Estudo piloto do efeito da estruturação temporal sobre a sincronização fala-metrônomo

Rafael Ésquines Cangiani; Pablo Arantes

Universidade Federal de São Carlos, Brasil

pabloarantes@gmail.com

Resumo

No presente trabalho apresentamos resultados de um experimento de sincronização entre a produção de participantes e estímulos sonoros externos cujo objetivo é mostrar que o desempenho na tarefa de sincronização é afetado pelo tipo de estruturação temporal dos estímulos. Foram geradas sequências de dez sílabas sintéticas /ba/ com intervalos entre os *onsets* vocálicos de quatro tipos: (i) isócronos, (ii) aleatórios, (iii) estruturados de acordo com um modelo de ritmo baseado em osciladores acoplados de forma a simular dois grupos acentuais e (iv) uma sequência isócrona com um alongamento súbito na quinta posição. A sincronia de fase entre os *onsets* da produção dos participantes e os *onsets* das sílabas sintéticas foi tomada como correlato do grau acoplamento entre a produção do participante e o estímulo externo. Os resultados iniciais indicam que as sequências do tipo (i) são as que mais favorecem a sincronia. As demais resultam em assincronia, embora os padrões difiram. Os observados nas de tipo (ii) sugerem que o desempenho é relativamente uniforme ao longo das dez sílabas da sequência, enquanto nas de tipo (iii) a sincronia é melhor nas proximidades do alongamento que corresponde à culminância do acento frasal. Os resultados favorecem a hipótese de que a organização temporal da produção dos participantes é afetada pela percepção da estruturação rítmica do estímulo externo. Se aprofundados em estudos posteriores, os resultados indicam que a relação entre produção e percepção do ritmo pode ser tratada de forma unificada por modelos inspirados em sistemas dinâmicos de osciladores acoplados.

Palavras-chave: Prosódia; Ritmo da Fala; Sincronização; Sistemas Dinâmicos.

1. Introdução

O objetivo deste trabalho é produzir evidências experimentais que confirmem a possibilidade de um entendimento integrado da produção e da percepção do ritmo da fala em um mesmo quadro teórico e metodológico. A inspiração para essa proposta de integração é a tentativa de aplicar os princípios da teoria dos sistemas dinâmicos ao estudo da cognição em geral e da fala em particular [1][2]. Uma das motivações para adotar uma teoria de sistemas dinâmicos como método para estudar a cognição é a possibilidade de tratar com o mesmo aparato formal aspectos do mundo que outras teorias tratariam como irreconciliáveis. A proposta para a integração da produção e da percepção do ritmo da fala é interessante do ponto de vista da parcimônia, uma vez que se a percepção puder ser vista como uma operação especular da produção, os mesmos recursos e princípios usados para explicar a produção podem ser usados também para explicar a percepção. A seção 1.1 apresenta o Modelo Dinâmico do Ritmo [3], que tomamos como base para entender a produção. Propomos que seja possível estendê-lo para modelar a percepção do ritmo. Um conceito importante para estabelecer essa ligação é a ideia de *resetting*, explicada a seguir. Um dos objetivos do experimento apresentado aqui é apresentar evidência para a ocorrência de *resetting* na percepção dos ouvintes. Nesse contexto, o uso da tarefa de sincronização nos pareceu interessante pois ela aciona o participante simultaneamente na dimensão da percepção e da produção.

1.1. Descrição do Modelo Dinâmico do Ritmo

A partir de estudos que mostram a necessidade de supor pelo menos dois níveis de organização temporal para explicar os padrões rítmicos do português, um que diz respeito a unidades do tamanho de sílaba e outro relacionado à recorrência de acentos no nível da frase, [3] elabora um modelo de produção do ritmo inspirado em um sistema de osciladores acoplados. O oscilador que representa a sequência de unidades do tamanho de sílabas num dado enunciado, chamado de oscilador silábico, é o oscilador a ser induzido; o outro é chamado de oscilador acentual, que representa a sequência de acentos frasais de um enunciado. A batida do oscilador silábico já induzido no acoplamento com o oscilador acentual interage com a pauta gestual, especificando o *onset* dos gestos vocálicos. O período do oscilador silábico induzido sofre a influência crescente do oscilador acentual, atingindo um valor culminante nas proximidades da batida do mesmo. Após a batida, ocorre o *resetting* do período, isto é, a duração do oscilador silábico tende a voltar ao que era antes da influência indutora do oscilador acentual, isto é, para seu período de repouso. O modelo, representado esquematicamente na Figura 1, gera padrões de duração semelhantes aos observados na produção de fala natural e reproduz com sucesso o comportamento da duração acústica observado em fenômenos linguísticos.

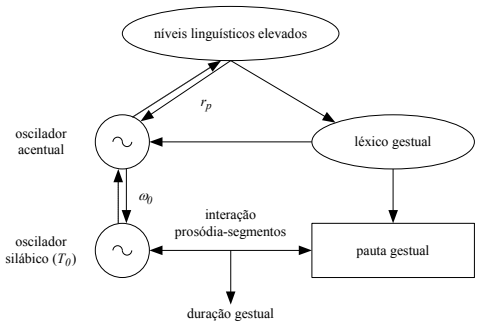

Figura 1: *Diagrama esquemático da proposta de arquitetura que integra produção e percepção em um modelo dinâmico tanto da produção quanto da percepção do ritmo*

No que tange à questão perceptual, propomos que o *resetting* de período é o evento que induz no ouvinte a percepção da ocorrência de um acento frasal ao marcar o a suspensão da influência do oscilador acentual sobre o silábico. Os resultados do teste de monitoramento de clique apresentados em [8] indicam que o nível de atenção do ouvinte aumenta ao longo do grupo acentual e apresenta uma piora local nas imediações da fronteira entre grupos acentuais, marcada pela ocorrência de um pico de duração no contorno de duração da frase. Pode-se interpretar este padrão como evidência para a ocorrência, na percepção, do processo de *resetting* de indução do oscilador silábico interno ao ouvinte pelo estímulo externo representado pela duração das unidades v-v produzidas pelo falante.

2. Objetivos

Relatamos neste trabalho a condução de um experimento piloto de sincronização fala-metrônomo na tradição daqueles realizados por Allen [11][12] e usados por Barbosa e colegas [13] para investigar o fenômeno do *Perceptual-center* ou *P-center* em português brasileiro. Nos experimentos pilotos de sincronização fala-metrônomo que realizamos, instruímos dois participantes a tentar sincronizar sua produção a um estímulo externo. Em experimentos desse tipo, a tendência dos falantes é fazer coincidir de forma aproximada os *onsets vocálicos* da sua produção com o estímulo externo, que muito comumente são batidas de um metrônomo. No experimento, verificamos qual é o comportamento da sincronia entre a fala do participante e o estímulo externo quando esses estímulos sonoros apresentam estruturação temporal semelhante à observada na fala. Nossa hipótese inicial é que a sincronia entre a fala e o estímulo externo deve ser afetada pela estruturação temporal desse estímulo. No caso em que a estrutura das durações for semelhante à produzida na fala natural, a sincronia deve aumentar ao longo do tempo e sofrer degradação em algum momento após o alongamento tipicamente associado à ocorrência de um acento frasal. Este resultado seria interpretado como evidência para a indução do oscilador interno ao ouvinte, que usaria esta informação para ajustar sua produção ao estímulo externo percebido. Comparamos o desempenho dos participantes quando sincronizaram sua produção com estímulos artificiais produzidos seguido as previsões do Modelo Dinâmico do Ritmo com outros três tipos de estruturação temporal não comparáveis à encontrada na fala, que serão descritos a seguir.

3. Procedimentos

3.1. Estímulos gerados para a tarefa de sincronização

Os estímulos externos com os quais os falantes estabeleciam sincronização foram sequências de sílabas [ba] sintetizadas. As sílabas sintéticas foram geradas por meio da função "Create KlattGrid" do Praat [14]. Os valores de F1, F2, F3, e F4 passados para a função foram 7500 Hz, 1250 Hz, 2300 Hz e 3600 Hz. Os valores de largura de banda para os mesmos formantes foram 50 Hz, 100 Hz, 200 Hz e 300Hz. Especificou-se um intervalo de 25 ms para cada transição entre o "locus" de F1 e F2 e os valores estáveis dos formantes. A intensidade da porção central é de aproximadamente 83 dB e o contorno de F_0 é descendente, começando em 140 Hz e caindo 3 semitons ao final da vogal. Para atenuar a amplitude no início e no final da vogal, a sílaba é filtrada por uma janela Hann. O espectograma e o oscilograma de duas sílabas sintetizadas segundo o procedimento descrito são mostrados na Figura 2.

Geramos sequências de dez sílabas sintetizadas nas quais a duração entre os *onsets* vocálicos seguiam quatro diferentes padrões:

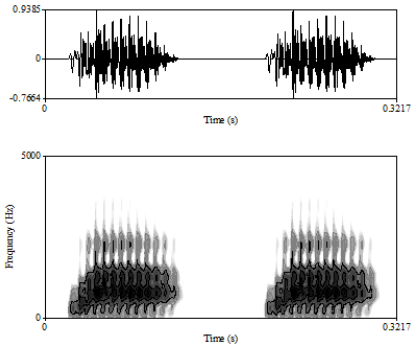

Figura 2: *O oscilograma e o espectrograma de duas sílabas sintetizadas segundo as especificações já descritas nesta seção.*

Isócrona: Durações dos intervalos entre os estímulos são constantes e correspondentes ao valor médio das unidades v-v do falante.

Modelo: Durações dos intervalos seguem as geradas a partir do modelo dinâmico do ritmo, baseado em osciladores acoplados. Um *script* do Praat que implementa computacionalmente o modelo gerou durações simuladas de unidades v-v correspondente a dois grupos acentuais. O valor do período de repouso, em princípio, deve ser estimado separadamente para cada participante, uma vez que pode ser considerada uma característica predominantemente individual. No experimento piloto escolhemos o valor de 180 ms a partir de dados a respeito da duração de unidades v-v [15].

Súbita: A sequência de durações é semelhante à de um grupo acentual, mas com uma estrutura simplificada, isto é, uma sequência de intervalos isócronos seguidos de um intervalo subitamente alongado, ou seja, está ausente o aumento gradiente da duração das unidades v-v típicas da fala natural gerada pelo modelo.

Aleatória: Durações dos intervalos entre os estímulos foram escolhidas aleatoriamente dentro de intervalos especificados. O limite inferior e superior dos intervalos é próximo respectivamente ao menor e ao maior valor definidos

na condição Modelo, a saber [180 ms até 280 ms]. Uma função do Praat faz a seleção dos dez valores a partir de uma função de probabilidade uniforme limitada entre o valor mínimo e máximo estabelecidos.

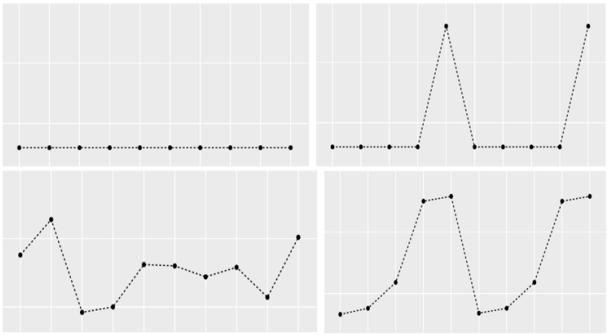

Figura 3: *Estrutura temporal dos estímulos a serem usados nos experimentos de sincronização fala-metrônomo. O eixo vertical representa a duração, em milissegundos, dos intervalos entre os onsets das vogais das sílabas sintetizadas. A figura superior esquerda representa a condição Isócrona, a figura inferior esquerda representa a condição Aleatória, a figura superior direita representa a condição Súbito e a figura inferior direita representa a condição Modelo.*

A estrutura dos estímulos usados no experimento foi uma sequência de intervalos que chamamos de *repetição*. Ao todo, os participantes ouviram 6 repetições, cada uma composta por 9 estímulos sonoros que definem 10 intervalos temporais. Para marcar o início e o fim de cada repetição e ajudar o participante, colocamos um tom puro com frequência de 440 Hz e duração de 400 ms. Os participantes foram orientados a ouvirem atentamente o primeiro grupo de intervalos para se acostumarem com a estrutura temporal dos intervalos e só produzir tentativas de sincronização nos 5 grupos seguintes. Essa forma de estrutura temporal foi usada nas quatro condições rítmicas. O procedimento foi repetido 5 vezes, de modo que para cada condição o participante produziu 25 sincronizações de cada sequência de 10 intervalos.

3.2. Arranjo da gravação

Duas pessoas se voluntariaram para participar do experimento de sincronização, um participante é do sexo masculino e outro do sexo feminino. Ambos os participantes são estudantes de linguística. Para gravar os dados gerados pelas produções dos dois participantes no experimento de sincronização fala-metrônomo, pedimos à ambos os participantes para produzirem a sílaba /ba/ em sincronia com as sílabas sintetizadas no programa Praat. Os participantes foram posicionados diante de um microfone de tipo condensador da marca Samson, modelo C01, ligado a uma placa de som externa da marca Behringer, modelo UMC202HD, conectada a um computador portátil. As sequências de sílabas sintetizadas foram gravadas em arquivos de som em formato wav. Esses arquivos de som foram executados a partir de um segundo computador portátil. Na saída de som desse segundo computador usou-se um plugue divisor de canais, que enviava o som simultaneamente a um fone de ouvido supra-auricular binaural da marca AKG, modelo K 240 MK II, usado pelo participante, e à placa de som conectada ao primeiro computador. Esse arranjo permitiu a captura simultânea do estímulo externo e das tentativas de sincronização do participante em um arquivo de som estéreo. A figura 5 mostra um exemplo de arquivo de som gravado segundo esse arranjo. O canal superior mostra as sílabas sintéticas e a inferior as sílabas produzidas pelo participante em sua tentativa de sincronia. Esta divisão em dois canais é útil por permitir a distinção entre a produção do participante e o estímulo externo, que podem ser carregados independentemente no programa Praat.

3.3. Medida de sincronismo

O desempenho dos participantes foi medido através dos dados de sincronia entre as produções de sílabas dos participantes com as condições especificadas acima. A variável dependente é a assincronia entre os *onsets* vocálicos produzidos pelos dois participantes e a sequência de sílabas artificiais que os sujeitos devem tentar seguir. A marcação das assincronias foi feita com o auxílio do programa de análise acústica Praat. As marcações dos *onsets* vocálicos na produção dos participantes e das sílabas artificiais correspondentes foram armazenadas em arquivos de metadados (objetos TextGrid do programa Praat) em camadas separadas para serem processadas posteriormente por meio de *scripts* escritos para essa finalidade.

Dois tipos de sincronia independentes foram levados em consideração, a sincronia de fase e a de período. A sincronia de período é medida pela razão entre as durações das unidades v-v produzidas pelos participantes e entre a duração das sílabas artificiais. As durações das unidades v-v estão esquematizadas na figura 5.

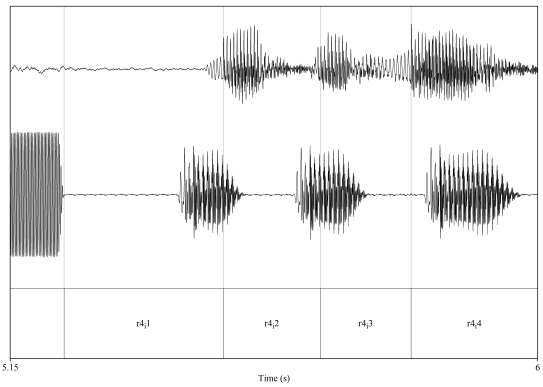

Figura 4: *O primeiro canal (superior) mostra três sílabas [ba] produzidas por um dos participantes. O segundo canal mostra três sílabas sintetizadas que são alvos da sincronização. As retas verticais pontilhadas marcam os intervalos entre* onsets *das sílabas produzidas pelos participantes. O intervalo entre cada reta nos dá a duração das unidades v-v consecutivas.*

A medida de sincronismo fase (φ) é calculada a partir do intervalo de tempo (Δ) entre os *onsets* da sílaba produzida e da sílaba artificial em razão da duração do intervalo entre os onsets das sílabas artificiais (t), ou seja: $\varphi = (\Delta/t) * 360$. O sinal da fase varia de acordo com a posição relativa entre os *onsets* das sílabas artificiais e das produzidas pelos participantes. Se o *onset* da produção do participante precede o da silaba artificial, então φ assume valor negativo e se ela sucede o da sílaba artificial, então φ assume valor positivo, conforme vemos na figura 6.

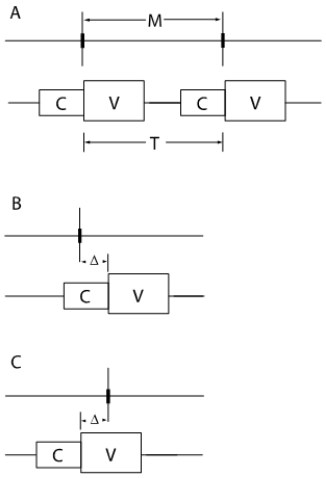

Figura 5 *Esboço da sincronia de fase. M representa o período da batida do metrônomo e T o período da unidade v-v, Δ representa o intervalo de tempo entre a transição CV e a batida do metrônomo. Por convenção, definiu-se que este intervalo é positivo em situações como a ilustrada em B e negativo em situações semelhantes a C.*

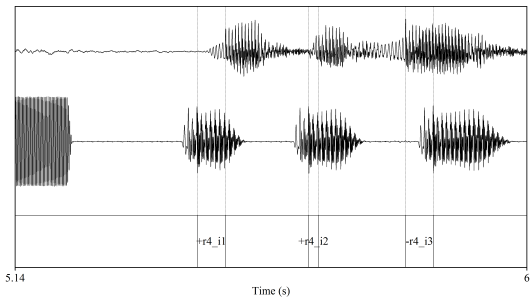

Figura 6 : *O primeiro canal (superior) mostra três sílabas [ba] produzidas por um dos participantes. O segundo canal mostra três sílabas sintetizadas que são alvos da sincronização. As retas pontilhadas marcam os* onsets *das sílabas. Os intervalos entre as linhas são as distâncias de fase entre as sílabas produzidas pelo participante e a sintetizada.*

4. Discussão

Por razões de espaço apresentaremos neste trabalho os resultados de um dos participantes. Embora haja algumas diferenças entre os dois, consideramos que no geral elas não levam a interpretações crucialmente diferentes. Também por razão de espaço restringimos a discussão aos dados de sincronia de fase, que se mostraram mais relevantes para a discussão da hipótese sendo testada aqui. A figura 7 exibe a média e a variabilidade (indicada pela amplitude dos intervalos de confiança de 95% em torno da média) da sincronia de fase em função da posição na sequência na condição ISÓCRONA. Nesta figura e nas próximas, os círculos indicam que a média da sincronia de fase não é significativamente diferente de zero (determinada por um teste-*t* de amostra única com $\alpha = 5\%$) e o símbolo "×" indica que há a média é diferente de zero, indicando que há assincronia de fase em uma determinada posição na sequência. O grau de acuidade do participante se mostra alta ao longo da tarefa de sincronização, uma vez que há sincronia para todas as posições com exceção da primeira. A precisão, indicada pela diminuição da amplitude dos intervalos de confiança, aumenta ao longo da sequência.

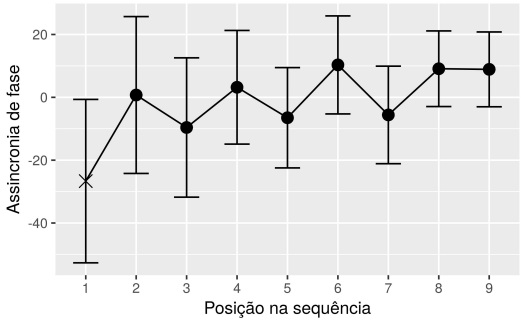

Figura 7: *Média e intervalo de confiança de 95% da sincronia de fase em função da posição na sequência de sílabas na condição de isocronia.*

A figura 8 exibe os resultados na condição MODELO. Ainda que o participante tenha mostrado baixa acuidade, uma vez que só há sincronia de fase apenas em duas posições da sequência, ela flutua: no início há sincronia, embora com grande variabilidade, nas posições seguintes ela é perdida e o grau de assincronia é negativa e aumenta bastante nas posições 2 e 3, quando a assincronia diminui tanto no que diz respeito à média quanto à variabilidade. Na posição 5 a média da assincronia é ligeiramente positiva e bem próxima a zero na posição cinco, que coincide com o alongamento que sinaliza a culminância do acento frasal. Na posição 5, a precisão da sincronia é alta, como indica o intervalo de confiança estreito.

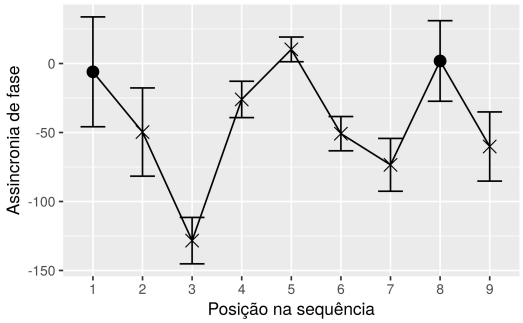

Figura 8: *Média e intervalo de confiança de 95% da sincronia de fase em função da posição na sequência de sílabas na condição do Modelo.*

A figura 9 exibe os resultados na condição ALEATÓRIA. Após o segundo intervalo o falante perde a sincronia de fase e, fora a oscilação observada na posição 4, mantem um patamar negativo pelo restante da sequência. Da quinta posição em diante, há melhora na precisão de sincronização, indicada pela diminuição da amplitude dos intervalos de confiança.

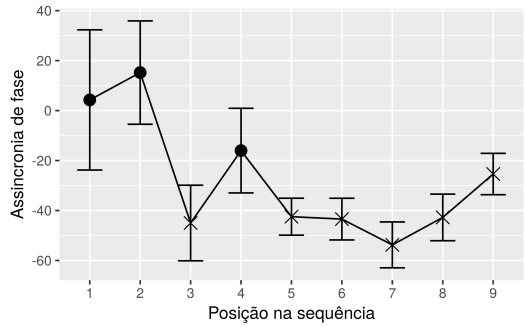

Figura 9: *Média e intervalo de confiança de 95% da sincronia de fase em função da posição na sequência de sílabas na condição aleatória.*

5. Conclusões

Os dados trazem evidências que corroboram a hipótese apresentada na Introdução segundo a qual a estrutura temporal do estímulo externo na tarefa de sincronização tem um papel importante no comportamento dos participantes. Mais especificamente, os resultados parecem mostrar que na condição MODELO parece haver uma flutuação da habilidade de sincronia que melhora nas proximidades do acento frasal, flutuação esta inexistente nas outras duas condições analisadas. Interpretamos esse resultado como indicação de que a estruturação temporal próxima à da fala natural gerada pelo modelo gera um comportamento de sincronismo que poderia ser entendido como resultado de um *resetting* no oscilador acentual interno ao falante que faz a sincronia. É necessário, no entanto, coletar dados de mais participantes para saber se as tendências observadas aqui se manterão.

6. References

[1] KELSO, J. A. S. *Dynamic Patterns:The Self-Organization of Brain and Behavior.* Cambridge, MA: MIT Press, 1995.

[2] MCAULEY, J. D. *Perception of tume as phase: towards an adaptative-oscillator model of rhythmic pattern processing.* Tese (Doutorado) – Indiana University, USA, Bloomington, IN, 1995.

[3] BARBOSA, P. A. *Incursões em torno do ritmo da fala.* Campinas: Pontes Editores, 2006.

[4] BARBOSA, P. A.; MADUREIRA, S. Toward a hierarchical modelo of rhythm production: evidence from phrase stress domains in Brazilian Portuguese. In: *Proceedings of the XIVth International Congress of Phonetic Sciences.* San Francisco, USA: [s.n.], 1999. V. 1, p. 297-300.

[5] BARBOSA, P. A. Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production. *Caderno de Estudos Linguísticos,* v. 43, p. 71-92, 2002.

[6] BARBOSA, P. A.; ARANTES, P. Investigation of non-pitch accented phrases in brazilian portuguese: no evidence favoring stress shift. In: *Proceedings of the Fifteenth International Congress of Phonetic Sciences;* [S.l.: s.n.], 2003. P. 135-143.

[7] BARBOSA, P. A.; ARANTES, P.; SILVEIRA, L. S. Unifying stress shift and secondary stress phenomena with a dynamical systems rhythm rule. In: *Proceedings of the Speech Prosody 2004 Conference.* Nara, Japan: [s.n.], 2004. P. 49-52.

[8] ARANTES, P.; BARBOSA, P. A. Production-Perception Entrainment in Speech Rhythm. In: *Proceedings of the 5th International Conference on Speech Prosody.* Chicago: [s.n.], 2010. p. 1-4.

[9] ARANTES, P. *Integrando produção e percepção de proeminências secundárias numa abordagem dinâmica do ritmo da fala.* Tese (Doutorado) – Universidade Estadual de Campinas, 2010.

[10] MEIRELES, A. R. *Reestruturações rítmicas da fala no português brasileiro.* Tese (Doutorado) – Universidade Estadual de Campinas, Campinas, 2003.

[11] ALLEN, G. D. The location of rhythmic stress beats in English: an experimental study I. *Language and speech,* v. 15, p 72-100, 1972.

[12] ALLEN, G. D. The location of rhythmic stress beats in English: an experimental study II. *Language and speech,* v. 15, p. 179-195, 1972.

[13] BARBOSA, P. A.; MEIRELES, A. R.; VIEIRA, J. M. Abstractness in speech-metronome synchronisation: P-centres as cyclic attractors. In: *Proceedings of the Interpeech 2005 Conference.* Lisbon: [s.n.], 2005. P. 1441-1444.

[14] KLATT, D. H. Synthesis by rule of segmental durations on English sentences. In: LINDBLOM, B.; OHMAN, S. (Ed.). New York: Academic Press, 1979.

[15] ARANTES, P; LIMA, V. Towards a methodology to estimate minimun sample lenght for speaking rate. Aceito para publicação no GEL.