ESTUDO PILOTO DA VARIAÇÃO RÍTMICA ENTRE FALANTES DO PORTUGUÊS BRASILEIRO À LUZ DE UM MODELO DINÂMICO DO RITMO

Pilot study of individual rythmic variability in Brazilian Portuguese in light of a dynamical rhythm model

ARANTES, Pablo
CANGIANI, Rafael Ésquines

Universidade Federal de São Carlos

Resumo: *O objetivo do trabalho é estimar o grau de variabilidade do ritmo da fala entre falantes do português brasileiro (PB) a partir da estimativa de valores para três parâmetros de um modelo que parte da proposta de que a produção do ritmo pode ser modelada como um sistema de osciladores acoplados que representam dois níveis de organização temporal: silabicidade e acentução. Os parâmetros são a taxa de indução do oscilador silábico pelo oscilador acentual* (α)*, a taxa de decaimento do oscilador silábico* (β) *e a força de acoplamento entre os dois osciladores* (w_0)*. Aplicamos um procedimento de otimização para extrair a combinação dos três parâmetros que minimizam a distância entre contornos de duração normalizada de enunciados produzidos por falantes nativos do PB e contornos produzidos pelo modelo de ritmo. Os eunciados são leituras de um texto com 144 palavras, lido por oito falantes do PB em três níveis de taxa de elocução (lenta, normal/habitual e rápida). Os resultados mostram que o parâmetro mais variável entre os falantes e entre os níveis da taxa é* w0*, o que contraria a expectativa de que esse parâmetro deveria ser uma característica relativamente invariante em uma variedade da língua.*
Palavras-chave*: Prosódia; Ritmo da fala; Modelamento do ritmo.*

Abstract: *Our goal is to estimate the degree of variability in speech rhtyhm in Brazilian Portuguese (BP). We do so by estimating three parameters of a coupled-oscillator-based model of speech rhythm – entrainment rate* (α), *reset rate* (β) *and coupling strength* (w_0)*. Model-generated duration contours are compared natural ones in order to derive parameter triplets that minimize the error between the two. Natural duration contours were obtainded from readings of a 144-word text by 8 native BP speakers in slow, normal and fast rates. Results show that* w_0 *is the most variable parameter across speakers and rate levels, which is counter to the notion that* w_0 *should be stable for a given speaker community*
Keywords: *prosody; speech rhythm; rhythm modeling.*

1 Introdução

O objetivo do trabalho é estudar a variabilidade individual na implementação do ritmo da fala no português brasileiro (PB a partir de agora) usando um modelo dinâmico do ritmo para quantificar e entender essa variação. O modelo do ritmo será apresentado em detalhes na seção 1.1. Os principais parâmetros do modelo que interessam para o trabalho são a taxa de indução do oscilador silábico pelo oscilador acentual (α), a taxa de decaimento do oscilador silábico (β) e a força de acoplamento entre os dois osciladores (w_0). A seção 1.2 explicará como os parâmetros se relacionam com a questão da chamada tipologia rítmica e em que medida os resultados obtidos podem ser entendidos como um avanço no entendimento dessa questão.

1.1 Modelo dinâmico do ritmo

A versão do modelo que será tomada como referência para o trabalho é aquela que se encontra descrita em Barbosa (2006). Os detalhes da implementação matemática são apresentados lá e serão omitidos aqui por brevidade. O modelo postula a existência de dois osciladores abstratos. O primeiro deles é o oscilador a ser induzido, chamado de *oscilador silábico*, que representa a sequência de unidades do tamanho da sílaba (ou unidades V-V) num dado enunciado; o outro é chamado de *oscilador acentual* e representa a sequência, inicialmente eurrítmica, de acentos frasais de um enunciado. A ação indutora do oscilador acentual tem como resultado o aumento progressivo do período do oscilador silábico e atinge seu grau máximo quando coincidem os inícios de um ciclo do oscilador acentual e do silábico. A batida do oscilador silábico já induzido interage com a pauta gestual, especificando o início (*onset*) dos gestos vocálicos. As quatro variáveis do modelo que nos interessam mais diretamente são: o período do oscilador silábico desacoplado (T_0); α, que modula quão rapidamente se dá a indução; β, que modula a volta do oscilador silábico a seu estado não induzido após a batida do oscilador acentual e w_0, que indica o grau de acoplamento entre os dois osciladores.

1.2 Tipologia rítmica

A tradição de análise do ritmo da fala baseia-se em uma intuição explicitada inicialmente por Lloyd-James e Pike, segundo a qual algumas línguas teriam seu ritmo caracterizado pela sucessão isócrona de sílabas acentuadas e outras pelo fato das sílabas apresentarem uma duração isócrona, independentemente de serem acentuadas ou não. Pike criou a nomenclatura línguas de ritmo acentual (*stress-timed*) e de ritmo silábico (*syllable-timed*) para fazer referência a esses dois tipos de organização temporal. Abercrombie atribuiu um caráter dicotômico a essa diferença, afirmando que todas as línguas do mundo são faladas com um ou outro tipo de ritmo. Barbosa (2000) faz uma revisão da literatura a respeito da busca pelos correlatos acústicos e perceptuais dessa hipótese de uma divisão dicotômica dos tipos rítmicos, na qual põe em relevo a falta de uma resposta definitiva a respeito do assunto. Em um trabalho posterior, Barbosa (2002) reinterpreta essa discussão à luz do seu modelo de ritmo, descrito na seção 1.1, sugerindo que os rótulos "ritmo acentual" e "ritmo silábico" podem ser entendidos como pontos em um contínuo definido pela predominância relativa de uma de duas tendências: a silabicidade e a acentuação.

Nos termos do modelo dinâmico de Barbosa, a força de acoplamento entre os dois osciladores (w_0) é o parâmetro que determina o grau de predomínio de um oscilador sobre o outro. Valores menores do parâmetro indicam menor influência do oscilador acentual sobre o silábico e o contrário é indicado por valores maiores. Barbosa (2002) sugere que silabicidade e acentuação são tendências universais nas línguas, representadas em seu modelo pelos osciladores silábico e acentual, mas o valor da força de acoplamento entre os dois seria uma propriedade específica de cada língua particular, devendo variar pouco entre os falantes de uma mesma língua.

A variável T_0 tem um papel importante para definir as características rítmicas de uma língua, segundo Barbosa (2000), uma vez que a taxa de elocução (que pode ser considerada uma estimativa aproximada de T_0) interage com w_0, como o autor demonstra a partir de dados do PB: taxas de elocução mais altas tendem a estar associadas com a manifestação de ritmo silábico e taxas mais baixas a manifestação de ritmo acentual.

O papel dos demais parâmetros livres do modelo, a taxa de indução (α) e decaimento (β), foi menos discutido no contexto da tipologia rítmica, portanto está menos claro que papel terão. Pode ser que variem mais entre os falantes do que Barbosa supõe que o parâmetro w_0 varie. Tomando por base Barbosa (2006), é possível dizer que valores maiores para os parâmetros α e β estão associados à implementação de um ritmo acentual: uma taxa de indução do oscilador silábico maior resulta em durações maiores das unidades V-V que precedem o acento frasal, violando uma possível sensação de isocronia entre as sílabas do grupo acentual; uma taxa de decaimento maior resulta em um volta mais rápida do oscilador silábico induzido ao seu período natural, o que resulta em um contraste maior entre a unidade V-V acentuada frasalmente e as unidades seguintes, pondo em relevo a força relativa do acento.

2 Objetivos

O objetivo do trabalho é estimar o grau de variabilidade do ritmo da fala entre falantes do português brasileiro (PB) a partir da estimativa dos valores α, β e w_0. A interpretação linguística desses parâmetros e a maneira como se relacionam com a questão da tipologia rítmica é explicada em detalhe na seção 1.2.

3 Materiais e métodos

3.1 Material de fala

O *corpus* de material de fala constitui-se da leitura da passagem "A menina do narizinho arrebitado", do escritor Monteiro Lobato. A leitura foi feita por oito locutores (cinco do sexo masculino), falantes nativos do PB, com idades em torno de 20 anos. A duração média das leituras é 33,5 segundos, com um mínimo de 21 s e máximo de 54 s.

Cada falante leu o texto designado em três níveis de taxa de elocução, começando por uma taxa normal, aquela habitual ou confortável para o falante, seguida de uma leitura em uma taxa mais rápida e outra em taxa mais lenta, ambas em relação ao habitual de cada falante.

3.2 Análise fonética

Uma etapa prévia à estimação dos parâmetros do modelo é a extração dos grupos acentuais a partir da duração acústica bruta de segmentos agrupados em unidades V-V para cada amostra de áudio. Essa extração é o produto final da aplicação de três procedimentos: normalização da duração bruta por *z-score* estendido, suavização do contorno de duração normalizado por meio da aplicação de uma função de média móvel de cinco pontos e detecção de picos no contorno de duração suavizada. A aplicação desses procedimentos a cada amostra de áudio do *corpus* forneceu a entrada para a fase seguinte, na forma da localização das fronteiras de grupo acentual, sua magnitude relativa e o cômputo do número de unidades V-V por grupo acentual.

3.3 Estimação dos parâmetros do modelo

A etapa descrita na seção anterior forneceu os dados necessários para a geração de contornos simulados com a mesma estrutura do contorno natural, gerados pelo modelo dinâmico do ritmo. A geração dos contornos simulados foi feita por uma implementação na linguagem R do algoritmo descrito na seção1.1.

O procedimento de otimização da extração dos parâmetros consistiu da geração de contornos simulados a partir de triplas de valores para os parâmetros α, β e w_0 que varreram todas as combinações possíveis dentro de uma gama de valores: entre 0,1 e 3 em passos de 0,1 para α e β e entre 0,05 e 1 em passos de 0,05 para w_0. Um contorno simulado será gerado para cada tripla de valores dos três parâmetros. Assumimos como valor de T_0 a média das unidades V-V não acentuadas frasalmente de cada contorno.

A combinação ótima de parâmetros foi aquela que minimizou a distância (ou erro) entre o contorno normalizado natural e cada um dos contornos simulados. Testamos quatro métricas comuns na literatura sobre otimização para medir a distância entre os contornos naturais e simulados: erro quadrático médio (*root mean squared error - RMSE*), soma dos quadrados dos resíduos (*sum of squared residuals - SSR*), erro absoluto médio (*mean absolute error - MAE*) e *dynamic time warping (DTW)*.

4 Resultados preliminares

Uma vez que adotamos quatro diferentes medidas de erro, começamos comparando o efeito do uso dessas medidas sobre a variabilidade das estimativas dos três parâmetros. A medida SSR gerou o menor coeficiente de variação para os três parâmetros, agrupando todos os falantes e os três níveis de taxa de elocução no cálculo. Em virtude da menor variabilidade, reportamos a seguir outros resultados, sempre considerando SSR como medida de minimização do erro.

A tabela 1 mostra o valor mediano e desvio mediano absoluto (MAD) dos parâmetros em função dos níveis da taxa de elocução. O parâmetro α tende a diminuir conforme a taxa fica mais rápida, indicando que a fala tende ao ritmo silábico conforme isso acontece; w_0 adota tendência contrária, aumentando na taxa rápida, possivelmente para manter as unidades V-V acentuadas frasalmente relativamente mais alongadas para compensar a diminuição em α; β fica estável nos extremos, com ligeira queda na taxa normal.

Tabela 1: Mediana e MAD dos parâmetros do modelo em função dos três níveis da taxa de elocução.

	lenta	normal	rápida
α	1,08 (0,15)	0,92 (0,37)	0,75 (0,19)
β	1,00 (0,11)	0,88 (0,15)	1,00 (0,11)
w_0	0,46 (0,22)	0,42 (0,20)	0,56 (0,09)

Considerando todos os falantes e todas os níveis da taxa de elocução, a estimativa do parâmetro β é a que apresenta a menor variabilidade, com coeficiente de variação de 24,7%, seguido dos parâmetros α, com 31,7% e w_0, com 41,5%. Uma análise do efeito dos falantes sobre a estimativa dos parâmetros indica que há uma grande variabilidade entre os 8 falantes, que pode ser visualizada na figura 1. Para α, o coeficiente de variação mediano é 27,6%; para β, 11,9% e para w_0, 38,2%. Na figura 1, é possível ver que

para alguns dos falantes, como *fa*, *ja* e *lr*, a estimativa de w_0 dos três níveis é muito próxima, mas em geral a dispersão é grande. A variabilidade alta de w_0 é um resultado inesperado, à medida em que Barbosa (2006) considera que esse parâmetro deve ser relativamente invariável para um mesmo falante e uma mesma comunidade linguística. Mantendo a interpretação de w_0 proposta por Barbosa, o resultado apresentado aqui poderia ser explicado por uma inadequação da função de sincronismo do modelo ou ser um artefato do procedimento de otimização adotado aqui. Uma variação do procedimento que poderia ser testada é manter o valor de w_0 fixo e reanalisar a variação dos demais parâmetros.

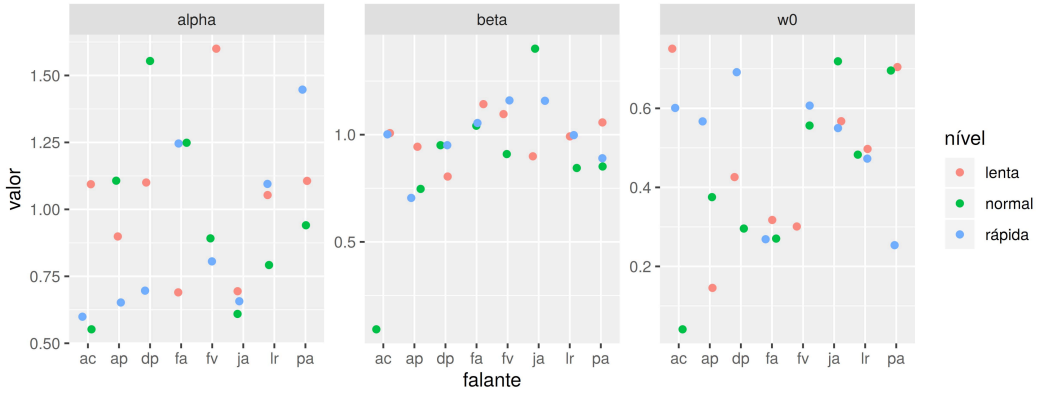

Figura 1: Valores dos três parâmetros para cada falante em função do nível da taxa

REFERÊNCIAS

1. BARBOSA, P. A. Incursões em torno do ritmo da fala. Campinas: Pontes, 2006.

2. BARBOSA, P. A. Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. Speech Prosody 2002. p. 163–166.

3. BARBOSA, P. A. “Syllable-timing in brazilian portuguese”: uma crítica a Roy Major.