

Implementação em *Praat* de algoritmos para descrição de correlatos acústicos da prosódia da fala

Pablo Arantes¹

¹Universidade Federal de Minas Gerais (Prodoc/Capes)

Abstract. *Algorithms for the description of acoustic correlates of prosody such as acoustic duration, fundamental frequency and spectral emphasis are presented. The algorithms are implemented as scripts for the open source acoustic analysis program Praat which are made available through a public repository.*

Resumo. *Este trabalho apresenta algoritmos para a descrição de três parâmetros acústicos que são considerados correlatos de características prosódicas: a duração acústica, a frequência fundamental e ênfase espectral. Todos os algoritmos foram implementados como scripts do programa de análise acústica Praat e estão disponíveis para o público em um repositório.*

1. Introdução

Tornou-se corrente no campo da linguística a percepção de que a prosódia, assim como a composição segmental, é um elemento indispensável para o entendimento integral da língua, tanto na sua face abstrata quanto na descrição de sua manifestação física, a fala. O interesse crescente na prosódia pode ser explicado pelo reconhecimento do seu papel como o nível de descrição e explicação linguística em que o componente sonoro da língua faz interface com outros componentes da gramática como a sintaxe, a semântica e a pragmática [Cutler et al. 1997].

Do ponto de vista da descrição linguística, analisar fenômenos prosódicos implica mostrar a presença de contrastes ou proeminências relacionais entre unidades linguísticas (que podem ser vogais, sílabas ou outras, a depender do quadro teórico adotado) organizadas hierarquicamente e projetadas no plano sintagmático. Este tipo de procedimento é característico tanto dos estudos de natureza fonológica [Lieberman and Prince 1977] quanto fonética [Nooteboom 1997].

Neste trabalho, abordaremos a questão da descrição da prosódia do ponto de vista fonético. Os algoritmos apresentados a seguir tem por objetivo facilitar a obtenção e processamento de dados que podem ser usados para estabelecer, para diferentes correlatos acústicos, possíveis padrões de proeminência relativa de unidades linguísticas que caracterizam diferentes fenômenos prosódicos. Nas seções a seguir são apresentados algoritmos para a descrição de três correlatos acústicos da prosódia: duração, frequência fundamental e ênfase espectral. Todos os algoritmos descritos foram implementados na forma de *scripts* do programa de análise acústica de código aberto Praat [Boersma 2001] e são disponibilizados sob a licença GNU/GPL no repositório Praat-tools, acessível pelo endereço <http://code.google.com/p/praat-tools/>.

2. Duração

O algoritmo descrito nesta seção permite delimitar de forma automática os grupos acen-tuais de um trecho arbitrário de fala em análise a partir da duração acústica bruta de

segmentos agrupados em unidades de vogal a vogal (unidades V-V)¹. Essa delimitação é o produto final da aplicação de três procedimentos: normalização da duração bruta por *z-score* estendido, suavização do contorno de duração normalizado por meio da aplicação de uma função de média móvel de cinco pontos e detecção de picos no contorno de duração suavizada. Em [Barbosa 2006] pode-se encontrar a formulação original e a motivação para cada um dos procedimentos que o compõem. No livro encontra-se também o código-fonte para implementação do algoritmo na forma de *scripts* do Praat. O autor do presente trabalho disponibiliza versões alternativas desses *scripts* no repositório Praat-tools.

A adoção do procedimento de normalização da duração por *z-score* no contexto de experimentos prosódicos é justificada em razão da natureza relativa das proeminências. Como o interesse é comparar a força relativa de unidades contíguas, se a duração das unidades for comparada sem nenhum tipo de transformação prévia, o resultado pode ser distorcido pela influência da duração intrínseca, que pode encobrir a influência do contexto prosódico sobre a duração dos fones. Os demais procedimentos são aplicados para minimizar ainda mais a influência de fenômenos duracionais locais, como a duração intrínseca, e enfatizar os movimentos duracionais motivados pela atribuição de proeminências de natureza prosódica. Fenômenos prosódicos, como o acento secundário [Arantes 2010], o encontro acentual [Barbosa et al. 2004], a influência da taxa de elocução sobre a organização prosódica [Meireles 2003] e a marcação prosódica de fronteiras sintáticas [Barbosa 2006] foram estudados com sucesso por meio da aplicação deste algoritmo.

A Figura 1 mostra a aplicação do algoritmo de análise duracional no trecho delimitado por colchetes em uma leitura da oração “[Os edessenos budistas] parecem menores”. A subfigura (a) mostra a duração bruta das oito unidades V-V existentes no trecho delimitado, [uz.ed.es.'en.ʊzb.uç.'ist.ɐs], medidas manualmente. As subfiguras (b) e (c) mostram, respectivamente, o mesmo contorno de duração após a aplicação do procedimento de normalização e suavização (c). O resultado da aplicação dos dois procedimentos ao contorno bruto de duração tem como resultado evidenciar a presença de duas saliências no contorno de duração analisado. O procedimento de detecção de picos identifica a presença de uma proeminência na quarta unidade V-V, que contém a vogal tônica da palavra “edessenos”, uma proeminência mais forte na unidade V-V de índice 7, que contém a vogal tônica da palavra “budistas”, além de informar que a primeira unidade V-V também é saliente. Os trechos de enunciado entre dois picos consecutivos de duração suavizada são chamados, na metodologia desenvolvida em [Barbosa 2006], de grupos acentuais.

3. Frequência fundamental

A frequência fundamental (F_0) é frequentemente apontada como um dos principais correlatos acústicos da prosódia [Nooteboom 1997]. Nesta seção serão brevemente descritos dois algoritmos para a descrição semi-automática de padrões de F_0 que podem ser usados para estudar uma gama variada de fenômenos.

3.1. Contornos normalizados temporalmente

A normalização temporal é uma técnica, empregada na literatura recente por [Xu 1999], que consiste em extrair um número fixo, definido pelo analista, de amostras de F_0 por in-

¹Unidade delimitada entre dois *onsets* vocálicos consecutivos. Ver [Barbosa and Madureira 1999] e [Barbosa 2006] para uma justificativa desta forma de agrupamento.

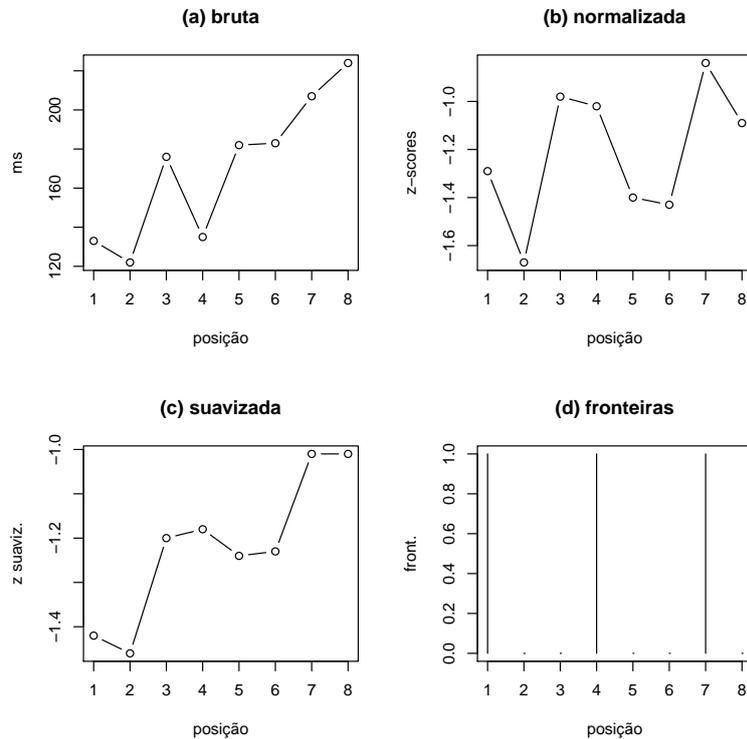


Figura 1.

intervalo de análise, que pode ser a sílaba, a unidade V-V, a palavra ou frase, por exemplo, de modo a permitir a comparação por sobreposição de curvas de F_0 com durações diferentes. Em cada intervalo de análise a F_0 é amostrada a intervalos isócronos e proporcionais à duração do intervalo. A sobreposição de curvas permite, por exemplo, que se avalie a magnitude e o alinhamento de movimentos de F_0 em diferentes condições experimentais. Entre os fenômenos prosódicos do português brasileiro que a técnica permitiu descrever estão o acento secundário [Arantes and Barbosa 2009] e a marcação de informação nova e dada [Arantes 2011].

A Figura 2 mostra a aplicação da técnica de normalização temporal em duas curvas de F_0 . No quadro superior e central, a curva de F_0 bruta é mostrada em cinza e as curvas contínuas em azul e verde são versões suavizadas pelo algoritmo *Smooth* do programa Praat. Os círculos em vermelho indicam a localização dos cinco pontos da curva amostrados em cada unidade V-V do sintagma nominal (SN) “a peregrinação”. A condição contrastada neste exemplo é o *status* informacional do referente expresso pelo SN analisado, isto é, a diferença entre informação nova (quadro superior) e informação dada (quadro central). Como há seis unidades V-V no trecho analisado e 5 amostras de F_0 por intervalo, a curva gerada pelo algoritmo é composta, neste caso, por trinta pontos. O quadro inferior mostra a superposição das duas curvas normalizadas temporalmente, o que permite a comparação do comportamento das duas em cada unidade V-V. No caso dessas duas curvas, a aplicação da normalização temporal permite observar que a principal diferença no padrão de F_0 entre as duas condições acontece no início do SN e não na unidade V-V que contém a vogal tônica do substantivo. Se várias curvas normalizadas forem obtidas para cada condição, o que frequentemente é o caso num cenário de pesquisa

linguística, recomenda-se fazer a média dos valores coletados para cada amostra de modo que sejam comparadas as curvas médias e não a nuvem de curvas individuais.

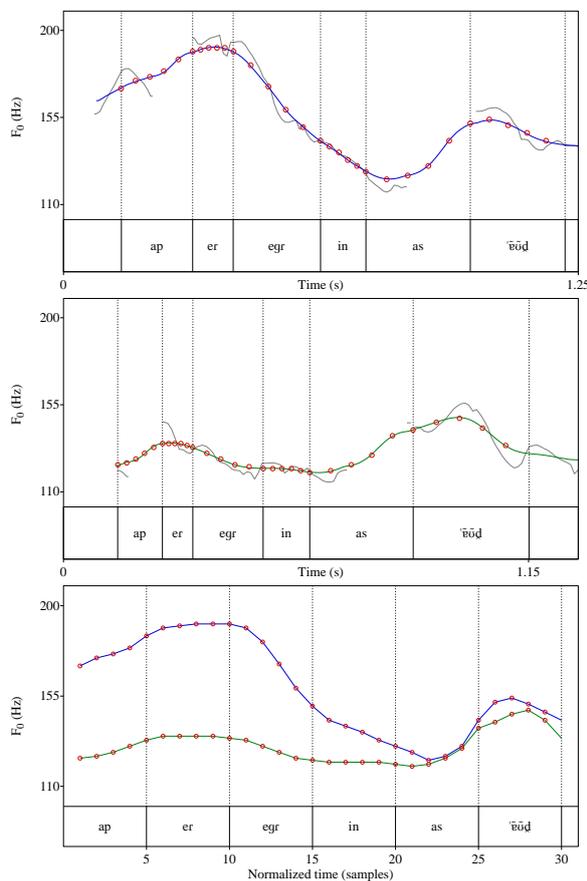


Figura 2.

3.2. Pontos máximos e mínimos

O algoritmo descrito nesta seção tem como objetivo encontrar os pontos de inflexão da curva de F_0 , isto é, seus máximos e mínimos locais ou globais, chamados, em conjunto, de pontos extremos. Os pontos extremos da curva de F_0 são considerados relevantes porque identificam os movimentos ascendentes e descendentes do contorno entoacional cuja produção pode ter sido linguisticamente motivada. No algoritmo descrito aqui, os máximos e mínimos são identificados por meio do teste da mudança de sinal da velocidade de F_0 (ou primeira derivada). O algoritmo também registra o alinhamento dos pontos identificados em relação ao material segmental (obtido por meio da anotação manual das fronteiras de sílabas, por exemplo) e seu valor em Hertz. Estas informações podem ser utilizadas para descrever os padrões de distribuição de movimentos ascendentes e descendentes de F_0 ao longo de trechos de fala de interesse.

Na Figura 3, o painel superior mostra a curva de F_0 suavizada da oração “O gonococo parece menor hoje”. No centro, pode-se ver as fronteiras das cinco unidades V-V delimitadas no SN “O gonococo”. O painel inferior mostra a curva da velocidade de F_0 . Nos painéis superior e inferior as linhas ascendentes (verdes) indicam a localização dos pontos máximos (os picos) e as linhas descendentes (vermelhas) indicam os pontos mínimos (os vales). A marcação dos agrupamentos segmentais (unidades V-V, aqui) permite que se identifique o alinhamento dos pontos máximos e mínimos com o material segmental e a contagem do número de picos de um determinado tipo.

Na descrição dos possíveis correlatos acústicos do acento secundário em português brasileiro apresentada em [Arantes and Barbosa 2009], a aplicação do algoritmo permitiu verificar que pode ocorrer ou não a presença de picos de F_0 na cadeia de pré-tônicas. Nas palavras com cadeia menor (duas sílabas) a ocorrência de picos alinhados com as pré-tônicas é menos frequente do que nas palavras com cadeia maior (3 ou 4 sílabas) e que são muitos raros os casos em que mais de um pico ocorre entre as pré-tônicas. Este achado desafia a descrição intuitiva do acento secundário em português que prevê a presença sistemática de mais de uma proeminência em palavras com cadeia de pré-tônicas com quatro sílabas.

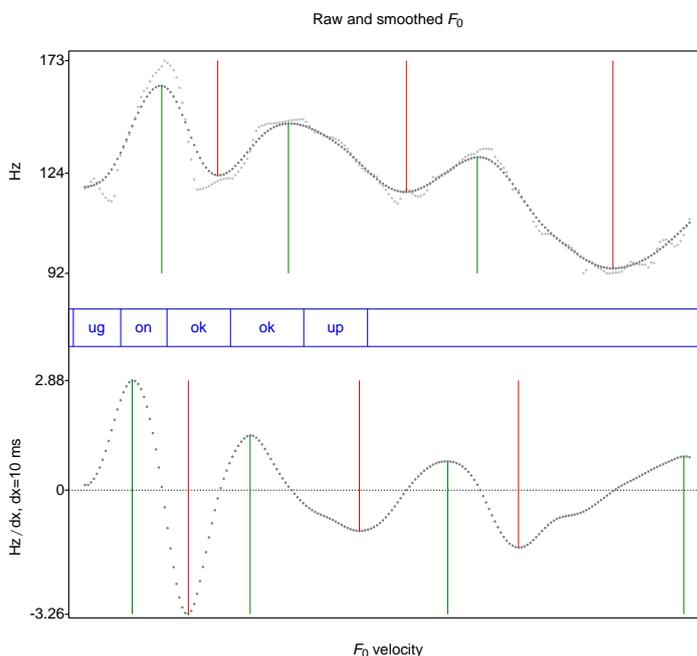


Figura 3.

4. Ênfase espectral

O interesse desta medida acústica para a descrição da prosódia se deve à observação experimental de que vogais em posições prosodicamente proeminentes tendem, quando comparadas às não proeminentes, a ser produzidas com mais esforço vocal, o que, do ponto de

vista acústico, resulta em uma distribuição da energia ao longo do espectro caracterizada pela presença de mais energia em faixas de mais alta frequência. Existem muitas formas de operacionalizar esta definição e a que foi adotada para implementação é a proposta em [Traunmüller and Eriksson 2000]. Nesta proposta a ênfase espectral é definida, para os sons vocálicos, como a diferença entre a intensidade acústica do sinal integral e a intensidade do sinal submetido a um filtro passa-baixas com um limite de banda superior definido pela expressão $1,5 \cdot F_0$, em que F_0 é a média da frequência fundamental na vogal sendo analisada. [Heldner 2003] sugere modificar o algoritmo de forma que a ênfase seja calculada em janelas de análise que avançam ao longo do sinal acústico e que, em cada janela, o limite superior do filtro seja definido dinamicamente, acompanhado as mudanças da F_0 no trecho analisado. O autor não disponibiliza publicamente uma ferramenta computacional que implementa sua proposta. O script para o cálculo de ênfase que o autor do presente trabalho disponibiliza no repositório Praat-tools é uma implementação independente do algoritmo descrito em [Heldner 2003].

Referências

- Arantes, P. (2010). *Integrando produção e percepção de proeminências secundárias numa abordagem dinâmica do ritmo da fala*. PhD thesis, Universidade Estadual de Campinas, Campinas–SP.
- Arantes, P. (2011). Correlatos prosódicos da estrutura informacional em português brasileiro. Trabalho apresentado no III Colóquio Brasileiro de Prosódia da Fala.
- Arantes, P. and Barbosa, P. (2009). Duration and fundamental frequency patterns in brazilian portuguese polysyllabic words. *Journal of the Acoustical Society of America*, 125(4):2573.
- Barbosa, P. A. (2006). *Incursões em torno do ritmo da fala*. Pontes, Campinas.
- Barbosa, P. A., Arantes, P., and Silveira, L. S. (2004). Unifying stress shift and secondary stress phenomena with a dynamical systems rhythm rule. In *Proceedings of the Speech Prosody 2004 Conference*, pages 49–52, Nara, Japan.
- Barbosa, P. A. and Madureira, S. (1999). Toward a hierarchical model of rhythm production: Evidence from phrase stress domains in brazilian portuguese. In *Proc. of the XIVth International Congress of Phonetic Sciences*, pages 297–300, San Francisco.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10):341–345.
- Cutler, A., Dahan, D., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40(2):141–201.
- Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in swedish. *Journal of Phonetics*, 31:39–62.
- Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2):249–336.
- Meireles, A. R. (2003). *Reestruturações rítmicas da fala no português brasileiro*. PhD thesis, Universidade Estadual de Campinas, Campinas–SP.

- Nooteboom, S. G. (1997). The prosody of speech: Melody and rhythm. In Hardcastle, W. J. and Laver, J., editors, *The Handbook of Phonetic Sciences*, pages 640–673. Blackwell Publishers, Oxford.
- Traunmüller, H. and Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America*, 107(6):3438–3451.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, 27:55–105.