

**RELACIONANDO A PRODUÇÃO E A PERCEPÇÃO DO RITMO DO
PORTUGUÊS BRASILEIRO POR MEIO DE UMA TÉCNICA SEMIAUTOMÁTICA**

Wellington da SILVA
Orientador: Plínio Almeida Barbosa

RESUMO: No estudo apresentado, foi investigada a relação entre as medidas de produção do ritmo do português brasileiro apontadas por três métodos de caracterização do ritmo da fala e sua percepção por uma bateria de ouvintes, através de um teste de discriminação com fala deslexicalizada. O método avaliado na produção foi uma técnica semiautomática fundamentada na teoria de osciladores acoplados de Barbosa (2006). Os outros métodos utilizados foram o índice de variabilidade pareada e o comprimento resultante em estatística circular. O método de osciladores acoplados foi capaz de captar diferenças no ritmo de fala entre os locutores, entre os dois estilos (leitura e narração) e também ao longo das gravações. Entretanto, pesquisas adicionais são necessárias para melhor se investigar a relação entre produção e percepção do ritmo da fala.

Palavras-chave: Fonética; Prosódia da Fala; Ritmo; Teste de Percepção.

INTRODUÇÃO

Diferenças no modo de falar das pessoas são facilmente notáveis, seja entre falantes de línguas distintas, entre estilos de fala diferentes, entre dialetos de uma mesma língua ou até mesmo entre falantes de uma mesma variante linguística. Um dos fatores que contribuem para essa diferença é o ritmo da fala, e é esse fator que será abordado no presente trabalho¹.

O ritmo da fala é estudado pela prosódia da fala, ramo da fonética que se ocupa das características suprasegmentais da fala. Entende-se por característica suprasegmental todo fenômeno da fala que acompanha mais de um fone (segmento) do enunciado, como entonação, acentuação e ritmo, por exemplo. Desde a antiguidade estudiosos da comunicação humana têm voltado seus olhares para o ritmo da fala (como Cícero, em sua obra “*De Oratore*”), mas foi somente nas últimas décadas que resultados mais expressivos foram alcançados, com o desenvolvimento de vários métodos para caracterizar o ritmo das línguas naturais através de medidas acústicas.

¹ Este trabalho é resultado da minha Iniciação Científica, intitulada “Caracterização automática da tipologia rítmica do português brasileiro”, que foi orientada pelo prof. Dr. Plínio A. Barbosa e realizada com bolsa concedida pela Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, Processo 2010/01273-2, no período de 01/04/2010 a 31/03/2011.

1 De acordo com o psicólogo Fraisse (1968), todo ritmo apresenta dois componentes: a periodicidade e a estruturação. No caso do ritmo da fala, a periodicidade seria a regularidade de ocorrência de uma unidade do tamanho da sílaba. A estruturação, por outro lado, seria a organização dessas unidades em estruturas (Barbosa, 2006). Como será discutido neste artigo, nem todos os métodos de caracterização do ritmo da fala consideram esses dois componentes, uma falha que impede o conhecimento do funcionamento do ritmo nas línguas naturais e sua função na fala (Kohler, 2009).

Houve um período no qual o ritmo era entendido como um fenômeno de regularidade temporal estrita (isocronia) e as línguas classificadas tipologicamente em ritmo acentual ou em ritmo silábico. Uma língua teria ritmo acentual se a duração dos grupos acentuais (formado por uma sílaba proeminente no enunciado e todas as átonas que a seguem, em línguas com cabeça à esquerda, ou que a precedem, em línguas com cabeça à direita) fosse a mesma (isocronismo acentual) e ritmo silábico se a duração das sílabas fosse a mesma (isocronismo silábico), vide Barbosa (2000) para uma revisão. Porém, tal isocronia absoluta nunca foi atestada empiricamente e o ritmo passou a ser visto como um ponto entre os dois extremos ideais: acentual e silábico, com a coexistência de ambos os tipos e a predominância de um deles (Barbosa, 2009), sendo a isocronia entendida como um fenômeno da percepção da fala (Lehiste, 1977).

No estudo que será apresentado neste trabalho, foi investigada a relação entre as medidas de produção do ritmo do português brasileiro apontadas por três métodos de caracterização do ritmo da fala e a sua percepção por uma bateria de ouvintes.

TRÊS MÉTODOS DE CARACTERIZAÇÃO DO RITMO DA FALA

Método fundamentado no modelo de osciladores acoplados

A teoria de osciladores acoplados foi aplicada por Barbosa (2006) para a caracterização do ritmo da fala, no seio de um programa de pesquisa que integra pesquisadores como Cummins, Port e O'Dell. De acordo com essa teoria, o ritmo das línguas naturais é determinado pelo acoplamento de dois osciladores: acentual e silábico, existentes no sistema cognitivo do falante. A força de acoplamento entre esses dois osciladores é o que indica se a língua tende ao ritmo acentual ou ao silábico. Os pulsos do oscilador silábico determinam os *onsets* vocálicos do enunciado e os pulsos do oscilador acentual, os locais de acento frasal.

Empiricamente atesta-se que a duração do grupo acentual tem uma componente linear que é função do número de sílabas nele contidas e, por isso, é possível obter uma regressão linear com essas duas variáveis. O'Dell e Nieminen (1999) mostraram, através da técnica matemática *Averaged Phase Difference*, que a força de acoplamento (c) entre os dois osciladores pode ser estimada através da razão entre o coeficiente de interceptação da reta de regressão com o eixo das ordenadas (a) e o coeficiente de inclinação da reta (b) da equação (1), em que I é a duração do grupo acentual e n o número de sílabas fonéticas ou fonológicas do grupo acentual correspondente ($c = a/b$). Se a força de acoplamento (c) for igual a 1, significa que ambos os osciladores se influenciam de maneira equivalente. Se $0 < c < 1$, tem-se uma dominância do oscilador silábico sobre o

acentual (tendência ao ritmo silábico), e se $c > 1$, o oscilador acentual domina o silábico (tendência ao ritmo acentual).

$$I = a + b \cdot n \quad (1)$$

Essa é a teoria adotada para este trabalho, pois ela é capaz de dar conta da estrutura hierárquica do ritmo da fala (Cummins e Port, 1998), além de considerar os dois componentes propostos por Fraisse (1968), a periodicidade e a estruturação.

O Índice de Variabilidade Pareada (PVI: *Pairwise Variability Index*)

Desenvolvido por Low *et al.* (2000), o índice de variabilidade pareada mostra, em termos de duração, o quanto uma unidade linguística silábica ou subsilábica difere da unidade imediatamente seguinte. É computado de acordo com a fórmula (2), em que d_k corresponde à duração bruta da unidade V-V (unidade delimitada por dois *onsets* de vogais consecutivos, incluindo a vogal à esquerda) e m ao número total de unidades VV.

As unidades V-V em posição de fronteira de grupo acentual não foram consideradas para o cálculo. Um PVI alto reflete grandes irregularidades de unidade a unidade (predominância de ritmo acentual) e um PVI baixo, regularidade (predominância de ritmo silábico).

$$nPVI_{vv} = 100 \times \left[\frac{\sum_{k=1}^{m-1} |d_k - d_{k+1}|}{(\sum_{k=1}^{m-1} (d_k + d_{k+1})/2)} / (m-1) \right] \quad (2)$$

Note que essa fórmula calcula a diferença do par levando em conta a média do mesmo. Isso é feito para atenuar os efeitos causados por diferentes taxas de elocução, os quais causam um aumento na diferença entre as unidades e leva a uma falsa impressão de aumento na irregularidade das durações. Por isso, ela é conhecida como PVI normalizado.

Esse método, entretanto, não considera o papel das proeminências (i.e. acentos frasais), elementos muito importantes na composição do ritmo das línguas naturais, pois, como diz Kohler (2009, p.34), “*we are dealing with data sorting on the basis of consonantal and vocalic, i.e. local segmental, durations in their point-to-point variability, not with global rhythmical patterns.*”

Caracterizando o ritmo da fala por estatística circular

A estatística circular é utilizada em vários ramos da ciência, como nas ciências da terra, meteorologia, biologia, física, psicologia, medicina, entre outras (Mardia e Jupp, 2000). Para a análise do ritmo das línguas naturais, recentemente Brady e Port (2007) usaram a estatística circular em um estudo sobre o ritmo do japonês.

Um dado circular consiste em um ponto em um círculo unitário ou um vetor unitário em um plano. Uma vez escolhidas uma direção inicial e uma orientação no círculo, cada observação circular pode ser especificada pelo ângulo formado entre o ponto zero e o ponto correspondente à observação (Mardia e Jupp, *ibid.*).

Para plotar as durações das unidades V-V no círculo, inicialmente é calculada a razão entre a duração de cada unidade V-V (d_{vv}) e a média das durações das unidades VV ($meand_{vv}$). Então o ângulo obtido é convertido em radianos, multiplicando-o por 2π .

O objetivo é chegar ao valor do comprimento resultante médio, \bar{R} . O comprimento resultante, R , é a soma dos vetores correspondentes aos pontos do círculo, calculado de acordo com a fórmula (3) e o seu valor médio é obtido ao dividi-lo pelo número total da amostra.

Pode-se deduzir que, se os pontos do círculo forem mais agrupados, seus vetores correspondentes terão aproximadamente a mesma direção e o comprimento resultante médio será próximo de 1, que é o valor normalizado da principal direção resultante.

Esse fato indica uma predominância de *ritmo silábico*, já que a característica dessa classe rítmica é de sílabas idealmente com a mesma duração. Se, por outro lado, os pontos do círculo estiverem muito dispersos, serão obtidos vetores correspondentes em várias direções, e, portanto, \bar{R} será relativamente pequeno, o que indica predominância de *ritmo acentual*.

$$R^2 = \left[\sum \text{sen} \left(\frac{d_{vv}}{meand_{vv}} \times 2\pi \right) \right]^2 + \left[\sum \text{cos} \left(\frac{d_{vv}}{meand_{vv}} \times 2\pi \right) \right]^2 \quad (3)$$

METODOLOGIA

Para esse estudo, três locutores paulistas, duas mulheres e um homem, pós-graduandos na faixa de 30 a 45 anos de idade sem problemas auditivos ou fonoarticulatórios, leram o texto “O monge desastrado”, com cerca de 1.500 palavras, (estilo leitura) e logo em seguida contaram sobre o que ele se tratava (estilo narração; semiespontâneo). As seis gravações foram segmentadas de forma automática em unidades V-V com um *script* para o programa PRAAT (Boersma e Weenink, 2010), o *Beatextractor* (Barbosa, 2006), seguido de correção manual. A escolha da unidade V-V se justifica pelas pesquisas do *perceptual-center* como ponto de ancoragem universal para a percepção de isocronismo, o qual se dá no *onset* de uma vogal (cf. Barbosa, 2000; Barbosa *et al.* 2005). Além disso, Barbosa (1996) mostrou que, no português brasileiro, a evolução das durações das unidades V-V sucessivas em frases isoladas indica pontos de culminância de crescendos duracionais que delimitam os grupos acentuais, fato não observável sistematicamente com as sílabas.

As unidades foram, então, etiquetadas manualmente. Em seguida, os grupos acentuais foram delimitados de forma automática, com o auxílio do *script* SGdetector (disponível em Barbosa, 2006, p. 459-468), que detecta as fronteiras dos grupos acentuais a partir das durações normalizadas e suavizadas das unidades V-V e fornece uma tabela com a duração desse grupo e o número de unidades V-V nele contidas, além de uma estimativa da força da fronteira que delimita o grupo à direita. O número de sílabas fonológicas contidas em cada grupo acentual foi informado após contagem manual, pois estávamos interessados em comparar os resultados obtidos com as unidades V-V e com as sílabas, para o método fundamentado na teoria de osciladores acoplados.

Com essas informações, foram construídos os gráficos para as regressões lineares e realizados os cálculos para o método PVI e o método fundamentado em estatística circular.

Assim como em Barbosa *et al.* (2009), optou-se por considerar o nível de proeminência de cada grupo acentual para explicar a variância da duração deste. Este nível de proeminência corresponde a uma medida normalizada (*z-score*) da duração da última unidade V-V do grupo acentual, a que carrega o acento frasal. A adição dessa variável resulta na regressão linear múltipla representada pela equação (4), em que z_p representa esse nível de proeminência, sendo os demais coeficientes os mesmos da equação (1). A força de acoplamento (c) continua sendo determinada por a/b . Essa variável, juntamente com o número de sílabas (fonológicas ou unidades V-V), explica a maior porcentagem da variância da duração do grupo acentual. Aplicando esse método, obtemos um coeficiente de determinação (r^2) em torno de 70 %, isto é, 70 % da variância da duração dos grupos acentuais provêm da variância do número de unidades (V-V ou sílaba) no mesmo e da proeminência expressa pela duração normalizada.

$$I = a + b. n + p. z_p \quad (4)$$

ANÁLISES ESTATÍSTICAS

Feitas as regressões lineares, foi realizado o cálculo para a identificação de *outliers* (pontos espúrios). O cálculo se baseia no *z-score*, uma medida estatística que mostra o afastamento de um ponto em relação à média, em unidades de desvios-padrão. É considerado *outlier* aquele ponto que tem um *z-score* maior ou igual a 3 (3 desvios-padrão afastado da média), pois tem menos de 1% de probabilidade de ocorrer. O motivo experimental que levou alguns grupos acentuais a serem considerados *outliers* quase sempre foi uma pausa anormal, sem explicação clara por conta da condição de narração ou leitura.

As gravações foram divididas em trechos, a fim de constatar mudanças no tipo rítmico dos sujeitos. Para fazer tal constatação, foram realizados testes de Análise de Covariância (ANCOVA) com todos os trechos para todas as gravações. Esses testes foram feitos com o programa R (R Development Core Team, 2010).

O critério adotado para a divisão dos trechos foi que cada um contivesse, no mínimo, 100 grupos acentuais, a fim de garantir um número suficiente de dados para os testes estatísticos. A leitura da locutora AG teve o quarto trecho dividido em dois, pois estávamos interessados em saber se a mudança do estilo do texto – de narrativo para explicativo – causaria uma mudança no ritmo, o que não aconteceu considerando a influência de z : a força de acoplamento c passou de 0,79 para 0,89, uma diferença rítmica não muito grande, apenas um leve aumento de ritmo acentual (vide tabelas 1 e 2). A narração do locutor FA, por ser muito curta, não foi dividida.

Para realizar a ANCOVA, primeiramente consideramos todos os coeficientes de inclinação (b e p) dos trechos. Quando os resultados para a influência do trecho considerando as duas inclinações não eram significativos, desconsideramos o efeito de trecho (inclinação indistinta do trecho 1) e refizemos a ANCOVA. No caso de intercepção de trecho não significativa, também foi considerada indistinta do trecho 1.

No caso de o coeficiente de intercepção (a), ter valor não significativo, aceitamos a hipótese nula e consideramos o mesmo como nulo (ritmo silábico).

Resultados

As tabelas 1 e 2 trazem as equações das regressões lineares para as sílabas fonológicas e unidades V-V, respectivamente, bem como as respectivas forças de acoplamento. Os locutores estão identificados pelas duas primeiras letras, sendo que a terceira corresponde ao sexo (F: feminino e M: masculino). Os estilos são identificados pelas siglas “Le” (leitura) e “Na” (narração; semiespontâneo). O número após a barra indica o trecho correspondente. São indicados apenas os valores de significância correspondentes ao coeficiente de intercepção, uma vez que os coeficientes de inclinação, quando diferentes de zero, são todos significativos.

A tabela 3 mostra os valores obtidos para os mesmos trechos com o índice de variabilidade pareada normalizado (nPVI) e com o método fundamentado em estatística circular (\bar{R} representa o comprimento resultante médio). Os locutores, trechos e estilos estão indicados da mesma maneira que nas tabelas 1 e 2.

TABELA 1: Equações das regressões lineares obtidas para as *sílabas fonológicas*. I corresponde à duração do grupo acentual, n ao respectivo número de sílabas fonológicas e z ao valor de z -score suavizado da respectiva proeminência. p corresponde à significância do coeficiente de intercepção (a) e ns indica não significativo. Coeficientes c significativos em itálico.

Locutor/Trecho	Regressão Linear	c
FAM – Le / 1	$I = 189 + 126n + 71z$ ($p < 0,04$)	<i>1,5</i>
FAM – Le / 2	$I = 0 + 166n + 50z$ (ns)	0
FAM – Le / 3	$I = 0 + 150n + 57z$ (ns)	0
FAM – Na	$I = 237 + 143n + 45z$ ($p < 0,07$)	<i>1,7</i>
AGF – Le / 1	$I = 0 + 170n + 59z$ (ns)	0
AGF – Le / 2	$I = 90 + 170n + 46z$ ($p < 0,05$)	<i>0,53</i>
AGF – Le / 3	$I = 196 + 170n + 41z$ ($p < 0,0004$)	<i>1,15</i>
AGF – Le / 4	$I = 135 + 170n + 42z$ ($p < 0,02$)	<i>0,79</i>
AGF – Le / 5	$I = 151 + 170n + 32z$ ($p < 0,05$)	<i>0,89</i>
AGF – Na / 1	$I = 366 + 137n + 43z$ ($p < 0,01$)	<i>2,7</i>
AGF – Na / 2	$I = -15 + 197n + 43z$ ($p < 0,04$)	0
AGF – Na / 3	$I = 0 + 137n + 43z$ (ns)	0
LLF – Le / 1	$I = 212 + 126n + 68z$ ($p < 0,03$)	<i>1,68</i>
LLF – Le / 2	$I = 0 + 161n + 41z$ (ns)	0
LLF – Le / 3	$I = -31 + 167n + 56z$ ($p < 0,02$)	0
LLF – Na / 1	$I = 0 + 184n + 47z$ (ns)	0
LLF – Na / 2	$I = 0 + 135n + 47z$ (ns)	0

TABELA 2: Equações das regressões lineares obtidas para as *unidades V-V*. *I* corresponde à duração do grupo acentual, *n* ao respectivo número de unidades V-V e *z* ao valor de *z-score* suavizado da respectiva proeminência. *p* corresponde à significância do coeficiente de interceptação (*a*) e ns indica não significativo. Coeficientes *c* significativos em itálico.

Locutor/Trecho	Regressão Linear	<i>c</i>
FAM – Le / 1	$I = 0 + 180n + 69z$ (ns)	0
FAM – Le / 2	$I = 224 + 180n + 50z$ ($p = 0,05$)	<i>1,24</i>
FAM – Le / 3	$I = 0 + 180n + 55z$ (ns)	0
FAM – Na	$I = 262 + 175n + 42z$ ($p = 0,05$)	<i>1,5</i>
AGF – Le / 1	$I = 0 + 223n + 42z$ (ns)	0
AGF – Le / 2	$I = 15 + 223n + 42z$ ($p < 0,1$)	<i>0,07</i>
AGF – Le / 3	$I = 79 + 223n + 42z$ ($p < 0,005$)	<i>0,35</i>
AGF – Le / 4	$I = 43 + 223n + 42z$ ($p < 0,05$)	<i>0,19</i>
AGF – Le / 5	$I = 0 + 223n + 42z$ (ns)	0
AGF – Na / 1	$I = 0 + 194n + 40z$ (ns)	0
AGF – Na / 2	$I = 0 + 266n + 40z$ (ns)	0
AGF – Na / 3	$I = 0 + 194n + 40z$ (ns)	0
LLF – Le / 1	$I = 0 + 190n + 59z$ (ns)	0
LLF – Le / 2	$I = 188 + 190n + 36z$ ($p < 0,05$)	<i>0,99</i>
LLF – Le / 3	$I = 0 + 190n + 59z$ (ns)	0
LLF – Na / 1	$I = -258 + 284n + 43z$ ($p < 0,1$)	0
LLF – Na / 2	$I = 305 + 185n + 43z$ ($p < 0,02$)	<i>1,65</i>

TABELA 3: Valores obtidos com o índice de variabilidade pareada normalizado (nPVI) e com o método fundamentado em estatística circular. Vide texto para mais informações.

Locutor/Trecho	nPVI	\bar{R}
AM – Le / 1	55,09	0,102
FAM – Le / 2	54,94	0,146
FAM – Le / 3	54,83	0,132
FAM – Na	55,39	0,098
AGF – Le / 1	45,75	0,178
AGF – Le / 2	50,45	0,152
AGF – Le / 3	50,41	0,191
AGF – Le / 4	53,14	0,183
AGF – Le / 5	52,10	0,186
AGF – Na / 1	53,37	0,160
AGF – Na / 2	53,86	0,235
AGF – Na / 3	53,52	0,204
LLF – Le / 1	50,47	0,135
LLF – Le / 2	48,82	0,169
LLF – Le / 3	48,00	0,157
LLF – Na / 1	55,46	0,185
LLF – Na / 2	57,62	0,172

O TESTE DE PERCEPÇÃO

Com o propósito de investigar se as pessoas são capazes de identificar as diferenças de ritmo assinaladas por esses métodos, elaboramos e aplicamos um teste de discriminação, inédito na literatura. O teste consistiu em apresentar aos sujeitos pares de trechos de fala deslexicalizados, extraídos das gravações utilizadas nas análises anteriores. O método usado para retirar a informação lexical dos enunciados foi o proposto por Vainio *et al.* (2009). O uso de fala deslexicalizada é importante por garantir que os ouvintes se concentrem apenas na prosódia dos enunciados.

Solicitou-se a eles que julgassem se o modo de falar nos dois trechos do par apresentado era semelhante ou diferente, selecionando um valor em uma escala graduada de 0 a 6. O valor 0 indicava modos de falar idênticos e o valor 6, modos de falar diferentes e era possível selecionar valores intermediários. Quando a pessoa clicava com o mouse no valor desejado, o próximo estímulo já era apresentado. Era permitido ao sujeito ouvir cada estímulo apenas mais uma vez, caso desejasse, clicando em um botão apresentado na parte inferior da tela.

O experimento foi montado com a ferramenta *ExperimentMFC*, através de um *script* para o programa PRAAT. Foram selecionados 36 trechos de 1 e de 2 segundos, pois também era de nosso interesse investigar se a duração dos trechos influenciaria nas respostas dos ouvintes. Cada estímulo (par) foi usado duas vezes, mas em ordem inversa, por exemplo: “AGreadA1,FAreadA1” e “FAreadA1,AGreadA1”. Isso foi feito para que se testasse a coerência dos julgamentos dos sujeitos. No total, 76 estímulos compuseram o experimento. As respostas dos sujeitos foram convertidas em uma escala de -1 (mesmo modo) a 1 (modos diferentes), com os valores 0, -0,67, -0,33, 0, 0,33, 0,67 e 1. O valor -1 foi posteriormente transformado em 0, por indicar “mesmo modo”, aplicando-se a mesma equação para todas as respostas.

Análises

O teste foi aplicado em 21 sujeitos, sendo 6 homens e 15 mulheres, de idade entre 18 e 25 anos e sem problemas auditivos ou fonoarticulatórios, por meio de um computador e com uso de fones de ouvido.

Das respostas obtidas, foram selecionadas aquelas que tiveram desvio-padrão menor que a “média dos desvios-padrão menos a metade de um desvio-padrão dos desvios-padrão”, pois são as respostas menos variáveis entre os sujeitos. Depois, verificamos a consistência das respostas dos ouvintes nesses estímulos. Esperava-se que as respostas não variassem muito quando a ordem dos pares de trechos era invertida. Tal verificação foi feita a partir da diferença entre a média das respostas para os dois estímulos (o original e o de ordem invertida). Foram descartados 2 estímulos que apresentaram essa diferença maior que o critério descrito acima. Ao todo, foram selecionados 20 estímulos para a análise posterior.

Em seguida, calculamos as diferenças dos valores obtidos com os métodos (*c*, PVI e R) entre os trechos de cada par, por exemplo: para o par “AGreadA1,FAreadA1”, tomamos o valor do *c* da parte 1 da leitura de AG e subtraímos o valor do *c* da parte 1 da leitura de FA, e assim por diante, para os três métodos (assim, se ambos tivessem um mesmo valor de *c* a diferença seria 0, indicando ritmos de fala idênticos, segundo essa técnica). Então construímos gráficos de dispersão entre essas diferenças e a média das respostas dos sujeitos.

Resultados

Para o método fundamentado na teoria de osciladores acoplados, quando consideramos as regressões lineares múltiplas, não foram alcançadas correlações satisfatórias para os gráficos ($r^2 = 0,002$ para as unidades V-V e $r^2 = 0,02$ para as sílabas fonológicas). A razão disso se deve ao grande número de coeficientes não significativos obtidos. Por isso, optamos por utilizar as regressões lineares simples (calculadas sem levar em consideração o *z-score* da proeminência dos grupos acentuais como variável), a fim de investigar se dessa forma esse método consegue pelo menos prever parte da percepção dos sujeitos. Foi obtido um resultado interessante: foi alcançado um coeficiente de determinação maior quando consideradas as unidades V-V ($r^2 = 0,18$), em comparação às sílabas fonológicas ($r^2 = 0,004$). O resultado obtido para as unidades VV também foi estatisticamente significativo: $p < 10^{-8}$ para o coeficiente de interceptação e $p < 0,06$ para o de inclinação.

Para o índice de variabilidade pareada (PVI) foi obtido um coeficiente de determinação de 0,12. No entanto, essa correlação foi significativa somente para o coeficiente de interceptação da reta: $p < 10^{-5}$.

Para o método fundamentado em estatística circular não foi obtido um coeficiente de determinação relevante ($r^2 = 0,08$).

Um teste T ($t = 0,55$; $\nu = 35$; $p \text{ H}^0 0,59$) mostrou que a duração dos trechos não influenciou na coerência das respostas dos sujeitos, ou seja, a diferença entre a média das respostas para os estímulos e seus equivalentes invertidos com a duração de 1 segundo foi estatisticamente a mesma com a duração de 2 segundos.

CONSIDERAÇÕES FINAIS

O uso das unidades V-V permite a automatização da técnica fundamentada na teoria de osciladores acoplados. No entanto, em alguns casos o uso das sílabas fonológicas se mostrou mais eficaz, resultando em um maior número de coeficientes significativos. Ao relacionar as medidas de produção com a percepção dos sujeitos, o uso das unidades V-V alcançou resultados mais expressivos, pelo menos quando consideradas as regressões lineares simples. O número de coeficientes não significativos foi grande, fato que pede uma investigação mais apurada, com maior número de dados. De fato, o número de dados necessários para se obter resultados significativos é uma das limitações desse modelo. Contudo, a possibilidade de ser submetido a uma análise estatística não deixa de ser uma vantagem.

O uso do nível de proeminência dos grupos acentuais para explicar a variância da duração destes juntamente com o número de unidades linguísticas contidas neles melhorou consideravelmente a capacidade explicativa do modelo.

O método de osciladores acoplados foi capaz de captar diferenças no ritmo de fala entre os locutores, entre os dois estilos (leitura e narração) e também ao longo das gravações.

É possível que a escala de 7 pontos utilizada no teste de discriminação exija dos sujeitos um grau de percepção do ritmo da fala bastante apurado, o que possivelmente eles não têm. Por isso, seria interessante utilizar em futuros experimentos uma escala com apenas 5 pontos, com o propósito de testar essa hipótese. Seria interessante também utilizar trechos com duração maior do que 2 segundos, pois talvez isso permita aos juízes avaliar o ritmo dos enunciados mais apropriadamente.

As correlações que explicam baixo índice de variância, obtidas entre as três medidas de produção do ritmo da fala e as respostas dos juízes no teste de discriminação, sugerem que estes usam outras pistas acústicas para avaliar diferenças no ritmo da fala. Portanto, esses métodos são capazes de descrever o ritmo da fala, mas não de explicar o julgamento dos sujeitos em um teste de discriminação.

BIBLIOGRAFIA

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

- BARBOSA, P. A. "Syllable-Timing in Brazilian Portuguese: uma Crítica a Roy Major". **DELTA**, vol. 16, n. 2, p. 369-402, 2000.
- BARBOSA, P. A. **Incursões em torno do ritmo da fala**. Campinas: Pontes/Fapesp, 2006.
- BARBOSA, P. A. "Measuring speech rhythm variation in a model-based framework." In Interspeech 2009 - Speech and Intelligence, Brighton. **Proc. of Interspeech 2009 - Speech and Intelligence**. Londres: Causal Productions, 2009, p. 1527 – 1530.
- BARBOSA, P. A., ARANTES, P., MEIRELES, A. R., VIEIRA, J. M. "Abstractness in Speech-Metronome Synchronisation: P-Centres as Cyclic Attractors." **Proceedings of the Ninth European Conference on Speech Communication and Technology (Interspeech 2005)**. Lisboa, Portugal, 2005, p. 1441-1444.
- BARBOSA, P. A.; VIANA, M. C.; TRANCOSO, I. "Cross-variety Rhythm Typology in Portuguese." In Interspeech 2009 - Speech and Intelligence, Brighton. **Proc. of Interspeech 2009 - Speech and Intelligence**. Londres: Causal Productions, 2009, p. 1011 – 1014.
- BOERSMA, P. e WEENINK, D. "**Praat: doing phonetics by computer (Versão 5.1.04)**". URL <<http://www.praat.org>>, acessado em março de 2010.
- BRADY, M. C. e PORT, R. F. "Quantifying Vowel Onset Periodicity in Japanese." **International Congress of Phonetic Sciences**. Saarbrücken, 2007, p. 337-342.
- CUMMINS, F. e PORT, R. "Rhythmic constraints on stress timing in English." **Journal of Phonetics**, 26, p. 145–171, 1998.
- FRAISSE, P. "Les Rythmes." **Journal Français d'Oto- Rhino-laryngologie**, S7, p. 23–33, 1968.
- KOHLER, K. J. "Rhythm in speech and language: a new research paradigm". **Phonetica**, 66, p. 29-45, 2009.
- LEHISTE, I. "Isochrony reconsidered". **Journal of Phonetics**. 5, p. 253-263, 1977.
- LOW, E. L., GRABE, E., NOLAN, F. "Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English". **Language and Speech**, 43, p. 377–401, 2000.
- MARDIA, K. V. e JUPP, P. E. "**Directional Statistics**". Chichester: Wiley, 2000.
- O'DELL, M. e NIEMINEN, T. "Coupled Oscillator Model of Speech Rhythm". In XIV International Congress of Phonetic Sciences, São Francisco. **Proc. XIVth ICPhS**. São Francisco, p. 1075-1078, 1999.
- R Development Core Team. "**R: A language and environment for statistical computing**". R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>. Acessado em novembro de 2010.
- VAINIO, M. *et al.* "New Method for Delexicalization and its Application to Prosodic Tagging for Text-to-Speech Synthesis". In Interspeech 2009 - Speech and Intelligence, Brighton. **Proc. of Interspeech 2009 - Speech and Intelligence**. Londres: Causal Productions, p. 1703 – 1706, 2009