

Efeitos da entoação e da duração na análise automática das manifestações emocionais

(Effects of pitch and duration in the automatic analysis of emotional expressions)

Waldemar Ferreira Netto¹, Marcus Vinícius Moreira Martins²,
Maressa de Freitas Vieira³

¹ Universidade de São Paulo/Conselho Nacional de Pesquisa (USP/CNPq)

² Universidade de São Paulo (USP)

³ Faculdade Sudoeste Paulista (FSO)

wafnetto@usp.br; marcusmartins@usp.br; maressafv@gmail.com

Abstract: Our purpose in this essay is to check if the change in expression of emotions through the prosody of Brazilian Portuguese may be restricted to intonation. We analyzed eight sound files classified as sadness or anger. For the automatic description done by the software ExProsodia, ten variables were defined and divided into two groups: one considering the ideal Middle Tom (TM) and other considering the duration. Quantitative analysis of this set of variables was performed with Cluster Analysis of multivariate statistics, with a confidence level of 0.05, for raw data. The dendrogram analysis showed that vertical variations of F0 are necessary for the automatic separation of files for the emotions assigned to it. However the same variations are not sufficient for a better separation correlated to the proposed emotional classification.

Keywords: prosody; phonetics; automatic analysis.

Resumo: Nosso objetivo neste ensaio é verificar se a variação da manifestação das emoções pela prosódia do português do Brasil pode restringir-se à entoação. Analisaram-se arquivos sonoros classificados como tristeza ou raiva. Para a descrição automática da fala pelo modelo ExProsodia, foram definidas dez variáveis, divididas em dois grupos: um considerando o Tom Médio ideal (TM) e outro considerando a duração. A análise quantitativa desse conjunto de variáveis foi feita com a estatística multivariada de Cluster Analysis. A análise dos dendrogramas obtidos mostraram que as variações verticais de F0 são necessárias para a separação automática dos arquivos pelas emoções que lhe foram atribuídas, mas que não são suficientes para uma separação mais bem correlacionada com a classificação emocional proposta.

Palavras-chave: prosódia; fonética; análise automática.

Introdução

A análise da manifestação das emoções associadas à fala tem sido objeto de especulação científica desde o século XIX. A esse tempo, Darwin afirmava que a tonalidade da voz tem relação com certos sentimentos, exemplificando que uma pessoa, delicadamente reclamando de maus-tratos, ou de um pequeno sofrimento, quase sempre fala com voz aguda (DARWIN, 2000). No mesmo período, Spencer (1890) também fazia associações semelhantes. Ele lembrava que era inegável que certos tons de voz e cadências que têm alguma semelhança com a natureza são espontaneamente usados para expressar tristeza, para expressar alegria, para expressar o afeto e para expressar o triunfo ou ardor marcial. Embora tais proposições se estendessem para além da preocupação com a manifestação das emoções, o reconhecimento da entoação como um fato comunicativo, voluntário ou não, teve seu início no século XIX. A análise da manifestação das emoções associadas à

fala passou a ser objeto de investigação científica com Weaver (1924a; 1924b; 1924c), que foi um dos primeiros a estabelecer parâmetros para descrever a frequência média da fala e as condições de produção. O autor definiu como parâmetros: frequência média normal (*average normal pitch*), extensão das variações tonais (*pitch range* ou *span*), soma dos desvios em relação à média tonal (*total deviation*), número de variações tonais (*total fluctuation*) e a área da curva gráfica que se formou (*curve area*). Os dados analisados foram extraídos de leituras feitas pelos sujeitos. A partir dos resultados obtidos, foram estabelecidas diferenças entre vozes femininas e vozes masculinas.

Numa das primeiras pesquisas que procurou descrever de forma mais sistemática a relação entre a variação de frequência e a manifestação das emoções na fala, Skinner (1935) verificou que a frequência média na fala provocada pela alegria (*happyness*) era mais aguda do que a provocada pela tristeza (*sadness*). Fairbanks e Pronovost (1938; 1939) procuraram estabelecer a relação entre as variações da entoação e a manifestação das emoções na fala, bem como o julgamento dessas emoções por sujeitos ouvintes. A partir da fala simulada com leituras feitas por atores, compararam-se variações de F0 em manifestações emocionais de desprezo (*contempt*), raiva (*anger*), medo (*fear*), tristeza (*grief*) e indiferença (*indifference*). Foram estabelecidos quatro parâmetros: o valor médio da frequência (*pitch level*), a variação tonal média (*wide mean inflectional range*), a extensão tonal (*wide total pitch range*) e a taxa de variação tonal (*pitch change*). As comparações basearam-se nos valores máximos e mínimos encontrados para esses parâmetros. Os resultados mostraram que raiva e medo ocorrem com frequência média mais aguda e que indiferença ocorre com a mais grave. No teste de avaliação dessas emoções, todas as leituras cuja frequência média fosse mais grave foram consideradas desprezo, tristeza e indiferença. Quanto à extensão tonal, desprezo e raiva apresentaram valores mais altos e indiferença, mais baixos. Tristeza apresentou a menor variação tonal e raiva, a maior. Quanto à taxa de variação tonal, a mais rápida foi a raiva e a mais lenta foi o medo. Fairbanks e Hoaglin (1941) analisaram a taxa de duração das mesmas emoções e verificaram que tristeza e indiferença apresentaram as menores taxas de duração, atribuindo esse fato aos prolongamentos das fonações e às pausas.

A partir dos anos 1960, a entoação emocional foi retomada (MARKEL, 1965; CONSTANZO; MARKEL; COSTANZO, 1969; WILLIAMS; STEVENS, 1972; para uma revisão desses trabalhos, cf. SCHERER, 1986; MURRAY; ARNOTT, 1993). Com exceção do trabalho de Williams e Stevens (1972), pesquisas de análise acústica das emoções realizavam-se a partir de dados de fala atuada. Scherer e seus colegas (1984) verificaram que, embora a entoação seja responsável pela manifestação de emoções, seu reconhecimento associa-se a traços gramaticais do texto. Os autores verificaram a necessidade de se distinguirem traços linguísticos e paralinguísticos de F0. Em 1998 (BATLINER et al., 2011), o trabalho de Slaney e McRoberts (1998), com dados de fala dirigida às crianças, inova utilizando dados de fala espontânea na análise automática. Recentemente, um grande número de pesquisas tem analisado fala espontânea (cf. BATLINER et al., 2011 para um levantamento mais detalhado). Bänzinger e Scherer (2005) verificaram que a variação global de F0 é afetada diretamente pelo estímulo emocional representado na fala e é a variação mais importante para a discriminação das categorias emocionais observadas.

Para a língua portuguesa, Vassoler e Martins (2012a; 2012b), analisando raiva, alegria, medo, tristeza e neutra na fala atuada, verificaram que as variações da frequência média e da dispersão, medidas pelo coeficiente de variação da frequência média, são pa-

râmetros que permitem a diferenciação entre as emoções de raiva e as neutras. Vassoler (2012), utilizando os mesmos dados, verificou que a variação da frequência média permite a diferenciação entre as emoções de tristeza e neutras.

Nosso objetivo neste ensaio é verificar se a variação da manifestação das emoções caracterizada pela prosódia da língua portuguesa falada no Brasil pode restringir-se à entoação — entendida como uma sequência de tons, iguais ou diferentes, produzidos pela voz durante a fala — ou se é necessário estender as análises a outros fatos prosódicos de F0, particularmente a duração.

Materiais e métodos

Para a realização desse propósito, analisamos oito arquivos sonoros, em formato mp3, tomados do Youtube. Os arquivos foram classificados como tristeza (T1-4) (DILMA, 2013; ENTREVISTA, 2103; PAIS, 2013) ou raiva (R1-4) (MINISTRA, 2013; CLIENTE, 2013; ATENDIMENTO, 2013; FUZILEIRO, 2013). Um dos arquivos (PAIS, 2013) foi dividido em dois arquivos por conter sujeitos distintos produzindo fala. Todas as extrações sonoras foram realizadas com o *software Soudtap Streaming Audio Recorder* v2.11 (NCH, 2007). Os arquivos sonoros foram editados e segmentados com o programa *Adobe Audition* (ADOBE, 2012) A edição realizada foi a filtragem das vozes de terceiros e de ruídos indesejáveis, mantendo exclusivamente a voz a ser considerada. Para isso foi aplicado o efeito *Dynamic EQ effect* com frequência zero sobre o trecho a ser eliminado. Assim, foi possível manter a duração original de todos os arquivos. A análise e a conversão da curva de frequência fundamental e da curva de intensidade para arquivos de texto foram realizadas pelo *software Speech Filing System* (HUCKVALE, 2008). Todas as demais análises foram feitas pelo aplicativo *ExProsodia* (FERREIRA NETTO, 2008; 2010; PERES, CONSONI; FERREIRA NETTO, 2011).

O programa ExProsodia de pesquisa objetiva a análise automática da entoação no português do Brasil, entendendo que entoação é uma sequência de tons, iguais ou diferentes, produzidos pela voz durante a fala. O desenvolvimento desse programa parte da hipótese de Xu e Wang (1997) de que alguns fatos prosódicos têm restrições mecânico-fisiológicas e outros decorrem das necessidades expressivas dos falantes. Ferreira Netto (2006) propôs que a entoação da fala pode ser decomposta em componentes estruturadoras e semântico-funcionais: finalização (F) e sustentação (S), de um lado, foco/ênfase (E), de outro.

Segundo Ferreira Netto (2006; 2008) e Peres, Consoni e Ferreira Netto (2009; 2011), a produção da fala exige esforço para sustentar a voz com uma frequência relativamente estável, definida aqui como tom médio ideal (TM) de F0, que se repete nos momentos Z(t) mensurados de F0. A supressão desse esforço desencadeia uma declinação pontual que exige a retomada da tensão inicial. A sustentação (S) é consequência do esforço que se acrescenta a cada um dos momentos da fala, incluindo-se o inicial, para compensar a declinação pontual de finalização (F). Ritmo tonal é consequência da ação dessas tendências que atuam em sentidos opostos, possibilitando a produção da fala. A componente F associa-se ao fato de que se trata do tom alvo da declinação pontual, estabelecida por um intervalo ideal decrescente de 7 semitons (st) do TM obtido até o momento Z(t). TM é a tendência central dos valores válidos de F0 calculada como a média

aritmética acumulada no tempo. Os valores válidos mensurados são os momentos de F0 que cumprem as restrições de altura, intensidade e duração. A componente foco/ênfase (E) decorre dos valores que ultrapassaram os limites superior e inferior do TM, definidos em 3 st acima e 4 st abaixo, no momento de chegada do RT. Assim, estabelecido um TM, espera-se que no momento seguinte esse valor se repita, aceitando-se uma variação previamente estabelecida. Esses valores servem especialmente para a língua portuguesa falada no Brasil e não foram testados para outras línguas. A série temporal se configura aditivamente como $Z(t)=S(t)+F(t)+E(t)$.

A seleção das unidades $Z(t)$ — ora chamadas de UBI (*Unit of Base of Intonation*) — para análise é feita pelo aplicativo ExProsodia. Três critérios são estabelecidos para essa definição: frequência maior do que 50 Hz e menor do que 700 Hz; intensidade suficiente para ser percebida e, garantidos os critérios anteriores, duração maior do que 20 ms. Esses valores podem ser modificados pelo usuário. Para esta análise, a UBI foi definida a partir de 50 Hz até o máximo de 700 Hz com duração mínima de 60 ms. Os valores máximos e mínimos da escala de frequência possibilitaram que não houvesse nenhuma restrição necessária quanto ao gênero dos sujeitos que produziram os registros entoacionais. A duração mínima de 60 ms está de acordo com os trabalhos de Schaeffer (1966) e Winckel (1967) que estabelecem como “constante de integração” ou “espessura do presente” a duração mínima de 50 ms, acrescentando uma margem de erro de 20%, tendo em vista a heterogeneidade da amostra (BOEMIO et al., 2005; FERREIRA NETTO, 2006).

Para a descrição automática da fala pelo modelo ExProsodia, foram definidas dez variáveis, divididas em dois grupos: um considerando o Tom Médio ideal (TM) — estabelecido pela técnica de convoluções gaussianas (FERREIRA NETTO et al., 2013), medido em Hz — e outro considerando a duração entre as porções analisadas da fala (UBI) e a duração intrínseca dessas mesmas porções, medidas em milissegundos (ms). Para o primeiro grupo foram definidas as seguintes variáveis: F/Epos_cv, F/Eneg_cv, F0_skew, TM-mUBI, F0_dp/TM; para o segundo: entreUBI, entreUBI_dp, intraUBI e intraUBI_dp.

Das variáveis referentes ao primeiro grupo, F/Epos_cv é o coeficiente de variação das frequências F0 que ultrapassaram o limite lateral superior de 3 st do TM; F/Eneg_cv é o coeficiente de variação das frequências de F0 que ultrapassaram o limite lateral inferior de 4 st do TM; F0_skew é a assimetria dos valores de F0 em relação a uma curva normal; TM-mUBI é a diferença entre o TM e o menor valor obtido na análise das UBI.

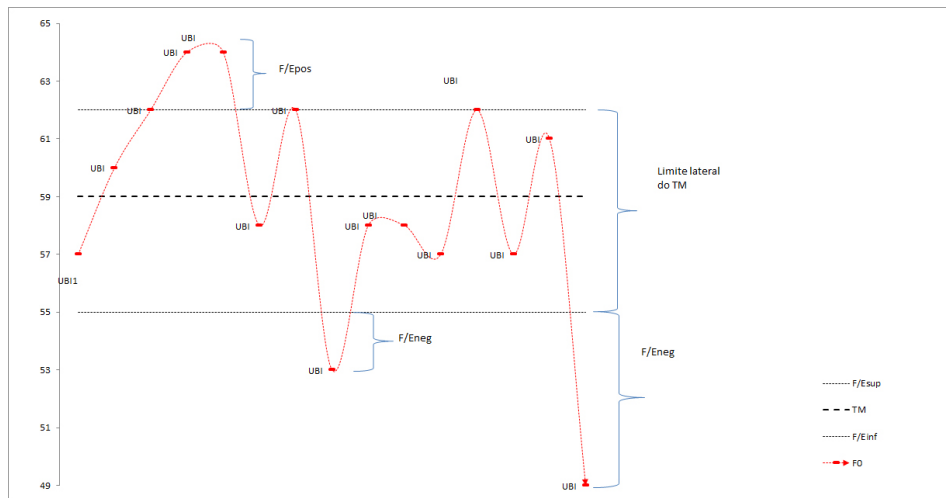


Gráfico 1: No gráfico acima, as chaves em azul mostram as variações de F/Epos e de F/Eneg, além do limite lateral do TM, padrões do sujeito com emoção R (raiva) e em vermelho com emoção T (tristeza). As linhas tracejadas em preto mostram o TM; as pontilhadas, os limites superior e inferior dos TM; e a linha tracejada em vermelho mostra F0.

Das variáveis referentes ao segundo grupo, entreUBI é a duração média do intervalo temporal entre as UBI; entreUBI_dp é a dispersão média, medida em unidades de desvio-padrão, desse intervalo temporal entre as UBI; intraUBI é a duração média das UBI e, finalmente, intraUBI_dp é a dispersão média, medida em unidades de desvio-padrão, da duração das UBI.

Por se ter como propósito compreender a importância da entoação na manifestação prosódica das emoções, reproduzimos abaixo os gráficos referentes ao TM obtidos com a análise automática feita pelo aplicativo ExProsodia. Trata-se especialmente dos arquivos R1 e T4, que foram produzidos pelo mesmo sujeito, em dois momentos distintos.

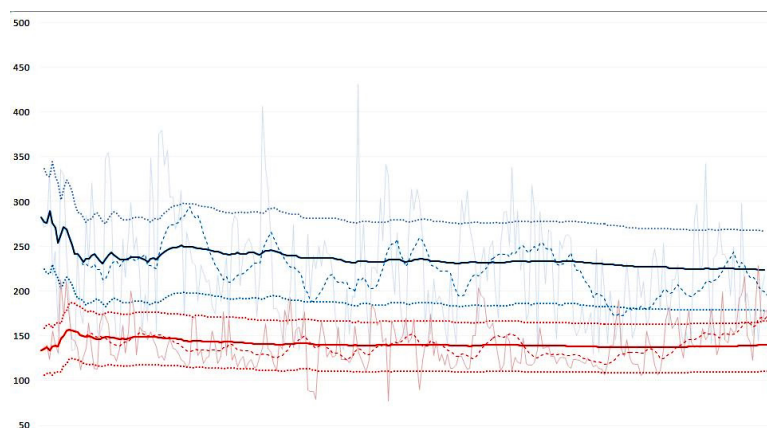


Gráfico 2: O gráfico mostra os padrões da análise automática ExProsodia da fala de um mesmo sujeito. Em azul, veem-se padrões do sujeito com emoção R (raiva) e em vermelho com emoção T (tristeza). As linhas contínuas escuras mostram os TM (tom médio); as pontilhadas, os limites superior e inferior dos TM; as linhas claras mostram F0 e as tracejadas, a média móvel de 15 pontos de F0. A escala à esquerda está em Hz e está correta para ambos os padrões.

No Gráfico 1, é perceptível que a variação do TM é responsável pela diferenciação entre as duas manifestações emocionais representadas pela curva entoacional. Os valores de F0 de R1 (em azul) e de T4 (em vermelho) confundem-se em vários momentos. Resultados semelhantes foram obtidos por Vassoler e Martins (2012a; 2012b). Os autores partiram da fala produzida por atores que procuraram simular manifestações emocionais a partir de um texto estímulo. A variação do TM para a simulação da raiva em relação a uma simulação de emoção neutra também foi perceptivelmente mais aguda em relação à neutralidade, mais grave.

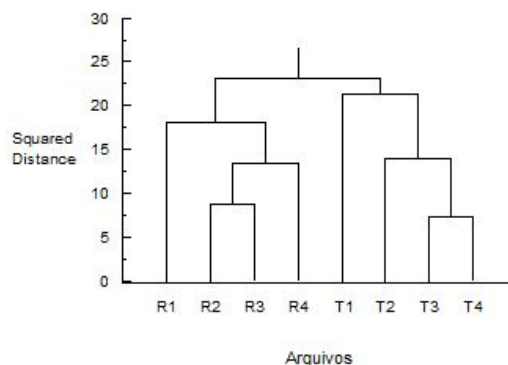
Resultados

Os resultados gerais podem ser vistos na Tabela 1, abaixo.

Tabela 1: Nas linhas são descritas as variáveis (ver o texto) e nas colunas, os arquivos analisados. As abreviaturas referentes aos nomes dos arquivos descrevem T para tristeza e R para raiva.

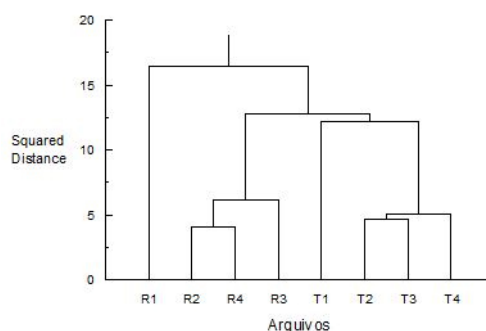
	R1	R2	R3	R4	T1	T2	T3	T4
F/Epos_cv	0,85	0,85	0,92	1,06	1,07	0,75	0,8	0,83
F/Eneg_cv	-0,55	-0,75	-0,67	-0,72	-0,72	-0,67	-0,67	-0,75
F0_skew	0,57	0,27	0,13	0,4	0,89	0,43	0,51	0,67
TM-mUBI	94	109	154	112	109	99	71	63
F0_dp/TM	0,27	0,24	0,22	0,22	0,18	0,2	0,18	0,18
TM_cv	0,05	0,04	0,05	0,04	0,04	0,02	0,04	0,02
entreUBI	52,37	53,87	61,3	56,6	54,7	81,4	57,1	63,35
entreUBI_dp	79,22	65,43	70,8	78,2	82,6	102,2	79,8	95,42
intraUBI	109,15	110,3	99,9	94,8	122,2	113,3	100	101,22
intraUBI_dp	75,9	77,28	59,7	72,4	81,2	77	76,9	74,59

A análise quantitativa desse conjunto de variáveis foi feita pelo *software* KyPlot (YOSHIOKA, 1997). A estatística utilizada foi a análise multivariada Cluster Analysis, com um nível de confiança de 0,05, para dados brutos. As medidas foram tomadas com a opção Standardized Euclidean, pelo Clustering Method de Group Average (Average Linkage). O dendrograma a seguir vai apresentado em Squared Distances.



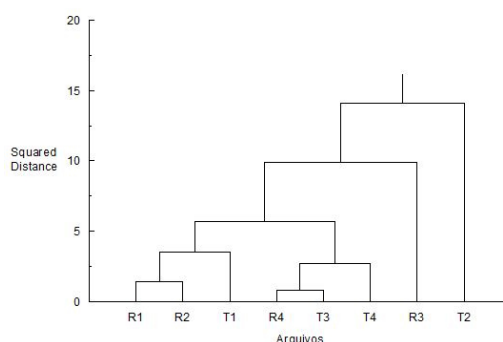
Dendrograma 1: No eixo horizontal do dendrograma vão os arquivos sonoros analisados, classificados como R_{1-4} para raiva e T_{1-4} para tristeza. No eixo vertical vão apontadas as distâncias em Square Distances tomadas em Standardized Euclidean.

No dendrograma podem-se observar que os arquivos contendo gravações classificadas como Raiva (R) ficaram organizados sob um mesmo nó à esquerda e que os classificados como Tristeza (T) ficaram organizados sob um mesmo nó à direita. A análise de *clusters* a partir das variáveis globais considerou as porções analisadas dos arquivos tanto do ponto de vista das variações de entoação quando das variações de duração entre elas e intrínseca a elas. O dendrograma 2 mostra os mesmos arquivos, analisados pela mesma estatística multivariada .



Dendrograma 2: No eixo horizontal do dendrograma vão os arquivos sonoros analisados, classificados como R_{1-4} para raiva e T_{1-4} para tristeza. No eixo vertical vão apontadas as distâncias em Square Distances tomadas em Standardized Euclidean.

No Dendrograma 2 acima, observa-se que a separação entre arquivos classificados como T ou como R não resultou em dois conjuntos correlacionados a essa classificação. A diferença entre ambas análises decorreu da ausência das variáveis entreUBI, entreUBI_dp, intraUBI e intraUBI_dp que consideravam a durações e as dispersões entre as porções analisadas e as intrínsecas a essas mesmas porções.



Dendrograma 3: No eixo horizontal do dendrograma vão os arquivos sonoros analisados, classificados como R₁₋₄ para raiva e T₁₋₄ para tristeza. No eixo vertical vão apontadas as distâncias em Square Distances tomadas em Standardized Euclidean.

No Dendrograma 3 acima, que considera exclusivamente as variáveis referentes à duração — entreUBI, entreUBI_dp, intraUBI e intraUBI_dp — podemos verificar que não houve separação dos arquivos de maneira a se correlacionar com as classificações T e R. Ao contrário da análise de *cluster* que considerou exclusivamente as variações de frequência, em que arquivos classificados como T foram isolados num subagrupamento, na análise exclusiva das durações não ocorreu nada parecido.

Considerações finais

As variações verticais de F0 mostram-se necessárias para a separação automática das emoções, no entanto essas mesmas variações não foram suficientes para uma separação mais bem correlacionada com a classificação semântica proposta para os arquivos. Embora esses resultados apontem para um fato semelhante ao dos resultados obtidos por Bänziger e Scherer (2005) para o alemão, podem-se salientar algumas diferenças. Os autores sugeriram que, para uma análise mais eficaz das manifestações emocionais, timbre (fonético) e sintaxe poderiam contribuir de forma mais significativa. Naquele trabalho os autores valeram-se de produções feitas por atores a partir de frases desprovidas de significado. Neste trabalho, os arquivos foram coletados de situações reais de produção de fala, o que, de certa maneira, acrescenta a análise fonética e a sintática para o esboço de uma interpretação automática das emoções. Considere-se, no entanto, que foi necessário o acréscimo das variações de duração das porções portadoras da entoação analisadas nos arquivos. Essa diferença pode decorrer não somente do fato de que a produção de fala deste trabalho foi tomada em contextos muito distintos daqueles de Bänziger e Scherer (2005), mas também do fato de se estar aqui tratando especialmente da língua portuguesa falada no Brasil.

REFERÊNCIAS

ADOBE Systems Incorporated. *Adobe Audition 3.0.1 build 8347.0* [Audition\3.0]. San Jose, 2012.

ATENDIMENTO SAMU Juiz de Fora... Disponível em: <<http://www.youtube.com/watch?v=l8wVeVT3uDw>>. Acesso em: 19 fev 2013

BÄNZIGER, T.; SCHERER, K.R. The role of intonation in emotional expressions. *Speech Communication*, n. 44, p. 252-267, 2005.

BATLINER, A. et alii. The automatic recognition of Emotions in Speech. In.: COWIE, R.; PELACHAUD, C.; PETTA, P. (Eds.) *Emotion-oriented Systems*. The Humaine Book. Berlin: Springer, 2011. p. 71-99.

BOEMIO, A.; FROMM, S.; BRAUN, A.; POEPEL, D. Hierarchical an asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, v. 8, n. 1, p. 389-395, 2005.

CLIENTE calminho. Disponível em: <<http://www.youtube.com/watch?v=53j7lZrp2pM>>. Acesso em: 19 fev. 2013.

COSTANZO, F. S.; MARKEL, N. N.; COSTANZO, P. R. Voice quality profile and perceived emotion. *Journal of Counseling Psychology*, v. 16, n. 3, 267-270, 1969.

DARWIN, C. *A expressão das emoções no homem e nos animais*. São Paulo: Companhia das Letras, 2000.

DILMA se emociona ao falar de tragédia... Disponível em: <<http://www.youtube.com/watch?v=trfM5uMH0mY>>. Acesso em: 19 fev. 2013.

ENTREVISTA com Ana Carolina. Disponível em: <<http://www.youtube.com/watch?v=3B5eGuG0C8Q>>. Acesso em: 19 fev. 2013.

FAIRBANKS, G.; HOAGLIN, L. W. An experimental study of the durational characteristics of the voice during the expression of emotion. *Speech Monographs*, v. 6, n. 1, p. 85-90, 1941.

FAIRBANKS, G.; PRONOVOST, W. Vocal pitch during simulated emotion. *Science*, v. 78, n. 2286, p. 382-383, 1938.

_____. An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech Monographs*, v. 6, n. 1, p. 87-104, 1939.

FERREIRA NETTO, W. Decomposição da entoação frasal em componentes estruturadoras e em componentes semântico-funcionais. In: CONGRESSO NACIONAL DE FONÉTICA E FONOLOGIA / CONGRESSO INTERNACIONAL DE FONÉTICA E FONOLOGIA, 10/4., Niterói, 2008. Disponível em: <http://www.academia.edu/2272651/Decomposicao_da_entoacao_frasal_em_componentes_estruturadoras_e_semantico-funcionais> Acesso em: 22 jan. 2013

_____. ExProsodia. *Revista da Propriedade Industrial – RPI*, 2038, pág. 167, item 120, em 26/out/2010. Disponível em: <<http://revistas.inpi.gov.br/pdf/PATENTES2038.pdf>> Acesso em: 22 jan. 2013.

_____. *Variação de frequência e constituição da prosódia da língua portuguesa*. Tese (Livre-Docência) – USP, 2006.

FERREIRA NETTO, W.; SOUZA, A. R.; VIEIRA, M. F.; PERES, D. O.; MARTINS, M. V. M. Desenvolvimento de análise automática da curva de frequência por meio de convoluções gaussianas do histograma de frequências. In: COLÓQUIO BRASILEIRO DE PROSÓDIA DA FALA-UFAL, 4., Maceió, Al, em 16 out. 2013.

FUZILEIRO naval – *Toca terror na delegacia*. Disponível em: <<http://www.youtube.com/watch?v=UlaVqH6SsXs>>. Acesso em: 19 fev. 2013.

HUCKVALE, M. A. *Speech Filing System v.4.7/Windows SFSWin Version 1.7*, em 17/02/2008. Disponível em: <<http://www.phon.ucl.ac.uk/resource/sfs>> . Acesso em: 22 jan. 2013.

MARKEL, N. N. The reliability of coding paralinguistic: pitch, loudness, and tempo. *Journal of Verbal Learning and Verbal Behavior*, n. 4, p. 306-308, 1965.

MINISTRA Dilma. Disponível em <<http://www.youtube.com/watch?v=z6-NjcE8t7g>>. Acesso em: 19 fev. 2013.

MURRAY, I. R.; ARNOTT, J. L. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of Acoustical Society of Americas*. v. 93, n. 2, p. 1097-1108, 1993.

NCH Software. *Soundtap Streaming Audio Recorder v2.11*. Greenwood Village, 2007. Disponível em: <<http://www.nch.com.au/soundtap/index.html>>. Acesso em: 19 fev. 2013.

PAIS de João Helio fazem depoimento emocionante. Disponível em: <<http://www.youtube.com/watch?v=GNaf7qXDGVA>>. Acesso em: 19 fev. 2013.

PERES, D. O.; CONSONI, F.; FERREIRA NETTO, W. Decomposição da entoação frasal em componentes estruturais e semântico-funcionais: um teste com análise da variação de gênero. In: *OSUCHIL – The Ohio State University Congress on Hispanic and Lusophone Linguistics*, 12, Ohio, 2009.

PERES, D.; CONSONI, F.; FERREIRA NETTO, W. A influência da cadeia segmental na percepção de variações tonais. *LLJournal*, v. 6, n. 1, 2011. Disponível em: <<http://ojs.gc.cuny.edu/index.php/lljournal/article/view/652>> Acesso em: 22 jan. 2013

SCHAEFFER, P. *Traité des objets musicaux*: essai interdisciplines. Paris: Seuil, 1966.

SCHERER, K. R. Vocal affect expression: a review and a model for futures research. *Psychological Bulletin*, v. 99, n. 2, p. 143-165, 1986.

SCHERER, K. R.; LADD, D. R.; SILVERMAN, K. E. A. Vocal cues to speaker affect: testing two models. *Journal of Acoustical Society of Americas*, v. 76, n. 5, p. 1346-1356, 1984.

- SIERTSEMA, B. Timbre, pitch and intonation. *Lingua*, v. 11, p. 388-398, 1962.
- SKINNER, E. R. A calibrated recording an analysis of the pitch, force and quality of vocal tones expressing happiness and sadness. *Speech Monographs*, v. 2, n. 2, p. 81-137, 1935.
- SLANEY, M.; McROBERTS, G. Baby ears: a recognition system for affective vocalization. In : INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECHS AND SIGNAL PROCESSING, 1998, Seattle. *Proceedings...* Vol. II: Features for automatic speech recognition. Seattle: The Institute of Electrical and Electronic Engineers Signal Processing Society, 1998. p. 985-988.
- SPENCER, H. The origin of music. *Mind*, v. 15, n. 60, p. 449-468, 1890.
- VASSOLER, A. M. O. A Prosódia das Emoções: Um Exercício a partir da Fonologia Prosódica. In: BAIA, F. A. et alii. (Org.) *Anais dos XII e XIII Encontros dos Alunos de Pós-Graduação em Linguística da USP*. São Paulo: Paulistana, 2012. p. 1-12.
- VASSOLER, A. M. O.; MARTINS, M. V. M. *A entoação em falas teatrais: uma análise da raiva e da fala neutra*. In: SEMINÁRIO DO GEL, 60., 2012, São Paulo. Seminário do GEL, 60. São Paulo, 2012a.
- _____. *O estudo da variabilidade da frequência fundamental intra-sujeito na fala emotiva*. In: ENCONTRO DO DINAFON, 5., 2012, Marília. Anais do 5º Encontro do Dinafon, 2012b.
- WEAVER, A. T. Experimental studies in vocal expression. *The Quarterly Journal of Speech Education*, v. 10, n. 3, p. 199-204, 1924a.
- _____. Experimental studies in vocal expression. *The Journal of Applied Psychology*, v. 8, n. 1, p. 23-51, 1924b.
- _____. Experimental studies in vocal expression. *The Journal of Applied Psychology*, v. 8, n. 2, p. 159-186, 1924c.
- WILLIAMS, C. E.; STEVENS, K. N. Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*, v. 52, n. 4, parte 2, p. 1238-1250, 1972.
- WINCKEL, F. *Music, sound and sensation*. A modern exposition. New York: Dover Publications, 1967.
- XU, Yi; WANG, Q. E. Component of intonation: what are linguistic, what are mechanical/physiological? Presented at International Conference on Voice Physiology and Biomechanics, Evanston Illinois, 1997.
- YOSHIOKA, K. *KyPlot*, version 2.0 beta 15 (32 bit). [sl], 1997. Disponível em: <<http://www.uv.es/piefisic/w3pie/castellano/serv/laboratorios/Kyp2b15.exe>>. Acesso em: 19 fev. 2013.