

É POSSÍVEL INTEGRAR O DISCRETO E O CONTÍNUO EM UM MODELO DE PRODUÇÃO DO RITMO DA FALA?

PLÍNIO ALMEIDA BARBOSA
(LAFAPE/IEL/UNICAMP)

A model of speech rhythm production is proposed that is able to explain the differences found between the durational patterns of stress- and syllable-timed languages from strict considerations of timing. This is possible from a two-parameter characterization of Brazilian Portuguese V-to-V duration patterns within stress groups. The parameters are speech rate (measured in terms of V-to-V units per unit of time) and the coupling strength between two oscillators (syllabic and stress group oscillators), which act as a cognitive pacemaker whose pace is the vowel onset succession entrained by the beats of the stress group oscillator. Differences in amount of coupling strength seem to explain differences in rhythmic typology. A link with lexical processing is also suggested here.

0. INTRODUÇÃO

A Dinâmica é a parte da Física que estuda os corpos em movimento, movimento este resultado da ação de forças externas ou internas ao sistema físico ao qual os corpos pertencem. O movimento de um corpo que faz parte de um sistema é expresso matematicamente por uma função que define os valores assumidos por uma ou mais variáveis ao longo do tempo (posição e velocidade de um ponto do corpo, por exemplo). Para sistemas físicos lineares, cujos parâmetros não variam com o tempo, essa função é determinada inequivocamente a partir da ordem da ou das equações diferenciais que modelam o sistema, do valor dos parâmetros da ou das equações e das condições iniciais do sistema. Uma equação estabelece uma relação fixa entre uma variável (posição, por exemplo) e suas derivadas (velocidade e aceleração, por exemplo), como consequência das forças que atuam sobre o sistema. O caso do sistema massa-mola é exemplar. Ele é usado para modelar o movimento de determinados pontos do trato vocal em direção a seus alvos (ou pontos de equilíbrio) submetido ao parâmetro rigidez da mola.

1. A REVOLUÇÃO INTRODUZIDA PELA FONOLOGIA ARTICULATÓRIA

O sistema massa-mola com amortecimento crítico é empregado pela Fonologia Articulatória, doravante FAR (Browman & Goldstein 1989, 1992), para modelar a produção de fala ao nível lexical. Trata-se de uma descrição abstrata comensurável com

características básicas do mundo físico: a magnitude espacial (o alvo, que pode ser a distância em relação à glote de uma constricção no trato vocal) e a extensão temporal (a duração) dos movimentos articulatorios.

Os gestos da FAR são abstratos o suficiente para corresponder a uma intenção gestual, intenção de mesma natureza – intenção de movimento – da intenção de pegarmos uma xícara ou de andarmos sobre a calçada. Por exemplo, o gesto labial, assim, intencional, se traduz na FAR em um par de variáveis de trato (*Lip Protrusion* ou LP e *Lip Aperture* ou LA), que são as variáveis de dois sistemas massa-mola abstratos. Os valores das variáveis correspondentes são em seguida distribuídos percentualmente para os verdadeiros articuladores para gerar, somente então, a trajetória de corpos físicos concretos (nesse caso, mandíbula, lábios superior e inferior). Esses passos podem ser esquematizados na figura 1, no caso da realização de um [b]. A porcentagem de movimento distribuída para cada articulador relevante é hipotética, mas é fundamentada em dados realísticos (visto que o lábio superior participa bem menos que o inferior e esse, menos que a mandíbula para a realização do gesto labial). Os retângulos ao lado das variáveis LA e LP têm dimensões diferentes: a magnitude, altura do retângulo, é distinta (maior envolvimento da abertura labial do que da protrusão), mas a extensão temporal, base do retângulo, é a mesma. Essa constitui uma hipótese forte da FAR, isto é, a de que as variáveis de trato ligadas a um mesmo gesto intencional têm mesma extensão temporal. As trajetórias das variáveis de trato são geradas por um modelo conhecido como *task dynamics* (dinâmica de pequenas tarefas) e indicadas dentro dos retângulos, de forma esquemática e para exemplificação. Um modelo articulatorio atribui a trajetória da constricção (dada pela variável de trato), sob a forma de participação percentual, a cada um dos articuladores pertinentes.

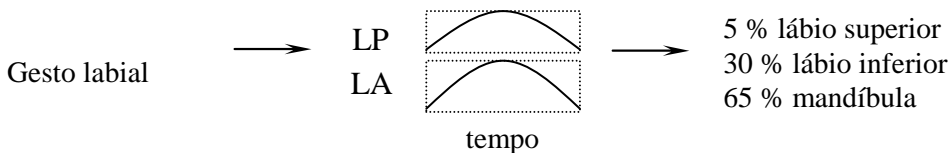


Figura 1: Do gesto labial de [b] à trajetória dos articuladores

Na FAR, dois gestos consecutivos estão submetidos a regras específicas e rígidas de coordenação intergestual, sobretudo entre gestos consonantais e gestos vocálicos (o gesto *Tongue Body* ou TB não fechado e não crítico de qualquer variável de trato) ou entre dois gestos consonantais (que estão associados a graus de constricção fechados ou críticos).

Em outra proposta de fonologia gestual, a Fonologia Acústico-Articulatória (FAAR) de Albano (2001), a abstração ainda é maior, pois as bordas esquerda, direita e o intervalo de ativação das variáveis de trato podem ser usados de forma discreta como pontos de referência para associação de informação morfológica. Na FAAR, a coordenação pode se dar de maneira mais vaga do que na FAR, uma maneira discretizável, sem perda alguma, pelo emprego da noção de borda gestual.

Mas apesar da abstração das fonologias gestuais, e mesmo que a coordenação intergestual possa ser especificada fonologicamente com valores discretizáveis como

borda esquerda, direita ou intervalo, a extensão no tempo e a magnitude no espaço de cada gesto devem ser especificadas a partir de valores contínuos. A primeira, via rigidez e faseamento (discretizável) entre gestos. A segunda, através de um número real representando o alvo de cada variável de trato e portanto, o grau de redução do gesto correspondente. A infinitude desse contínuo não é, no entanto, problemática pois ela se dá intersubjetivamente ou mesmo intrasubjetivamente pela interação com outras variáveis discretizáveis. Em um indivíduo, um valor real *default* para magnitude e extensão temporal, resultado do final do processo de aquisição de uma determinada língua, pode ser armazenado no léxico.

Esse tipo de aparato exige evidentemente uma fonética lingüística, para a qual as extensões temporais e magnitudes de uma comunidade de falantes definem regiões, isto é, classes de valores que são significativamente distintos de língua para língua ou mesmo de dialeto para dialeto.

O contínuo da chamada implementação fonética, a articulação, se dá quando da enunciação, pela interação que ocorre entre as pautas gestuais e um sistema duplamente oscilatório, o rítmico, que define uma grade em torno da qual se associam gestos segmentais e gestos prosódicos em geral. Tal sistema é aqui considerado paralingüístico e comum a outros movimentos corpóreos oscilatórios (braços e pernas no andar, no correr, no remar).

Mas como se dá a interação entre o fônico lexical e as oscilações do sistema rítmico? Antes de responder a essa questão precisamos mostrar a diferença entre os tempos intrínseco e extrínseco na FAR e como as duas oscilações do sistema rítmico interagem entre si.

2. O SISTEMA RÍTMICO

Apesar da referência à camada rítmica em outro artigo de Browman & Goldstein (1990), nada é dito sobre a coordenação entre gestos vocálicos. Nossa hipótese é a de que não poderia dizê-lo, visto que a coordenação V-V é da ordem do tempo extrínseco, dada por um oscilador externo ao léxico, que acompanha os eventos de *onset* da vogal. Esse oscilador é o oscilador silábico (por referência à silabidade e não à sílaba). Enquanto oscilador ele é, como o gesto, ao mesmo tempo discreto (o ciclo) e contínuo (o período em segundos).

Que ele acompanhe os eventos vocálicos de *onset* é confirmado por estudos de percepção sobre a importância da transição CV para o processamento auditivo (Chistovich & Ogorodnikova 1982), processamento que varre pontualmente as transições CV, varredura cujo resultado é a delimitação de segmentos VC menos analisáveis internamente que os CV. Isso também é confirmado pelos estudos posteriores sobre o *p-center* (Marcus 1976, Pompino-Marschall 1989, 1991, Janker 1995). Estudos de comparação interlingüística (Vaissière 1983) mostram ainda a coesão da rima (VC) e sua universalidade como veículo do acento e da prosódia. Análises estatísticas de padrões articulatórios ^lV.C, em palavras como *leper* (Turk

1994), ou acústicos em padrões V.C e V#C (Barbosa & Madureira 1999) mostram a forte coesão dessas seqüências.

Os efeitos da acentuação evidenciados nesses resultados, confirmando que a unidade sobre a qual culmina o mecanismo de acentuação é a V(#)C, se superpõem evidentemente à oscilação silábica, constituindo portanto uma dupla oscilação. A presença desse duplo mecanismo é universal. A maneira precisa como se dá a interação entre os dois osciladores varia de língua para língua (se pensarmos em classes de interação), de indivíduo para indivíduo (se pensarmos que as interações são o resultado singular da aquisição de uma língua para cada indivíduo) e mesmo em um mesmo indivíduo, dependendo da taxa e estilo de elocução por ele usadas (ver Barbosa submetido, para mais detalhes em torno do universal e do específico à língua em um modelo de produção do ritmo da fala).

Essa interação entre osciladores é conhecida como acoplamento. Um oscilador influencia o outro através de um mecanismo de indução (*entrainment*) que pode variar de magnitude, dependendo da força de acoplamento entre os dois osciladores. Na natureza há inúmeros exemplos de acoplamento entre osciladores intra- ou interorganismos, como dois relógios em uma mesma parede, neurônios em fase em uma rede de neurônios, o som uníssono dos grilos, ou ainda o piscar luminoso uníssono das luciolas (Strogatz & Stewart 1994). Vejamos as conseqüências, para um oscilador silábico, das diferentes forças de acoplamento de um oscilador acentual. Na figura 2 pode-se verificar o efeito de pouco (menos *stress-timing* ou mais *syllable-timing*) ou muito acoplamento (mais *stress-timing* ou menos *syllable-timing*) sobre osciladores silábicos em grupos acentuais de quatro e oito sílabas. Observa-se assim, em línguas ditas *stress-timed*, que o quarto pico, marcado com uma seta, ocorre mais tardiamente (coluna da direita, acima) do que na língua dita *syllable-timed* (à esquerda, acima). As escalas temporais nas oscilações inferiores (grupo acentual de oito sílabas) são o dobro daquelas das oscilações superiores (grupo acentual de quatro sílabas). Notar que nas línguas *syllable-timed* (coluna da esquerda), os picos tanto para quatro quanto para oito sílabas estão próximos na escala, o que indica que demora-se quase o dobro do tempo para pronunciar oito sílabas em relação a quatro nesse tipo de língua. Já na língua dita *stress-timed*, demora-se menos que o dobro do tempo na mesma situação, indicando algum tipo de compressão temporal. Todos esses efeitos são relatados na literatura sobre tipologia rítmica das línguas (cf. por exemplo Bertinetto 1989, Lehiste 1977, Roach 1982).

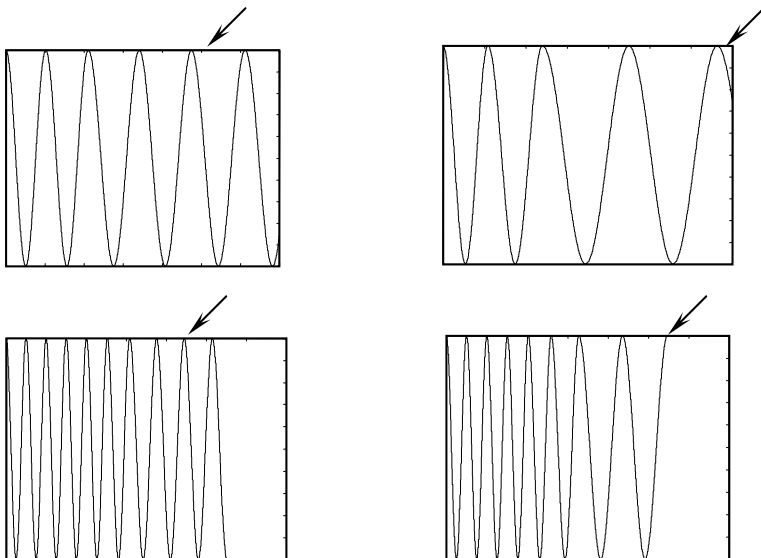


Figura 2: Osciladores acoplados em línguas *syllable-timed* (esquerda) e *stress-timed* (direita) para grupos acentuais com quatro (acima) e oito sílabas (abaixo)

As conseqüências sobre o oscilador silábico do acoplamento com o acentual esquematizado acima explicam claramente os padrões das línguas ditas de ritmo acentual (*stress-timed*), isto é, em que os acentos tendem a ocorrer a intervalos regulares no tempo (isocronismo acentual) ou ditas de ritmo silábico (*syllable-timed*), isto é, em que as sílabas tendem a ocorrer a intervalos regulares no tempo (isocronismo silábico). Os dois tipos rítmicos são obtidos através de um mesmo dispositivo: a maior força de acoplamento do oscilador acentual em relação ao silábico gera uma tendência a *stress-timing* e uma menor força de acoplamento em relação ao silábico, uma tendência a *syllable-timing*.

3. TIPOLOGIA RÍTMICA

Ao forjar os termos de línguas *stress-timed* e *syllable-timed* em 1945, a partir da distinção assinalada por Lloyd-James (1940), e percebida pelo menos desde o século XVIII, Pike também afirmou que uma mesma língua pode exibir características de ambos os ritmos. Discursos de políticos de língua inglesa (Luther King, por exemplo) estão cheios de exemplos desse fato. Abercrombie (1967) propõe em *Elements of General Phonetics*, no entanto, que “tanto quanto se conhece, as línguas do mundo se dividem em línguas de ritmo acentual, como o inglês, o russo e o árabe e línguas de ritmo silábico, como o francês, o telugu e o iorubá” (p. 97). A busca por isocronismo absoluto em produção de fala prosseguiu e terminou com críticas severas (Lea 1974, Nootboom 1991) a qualquer tipo de tendência a isocronismo no domínio da produção,

embora uma tendência se mantenha em termos de percepção (Lehiste 1977). Mas o que se verificava nos estudos de produção? Vejamos os tipos possíveis de relação entre o número de sílabas do grupo acentual, n (delimitado por dois acentos sucessivos), e a duração do mesmo grupo, I , na figura 3.

DURAÇÃO DO GRUPO ACENTUAL

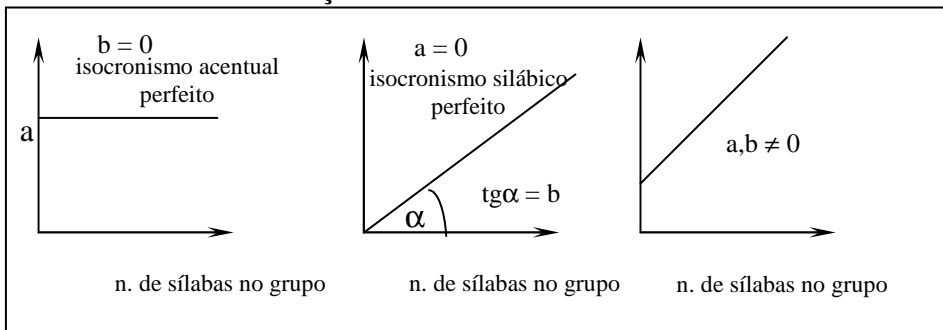


Figura 3: Os três Padrões de variação da duração dos grupos acentuais em relação ao número de sílabas contidas neles

Os primeiro e segundo padrões da figura acima são casos ideais: não há isocronismo acentual nem silábico perfeitos, apenas uma tendência para um tipo ou outro, isto é, graus diferentes de *stress-timing*, em um contínuo, isto é apenas o caso do terceiro padrão, que é um tipo misto entre os dois primeiros. Como descrever matematicamente, via osciladores acoplados, os diferentes graus dessa tendência para *stress-timing*?

O trabalho de O'Dell & Nieminen (1999) mostra através de uma técnica matemática, APD (*Averaged Phase Difference*), como fazê-lo, a partir do efeito médio, para cada ciclo, da diferença de fase entre os osciladores. Como resultado da aplicação dessa técnica, eles propõem a realização de uma regressão linear para a obtenção de um padrão semelhante ao terceiro da figura 3, isto é, expresso pela equação $I = a + b.n$ em que a razão a/b representa a força de acoplamento r entre os osciladores acentual e silábico: quanto maior o valor de r , maior o grau de *stress-timing* do *corpus* analisado.

Três questões precisam, no entanto, ser respondidas antes de qualquer investigação experimental para se conseguir condições de comparabilidade adequadas entre diferentes *corpora* e entre línguas, tal como explicado detalhadamente em Barbosa (2000). Respostas preliminares a essas questões são propostas logo em seguida.

- (1) Definição dos grupos acentuais. Como determinar quais palavras foram acentuadas, são proeminentes? Via produção (duração em PB) e testes de percepção de acentos dos grupos de força;
- (2) Cabeça do grupo. O acento termina ou inicia o grupo acentual? Pela Fonologia Prosódica o pé binário é trocaico, mas os demais níveis têm dominância à direita. Assim, adotamos que o acento termina o grupo;

(3) Como delimitar os grupos acentuais? Unidades V(#)C ou sílabas fonológicas? Unidades V(#)C, pois é sobre elas que o acento culmina.

A partir dessas primeiras decisões é possível realizar as primeiras avaliações do PB.

3.1 TIPOLOGIA RÍTMICA DO PB: CARACTERIZAÇÃO BIPARAMÉTRICA

Para o caso do PB, um *corpus* de 36 frases lidas em três taxas de elocução por um locutor pernambucano permitiu obter os primeiros resultados de forças de acoplamento para essa língua. Verifica-se pelos pontos na figura 4 representando o PB (PB1, PB2 e PB3) que essa força de acoplamento depende da taxa de elocução!

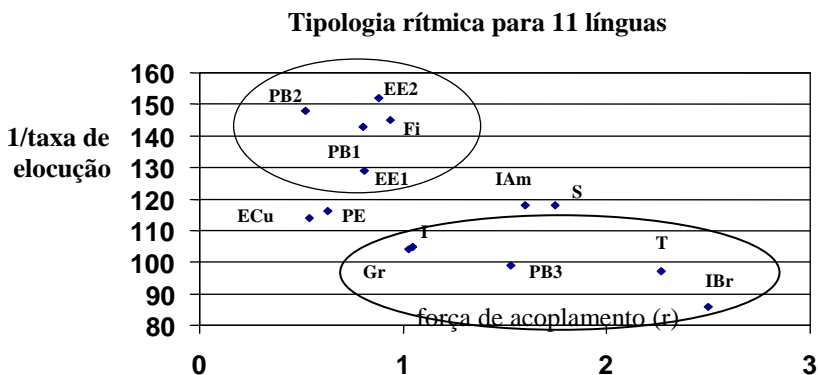


Figura 4: Caracterização Rítmica Biparamétrica de 11 Línguas. Forças de acoplamento e inclinações da reta de regressão. As abreviauras representam o inglês britânico (IBr), e o americano (IAm), o thai (T), o grego (Gr) e o italiano (I), espanhol europeu (EE1 e EE2) e cubano (ECu), a partir de regressões feitas ou refeitas por mim, com os dados de Dauer (1983). Os dados do sueco (S) provêm da fórmula de Fant & Kruckenber (1989). Os dados do português europeu (PE) e brasileiro (PB2) são de um estudo de Barbosa (em preparação), a partir de um *corpus* de 20 frases apresentados em Frota & Vigário (1999). Os demais dados do PB (PB1 e PB3) são do estudo apresentado aqui, para grupos acentuais com menos de 5 sílabas. Os dados do finlandês (Fi) são de O'Dell & Nieminen (1999).

O valor da força de acoplamento dependerá assim da possibilidade de mantermos condições *ceteris paribus* interlingüísticamente e de separar as variedades regionais de uma mesma língua pois não se sabe a priori se possuem mesma tipologia rítmica. Muitas vezes, essas variedades foram gravadas em taxas de elocução distintas, o que é problemático, como nos dados de Dauer (1983).

4. DISCUSSÃO

É possível integrar a pauta gestual ao sistema rítmico? Nossa resposta consiste em dizer que a integração pode ser feita via bordas esquerdas do gesto TB de grau de

construção não crítico e não fechado. É dessa forma que o sistema rítmico impõe ao componente fonológico gestual o seu *timing*, e portanto a taxa de elocução.

As batidas (*beats*) do oscilador silábico (correspondendo aos picos positivos do mesmo) já induzido pelo acentual completam o que faltava às fonologias gestuais para que possam ter a possibilidade de implementação fonética, o tempo extrínseco. Gestos reduzidos, como as vogais pós-tônicas em PB e consoantes de coda como /s/ e /r/, que têm uma duração curta quando não seguidas de pausa silenciosa, são reduzidas devidas ao seu tempo intrínseco na pauta gestual lexical e não por uma suposta aceleração local da taxa de elocução. Uma representação esquemática da integração pode ser vista na figura 5.

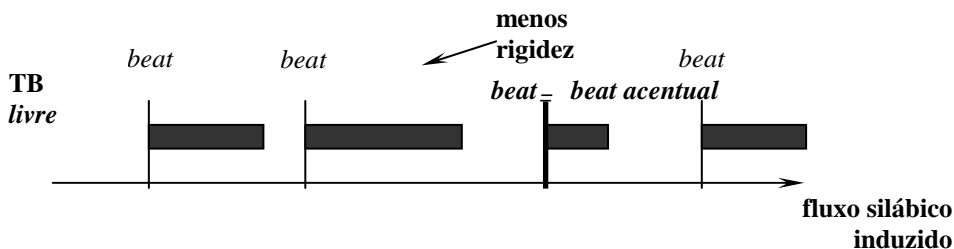


Figura 5: Pauta gestual integrada. Detalhes no texto

Em relação à grade imposta pelo oscilador induzido à pauta gestual lexical, uma vez que os gestos consonantais se coordenam primeiramente em relação aos gestos vocálicos, pode-se pensar que o intervalo entre duas batidas do oscilador silábico define um ciclo completo (a parte positiva correspondendo à abertura vocálica, e a negativa correspondendo ao fechamento vocálico), um quadro dentro do qual os gestos consonantais gravitam em torno do vocálico.

Torna-se claro aqui que os segmentos acústicos, sobretudo as consoantes, são subprodutos de uma organização eminentemente silábica.

Em relação ao discreto vs contínuo na pauta para implementação fonética, do *ponto de vista* lingüístico, abstrato, as batidas do oscilador definem ciclos, unidades discretas porém dotadas de extensão. Do *ponto de vista* paralingüístico, para implementação motora, essas batidas definem a taxa de elocução, cujo valor *default* é adquirido pelo locutor, valor esse que pode ser representado por um número real. Mais uma vez a variabilidade dar-se-á em função da interação prosódia-segmentos, condicionada por fatores lingüísticos, para- e extralingüísticos ligados a aspectos situacionais.

A interação prosódia-segmentos é implementada em nosso modelo por uma rede connexionista, tendo em vista a hipótese de um sistema cognitivo realista e eminentemente dinâmico (Barbosa 2001). Os reflexos dessa interação fazem-se sentir sobre o articulador portador da oscilação silábica, a mandíbula. O *timing* desse oscilador, com relação a efeitos de primeira ordem seria resultado da indução do oscilador silábico abstrato, cognitivo, sobre esse oscilador físico, motor (que possui uma frequência natural de vibração).

A proposta aqui delineada questiona seriamente toda Fonética não lingüística (isto é, que não seja parametrizada de língua para língua). Ela também constitui um questionamento e uma tentativa de lidar com a vagueza que existe entre representações discretas simbolizáveis (em correspondência biunívoca com os números naturais) e representações do contínuo como aquela dada pelo grau e abstração dos números reais (sem correspondência biunívoca com os números naturais).

REFERÊNCIAS

- ABERCROMBIE, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- ALBANO, E. C. (2001). *O Gesto e suas Bordas: por uma fonologia acústico-articulatória do português brasileiro*. Campinas: Mercado de Letras.
- BARBOSA, P. A. (2000). "Syllable-timing in Brazilian Portuguese": uma crítica a Roy Major. *D.E.L.T.A.*, 16 (2), 369-402.
- _____. (2001). "Generating Duration from a Cognitively Plausible Model of Rhythm Production". *Proceedings of the Seventh European Conference on Speech Communication and Technology (Eurospeech 2001)*, Aalborg, Dinamarca, 3-7 setembro, v. 2, 967-970.
- _____. (submetido). O Universal e o Específico a Língua em um Modelo Dinâmico de Produção do Ritmo (Fonético e Fonológico) da Fala. *Anais do II Congresso Internacional da Abralin*, 14 a 16 de março, Fortaleza, Ceará.
- BARBOSA, P. A. & S. MADUREIRA. (1999). Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1: 297-300.
- BERTINETTO, P.M. (1989). Reflections on the dichotomy "stress-" vs "syllable-timing". *Revue de Phonétique Appliquée*, 91-92-93: 99-130.
- BROWMAN, C. & GOLDSTEIN, L. (1989). Articulatory Gestures as Phonological Units. *Phonology*, 6: 201-251,.
- _____. (1990). Tiers in Articulatory Phonology with some implications for casual speech. In: KINGSTON, J. & BECKMAN, M.E. (Eds.) *Papers in Laboratory Phonology I*. Cambridge: Cambridge University Press, 341-376.
- _____. (1992). Articulatory Phonology: an overview. *Phonetica*, 49: 155-180.
- CHISTOVICH, L.A. & E.A. OGORODNIKOVA. (1982). Temporal processing of spectral data in vowel perception. *Speech Communication*, 1: 45-54.
- DAUER, R.M. (1983). Stress-timing and syllable-timing re-analysed, *Journal of Phonetics*, 11: 51-62.
- FANT, G. & A. KRUCKENBERG. (1989). Preliminaries to the study of Swedish prose reading and reading style. *Quarterly Progress and Status Report, Speech Transmission Laboratory, KTH*, 2: 1-83.
- JANKER, P. (1995). On the influence of the internal structure of a syllable on the P-center perception. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 2: 510-513.
- LEA, W.A. (1974). Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding. *Univac Report PX10791*, Sperry Univac, DSD, St. Paul, Minnesota, EUA.
- LEHISTE, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5: 253-263.

- LLOYD-JAMES, A. (1940). *Speech signals in Telephony*. London.
- MARCUS, S. M. (1976) *Perceptual-centres*. Tese de Doutorado Inédita. Cambridge: Cambridge University.
- NOOTEBOOM, S.G. (1991). Some observations on the temporal organisation and rhythm of speech. *Proceedings of the XIIth International Congress of Phonetic Sciences*, 1: 228-237.
- O'DELL, M. & T. NIEMINEN. (1999). Coupled oscillator model of speech rhythm. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 2: 1075-1078.
- PIKE, K. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- POMPINO-MARSCHALL, B. (1989). On the psychoacoustic nature of the p-center phenomenon. *Journal of Phonetics*, 17: 175-192.
- _____. (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.
- 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*. Londres: Edward Arnold, 73-79.
- STROGATZ, S. & I. STEWART. (1994). Oscillateurs couplés et synchronisation biologique. *Pour la Science*, 196: 40-46.
- TURK, A. (1994). Articulatory phonetic clues to syllable affiliation: gestural characteristics of bilabial stops. In: KEATING, P. A. (Ed.) *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge, UK: Cambridge University Press.
- VAISSIÈRE, J. (1983). Language-independent prosodic features. In: *Prosody: models and measurements*. Cutler, A. and Ladd, D.R. (Eds.) Berlin: Springer-Verlag, 53-66.