

# A variação tonal em português brasileiro a partir de frases espontâneas e textos lidos

Waldemar Ferreira Netto<sup>1</sup>, Daniel Oliveira Peres<sup>1</sup>

<sup>1</sup>Departamento de Letras Clássicas e Vernáculas – Universidade de São Paulo (USP)  
[wafnetto@usp.br](mailto:wafnetto@usp.br) [danielperes@usp.br](mailto:danielperes@usp.br)

**ABSTRACT:** *We verify the tonal variation in spontaneous speech (ES) and reading speech (RS). The analysis presented in this work is based on principles of the musical analysis. These principles claim that the manifest cultural particularities in the musical tradition of people are reflected in the prosody (Glaser, 2000; Schellenberg; Treh, 1999). The statistical results demonstrated that RS differs significantly from ES in relation to the occurrence of tonal rhythm (R) ( $P < 0,001$ ), but the same does not occur between them in relation to the intensity rhythm (I). In ES, the difference among the occurrences of (I) and (R) was not significant either, but it was in RS ( $P < 0,001$ ). On the one hand, the presence of (R) in ES points out a bigger control of the phrasal declination, which makes the speaker keep the tone of his/her speech steady. On the other hand, in ES, the necessity of the lexical election seems to concur with intonation in the sentence construction.*

**Keywords:** *Phonetics, prosody, tonal variation*

**RESUMO:** *Verificamos como ocorre a variação tonal em textos de fala espontânea (FE) e de leitura (TL). A análise apresentada neste trabalho baseia-se nos princípios da análise musical, em que as particularidades culturais manifestas na tradição musical de um povo refletem-se na prosódia (Glaser, 2000; Schellenberg; Treh, 1999). Os resultados estatísticos demonstraram que TL difere significativamente de FE quanto à ocorrência de ritmo tonal (R) ( $P < 0,001$ ), mas o mesmo não ocorre entre elas quanto ao ritmo de intensidade (I). Em FE, a diferença entre as ocorrências de (R) e (I) também não foi significativa, mas foi em TL ( $P < 0,001$ ). De um lado a presença de (R) em FE aponta para um maior controle da declinação frasal, fazendo com que o falante procure sempre manter o tom de sua fala estável. Por outro lado, em FE, a necessidade da seleção lexical pode estar concorrendo com a entoação na construção da frase.*

**Palavras-chave:** *Fonética, prosódia, variação tonal*

## Introdução

O propósito deste trabalho é realizar uma análise prosódica a partir dos princípios básicos da análise musical, tomando a hipótese de que as particularidades culturais manifestas na tradição musical de um povo refletem-se na prosódia (GLASER, 2000; SCHELLENBERG; TREH, 1999). Uma das características próprias da interpretação

musical é a possibilidade da definição de uma coleção de tons cuja harmonia se estabelece a partir de um tom fundamental (SCHOENBERG, 2001).

Segundo esse princípio, uma sequência melódica é formada por diferentes trajetórias que envolvem afastamentos e aproximações do tom fundamental e criam sempre a expectativa de que o ponto final dessa trajetória seja esse mesmo tom fundamental (LONGACRE; CHENOWETH, 1986). Essa coleção de sons harmônicos definida a partir do tom fundamental se fixa pela adaptação mútua dos movimentos moleculares que formam as ondas sonoras desses mesmos sons. Movimentos harmônicos cuja periodicidade estabeleça ciclos em que a superposição adequa-se aos movimentos moleculares do fundamental, criando nodos em posições idênticas aos dele (ROEDERER, 2002), sendo mais harmônicos do que aqueles em que essa superposição seja menos adequada. Assim, frequências dobradas, cuja superposição é ótima, são as que têm maior harmonia com o fundamental; frequências três, quatro, cinco, seis vezes, e assim sucessivamente, estabelecem os diferentes graus de semelhança tonal.

Para analisar o produto dessas variações, partiremos da proposta de Waldemar Ferreira Netto (2006) para a análise prosódica do português, a saber, a prosódia como série temporal. Caracteriza-se uma série temporal como um conjunto de observações sequenciadas e interdependentes, sendo o resultado da observação feita no momento  $t+1$  condicionada ao momento  $t$  (FERREIRA NETTO, 2006). Assim, uma onda sonora forma uma série temporal de modo que os valores medidos em um dado momento são dependentes do valor imediatamente anterior.

## **O tratamento dado à entoação frasal**

A análise da entoação frasal tem sido objeto de diversos trabalhos. Troubetzkoy (1964 [1939]) propôs que a sintaxe se relacionasse diretamente com a variação de frequência no correr da frase, estabelecendo sentidos suplementares e atuando com função conclusiva, de continuidade e enumerativa. Pike (1954; 1964) propôs que a entoação frasal no inglês norte-americano fosse interpretada por uma escala de 4 graus contínuos. A proposição de Pike (1964) foi retomada por diversos autores: Maeda (1971), Pierrehumbert (1980; 1990), Cruttenden (1986) e Ladd (1996), dentre outros. Pierrehumbert (1980) propôs uma revisão dos quatro níveis de Pike (1954), especificando H(igh) e L(ow) com possibilidades combinatórias diversas, além de marcas morfológicas e sintáticas complementares.

Martinet (1970) diz que os fatos prosódicos não possuem valor quanto à sua presença ou ausência, tornando-os não-analisáveis paradigmaticamente:

Incluem-se na prosódia todos os fatos de discurso que não entram na fonemática, quer dizer, os que de qualquer modo escapam à segunda articulação [...] nestas condições, não é difícil compreender que tais fatos não possam valer linguisticamente pela sua presença ou ausência em determinado ponto, mas sim pelas suas modalidades, variáveis de parte para parte dos enunciados, por

consequência, não podem caracterizar unidades discretas (MARTINET, 1970, p. 82).

Para Hjelmslev, o texto pode ser fracionado em unidades cada vez menores, chegando-se ao segmento. Há elementos que não se limitam a um único segmento, são os constituintes prosódicos, pois são caracterizadores do texto, mas não o constituem. Tais elementos são chamados, por Troubetzkoy (1964 [1939]), de *prosodemas*, divididos em modulações (ao longo do sintagma) e acento (no interior da palavra).

Na década de 70, Liberman & Prince (1977) propõem a *Fonologia Métrica*, considerando uma análise binária para as sílabas. Diferentemente do modelo de Chomsky & Halle (1968), o acento passou a ser tratado na Fonologia Métrica como uma propriedade relacional das sílabas, tendo a sua representação dada por árvores métricas. A fonologia métrica recebeu algumas alterações de Halle & Vergnaud (1987), que descartaram a dupla representação (arbórea e numérica); e Hayes (1994) que propõe uma teoria métrica do acento, aplicável às palavras e às sentenças (marcando *stress* e *accent* respectivamente).

A Fonologia Prosódica, modelo proposto por Nespor & Vogel (1986), traz consigo a preocupação, antes não considerada pela Fonologia Métrica, de relacionar o acento e o ritmo com outros componentes fonológicos. Para Nespor & Vogel (1986) os princípios definidores dos constituintes prosódicos são estudados por meio de uma interface entre a fonologia e outros componentes gramaticais, mas sem que haja uma correspondência biunívoca entre os constituintes prosódicos e os gramaticais. As autoras propõem sete domínios numa escala hierárquica prosódica: **U** *Phonological Utterance* (enunciado fonológico), **I** *Intonational Phrase* (sintagma entoacional), **Φ** *Phonological Phrase* (sintagma fonológico), **C** *Clitic Group* (grupo clítico), **ω** *Phonological Word* (palavra fonológica), **Σ** *Foot* (pé) e **σ** *Syllable* (sílabas).

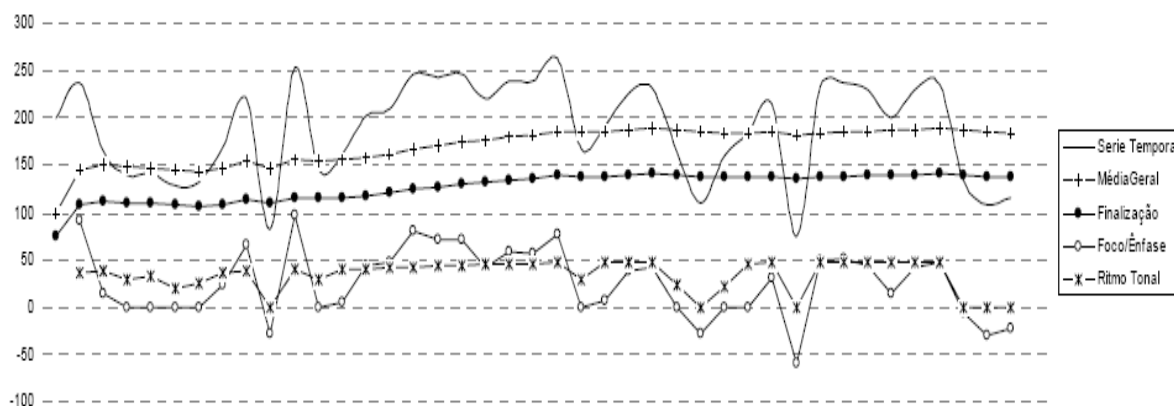
Para Fónagy (2003), a articulação prosódica da mensagem é a condição fundamental da interpretação, pois a leitura de manuscritos lineares, sem intervalos e pontuação, atesta a importância da prosódia pela sua ausência. Os signos prosódicos são organizados por Fónagy (2003) como portadores de funções. Dentre elas estão: *função demarcativa, função enfática, função sintática, função modal, função imitativa, função apelativa, função preventiva, função alusiva, função estética e função expressiva*. Fónagy estipula uma hierarquia das funções que varia de acordo com a tipologia discursiva.

O português do Brasil teve sua entoação interpretada pelos trabalhos de Cagliari (1982; 2007), Barbosa (1995), Moraes (1998), dentre outros. Todas as propostas assumem a hipótese de Troubetzkoy (1964 [1939]) de que a entoação seja uma função direta das variações globais de F0. No presente trabalho, a abordagem proposta para a prosódia do português brasileiro leva em conta o fato de que uma série temporal constitui-se por fenômenos simultâneos e não-hierarquizados, sendo de suma importância a manutenção dos valores relativos aos pontos tomados para avaliação.

## A abordagem da entoação frasal como série temporal

As variações de frequência ocorridas na frase, ou seja, a sequência de medições obtida, além de ser um conjunto de posições sucessivas temporalmente, pode ser fracionada em outras componentes que somadas seriam a própria prosódia. Tomando-se, por exemplo, um determinado valor da série temporal, teremos uma soma das componentes que juntas concorreram para a sua formação e valor. As componentes estão divididas em dois grupos: *componentes estruturadoras* e *componentes semântico-funcionais*. A primeira delas é subdividida em *declinação* (D) e *ritmo tonal* (R), sendo essas subdivisões caracterizadas como o esforço que o falante despende para a produção da fala. Na produção da fala, o falante tende a ter um esgotamento da energia necessária para a fala, fazendo com que ele reaja a essa tendência retomando o tom anterior.

A segunda das componentes é composta por *foco/ênfase* (E) e *acento lexical* (A)<sup>1</sup>, tendo um caráter mais linguístico do que composicional. A componente (A) aparentemente não é importante para a formação da série temporal, levando-se em conta que a variação de frequência não é correlato na marcação do acento lexical em português (MASSINI-CAGLIARI, 1992). Por outro lado, a componente *foco/ênfase* (E) é a maior portadora de significado, pois está diretamente condicionada às necessidades expressivas do falante, tendo como característica a não-previsibilidade, apesar de manter-se dentro de limites que possam ser apreciados numa análise automática.

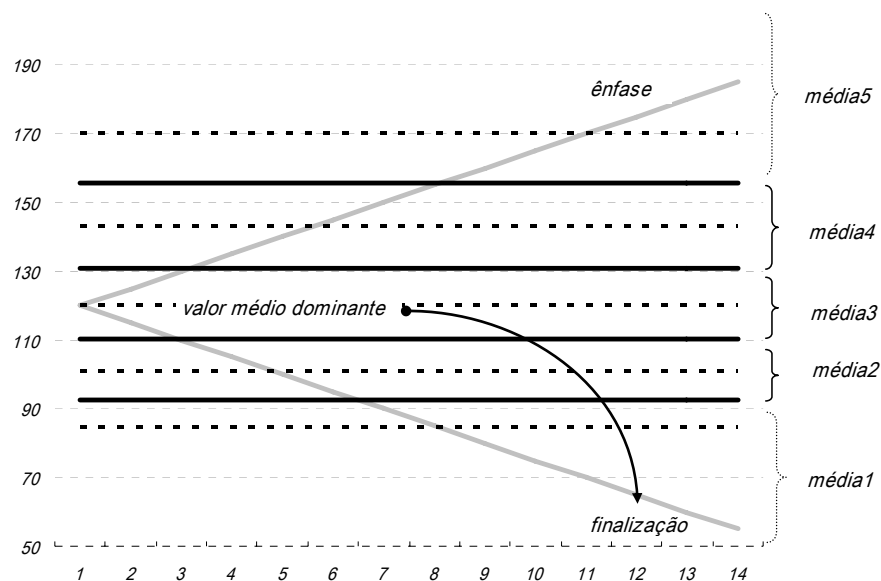


**Figura 1:** A linha contínua sem marcadores mostra a série temporal principal da fala (Z); a linha com marcadores como mais mostra o tom médio; a linha com marcadores circulares mostra a componente declinação (D) ou tom final; a linha com marcadores de asterisco mostra da componente ritmo tonal (R) e, finalmente, a linha, com marcadores de círculos vazados mostra a componente foco/ênfase (E)

<sup>1</sup> Há algumas mudanças entre a proposta original de Ferreira Netto (2006) e a atual, na medida em que a componente entoação ora substitui-se pela componente acento lexical.

Inspirando-se no estudo de Patrícia Kuhl (KUHL; IVERSON, 1995; KUHL, 2000; KUHL *et al.*, 2001) que postula uma forma prototípica e não-abstrata fixa que sirva de parâmetro para as demais formas percebidas, estabeleceu-se uma organização tonal da prosódia que leva em conta uma categoria medial a partir do modelo de *perceptual magnet effect* (PME). A forma prototípica do falante, segundo a autora, seria algum dos sons ouvidos em alguma etapa da vida do falante que, memorizada ao todo, serve como pólo de atração de todos os estímulos que atinjam regiões próximas. Kuhl (2001) faz referência quanto ao grau de aceitação dessas formas para o estabelecimento das formas prototípicas, propondo que o fenômeno de dispersão das vogais não é gestáltico, mas fruto do PME.

A partir dessa visão, imagina-se que a melhor frequência ou a melhor intensidade é a média de tudo o que acontece, supondo-se que as médias de frequência e intensidade sejam a intenção do falante, ou seja, a categoria *ótima* (goodness) e menos marcada. Assim, postulamos uma escala de cinco tons baseada numa categoria medial de frequência encontrada na voz do falante ao longo da frase.



**Figura 2:** Gráfico em que se mostram as bandas estabelecidas pelas médias de frequência, a partir de um valor médio dominante de 120Hz. As linhas pretas pontilhadas mostram os valores médios de cada banda (FERREIRA NETTO, 2006)

A divisão das bandas de frequência foi baseada na hipótese de T'Hart (1981) sobre a percepção das variações de frequências produzidas na fala, a qual se manteve em torno de 1,5 tom ou 3st (semitons). Justifica-se o uso de bandas de frequência por existir na fala grandes variações não-pontuais que dificultariam uma possível análise baseada na

divisão tonal da escala musical. O fato de as variações *microtonais* serem descartadas é parte da intenção de fazer dessa escala algo que possa abarcar de maneira geral as variações tonais da fala, não interessando, a princípio, sua variação intrabanda. Seguindo de maneira análoga o que foi previsto para a frequência, a intensidade será analisada a partir de um valor medial, em que haja dois valores para a sua variação, ou seja, um valor de intensidade menor e outro maior que o central.

A definição dos parâmetros ligados aos limites de frequência adotados para análise automática tem como base o trabalho de Russo e Behlau (1993) que afirmam que falantes masculinos do português brasileiro possuem uma frequência fundamental em torno de 105 Hz.

## Metodologia

A análise dos dados foi feita a partir da rotina de análise automática **ExProsodia** desenvolvida para o projeto *Análise automática da entoação na fala em língua portuguesa*. O levantamento dos dados foi feito por meio do software *Speech Filing System*, desenvolvido por Marc *Huckvale* da University College London. Por meio dessa ferramenta, extraímos arquivos em *txt* com medições de frequência e intensidade num intervalo de 5ms. Esses dados são analisados pela rotina que os categoriza segundo critérios pré-estabelecidos. A rotina inicia suas operações, estabelecendo a média geral das frequências para os valores válidos para os candidatos a pico silábico. Os valores válidos são pré-definidos como:

Limiar inferior de frequência: 50 Hz

Limiar superior de frequência: 350 Hz. Opção do usuário.

Limiar inferior de duração: 4 frames ou 20 ms (1 frame = 5 ms).

Limiar superior de duração: 10 frames ou 50 ms (1 frame = 5 ms)

Limiar de intensidade: 2000 RMS. Opção do usuário.

*Valores utilização para a elaboração da escala de cinco tons:*

Limite superior das frequências médias => valor médio \* 1,09)

Limite inferior das frequências média => valor médio / 1,09)

Distância entre cada média (3st =  $1,0595^3=1,19$ )

*Valores utilizados para a elaboração da escala de intensidade:*

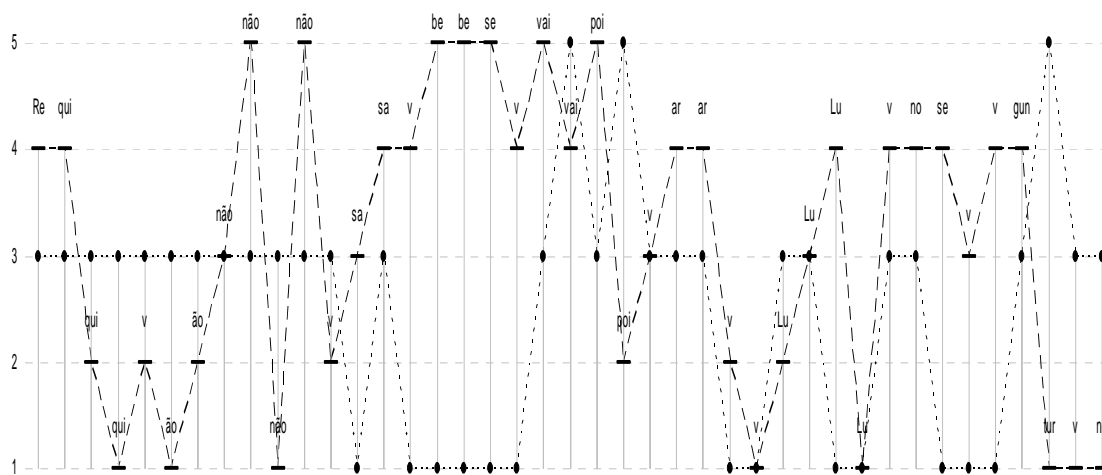
Limite superior do valor médio de intensidade (1,50)

Limite inferior do valor médio de intensidade (0,5)

Limite entre cada valor de intensidade= 1,25 sup e 0,5 inf

Categorização da intensidade na escala (= 1 ou 3 ou 5)

O processo de análise automática parte de um valor medial, como dito acima, de cada frase ou segmento de fala, sendo este a média aritmética acumulada ao longo do tempo em todas as frequências válidas, isto é, que estejam de acordo com os parâmetros estipulados para a análise automática. Todos os demais valores relativos aos limites de cinco tons serão baseados no valor medial dominante. Com a série de cálculos realizados pela **ExProsodia** gera-se, dentre outros, o seguinte gráfico que ilustra o comportamento tonal da fala juntamente com as variações de intensidade:



**Figura 3: Gráfico produzido pela rotina ExProsodia v1.03. Os marcadores pretos de tracinhos mostram a entoação na escala de cinco tons e os marcadores cinza, circulares a intensidade, na escala de graus**

O *corpus* foi formado por 300 frases assertivas divididas igualmente entre as modalidades *texto lido* (TL) e *fala espontânea* (FE) obtidas em gravações radiofônicas retiradas da Internet<sup>2</sup>. Para esta análise, foram coletadas cinco frases de 60 sujeitos. A avaliação da presença ou não de ritmo foi feita pelo teste de subsequências (LEVIN, 1985), em cada uma das frases obtidas para o corpus. Esse procedimento já foi testado em outro trabalho que realizamos (FERREIRA NETTO, 2006).

<sup>2</sup>Dados coletados em: <http://www.radiojornalismo.com/midia.htm>.

## Resultados

A partir dos resultados obtidos, considerou-se a ocorrência de ritmo tonal ou de ritmo de intensidade em relação ao conjunto de frases de cada sujeito analisado. Na Tabela abaixo, estão dispostas as quantidades de frases em que se constatou a presença de ritmo de intensidade e ritmo tonal por sujeito.

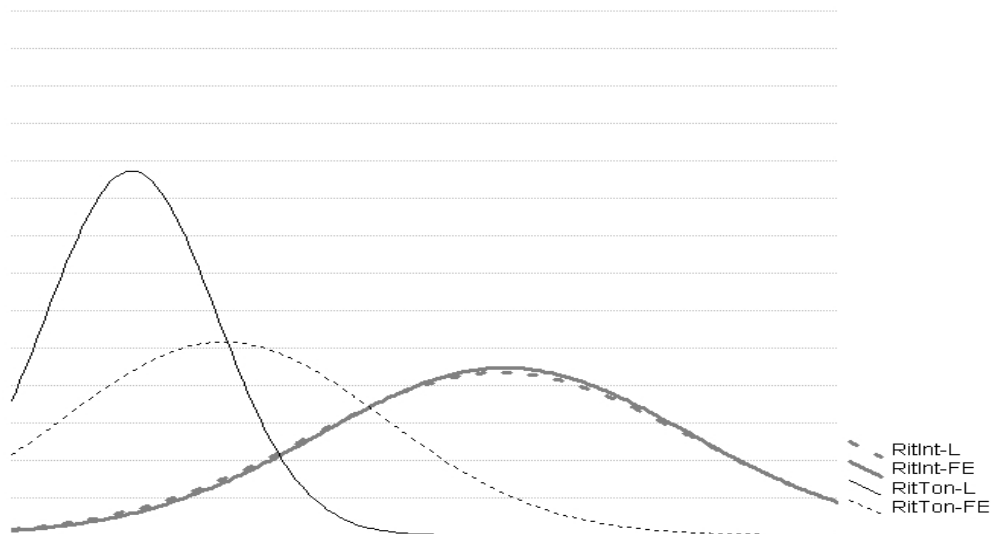
**Tabela 1: Ocorrências de Ritmo de Intensidade e Ritmo Tonal por sujeito**

Ritmo de Intensidade		Ritmo Tonal	
TL	FE	TL	FE
3	4	0	0
5	2	0	0
1	0	0	0
3	2	2	2
3	4	0	2
4	3	0	3
5	2	1	0
3	4	0	2
3	5	1	1
4	3	1	1
5	5	0	2
1	1	0	0
3	5	0	0
4	4	0	0
4	3	0	2
3	2	0	3
1	3	0	0
4	5	0	2
5	2	0	2
4	3	1	0
1	2	0	1
1	4	0	0
2	5	1	4
1	3	1	1
3	3	2	0
4	2	0	0
2	4	0	0
4	4	0	2
1	1	0	0
4	2	1	1

A comparação entre esses dados apresenta resultados significativos para TL em relação a ritmo de intensidade e ritmo tonal ( $F_o(5,01) > F_c(1,86)$ ,  $P < 0,001$ ), e para o ritmo

tonal em relação à TL e FE ( $F_o(0,28) < F_c(0,53)$ ,  $P < 0,01$ ). Para este caso, cuja variação manteve-se no limite da aleatoriedade, realizamos um teste t, com amostra em par, que apresentou resultados significativos ( $t_o(2,82) > t_c(2,05)$ ,  $P < 0,01$ ).

**Gráfico 1: Ritmo Tonal e de Intensidade nas categorias FE E TL**



Tendo em vista que nas frases de TL a ocorrência de ritmo tonal não é significativa, entendemos haver um maior controle da entoação frasal. No tocante à FE, a presença de ritmo tonal de forma mais significativa parece apontar para um maior controle da declinação frasal, fazendo com que o falante procure sempre manter o tom de sua fala estável. Por outro lado, na fala espontânea, a necessidade da seleção lexical poderia estar concorrendo com a entoação na construção da frase.

## **Bibliografia**

BARBOSA, Plínio Almeida (1995). Estrutura rítmica da frase revelada por aspectos de produção e percepção de fala. XLIII Seminário do GEL. São Paulo. (disponível em [www.lafape.iel.unicamp.br/Publicacoes/gel1995.pdf](http://www.lafape.iel.unicamp.br/Publicacoes/gel1995.pdf)).

CAGLIARI, L.C. Elementos de fonética do português brasileiro. Tese de Livre-Docência apresentada à UNICAMP em 1982. São Paulo: Paulistana, 2007.

CHOMSKY, N. e HALLE, M. *The sound pattern of English*. New York: Harper & Row, 1968.

- CRUTTENDEN, Alan. *Intonation*. Cambridge, Cambridge University Press, 1986.
- FERREIRA NETTO, W. *Variação de frequência e constituição da prosódia da língua portuguesa*. Tese de Livre-Docência. Universidade de São Paulo, 2006.
- FÓNAGY, I. Des fonctions de l'intonation: *Essai de synthèse*. *Flambeau*, 29, p.1 –20, 2003.
- GLASER, S. The missing link: Connections between musical and linguistic prosody. *Contemporary Music Review*, n. 19, v. 3, p. 131-154, 2000.
- HALLE, M. & VERNAUD, J-R. *An essay on stress*. Cambridge, Mass.: MIT Press, 1987.
- HAYES, B. *Metrical Stress Theory – Principles and Case Studies*. UCLA, 1994.
- KUHL, P.K.; IVERSON, P. Linguistic experience and the “perceptual magnet effect”. In: STRANGE, W. (ed.). *Speech perception and linguistic experience: issues in cross-language research*. Baltimore: York Press, 1995. p. 121-154.
- LADD, R. *Intonation Phonology*. Cambridge: Cambridge University Press, 1986.
- LEVIN, Jack. *Estatística aplicada a ciências humanas*. 2.ed. São Paulo; Harbra, 1985.
- LIBERMAN, M. & PRINCE, A. “On stress and linguistic rhythm”. *L. In.* 8: 249 – 336, 1977.
- LONGACRE, Robert E.; CHENOWETH, Vida. Discourse as music. *Word*, n. 37, v. 1-2. p. 125-134, 1986.
- MAEDA, S. A Characterization of American English Intonation. Doctoral Dissertation, MIT, 1976.
- MARTINET, A. *Elementos de linguística geral*. São Paulo: Martins Fontes, 1975.
- MASSINI-CAGLIARI, G. *Acento e ritmo*. São Paulo: Contexto, 1992.
- MORAES, J. A. Intonation in Brazilian Portuguese. In: HIRST, D.; DI CRISTO, A. (Eds.) *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998. p. 179-194.
- NESPOR, M. & VOGEL I. *Prosodic Phonology*, Foris Publications. Dordrecht, Holland, 1986.
- PIERREHUMBERT, J. The Phonology and Phonetics of English Intonation. Tese de Doutorado, Massachusetts Institute of Technology, 1980.
- PIKE, Kennet L. *Intonation of English Language*. Ann Arbor, 1954.
- PIKE, Kenneth L. On the grammar of intonation. In: WIRNER E.; BETHGE, W. (Eds.) *Proceedings of th fifth International Congress of Phonetic Sciences*. Basel: s.ed., 1964.
- ROEDERER, Juan G. *Introdução à física e à psicofísica da música*. Trad. Alberto Luis da Cunha. Edição original de 1975. São Paulo: Edusp, 2002.

RUSSO, I.; BEHLAU, M. *Percepção da fala: análise acústica do Português Brasileiro*. São Paulo: Lovise Científica, 1993.

SCHELLENBERG, Glenn E.; TREHUB, Sandra E. Culture-general and culture-specific factors in the discrimination of melodies. *Journal of Experimental Child Psychology*, n. 74, p. 107–127, 1999.

SCHOENBERG, Arnold. (2008). *Harmonia*. Trad. de Marden Maluf do original alemão *Harmonielehre*, publicado em 1949. São Paulo: Editora Unesp, 2001.

T'HART, J. Differential sensitivity to pitch distance, particularly in speech. *Journal of Acoustical Society of the America*, n. 69, v. 3, p. 811-821, 1981.

TROUBETZKOY, N.S. *Principes de phonologie*. Trad. de J. Cantineau. Paris: Klincksieck, 1964.

