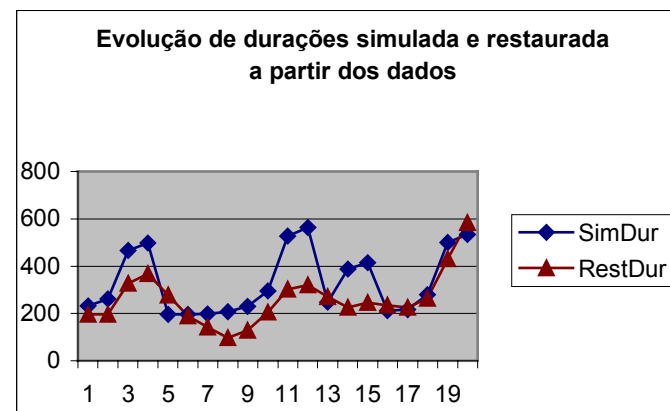
Para as simulações apresentadas em seguida, os valores dos parâmetros do modelo empregados serão informados. A evolução das durações observada e predita do trecho “Em seguida apareceu um papagaio real que tinha fama de orador.” será comparada graficamente acompanhada do coeficiente de correlação para avaliar o grau de coesão entre simulação e observação.

A primeira simulação (figura 5) foi realizada para a taxa normal utilizando os valores de  $(w_0, \alpha, \beta, T_0) = (0,70; 1,0; 1,0; 195 \text{ ms})$ . O valor do período do oscilador silábico não-acoplado ( $T_0$ ) foi calculado pela média das durações V-V não coincidentes com acentos frasais, o que resultou no valor de 195 ms para a taxa normal, 243 ms para a taxa lenta e 167 ms para a taxa rápida. Os valores da taxa de indução bem como o da força de acoplamento relativa empregados foram aqueles estimados

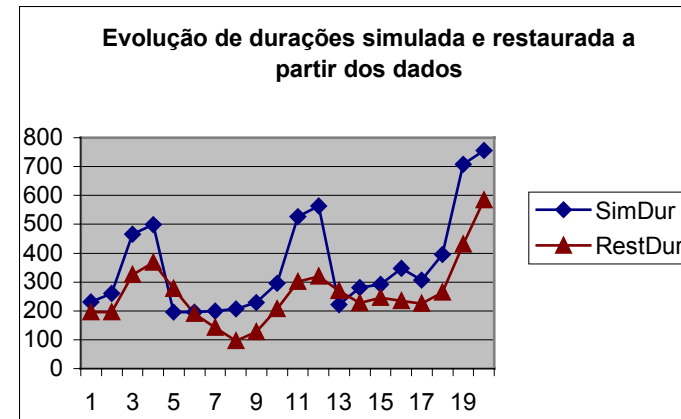
anteriormente para AP (tabela 10, para a taxa normal. O valor da taxa de decaimento  $\beta$  foi fixado em 1.

A detecção automática de grupos acentuais delimitou quatro grupos acentuais para o trecho considerado na taxa normal: [Em seguida a][pareceum papagaio real qu][e tinha f][ama de orador.], com respectivamente 4, 8, 3 e 6 V-V. Os valores das amplitudes dos pulsos do oscilador acentual foram todos fixados previamente em 1 (não há critério sintático-semântico para essa escolha, a não ser de uniformidade na produção de acentos frasais).

Após simular as durações V-V pela aplicação das fórmulas (1) e (2) acima, compararam-se as mesmas com as durações observadas e modificadas pelo procedimento tripla referido acima. O coeficiente de correlação entre as durações simulada e observada (usando-se  $\mu_r = 195$  ms e  $\sigma_r = 52$  ms para obter uma duração de V-V canônico) é de 56 %. Note que o terceiro pico da duração observada (RestDur) está bastante reduzido (figura 5). Para simular esse resultado experimental, o valor da amplitude do pulso do oscilador acentual correspondendo ao terceiro grupo acentual deve certamente ser menor que 1. Quando a amplitude do pulso do oscilador acentual nessa terceira posição é substituído pelo valor 0,5, obtém-se um coeficiente de correlação de 88 %, como se pode verificar pela melhor adequação entre as durações observada e simulada naquele trecho (figura 6).

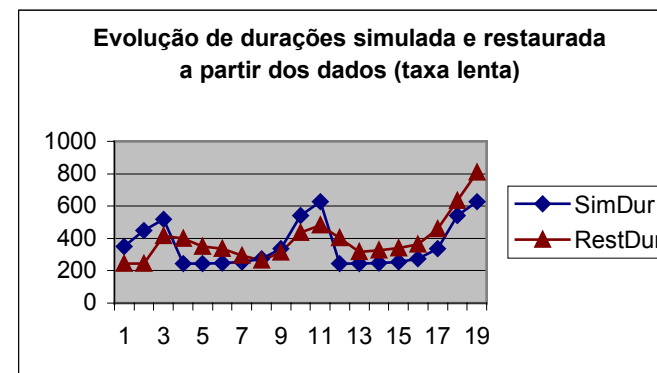


**Figura 5** – Durações V-V simuladas (SimDur) e restauradas (RestDur) a partir dos dados do locutor AP (em milissegundos), da primeira frase do texto de Lobato, anexo 1. Parâmetros do modelo usados no texto. Pulsos do oscilador acentual = (1,0; 1,0; 1,0; 1,0).



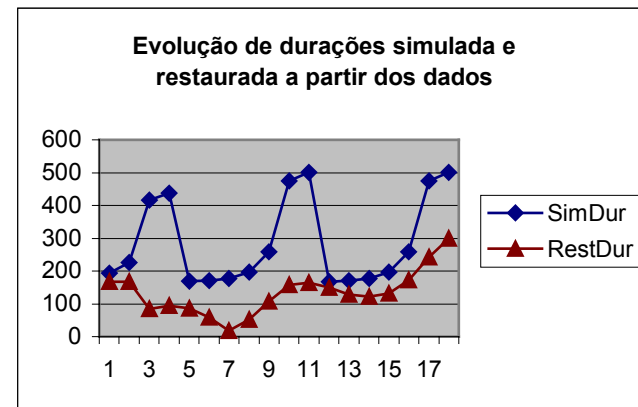
**Figura 6** – Durações V-V simuladas (SimDur) e restauradas (RestDur) a partir dos dados do locutor AP (em milissegundos), da primeira frase do texto de Lobato, anexo 1. Parâmetros do modelo usados no texto. Pulsos do oscilador acentual = (1,0; 1,0; 0,5; 1,0).

A simulação para a taxa lenta foi obtida com os valores  $(w_0, \alpha, \beta, T_0) = (0,46; 1,06; 1,0; 243 \text{ ms})$ , com três grupos acentuais de 3, 8 e 8 V-V, correspondendo aos trechos [Em seguid], [a apareceu um papagaio real] e [que tinha fama de orador], fixando-se as amplitudes dos pulsos do oscilador acentual em 1. Para restaurar os valores de duração a partir da duração normalizada e suavizada usaram-se os valores de  $\mu_r = 243 \text{ ms}$  e  $\sigma_r = 64 \text{ ms}$ , conforme explicado acima. O coeficiente de correlação obtido é de 77 %. A coerência entre simulação e observação pode ser vista na (figura 7).



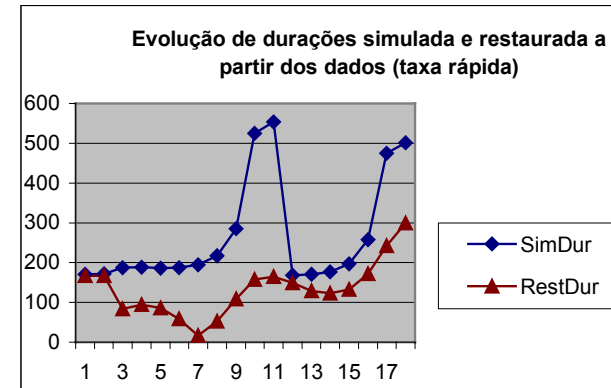
**Figura 7** – Durações V-V simuladas (SimDur) e restauradas (RestDur) a partir dos dados do locutor AP (em milissegundos), da primeira frase do texto de Lobato, anexo 1. Pulsos do oscilador acentual = (1,0; 1,0; 1,0), para a taxa lenta.

Para a taxa rápida, procedeu-se também a uma simulação com valores dos parâmetros  $(w_0, \alpha, \beta, T_0) = (0,58; 1,22; 1,0; 167 \text{ ms})$ , com três grupos acentuais de 4, 7 e 7 V-V, correspondendo aos trechos [Em seguidap], [areceu um papagaio real] e [que tinha fama de orador], com amplitudes dos pulsos do oscilador acentual fixadas inicialmente em 1. Para restaurar os valores de duração a partir da duração normalizada e suavizada usaram-se os valores de  $\mu_r = 167 \text{ ms}$  e  $\sigma_r = 44 \text{ ms}$ , conforme explicado acima. O coeficiente de correlação obtido foi de 60 %. A coerência entre simulação e observação pode ser vista na (figura 8).



**Figura 8** – Durações V-V simuladas (*SimDur*) e restauradas (*RestDur*) a partir dos dados do locutor AP (em milissegundos), da primeira frase do texto de Lobato, anexo 1. Parâmetros do modelo usados no texto. Pulsos do oscilador acentual = (1,0; 1,0; 1,0), para a taxa rápida.

Tendo em vista que a comparação entre as evoluções de duração simulada e observada revela um nível acentual bem menor para o primeiro grupo acentual, resolveu-se simular dados de duração nessa taxa rápida com o valor de 0,1 para a amplitude do primeiro pulso do oscilador acentual, obtendo-se a curva apresentada na (figura 9). O coeficiente de correlação obtido passa a ser de 69 %.



**Figura 9** – Durações V-V simuladas (*SimDur*) e restauradas (*RestDur*) a partir dos dados do locutor AP (em milissegundos), da primeira frase do texto do anexo 1. Pulsos do oscilador acentual = (0,1; 1,0; 1,0), para a taxa rápida.

Pelo exposto, para todas as taxas consegue-se explicar pelo menos cerca de metade das variância (coeficiente de correlação de 70 %) dos dados de duração observados através do modelo dinâmico de ritmo da fala. É preciso salientar que os resultados aqui apresentados foram obtidos a partir de modificações fundamentadas em considerações numéricas bem simples, sem uma análise sintático-semântica apurada. Essa última permitiria, muito provavelmente, a obtenção de coeficientes de correlação ainda mais elevados. Um estudo das conseqüências da manipulação da taxa de decaimento fundamentado na dinâmica dos padrões acentuais da fala natural também possibilitaria uma correlação maior, já que padrões descendentes de duração observada ao longo do segundo grupo acentual podem ser constatados nas figuras 5 ou 6 e 8 ou 9. Note que o valor desse parâmetro foi sempre fixado em 1 para as simulações aqui apresentadas.

Uma primeira tentativa de manipulação desse parâmetro já foi realizada, a partir da reprodução de dados comportamentais de outros locutores paulistas, dados esses relacionados à implementação de dois processos fônicos, aquele que chamamos de acentuação secundária (devido a evidência experimental de que seja um fenômeno prosódico e não exclusivamente lexical), bem como aquele relacionado à constituição de padrões acentuais na situação de adjacência de sílabas acentuadas, isto é, aquilo que a literatura fonológica tradicional chama de choque acentual (*stress clash*).

Barbosa *et al.* (2004) mostram que, no primeiro caso, não há indício de alternância binária nos dados comportamentais do PB, mas apenas uma proeminência inicial, implementada pelo modelo dinâmico de ritmo pela manipulação do parâmetro taxa de decaimento. Quando o valor desse parâmetro é elevado em relação ao da taxa de indução, é possível gerar durações maiores para a primeira unidade V-V em relação à segunda, aproximando-se assim do padrão duracional observado no *corpus* estudado.

Quanto à situação de aparente choque acentual, Barbosa e Arantes (2003) mostraram que não há tentativa alguma de retração (ou desfazimento) do choque aparente, pelo contrário, um reforço do caráter iâmbico na situação alegada como de choque, que exhibe valores maiores de duração na sílaba imediatamente anterior à sílaba frasalmente acentuada (em comparação com a mesma sílaba numa frase em que a mesma palavra, nas mesmas condições prosódicas, é seguida por uma sílaba não acentuada). Essa maior duração para a unidade mais próxima do acento frasal é exatamente aquela esperada por simulações com o modelo, como se pode depreender do exposto neste artigo. Assim, segundo o modelo, não pode haver choque acentual algum, pelo menos no que diz respeito à duração. No modelo, a alternância rítmica é assegurada a priori, pelo posicionamento dos pulsos do oscilador acentual, que obedece princípios de eurritmia.

## 7. Discussão

O modelo dinâmico de produção do ritmo da fala apresentado aqui é capaz de dar conta de dados de duração natural, predizendo com alto valor de correlação o padrão global de acentuação ligado à implementação da duração encontrado em frases naturais sujeitas a variação da taxa de elocução. A língua em estudo aqui é o português brasileiro. A aplicação das técnicas apresentadas aqui para o italiano, espanhol americano e europeu e o francês padrão deve poder ser feita de forma transparente a partir do exposto neste trabalho. Para línguas que constroem grupos acentuais com cabeças à esquerda, a aplicação não é imediata, pois envolveria uma mudança nas fórmulas de acoplamento de período que, na forma em que foram sugeridas, tende a fornecer durações cada vez maiores da esquerda para a direita. Porém acreditamos que a constituição de grupos com cabeça à direita com base em critérios duracionais apenas, mesmo nessas línguas, para a estimativa da força de acoplamento relativa deve fornecer resultados que valem a pena ser explorados. Em seguida, serão tecidos alguns comentários sobre o modelo dinâmico de ritmo, do ponto de vista de seu emprego para a geração automática de durações, bem como sobre o uso das técnicas aqui apresentadas para a pesquisa tipológica do ritmo da fala.

### 7.1. Do modelamento dinâmico do ritmo da fala

O modelo de osciladores acoplados gera os padrões acentuais prosodicamente mais relevantes de enunciados do PB porque integra três características fundamentais para o controle prosódico da duração, possivelmente universais (embora parametrizadas segundo a língua e a relação indivíduo-contexto de enunciação): (1) a tendência a aumentar a duração ao longo do grupo acentual, da esquerda para a direita, se dá com a proximidade do acento frasal com taxa de aumento guiada por uma função exponencial (função de sincronismo), parametrizada linguisticamente pela força de acoplamento, até a penúltima unidade V-V do grupo acentual. (A última unidade V-V sofre um aumento muito menor em relação à unidade anterior, dada por uma porcentagem fixa parametrizada por  $T_0$ , como pode ser visto na fórmula (1a) para a posição N-1.); (2) duas tendências contrárias convivem durante os dois primeiros ciclos do oscilador silábico em cada grupo acentual: uma tendência a aumentar a duração via indução e outra tendência a fazer a duração retornar a seu

valor na condição de não-acoplamento. Esse efeito conjunto gera proeminências iniciais. A parametrização dessas tendências respectivamente pelas taxas de indução e decaimento é dependente do indivíduo e de suas experiências com a enunciação (que começa durante a aquisição da linguagem), bem como da relação desse indivíduo com o contexto de sua enunciação, a finalidade do ato comunicativo, além de outros fatores discursivos; (3) a duração das unidades V-V ao longo do grupo acentual é modificada pela amplitude do pulso do oscilador acentual, ela mesma resultado de especificação sintático-semântica o que garante a conexão do nível prosódico com níveis superiores da gramática. A amplitude do pulso também dependerá do indivíduo e de sua relação com o contexto comunicativo. Toda relação com o contexto de produção é fundamental para a Teoria de sistema Dinâmicos, como se pode ver no trabalho de Kelso (1995) ou no trabalho sobre aquisição de habilidades motoras do ponto de vista dinâmico em Thelen e Smith (1994).

Para que se torne um modelo completo de produção do ritmo da fala, é necessário ainda integrar a geração de padrões de frequência fundamental àqueles de duração. Isso poderá ser feito considerando ainda um oscilador glotal, a ser controlado por parâmetros específicos (ligados a níveis superiores da gramática), sem impedir seu acoplamento com os osciladores acentual e silábico. Dado que o número de acoplamentos bidirecionais possíveis passa a ser três, é necessário primeiramente recolher evidência experimental para um tal modelo, sobretudo de *corpora* e análises que permitam avaliar a covariação entre duração e frequência fundamental. Uma das questões relacionadas ao tema diz respeito à relação de causa e efeito entre implementação de duração e de frequência fundamental, pois em PB, unidades lingüísticas submetidas a ênfase de contraste, por exemplo, são implementadas com um movimento tonal do tipo *rise/fall* sempre acompanhado de maiores valores de duração para as mesmas unidades.

Explicar a variabilidade dos dados de duração ainda exige a interação com a especificação da duração intrínseca no léxico gestual. Esse trabalho de longo fôlego está sendo iniciado em nosso laboratório, e exigirá a análise de dados articulatorios, num primeiro momento em colaboração com outros laboratórios. Duas possibilidades se abrem para essa interação. Do ponto de vista dos efeitos da pauta gestual sobre o oscilador silábico induzido, a implementação pode se dar sob a forma de perturbações locais efetuadas sobre os padrões do oscilador silábico induzido, a partir de parâmetros fornecidos pela pauta (rigidez própria a um determinado gesto, presença ou ausência de gestos na pauta). Do ponto de vista dos efeitos do oscilador silábico induzido sobre a pauta gestual, os picos de seu ciclo especificam o início de gestos de Corpo de Língua (gesto TB nos trabalhos de Browman e Goldstein citados aqui) com constrição livre (não crítico e não fechado), para a pauta gestual. A duração do ciclo induzido pode ser usada para modificar os parâmetros da pauta gestual, que produzirá a trajetória dos gestos articulatorios concretos. Quanto maior o valor da duração (por exemplo, ao se aproximar do acento frasal), menor o valor da rigidez dos gestos da pauta, o que produz durações maiores para os mesmos gestos. Um procedimento semelhante, porém introduzindo uma noção local de modificação prosódica (no domínio dos chamados  $\pi$ -*gestures*) foi implementada por Byrd e Saltzman (2003). A duração acústica segmental seria inferida a partir da projeção da extensão desses gestos em suas camadas respectivas sobre o eixo temporal.

## 7.2. Das técnicas de investigação da tipologia rítmica

A técnica de detecção de grupos acentuais com base em dados de duração (para línguas que controlam principalmente esse parâmetro para assinalar acentos frasais) bem como a técnica para estimação de três parâmetros do modelo (força de acoplamento relativa, taxa de indução e período do oscilador silábico na condição de não-acoplamento) possibilita a automatização da análise de *corpora* para estudo tipológico do ritmo da fala. O valor da força de acoplamento é um dos candidatos para especificar distinções na implementação do ritmo entre as línguas. Seu uso deve ser feito de forma estatisticamente orientada. Para se atingir uma certa estabilidade na estimativa desse parâmetro a partir de um *corpus* a ser analisado, o número mínimo de grupos acentuais no mesmo não pode ser inferior a cerca de vinte (no caso do texto do anexo 1, esse número não foi inferior a 22 para as três taxas e para os dois locutores analisados).

Se o valor do parâmetro força de acoplamento relativa for empregado para uma avaliação de diferenças interlingüísticas de implementação rítmica, um número bem maior de locutores do que os dois apresentados aqui deve ser considerado para análise, tendo em vista que uma certa variabilidade entre os locutores deve ser esperada. Além disso, recomenda-se que as comparações entre línguas controlem a taxa de elocução, isto é, elas devem ser as mesmas entre grupos de falantes de línguas ou variedades lingüísticas distintas (Barbosa, 2000b). Tendo-se assegurado esse controle, tomam-se grupos interlingüísticos que apresentem uma homogeneidade em termos de gêneros, faixa etária e nível de escolaridade. Emprega-se em seguida um teste estatístico que avalie diferenças entre as amostras por língua (como por exemplo, a ANOVA). Para uniformização do material a ser lido, pode-se começar estudando textos traduzidos para as línguas em estudo. Neste artigo traduções para o francês e o inglês do texto de Lobato são sugeridas.

Como se depreende do exposto, este artigo tem um duplo interesse. De um lado provocar uma série de questões sobre a relação entre elementos discretizáveis e contínuos na pesquisa prosódica, por propor uma noção de abstração da implementação do ritmo da fala ligada a um modelo funcional que manipula um conjunto reduzido de parâmetros com a finalidade de reduzir os graus de liberdade para o controle da duração. De outro lado, sugerir uma metodologia de investigação da tipologia rítmica menos fundamentada em relações dicotômicas como aquelas de isocronismo acentual ou silábico (isocronismo jamais comprovados experimentalmente) e mais afeita ao contínuo, que impõe uma visão complementar do ritmo das línguas a partir de uma abordagem fundamentada num modelo de osciladores acoplados que gera padrões duracionais complexos a partir de sua dança entre dois atratores inatingíveis em situações de enunciação com vistas à comunicação: a acentuação pura e a silabidade pura.

## 8. Agradecimentos

Este trabalho se insere no Projeto Temático "Integrando Parâmetros Contínuos e Discretos em Modelos do Conhecimento Fônico e Lexical", nº 01/00136-2. Agradeço ao CNPq (Bolsa de Produtividade em Pesquisa nº 350382/98-0), e aos locutores. Sou grato também às discussões levantadas no LAFAPE, no âmbito do Temático e

dos seminários (especialmente a Aglael Gama Rossi, Ana Luíza G. P. Navas, Edson França e Sandra Madureira). Um agradecimento especial também dedico a colaboradores e alunos que testaram e discutiram o emprego das técnicas aqui descritas, a saber: Ana Cristina F. Matte, Alexandre Meireles, Jussara M. Vieira, Pablo Arantes e Sandra Merlo. Também agradeço os comentários de César Reis e Sandra Madureira a este artigo.

### **Anexo 1: Corpus “Narizinho”**

“Em seguida apareceu um papagaio real que tinha fama de orador. Subiu a tribuna de um poleiro de ouro e fez um belo discurso a respeito da arte de falar. Nesse discurso provou que os homens tinham aprendido a falar com os papagaios, e não os papagaios com os homens, como diz a ciência destes. Uma chuva de palmas acolheu suas palavras.

O mesmo não aconteceu, porém, com a poetisa Lagartixa, que principiou a recitar uma longa poesia e engasgou no meio, acabando o recitativo em choro e faniquito. Para destruir essa má impressão vieram três vagalumes mágicos que fizeram várias sortes, sendo muito apreciada a sorte de comer fogo.” (Lobato, 1920)<sup>4</sup>

---

<sup>4</sup> Versões do mesmo texto, em inglês britânico e em francês padrão, foram preparadas com o concurso de falantes nativos, para possibilitar uma maior uniformização de trabalhos na área de tipologia rítmica. Elas seguem abaixo.

“After that a royal parrot entered, famous for his rhetorical qualities. He stepped up onto a platform of a golden roost and made a wonderful speech about the art of speaking. During the speech he proved that the men had learned to talk with the parrots and not the parrots with the men, as says the science of man. A rain of claps welcomed his words.

The same did not happen, however, with Gecko the poetess, who started to recite a lengthy poem and choked in the course of her speech, closing the recital with cries and hysterics. To counter the bad impression created, three magical firebugs came by and did several tricks, with the fire eating trick being the most appreciated.” (Tradução de: P. Barbosa, M. Salmon, M. Fontes, P. Arantes e J. Santos).

« Ensuite entra un perroquet royal qui avait une grande renommée en tant qu’orateur. Il monta la tribune d’un perchoir en or et fit un beau discours à propos de l’art de parler. Dans ce discours-là il démontra que les hommes eurent appris à parler avec les perroquets et non l’opposé, comme le veut la science des hommes. Une pluie d’applaudissements accueillit ses paroles.

Cette situation ne se reproduisit pas avec la poétesse Mlle Léopard, qui, cependant, débuta la récitation d’un long poème, presque suffoca en plein milieu et acheva sa présentation par une crise de nerfs. Pour atténuer cette mauvaise impression vinrent trois lucioles magiques qui firent plusieurs numéros, entre lesquels, celui d’avalier du feu fut fort apprécié. » (Tradução de: P. Barbosa e P. Nüller).

## Anexo2: Corpus de referência para cálculo de z-score

**Tabela 2** - Durações (e desvios-padrão) em milissegundos dos fones do PB para um locutor pernambucano (segmentos transcritos com o A. F. I.)

VOGAIS						CONSOANTES		
seg	ex.	μ (σ)	seg	ex.	μ (σ)	seg	ex.	μ (σ)
i	sico	87 (19)	ẽ	seixo	126 (13)	p	pata	80 (16)
e	seco (adj.)	110 (19)	ẽ	papéis	134 (13)	t	tata	81 (20)
ε	seco (verbo)	113 (17)	ã	caixa	135 (12)	k	cata	80 (22)
a	saco	132 (45)	ɔ̃	sóis	133 (12)	b	bata	59 (18)
ɔ	soco (verbo)	119 (18)	õ	moita	131 (14)	d	data	58 (20)
o	soco (subst.)	111 (20)	ũ	cuido	122 (13)	g	gata	52 (20)
u	suco	103 (14)	'ɸ̃	cãibra	143 (18)	f	facas	89 (28)
'ɪ	cinto	128 (16)	õ̃	põe	120 (10)	s	saca	96 (24)
'ẽ	sento	134 (15)	ị	rio	145 (23)	ʃ	chaga	104 (19)
'ã	santo	134 (15)	ẹ	eu	134 (17)	v	vaca	57 (19)
'õ	sondo	137 (18)	ε̣	céu	143 (15)	z	zaca	64 (25)
'ũ	sunto	128 (10)	ạ	mau	145 (15)	ʒ	jaca	65 (17)
ɪ	saque	55 (14)	ɔ̣	sol	138 (19)	S <sub>coda</sub>	casca	74 (31)
ẽ	ópera	51 (13)	ọ	pouco	133 (14)	m	mata	61 (13)
ɐ	saca	59 (15)	'ɸ̣	pão	140 (13)	n	nata	51 (17)
õ	cômodo	51 (14)	ị	hável	61 (16)	ɲ	ganho	84 (27)

U	saco	47 (15)	u̯	cônsul	63 (18)	r	carro	78 (24)
ĩ	interim	89 (25)	ɪ̃	série	49 (14)	r	caro, prato, carta	30 (10)
ẽ	hífen	82 (26)	ɨ̃	tênuê	106(14)	l	galo	47 (13)
ẽ	ímã	93 (34)	ɨ̃	ânsia	80 (17)	ʎ	galho	72 (24)
õ	fóton	88 (19)	ɔ̃	água	70 (16)	L	sal	41 (11)
ũ	álbum	98 (16)	ẽ̃	bênção	96(26)			

### Anexo 3: Beat Extractor for Praat

O *script* do Praat (recomenda-se o uso da versão 4.1 ou superior do mesmo, pois algumas funções implementadas podem não funcionar em versões anteriores. A última versão do Praat está disponível na URL <http://www.praat.org>) que realiza a detecção automática de *onsets* de vogal foi fundamentado no programa de Cummins e Port (1998), esse mesmo guiado pelo algoritmo sugerido por Scott (1993) em seu trabalho sobre *perceptual-centres*. A detecção é feita em quatro etapas listadas em seguida.

- filtragem do sinal de fala pelo uso de um filtro passa-faixas com valores *default* das frequências de corte para um locutor masculino fixadas entre 1000 e 1800 Hz. Os valores *default* para um locutor feminino são 1150 e 2100 Hz. Essa filtragem preserva a energia na faixa entre o primeiro e segundo formantes. O filtro *default* é o filtro Butterworth de segunda ordem, mas a escolha por um filtro Hanning com faixas de transição mais abruptas do que a do Butterworth pode ser escolhida opcionalmente;
- retificação do sinal previamente filtrado;
- filtragem passa-baixas num valor *default* da frequência de corte de 20 Hz (preserva transições mais bruscas de amplitudes no PB – por exemplo, no caso de “tap” [ɾ] seguido de vogal);
- localização das fronteiras pela recuperação dos pontos que satisfazem duas restrições: (1) ponto de derivada máxima local positiva superior a 15 % do valor máximo de derivada do sinal obtido após a aplicação da etapa anterior, e (2) ponto de amplitude superior a 12 % da amplitude máxima obtida após a aplicação da etapa anterior.

As fronteiras assim detectadas são gravadas em arquivo com extensão *TextGrid*, que é um arquivo Praat para registro de marcação de fronteiras e rótulos de segmentos. A detecção é semi-automática porque há necessidade de correção

manual das predições do algoritmo. O programa está disponível via solicitação ao autor.

## Referências bibliográficas

**[Albano, 2001]** Albano E. C. (2001). *O Gesto e suas Bordas: esboço de uma fonologia acústico-articulatória do português brasileiro*. Mercado de Letras: Campinas.

**[Albano e Moreira, 1996]** Albano E. C., Moreira A. A. (1996). Archisegment-based letter-to-phone conversion for concatenative speech synthesis in Portuguese. *Proceedings of the International Congress of Spoken Language Research*. Filadélfia. 3, 1208-1711.

**[Albano et al., 1995]** Albano E. C., Moreira A., Aquino P., Silva A., Kakinohana R. (1995). Segment frequency and word structure in Brazilian Portuguese. *Proceedings of the XIII<sup>th</sup> International Congress of Phonetic Sciences*. Estocolmo, Suécia. 3, 346-349.

**[Allen, 1972]** Allen G. D. (1972). The location of rhythmic stress beats in English I & II. *Language & Speech*. 15, 72-100 e 179-195.

**[Barbosa, 1994]** Barbosa P. A. (1994) *Caractérisation et génération automatique de la structuration rythmique du français*, Thèse de doctorat de troisième cycle. ICP/INP de Grenoble, França.

**[Barbosa, 1996]** Barbosa P. A. (1996). At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation. *Cadernos de Estudos Lingüísticos*. 31, 33-53.

**[Barbosa, 2000a]** Barbosa P. A. (2000). "Syllable-Timing in Brazilian Portuguese": uma Crítica a Roy Major. *D.E.L.T.A.* 16 (2), 369-402.

**[Barbosa, 2000b]** Barbosa P. A. (2000). Illuminating some Methodological Issues Concerning Speech Timing Research from a Comparison between European and Brazilian Portuguese. *Cadernos de Estudos Lingüísticos*. 39, 41-50.

**[Barbosa, 2001]** Barbosa P. A. (2001). Generating Duration from a Cognitively Plausible Model of Rhythm Production. *Proceedings of the Eurospeech 2001*, Ålborg, Dinamarca. 2, 967-970.

**[Barbosa, 2002a]** Barbosa P. A. (2002). Explaining Cross-Linguistic Rhythmic Variability via a Coupled-Oscillator Model of Rhythm Production. In Bel B. e Marlien I. (eds.) *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence, França. 163-166.

**[Barbosa, 2002b]** Barbosa P. A. (2002). Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production. *Cadernos de Estudos Lingüísticos*. 43, 71-92.

**[Barbosa e Arantes, 2003]** Barbosa P. A., Arantes P. (2003). Investigation of non-pitch-accented phrases in Brazilian Portuguese: no evidence favoring stress shift. *Proceedings of the XV<sup>th</sup> International Congress of Phonetic Sciences*. Barcelona. 475-478.

**[Barbosa et al., 1999]** Barbosa P.A., Violaro, F., Albano, E.C., Simões, F.O., Aquino, P. A., Madureira, S. and Françoço, E. (1999). Aiuruetê: a High-Quality Concatenative Text-to-Speech System for Brazilian Portuguese with Demisyllabic Analysis-Based Units and a Hierarchical Model of Rhythm Production. *Proceedings of the Eurospeech 99*, Budapeste. 5, 2059-2062.

**[Barbosa et al., 2004]** Barbosa P. A., Arantes P., Silveira, L. S. (2004). Unifying Stress Shift and Secondary Stress Phenomena with a Dynamical Systems Rhythm Rule. *Proceedings of the 2<sup>nd</sup> International Conference on Speech Prosody*, Nara, Japão. 49-52.

**[Barbosa e Bailly, 1994]** Barbosa P., Bailly G. (1994). Characterisation of rhythmic patterns for text-to-speech synthesis. *Speech Communication*. 15 (1-2), 127-137.

**[Barbosa e Madureira, 1999]** Barbosa P. A., Madureira S. (1999). Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese. Proceedings of the XIV<sup>th</sup> *International Congress of Phonetic Sciences*, San Francisco. 1, 297-300.

**[Browman e Goldstein, 1986]** Browman C., Goldstein L. (1986). Towards an articulatory phonology. *Phonology Yearbook*. 3, 219-252.

**[Browman e Goldstein, 1989]** Browman C., Goldstein L. (1989). Articulatory Gestures as Phonological Units. *Phonology*. 6, 201-251

**[Browman e Goldstein, 1990]** Browman C., Goldstein L. (1990). Tiers in Articulatory Phonology with some Implications for casual speech. In Kingston J. I., Beckman M. E. (eds.), *Papers in Laboratory Phonology I*. Cambridge University Press: Cambridge. 341-376.

**[Browman e Goldstein, 1992]** Browman C., Goldstein L. (1992). Articulatory Phonology: an overview. *Phonetica*. 49, 155-180.

**[Byrd e Saltzman, 2003]** Byrd D., Saltzman E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*. 31, 149-180.

**[Campbell e Isard, 1991]** Campbell N. W., Isard S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics*. 19, 37-47.

**[Classé, 1939]** Classé A. (1939). *The Rhythm of English Prose*. Blackwell: Oxford.

**[Cummins, 2002]** Cummins F. (2002). Entraining speech with speech and metronomes. *Cadernos de Estudos Lingüísticos*. 43, 55-70.

**[Cummins e Port, 1998]** Cummins F., Port R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*. 26, 145-171.

**[Dogil e Braun, 1988]** Dogil G., Braun G. (1988). *The PIVOT model of speech parsing*. Verlag: Viena.

**[Eco, 1995]** Eco U. (1995). *The search for the perfect language*. Blackwell : Oxford.

**[Fernandes, 1976]** Fernandes N. H. (1976). *Contribuições para uma análise instrumental da acentuação e da intonação do português*. Dissertação de Mestrado inédita. USP. São Paulo.

**[Fowler, 1980]** Fowler C. A. (1980) Coarticulation and theories of extrinsic timing. *Journal of Phonetics*. 8, 113-133.

**[Fujimura, 1995]** Fujimura O. (1995). Prosodic organization of speech based on syllables: the C/D model. Proceedings of the XIII<sup>th</sup> *International Congress of Phonetic Sciences*, 3, 10-17.

**[Fujimura, 2002]** Fujimura O. (2002). Temporal Organization of Speech Utterance: A C/D Model Perspective. *Cadernos de Estudos Lingüísticos*. 43, 9-35.

**[Janker, 1995]** Janker P.M. (1995). On the influence of the internal structure of a syllable on the p-center perception. Proceedings of the XIII<sup>th</sup> *International Congress of Phonetic Sciences*. Estocolmo. 2, 510-513.

**[Kelso, 1995]** Kelso, J. A. S., 1995. *Dynamic patterns: the self-organization of brain and behavior*. MIT Press: Cambridge, EUA.

**[Lieberman e Prince, 1977]** Liberman M., Prince A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*. 8 (2), 249-336.

**[Lobato, 1920]** Lobato M. (1920). *A Menina do Narizinho Arrebitado*. Revista do Brasil: São Paulo.

**[Massini, 1991]** Massini G. (1991). *A duração no estudo do acento e do ritmo em português*. Dissertação de Mestrado inédita. Universidade de Campinas: Campinas.

**[Mcauley, 1995]** Mcauley J. D. (1995). *Perception of time as phase: toward an adaptive-oscillator model of rhythmic pattern processing*. Tese de Doutorado Inédita. Indiana University: Indiana, EUA.

**[Morton et al., 1976]** Morton J., Marcus S., Frankish C. (1976). Perceptual centers (p-centers). *Psychological Review*. 83 (5), 405-408.

**[O'Dell e Nieminen, 1999]** O'Dell M., Nieminen T. (1999). Coupled oscillator model of speech rhythm. Proceedings of the *XIV<sup>th</sup> International Congress of Phonetic Sciences*. San Francisco. 2, 1075-1078.

**[Öhman, 1966]** Öhman S. (1966) Coarticulation in VCV utterances: spectrographic measurements. *J. Acoustic. Soc. Am.* 39, 151-168.

**[Pompino-Marschall, 1989]** Pompino-Marschall B. (1989). On the psychoacoustic nature of the p-center phenomenon. *Journal of Phonetics*. 17, 175-192.

**[Pompino-Marschall, 1991]** Pompino-Marschall B. (1991). The syllable as a prosodic unit and the so-called p-centre effect. *Forschungsberichte der Institut für Phonetik und Sprachliche Kommunikation der Universität München*. 29, 66-124.

**[Port et al., 1995]** Port R., Cummins F., Gasser M. (1995). A Dynamic Approach to Rhythm in Language: Toward a Temporal Phonology. In Luka B., Needs B. (eds.), *Proceedings of the Chicago Linguistics Society*. 375-397.

**[Potter et al., 1947]** Potter R. K., Kopp G. A., Green H. C. (1947). *Visible speech*. van Nostrand: Nova York.

**[Roach, 1982]** Roach P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In Crystal D. (ed.), *Linguistic controversies. Essays in linguistic theory and practice in honour of F. R. Palmer*. Edward Arnold: Londres. 73-79.

**[Saltzman e Munhall, 1989]** Saltzman, E., Munhall K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*. 1 (4), 333-382.

**[Scott, 1993]** Scott S. K. (1993). *Perceptual centres in speech: an acoustic analysis*. Tese de doutorado inédita. University College London: Londres.

**[Strogatz e Stewart, 1994]** Strogatz S., Stewart I. (1994). Oscillateurs couplés et synchronisation biologique. *Pour la Science*. 196, 40-46.

**[Thelen e Smith, 1994]** Thelen E., Smith L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press: Cambridge, EUA.



**Plínio A. Barbosa** doutorou-se na área de Ciência da Fala em 1994 pelo *Institut de la Communication Parlée*, em Grenoble, França. É professor-doutor do Departamento de Lingüística do Instituto de Estudos da Linguagem na Universidade Estadual de Campinas (Brasil), onde desenvolve pesquisa sobre Análise e Modelamento Prosódicos da Fala, a partir do estudo do português brasileiro. Publicou mais de 30 artigos em congressos nacionais, internacionais (*ICPhS*, *Eurospeech* e *Speech Prosody*) e periódicos especializados. É um dos primeiros a propor um modelo dinâmico de osciladores acoplados para dar conta da produção do ritmo da fala. Seus interesses cobrem as áreas de Fonética e Fonologia, Fundamentos Cognitivos do Ritmo, Produção e Percepção de Fala e Bases Fonéticas para Síntese de Fala. É coordenador do grupo de pesquisa “Análise e Modelamento Prosódicos da Fala” no CNPq, do qual é pesquisador 1D. É enfim redator-chefe para a língua portuguesa da *Revista Românica em Ciências Cognitivas - In Cognito*, bem como membro da comissão editorial do *International Journal of Speech Technology*. Membro das sociedades científicas IPA, Abralín e ISCA (incluindo a SProSIG).