

O Lugar do Pé Métrico e do Acento no Modelamento Dinâmico do Ritmo

Plínio Almeida BARBOSA
(LAFAPE/IEL/UNICAMP)
plinio@iel.unicamp.br

ABSTRACT: The reassessment of some key linguistic notions such as (metrical) foot and stress and stressing are viewed under the framework of a dynamical model of speech rhythm production. This model has two main components: a two-coupled-oscillator model (a syllabic and a phrase stress oscillator) and an integrated gestural score specifying acoustic-articulatory coordinations between abstract gestures. The phrase stress oscillator deals with the metrical component of the model by delimitating stress groups (feet), the syllabic oscillator imposes a vowel flow continuum along the entire utterance, and the integrated gestural score it is the domain where non-derivational lexical contrasts take place.

KEY WORDS: Rhythm Typology, Coupled Oscillators, Duration, Brazilian Portuguese.

RESUMO: O modelamento dinâmico do ritmo permite reavaliar questões-chave como a definição e a delimitação do pé métrico e o lugar do acento na economia do modelo. O modelo sendo implementado tem dois componentes: um modelo rítmico com dois osciladores acoplados (um oscilador silábico e um acentual) e uma pauta gestual integrada que retém as coordenações entre os gestos acústico-articulatórios. O oscilador acentual dá conta da métrica delimitando grupos acentuais (pés), o oscilador silábico induzido impõe o timing vocálico do enunciado e a pauta gestual implementa não-derivacionalmente o chamado acento lexical.

PALAVRAS-CHAVE: Tipologia rítmica, Osciladores acoplados, Duração, Português brasileiro.

1. Introdução

Neste trabalho apresentamos um modelo dinâmico de produção do ritmo da fala em desenvolvimento no Laboratório de Fonética e Psicolinguística do Instituto de Estudos da Linguagem (Unicamp) fundamentado na Teoria de Osciladores Acoplados. Tal modelo visa a explicar a duração acústica em línguas que usam a duração como parâmetro independente para assinalar tanto o acento lexical quanto o frasal, pela consideração de dois níveis de organização temporal.

O nível intrínseco ou micro-rítmico dá conta da especificação de durações isomorfas a gestos segmentais de um léxico gestual, segundo a proposta de Albano (2001). O nível extrínseco, macro-rítmico, ou prosódico propriamente dito, controla a duração por meio da ação de um modelo de dois osciladores acoplados. O oscilador silábico se ocupa da sucessão de unidades do tamanho da sílaba e portanto está relacionado com a noção de silabidade entendida como um fluxo contínuo de vogais interrompido pelos gestos consonantais (Öhman;1966, Fujimura;1995, Browman & Goldstein;1990). O oscilador acentual se ocupa

da sucessão dos acentos frasais, isto é, das proeminências quase-periódicas ao longo de um enunciado e delimitadoras dos grupos acentuais (*stress groups*). Somente o nível extrínseco será considerado neste artigo.

Em outro lugar (Barbosa;2000) discorreremos sobre a utilidade da teoria dos osciladores acoplados para o estudo da duração. Limitar-nos-emos aqui a apresentar o modelo com que trabalhamos e as redefinições das noções de acento, pé métrico, acentuação e até mesmo do lingüístico que aquele impõe.

2. O modelo dinâmico de produção do ritmo da fala

A figura 1 apresenta o modelo dinâmico de produção do ritmo da fala em desenvolvimento. O modelo de osciladores acoplados pode ser visto à sua esquerda e é sobre ele que falaremos mais amiúde neste artigo. O oscilador acentual é implementado por um trem de pulsos e o silábico, por uma cossenóide levantada. Depreende-se da figura que o oscilador acentual recebe como entrada informação de níveis lingüísticos mais elevados e do léxico gestual. Além disso, pela sua característica oscilatória, ele impõe restrições eurrítmicas à ocorrência de seus pulsos, o que lhe confere uma oscilação quase periódica.

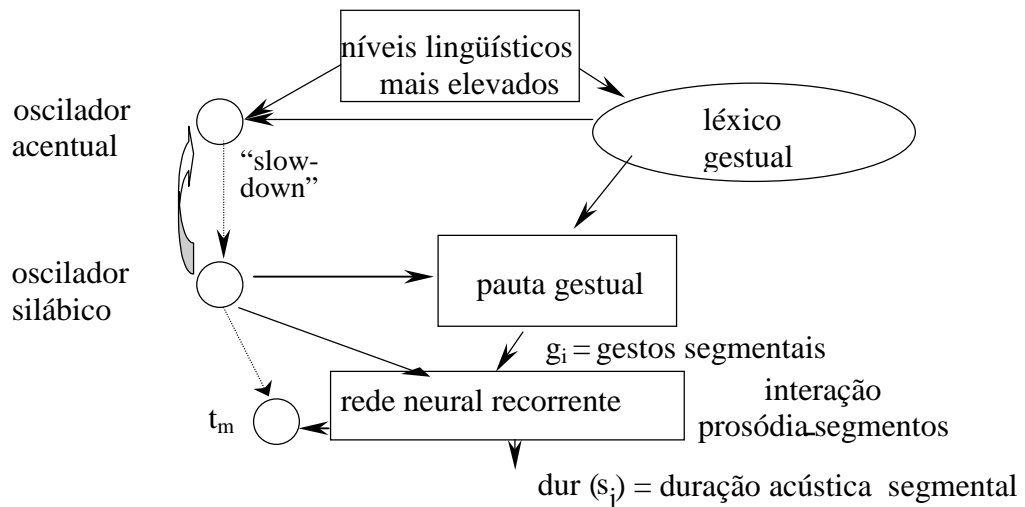


Figura 1: Modelo dinâmico de produção do ritmo da fala

Embora a informação de nível lingüístico superior diga respeito a descrições sintáticas e semânticas do enunciado, somente a posição de palavras funcionais na frase está sendo levada em conta atualmente, pois uma descrição sintático-semântica da frase exigiria um *parser*, de que não dispomos no momento. A posição dos acentos lexicais, a partir de sua especificação no léxico gestual, é levada em conta e desvia os pulsos, hipotetizados regulares, do oscilador

acentual para coincidirem com o *onset* de uma vogal lexicalmente acentuada. A sucessão de pulsos do oscilador acentual implementa a característica universal da acentuação nas línguas.

O oscilador silábico tem os máximos de seus pulsos coincidentes com os *onsets* das vogais do enunciado. Após o acoplamento por sua interação com o oscilador acentual, o oscilador silábico determina para a pauta gestual a posição dos *onsets* vocálicos ao longo do enunciado, extrinsecamente. Esta solução está relacionada à importância da transição CV para a produção e percepção da fala, conforme discutido extensamente em Barbosa (2000) a partir dos estudos de Allen (1972), Dogil & Braun (1988) e da literatura sobre o perceptual-center (Morton, Marcus & Frankish;1976, Pompino-Marschall;1989, 1991, Janker;1995, entre outros). A sucessão de transições CV funciona como ponto de ancoramento para a produção e percepção da fala, implementando um fluxo vocálico contínuo diretamente correlato à oscilação mandibular (oscilador de período t_m que se vê no modelo acima). A hipótese de que a enunciação se dá por um fluxo vocálico interrompido por perturbações consonantais foi proposta por Öhman (1966) e é defendida e implementada a partir da noção de *base function*, por Fujimura (1995). O oscilador silábico implementa assim a silabidade entendida nos termos aqui explicitados e, portanto, também uma característica universal.

O primeiro sentido do lingüístico no modelo dinâmico (modelo de osciladores, pauta gestual e rede neural recorrente) é a maneira pela qual os pulsos do oscilador acentual se alinham com os *onsets* de vogais tônicas do enunciado, isto é, com material lingüístico. O segundo sentido do lingüístico neste modelo refere-se à força de acoplamento entre os dois osciladores que, embora possa variar de indivíduo para indivíduo, a variabilidade intra-individual seria bem menor que aquela entre línguas distintas, da mesma forma que a variabilidade entre duas amostras provindas de uma mesma população estatística (variabilidade intra-individual) é bem menor que aquela provinda de duas populações distintas (variabilidade inter-lingüística). O terceiro sentido do lingüístico refere-se ao léxico gestual de cada língua, contrastando gestos plenos e reduzidos, conforme a proposta de Albano (2001), que considera a posição do acento lexical como ocorrendo no último conjunto de gestos plenos do tamanho da sílaba na palavra. Um quarto sentido do lingüístico refere-se à maneira como se dá a interação prosódia-segmentos, por enquanto sendo implementada por uma rede neural recorrente, conforme demonstrado em Barbosa (2001).

Neste tipo de modelo, o acento lexical é portanto especificado no léxico gestual. Os acentos lexicais marcados frasalmente durante a enunciação o são pelos pulsos de um oscilador acentual, que desencadeará um mecanismo de acentuação via desaceleração

progressiva (“slow-down”) ao longo do grupo acentual, delimitado por dois pulsos sucessivos do oscilador acentual. O pé métrico no modelo é justamente o grupo acentual: o domínio compreendido entre a primeira unidade VV (esta se estende do *onset* de uma vogal ao *onset* da próxima, incluindo a primeira vogal e excluindo a segunda) após um pulso do oscilador acentual até a última unidade VV, que tem seu *onset* de vogal alinhado com o pulso do oscilador acentual seguinte. Definido assim, o pé métrico tem sua extensão variada, dependendo da taxa de elocução, que é especificada pelo período do oscilador silábico antes do acoplamento com o acentual. Em uma taxa de elocução lenta, os pé métricos do enunciado “Maria Rita ama João” podem ser: Maria Rit|a am|a Joã|o (anacruse, dois grupos acentuais, e a sílabas cataléctica /o/); numa taxa mais rápida: Maria Rit|a ama Joã|o (com apenas um grupo acentual central).

O acoplamento entre os dois osciladores é realizado por um mecanismo de indução (*entrainment*) a partir de uma equação de diferenças, conforme explicado a seguir.

2.1. Implementação do modelo dinâmico de produção do ritmo da fala

Para a implementação do modelo de acoplamento entre os dois osciladores nos inspiramos no modelo de osciladores adaptativos para a percepção de intervalos de tempo de McAuley (1995), tendo introduzido várias modificações quanto à maneira como se dá o acoplamento e a relação com variáveis lingüísticas. O primeiro componente comum, é a função de sincronismo entre os osciladores, $s(\cdot)$ (*output function* em McAuley, op. cit.), que em nosso modelo é determinada empiricamente a partir de um *corpus* de frases lidas em três taxas de elocução por um locutor pernambucano (35 anos na época da gravação). A função é dada pela fórmula (2) e foi determinada por regressão não-linear (via regressão linear dos valores transformados da variável em questão) das durações VV ao longo das frases do *corpus*. A evolução da modificação relativa da duração da unidade VV ($\Delta VV_i/VV_i$) ao longo de cada frase foi computada e transformada por meio de algumas funções não-lineares, sendo a melhor regressão obtida com a exponencial ($R^2 = 0.999$).

$$s(0) = w_0 \cdot \exp(-N + 2), \text{ and } s(N-1) = \exp(-5.81 + 0.016 \cdot T_0) \quad (2a)$$

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

Na fórmula 2, N é o número de unidades VV no grupo acentual, e n é o índice da unidade VV corrente (começando de 0 para a primeira VV do grupo acentual). A força de acoplamento relativa entre os osciladores, w_0 , é um parâmetro contínuo, específico a uma

língua e representando o grau de acoplamento entre os osciladores acentual e silábico. T_0 é o período do oscilador silábico na condição de não-acoplamento.

O *reset* de fase no modelo de osciladores acoplados ocorre a cada máximo do oscilador silábico, isto é, no *onset* de cada vogal. O acoplamento de período no oscilador silábico é realizado via indução progressiva a partir do cálculo da parcela de alteração de período, ΔT , na fórmula (3).

$$\Delta T = \alpha.T.s(n).i(m) - \beta.(T-T_0).i(m-1) \quad (3)$$

Os parâmetros α , β , e T_0 são respectivamente a taxa de indução, a taxa de decaimento e o período do oscilador silábico livre de acoplamento. As funções $s(.)$ e $i(.)$ são respectivamente a função de sincronismo (dada pela fórmula 2) e a amplitude corrente do próximo pulso do oscilador acentual. T é o período corrente do oscilador silábico. O *reset* de período é realizado via decaimento, pelo segundo termo da equação acima, $-\beta.(T-T_0).i(m-1)$, e só está presente como primeira aproximação, durante os dois primeiros VV do grupo acentual, com a taxa especificada por β . Este termo não está presente ao final do enunciado, pois não há um próximo pulso do oscilador acentual. Esta característica do modelo explica o alongamento final, ao final do enunciado, incluindo aquele dos segmentos pós-tônicos. O período corrente T é atualizado a cada ciclo do oscilador silábico, iterativamente, segundo (4).

$$T(k) = \text{período da VV} = T(k-1) + \Delta T \quad (4)$$

Os valores de w_0 , α , e β foram otimizados de forma a minimizar o erro entre os valores do período VV do oscilador silábico induzido (pelo oscilador acentual) e a média das durações VV para os grupos acentuais do *corpus* citado, nas três taxas de elocução, o que resultou nos seguintes valores: $w_0 = 0.78$, e $(\alpha, \beta) = (0.31, 1.13)$ para uma taxa de elocução rápida (T_0 até 128 ms), $(\alpha, \beta) = (0.38, 1.13)$ para uma taxa normal ($T_0 = 129$ ms a 176 ms) e $(\alpha, \beta) = (0.43, 1.04)$ para a taxa lenta (T_0 acima de 177 ms).

De posse desses valores, foi possível simular os valores dos períodos VV para frases quaisquer. Vejamos os resultados para os dois pares de frases abaixo, em conexão com o fenômeno do choque acentual em português brasileiro (PB).

3. Predição da duração da unidade VV em frases contrastando situação com presença e ausência de choque acentual

As frases da tabela 1 abaixo, extraídas de um *corpus* maior (ver Barbosa;aceito e Madureira;aceito para uma análise mais ampla) contrastam uma situação de choque acentual (A) com uma situação sem choque acentual (B), de acordo com a proposta de Liberman &

Prince (1977). Segundo esta proposta inauguradora da Fonologia Métrica, nas situações com choque acentual, a Regra do Ritmo (*Rhythm Rule*) pode ser opcionalmente usada para desfazer a sucessão de sílabas proeminentes, via inversão iâmbica (o padrão oxítono vira paroxítono). Estudos fonéticos tradicionais (veja Grabe & Warren;1995, Vogel, Bunnell & Hoskins;1995 e a crítica de Shatuck-Hufnagel;1995) mostraram que a inversão iâmbica não precisa necessariamente implicar num aumento dos valores dos parâmetros acústicos correlatos do acento na primeira posição (bor- em 5A e co- em 6A); mas, uma vez que o acento é uma propriedade relacional, a inversão iâmbica pode se dar pela diminuição dos valores dos parâmetros correlatos na segunda posição (-deaux em 5A e -mi em 6A). Em PB, a duração de uma unidade do tamanho da sílaba é o principal parâmetro correlato do acento (cf. Massini;1991, Barbosa;1996). O que uma avaliação das frases abaixo, produzidas por um locutor, e a comparação com a predição do modelo poderia revelar sobre o fenômeno?

Tabela 1: Frases para avaliação de choque acentual segundo o modelo dinâmico. Durações medidas em negrito. Acento frasal na primeira metade da frase sublinhado.

O **bordeaux** xucr|o derramou-se pela me|sa. (5A)

O **bordeaux** chinês d|erramou-se pela me|sa. (5B)

Eu **comi** bol|o sexta-feira à noite. (6A)

Eu **comi** bolor s|exta-feira à noite. (6B)

Na tabela 2 são comparadas ortogonalmente as médias das durações das unidades VV dos trechos em negrito da tabela 1, com as frases pronunciadas por um locutor do PB (cerca de 25 anos) a partir de dez repetições das mesma, bem como com os correspondentes períodos VV do oscilador silábico induzido gerados pelo modelo a partir das fórmulas acima. É preciso ter em mente que as durações preditas pelo modelo não devem coincidir com as naturais e apenas revelar um padrão geral de acentuação, tendo em vista que as demais partes do modelo dinâmico (pauta gestual integrada e interação prosódia-segmentos) não são consideradas aqui por estarem em fase de implementação.

Tabela 2: Médias de durações naturais e períodos VV preditos para os trechos em negrito das frases da tabela 1, em milissegundos. O asterisco indica diferença significativa (ANOVA, nível de significância = 5%) entre VV na mesma posição 5A vs 5B e 6A vs 6B.

	NATURAL	PREDITO
5A	173 [ord], 253* [of], 188 [ukr]	161, 212, 214
5B	165 [ord], 195* [of], 108 [in]	146, 163, 216
6A	101 [om], 164* [ib], 163 [ol]	161, 212, 214
6B	89 [om], 148* [ib], 160 [ol]	146, 163, 216

Centrando a atenção sobre as duas primeiras unidades VV da palavra-alvo, isto é, aquela que sofre as modificações sugeridas pela Regra do Ritmo (“bordeaux”, em 5; “comi”, em 6), observa-se que a duração é significativamente maior na segunda posição, para a condição de choque acentual (frase natural), exatamente o oposto do sugerido pela regra da Fonologia Métrica. Isto se dá tanto no contraste entre as frases 5A e 5B (respectivamente 253 e 195 ms) quanto no contraste entre 6A e 6B (respectivamente 164 e 148 ms). Observe que o padrão é o mesmo que aquele previsto pelo modelo (diferença de cerca de 30 %: 212 vs 163 ms para as frases 5 e 6). A explicação é que, em PB, a duração VV aumenta gradualmente, pelo mecanismo de acentuação (implementado no modelo via indução) até o VV frasalmente acentuado. No caso da situação de choque acentual (em termos tradicionais), o acento frasal está mais próximo da palavra-alvo, induzindo durações maiores para as unidades VV da mesma palavra. A situação é menos clara se a sílaba fonológica é a unidade envolvida (visto que não é uma unidade adequada para a observação de fenômenos prosódicos: Barbosa;1996), mas também não ocorre nenhuma das duas situações previstas pela Regra do Ritmo quando a duração da sílaba é estudada (cf. Barbosa;aceito).

Somente em situação de manipulação de acento tonal (*pitch accent*) parece haver algum indício de aplicação de uma regra semelhante à *Rhythm Rule*, pela inversão habitual das proeminências tonais em PB, na palavra oxítone, em apenas um dos nove pares de frases estudados por Barbosa (aceito) com quatro locutores. A compreensão completa do fenômeno passaria por uma análise mais minuciosa da entoação do PB do que aquela que tem sido feita até agora. Esse fenômeno que parece ser completamente opcional e relativamente raro na frase lida de laboratório não contradiz o modelo, que se ocupa apenas em prever os valores de duração. Porém sugere a necessidade da integração da previsão da frequência fundamental em um modelo de ritmo mais abrangente.

4. Conclusões e perspectivas

O modelo dinâmico de produção do ritmo da fala que estamos propondo é capaz de dar conta dos dados de duração natural, predizendo com precisão (mesmo sem estar completo) o padrão global de acentuação encontrado em frases naturais em condição de suposto choque acentual. O que se depreende é que, línguas que manipulam duração para assinalar acentos lexical e frasal não desfariam este aparente choque acentual, pois a duração tende a aumentar continuamente até a posição do acento frasal. Se numa determinada frase o acento frasal está mais próximo de uma palavra do que de outra em contraste *ceteris paribus*, a taxa de aumento de duração na palavra mais próxima deve ser maior.

É claro que análises de duração em outras línguas que controlam a duração para assinalar acentos lexicais e/ou frasais (como o francês e o castelhano) precisam ser efetuadas para verificar a sustentabilidade das hipóteses aqui apresentadas. Além disso, um estudo da acentuação secundária em PB fornecerá subsídios adicionais para testar a capacidade explicativa do modelo. Este estudo já foi iniciado (Arantes & Barbosa;aceito).

Agradecimentos

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

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBANO, E. C. (2001) *O Gesto e suas Bordas: por uma fonologia acústico-articulatória*. Campinas: Mercado de Letras.
- ALLEN, G. D. (1972) The location of rhythmic stress beats in English I & II. *Language & Speech*, **15**: 72-100, 179-195.
- ARANTES, P. & P. A. BARBOSA (aceito) Acentuação secundária em português brasileiro à luz de um modelo dinâmico do ritmo: um estudo-piloto. *Anais do I Congresso Internacional de Fonética e Fonologia/VII Congresso Nacional de Fonética e Fonologia*. 28-30 de outubro de 2002. Belo Horizonte.
- BARBOSA, P.A. (1996) At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation. *Cadernos de Estudos Linguísticos*, **31**: 33-53.
- _____. (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 Eurospeech 2001*, Ålborg, Denmark. v. 2, 967-970.

- _____ (2002) Explaining Cross-Linguistic Rhythmic Variability via a Coupled-Oscillator Model of Rhythm Production. In: BEL, B. & I. MARLIEN (eds.) *Proceedings of the Speech Prosody 2002 Conference*, 11-13 April, Aix-en-Provence: Laboratoire Parole et Langage. 163-166.
- _____ (aceito) Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production. In: Barbosa, P. A. (org.) *Prosody analysis and modeling*.
- _____ & 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.
- BROWMAN, C. & L. GOLDSTEIN (1990) Tiers in Articulatory Phonology with some Implications for Casual Speech. In *Papers in Laboratory Phonology I*, Kingston, J. & Beckman, M. E., eds. Cambridge: Cambridge University Press, 341-376.
- DOGIL, G. & G. BRAUN (1988) *The PIVOT model of speech parsing*. Wien: Verlag.
- FUJIMURA, O. (1995) Prosodic organization of speech based on syllables: the C/D model. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, **3**: 10-17.
- GRABE, E. & P. WARREN (1995) Stress Shift: do speakers do it or do listeners hear it? In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, Connell, B. & Arvaniti, A., eds. Cambridge: Cambridge University Press, 95-110.
- 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.
- LIBERMAN, M. & A. PRINCE (1977) On stress and linguistic rhythm. *Linguistic Inquiry* **8.2**: 249-336.
- MADUREIRA, S. (aceito) An Acoustic Study of Phonological Phrases Containing Sequences of Words with Adjacent Primary-Stressed Syllables: does stress shift occur in Brazilian Portuguese? In: Barbosa, P. A. (org.) *Prosody analysis and modeling*.
- MASSINI, G. (1991) *A duração no estudo do acento e do ritmo em português*. Dissertação de Mestrado inédita. Campinas: Universidade de Campinas
- MCAULEY, J.D. (1995) *Perception of time as phase: toward an adaptative-oscillator model of rhythmic pattern processing*. Tese de Doutorado Inédita. Indiana, EUA: Indiana University.
- MORTON, J., S. MARCUS & C. FRANKISH (1976) Perceptual centers (p-centers). *Psychological revue*, **83.5**: 405-408.
- ÖHMAN, S. (1966) Coarticulation in VCV utterances: spectrographic measurements. *J. Acoustic. Soc. Am.*, **39**: 151-168.
- 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.
- SHATTUCK-HUFNAGEL, S. (1995) The importance of prosodic transcription in empirical approaches to “stress shift” versus “early accent”: comments on Grabe and Warren, and Vogel, Bunnell and Hoskins. In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, Connell, B. & Arvaniti, A., eds. Cambridge: Cambridge University Press, 128-140.
- VOGEL, I.; H. T. BUNNELL & S. HOSKINS (1995) The Phonology and Phonetics of the Rhythm Rule. In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, Connell, B. & Arvaniti, A., eds. Cambridge: Cambridge University Press, 111-127.